A5

100 mm

1.0 mm

1.5 mm

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890

2.0 mm

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

DOCUMENT RESUME

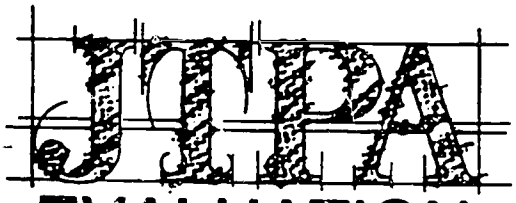
ED 278 844 CE 046 392

AUTHOR Stromsdorfer, Ernst; Blalock, Ann Bonar
TITLE JTPA Evaluation at the State and Local Level. Volume VII: Issues Related to Net Impact Evaluations. A. Issues in Evaluating Costs and Benefits. B. The Debate over Experimental vs. Quasi-Experimental Approaches.
INSTITUTION Washington State Dept. of Employment Security, Olympia.
SPONS AGENCY International Business Machines Corp., Armonk, N.Y.; National Commission for Employment Policy (DOL), Washington, D.C.
PUB DATE Mar 86
NOTE 106p.; A product of the JTPA Evaluation Design Project. For related evaluation materials, see CE 046 385-393.
PUB TYPE Guides - Non-Classroom Use (055)

EDRS PRICE MF01/PC05 Plus Postage.
DESCRIPTORS Comparative Analysis; *Cost Effectiveness; Cost Estimates; Educational Benefits; Educational Legislation; Employment Programs; Evaluation Criteria; *Evaluation Methods; Federal Legislation; Local Issues; Program Costs; *Program Evaluation; *Quasiexperimental Design; *Research Design; Research Methodology; *Statewide Planning
IDENTIFIERS Impact Studies; *Job Training Partnership Act 1982; Service Delivery Areas

ABSTRACT

This guide is intended to assist states and service delivery areas (SDAs) in addressing the new oversight responsibilities and opportunities stipulated by the Job Training Partnership Act (JTPA) with respect to net impact evaluations. It is divided into two main parts. The first part, which deals with issues in evaluating costs and benefits, covers the following topics: the distinction between benefit-cost and effectiveness-cost evaluation; identification and measurement of benefits and costs (general taxonomy of benefits, general taxonomy of costs, practical problem of cost estimation, introduction to economic cost theory, cost estimation at the service delivery area and state levels); and basic methods and issues in benefit-cost analysis. Examples of total, average, and marginal cost functions and a sample federal report for obtaining new data are appended. The second part of the guide, which examines the debate over experimental versus quasiexperimental approaches to assessing net impact, discusses the context of the debate, scientific trade-offs in the debate, and organizational and political trade-offs. Appendixes include definitions of the major concepts involved, some sources and kinds of bias, and a research design continuum. (MN)



EVALUATION
DESIGN PROJECT

JTPA Evaluation at the State and Local Level

Volume VII: Issues Related to Net Impact Evaluations
A. Issues in Evaluating Costs and Benefits
B. The Debate Over Experimental
vs.
Quasi-Experimental Approaches

March 1986

Special appreciation is expressed to the National Commission for Employment Policy, for serving as the project's national sponsor and contributing substantial staff consultation to the project as it developed.

Project Development and Coordination:

Washington State Employment Security
Isiah Turner, *Commissioner*

Project Funding:

National Commission for Employment Policy
IBM Corporation
Washington State Employment Security

CONTEXT OF THIS VOLUME

This is one in a series of volumes produced by the JTPA EVALUATION DESIGN PROJECT.

PURPOSE AND PHILOSOPHY

The purpose of this project has been to develop a set of evaluation tools that are useful to states and local service delivery areas (SDAs) in judging the way their JTPA programs are being managed and the impact they are having. The intent has been to base these analytic and managerial tools on sound program concepts and research methods, and to adapt them such that the information obtained is of practical and direct use in improving JTPA policies and programs at the state and local level. This kind of information is also expected to make a unique contribution to national training policy and Federal oversight of JTPA.

It is hoped that these volumes will stimulate and support state and local evaluation efforts in JTPA, and promote more consistency than in previous programs with respect to the issues studied and the methods used to investigate them. An important goal is to encourage the generation of complementary information on program implementation and impact that is comparable across states and SDAs. Comprehensive, comparable information is essential to the development of a valid and reliable knowledge base for resolving problems and improving programs. It is also required for adjusting national training strategies to changing needs and priorities at the state and local level.

PRODUCTS

Consistent with this purpose and philosophy, the project has produced a set of materials to assist states and SDAs in evaluating their programs. These are to be useful in planning, designing and implementing evaluation activities. As an integrated collection, each set is developed to support comprehensive evaluations over the JTPA planning cycle.

The careful tailoring of these materials to state and local users is appropriate. JTPA represents a new employment and training policy shaped not only by the experience of managers and the perspectives of employers, but by scientific assessments of previous approaches for addressing unemployment, poverty and other barriers to economic security. In this context, the value of JTPA programs is also expected to be judged. In fact, the Act's assessment requirements are more explicit and sophisticated than those of any employment and training legislation to date. It clearly distinguishes between *monitoring* activities, whose purpose is to determine compliance (such as with performance standards) and *evaluation* activities, whose purpose is to determine how a program is being managed and implemented, and the kinds of effects it is having on recipients and relevant others. Equally significant, new constitutencies are expected to make these more rigorous assessments. States and SDAs now have this important responsibility. It is the first time in the history of employment and training programs that the Federal government's evaluation role has been significantly reduced.

This change affords states and local areas opportunities to influence public policy. It also requires them to assume new oversight responsibilities. Program evaluation is expected to become an integral part of the management of organizations administering, planning and delivering public training services. This is as it should be. The more information available at these levels, where changes in organizations can most readily be made, the more effective the management of JTPA programs. This project was undertaken in that context.

The evaluation tools produced by the project have been developed with a sensitivity to the differing needs, interests and resources of state and local users. They have been packaged into a single comprehensive and integrated set of volumes called *JTPA Evaluation at the State and Local Level*. The set contains planning and evaluation *guides* and *issue papers*. The following volumes are available in the set:

Volume	Author
I: Overview	Project Team
II: A General Planning Guide	Deborah Feldman
III: A Guide for Process Evaluations	David Grembowski
III Supplement: Some Process Issues at the State Level	David Grembowski
IV: A Guide for Gross Impact Evaluations	Carl Simpson
V: A Guide for Net Impact Evaluations	Terry Johnson
VI: An Implementation Manual for Net Impact Evaluations	Terry Johnson
VII: Issues Related to Net Impact Evaluations	
A. Issues in Evaluating Costs and Benefits	Ernst Stromsdorfer
B. The Debate Over Experimental vs. Quasi-Experimental Approaches	Ann Blalock
VIII: MIS Issues in Evaluating JTPA	David Grembowski

NOTE: Although each of the discrete products listed above is the responsibility of a single author, each seeks to incorporate the results of professional peer review, the many excellent recommendations of the advisory group, and the ideas and suggestions of the numerous practitioners interviewed in the process of developing these materials.

To further qualify these volumes, Volume III is accompanied by a supplement for state users. This is consistent with the significant differences between states and SDAs in the kinds of process issues that are most essential to study. The volume on net impact evaluations is sufficiently technical, because of the statistical methods involved, that a practical manual has been written to accompany it. This guide and manual tend to be more appropriate for states, since relatively large sample sizes are required for analysis. However, they are equally useful to larger SDAs and consortia of smaller SDAs which may want to jointly study the net impact of their programs. Regional evaluations, for example, can be very productive in providing management information relevant to regional labor markets. Although there is a separate issue paper on evaluating costs and benefits, this issue is also covered in the gross impact and net impact guides. In this respect, the user benefits from three related but different approaches to this important element of program evaluations. Also, the user should be aware that the Appendix of Volume II includes *A Report on a National/State Survey of Local JTPA Constituencies*. This survey was carried out by Bonnie Snedeker, with the assistance of Brian O'Sullivan, to provide additional input from practitioners to the development of the planning and process evaluation guides.

In conclusion, several expectations have directed the development of these volumes:

THE GUIDES

The General Planning Guide

This guide is to assist users in *planning, funding and developing an organizational capacity* to carry out process, gross outcome, and net impact evaluations and to *utilize their results*. Separate state and local versions are available.

The Evaluation Guides

These volumes are to have the following characteristics:

☐The guides are to *complement* one another.

- •They are to provide information on program management and other characteristics of program implementation, which can:
  - —Describe the way in which administrative, managerial and service delivery policies and practices operate to affect outcomes, as a set of interventions separate from the program's services.
  - —Pinpoint the source, nature and extent of errors and biases for which adjustments must be made in gross and net impact evaluations.
  - —Help explain the results of gross and net impact evaluations.
- •They are to provide information on aggregate gross outcomes, and outcomes differentiated by type of service and type of recipient, which can:
  - —Describe relationships between certain implementation modes and service strategies, and a broad array of client and employer outcomes.
  - —Help explain the results of net impact evaluations.
  - —Suggest the more important outcomes that should be studied in net impact evaluations.
  - —Help sort out those aspects of implementation that may be most critical to study in process evaluations.
- •They are to provide information on net impact (the program's return on investment), which can:
  - —Closely estimate the effect of the program's services on clients.
  - —Suggest which services and client groups are most important to study in broader but less rigorous gross impact studies.
  - —Help identify the decision points in program implementation (particularly service delivery) which may be most important to study in process evaluations.

☐The guides are to enable the user to carry out *comprehensive assessments* of JTPA programs.

- •They are to allow the user to acquire several different perspectives on the same program within a particular time period: on program implementation, on outcomes for clients and employers and on net impact.
- •They are to permit the user to interrelate these different kinds of information to gain a wider understanding of what is happening in a program and why.

☐The guides are to describe approaches and methodologies as consistently as possible, to achieve *comparability*.

- •They are to define variables and relationships as similarly as possible.
- •They are to define research designs, and methods of data collection and analysis using as similar concepts as possible.

☐The guides are to *draw from past research* on employment and training programs, as well as seek *new* approaches and methods of specific value in evaluating JTPA at the state and local level.

- •They are to replicate, to the extent possible and feasible, the issues and measures reflected in Federal monitoring and evaluation decisions.
- •They are to make selective use of the results of relevant CETA studies, national studies of JTPA, and issue papers on JTPA evaluation by national public interest organizations in the employment and training area.
- •They are to rely on the professional literature in applied social research.

THE ISSUE PAPERS

Volume VII contains two issue papers which serve as companion pieces to the preceding volumes on net impact evaluation. The first paper on cost-benefit issues is designed to help users identify, measure and analyze relationships between monetary and nonmonetary costs and benefits in determining the program's return on investment. The second paper examines the pros and cons of different research strategies associated with the net impact approach. The final volume on MIS issues is to assist users in better understanding how JTPA and other employment and training *management information systems* can efficiently support the evaluation of program implementation and impact.

THE SET OF VOLUMES

The set is *integrated*, but affords *flexible use*. The user can utilize the entire set for comprehensive evaluations over a two-year planning cycle or longer planning period, or the user can apply the information in each volume independently, based on the most pressing evaluation priorities and timeframes and given the extent of resources, during a particular fiscal year or biennium.

It should be understood that although evaluation products have been developed for JTPA, their basic principles and methods can be applied more broadly by states and local areas to evaluate other employment and training programs and other social programs.

GENERAL ACKNOWLEDGMENTS

The JTPA EVALUATION DESIGN PROJECT was developed and carried out based on the partnership philosophy that underlies the JTPA legislation. Several partnerships should be recognized for their substantial contributions to the products previewed here: the project development and coordination partnership; the public-private funding partnership; the interdisciplinary design partnership; and the advisory partnership.

The Development and Coordination Partnership: Washington Employment Security Department

Project Coordinators: Gary Bodeutsch, Ann Blalock

Assistant Coordinators: Deborah Feldman, Chris Webster

Interdivisional Consultants:

Training Division: Martin McCallum, Brian O'Neall, Ross Wiggins

Research and Information Systems: Jeff Jaksich, Mike Gioimo, Irv Lefberg

The Public/Private Funding Partnership

Funders:

IBM Corporation: Corporate Support Program, Armonk, NY

National Commission for Employment Policy, Washington, D.C.

Washington State Employment Security Department, Olympia, WA

Contributors:

Safeco Insurance Company of America, Seattle, WA

Seattle/King County Private Industry Council, Seattle, WA

SPSS, INC., Chicago, IL

The Interdisciplinary Design Partnership

Consultant:	*Area of Expertise:*	*Affiliation:*
Burt Barnow	Labor Economics	ICF, INC., Washington, D.C.
David Grembowski	Planning	University of Washington, Seattle, WA
Terry Johnson	Labor Economics	Battelle Memorial Institute, Seattle, WA
Brian O'Sullivan	Planning	Seattle/King County Private Industry Council, Seattle, WA
Carl Simpson	Sociology	Western Washington University, Bellingham, WA
Bonnie and David Snedeker	Management	Snedeker Scientific, Inc., Seattle, WA
Ernst Stromsdorfer	Labor Economics	Washington State University, Pullman, WA

The Advisory Partnership

The National Advisory Committee:

State of California: Tim Taormina
San Francisco SDA: Greg Marutani

State of Kansas: Richard Hernandez
Pittsburg SDA: John Gobetz

State of Massachusetts: Kay Stratton
New Bedford SDA: Paul Vigeant

State of New York: Robert Bernstein
Dunkirk PIC: Bruce Ritenberg III

State of Texas: Rick Mackay
Fort Worth PIC: Judge Michael Moncrief

State of Washington: Ernest La Palm
Seattle/King County PIC: Al Starr

Additional Advisors to the Project:

National Governors Association: Kay Albright, Washington, D.C.

HRDI: AFL-CIO: Candace Brown, Helena, MT

U.S. Department of Labor: Hugh Davies, Washington, D.C.

National Association of Counties: Kathy DuChene, Washington D.C.

State of Ohio Impact Assessment Unit: Gail James, Columbus, OH

Berkeley Planning Associates: Deborah Kogan, Berkeley, CA

Abt Associates: Jeff Zornitsky, Cambridge, MA

Special appreciation is expressed to the Commissioner and Deputy Commissioner of the Employment Security Department, Irich Turner and Ernest La Palm, for supporting the need for evaluation at the state and local level and intitiating this project to address that need. They provided the strong encouragement and the resources necessary for designing the project, seeking a funding base for carrying it out and coordinating its development.

Other individuals made unique contributions to the development of the project: John Wallace and Ann Donohue of the National Commission for Employment Policy; Kay Albright of the National Governors Association; Dan Kelly of IBM's Corporate Support Program and Jim Ward of IBM's Olympia, WA office; Mark Cooper of the Safeco Insurance Company; and Steve Ballowe of SPSS, Inc.

A. Issues in Evaluating Costs and Benefits

By Ernst Stromsdorfer

March 1986

AUTHOR'S ACKNOWLEDGEMENTS

A number of my colleagues have helped me in the design and writing of this volume. Foremost among these is Ann Blalock, who conceived the project and has been its driving force. Her sound ideas and astute management of this project have helped guarantee the success of all our efforts. Gary Bodeutsch took over the shepherding of this project and his steady guidance and advice have improved my effort considerably. Terry Johnson, Dave Grembowski, Gary Holman and Carl Simpson have all served as excellent critics of this work. Their help is very much appreciated. Finally, the assistance of the many individuals in the Washington Employment Security Department and the Seattle/King County Private Industry Council is appreciated. This was an effective team effort in every way. Anything useful in this volume I owe to their help. Any mistakes or errors in reasoning are mine.

Ernst Stromsdorfer

ISSUES IN EVALUATING COSTS AND BENEFITS

INTRODUCTION

This volume contributes to the analysis of the JTPA program impacts by describing the rationale and procedures for estimating JTPA program costs and by showing how costs and benefits are related in a human capital investment framework. This investment framework then becomes a management tool to increase the efficiency of JTPA operation. The justification for this approach is contained in Sec. 106 of the JTPA which states that "..., the Congress finds that--

(1) it is essential that criteria for measuring the return on this investment be developed; and

(2) the basic return on the investment is to be measured by the increased employment and earnings of participants and the reduction in welfare dependency."

Thus, Congress treats the JTPA as a major effort to enhance the labor market productivity of program-eligible individuals by increasing their human capital. For Title IIA, this is to be achieved by a variety of means as set forth in Sec. 204, but all of them have the intended effect of either directly or indirectly increasing the human capital of program participants and thereby maintaining, improving or restoring their capacity to function effectively in the labor market.

It must be recognized, however, that the benefits that are intended to flow from the program services or treatments described in Sec. 204 are achieved only by the well-considered and effective expenditure of society's resources. In short, for every benefit, there is a cost. And, in general terms, for the cost outlay of the treatment to have value and purpose for society, benefits must equal or exceed costs.

Benefit-Cost Analysis versus Effectiveness-Cost Analysis.

The primary benefits of the JTPA are of two kinds: monetary and non-monetary. The monetary benefits, such as increased before-tax earnings or reductions in welfare payments due to successful integration into the labor market, are compared against monetary costs in a benefit-cost framework. Non-monetary benefits, such as increased self-esteem measured by some psychological scale or decreased welfare dependency measured either as a percent, a probability or total weeks on welfare, for instance, are compared against monetary costs in an effectiveness-cost analysis.

The two approaches are conceptually equivalent and the technical approaches required for a valid estimate of either are in principle the

same. Neither is, in principle, technically superior to the other; neither is necessarily more practical or easier to implement. As a final observation we should stress that data limitations, such as an absence of crucial cost or benefit data, do not, a priori, cause one method to dominate the other. Each, for instance, ideally requires money cost estimates since there is no other common denominator of costs. Using the benefit-cost or the effectiveness-cost ratio, the measures of net human capital investment return take the following general form:[1]

1. Benefit-Cost:

   $$\frac{\text{Present Value of Total Money Benefits}}{\text{Present Value of Total Money Costs}} = \text{Total Benefit-Cost Ratio}$$

2. Effectiveness-Cost:

   $$\frac{\text{Any Effectiveness Measure Expressed as a Total}}{\text{Present Value of Total Money Costs}} = \text{Total Effectiveness-Cost Ratio}$$

It is mostly a matter of policy focus or need that will determine which of the two approaches to use. For instance, the JTPA speaks specifically of "return on investment." To estimate a return on investment requires that one estimate a benefit-cost ratio, an internal rate of return on investment or the net present value of an investment--the sum of discounted benefits minus the sum of discounted costs. However, JTPA also has as a goal the reduction of the number of individuals who are dependent on cash welfare payments. An effectiveness-cost ratio would be employed in this case to measure for example, the decreased probability of receiving any cash welfare payment during the year after end of program treatment per $1,000 of expenditure of JTPA funds for, say, classroom training.

However, we recommend that benefit-cost analysis always be attempted. Then, one can expand the analysis, time and resources permitting, with effectiveness-cost analysis. The overall summary measure of program impact--the increase in total before-tax earnings--gives policy precedence to the use of the benefit-cost approach.

THE BENEFIT-COST EVALUATION FRAMEWORK

As noted above, benefit-cost analysis is a technique for evaluating social investments and judging the relative desirability of competing social policies. In its typical application, the technique would be used by a government agency to judge a program's efficiency or effectiveness to help decide whether the program should be expanded or reduced in size and scope, redirected toward goals it might achieve more easily, or discontinued altogether. As a management technique, benefit-cost analysis is not so much a fixed course of action as a broad evaluative approach to addressing complex questions of resource

---

[1] See Exhibit 7 for a precise mathematical definition of these concepts.

allocation, including the understanding of the structure and size of benefits and costs. Properly executed, such analysis can aid significantly in program management. There are certain general procedures that are common to all benefit-cost evaluations.

The first step in an evaluation is to define the set of program benefits and attach a dollar value or some non-monetary value to them. The benefits will, of course, be program-specific, and frequently will be identified on the basis of a program's legislatively established purposes and objectives. In the case of government-sponsored employment and training programs, the measures specified in the JTPA as program outcomes are usually specified as a program's main benefits, but other economic and non-economic benefits can also be considered as part of a more comprehensive evaluation. As noted further below, a major concern here is to avoid double-counting. For instance, hours worked are a component of earnings (hourly wage rate x total hours worked = total before-tax earnings), so in a program evaluation one cannot claim that an increase in hours worked and an increase in earnings can be simultaneously counted as benefits.

The second step is to identify and estimate program costs. For an employment and training program, the most obvious costs are the direct costs of operating and administering the program. Most of these are incurred at the SDA level or among its subcontractors. Some costs, mainly administrative, will occur at the state level. In addition, a comprehensive calculation of costs should also include the costs that an individual incurs when he or she participates in the program. These include the direct costs of participating in a program treatment such as the purchase by the participant of books, tools or special clothing. A potentially large component of participant costs are the earnings the person may have forgone by entering the program rather than remaining in the labor market and working at some job.

The third step is to recognize that the costs and benefits to investment in human resources occur in different time periods; it is therefore necessary to adjust the dollar values of each into present terms using an appropriate discount rate; compare their total values; and determine the social efficiency of the program, based on the difference in the total dollar value of these costs and benefits. If the present value of money benefits exceeds that of costs, then the program would be judged worthwhile from the perspective of economic efficiency; in the opposite case, the program would be judged inefficient, if all costs and benefits that exist have been accounted for and they are all measurable in money terms.[2]

Discounting. As noted above, the nature of investments in human capital is that their costs and benefits occur over time. Before-tax earnings benefits to adult female participants in the original Manpower Development and Training Act of 1962 were shown to be large and positive for more than five years after the participants left the

---

[2] Clearly, as already noted, costs and benefits occur in both money and non-money components as specified in the JTPA. The above money benefit-cost approach, taken alone, therefore is subject to error.

program (Ashenfelter, 1978). Similar post-program benefit patterns exist for adult women who participated in the Comprehensive Employment and Training Act. The dispersal of costs and benefits over time requires that they be made commensurable in terms of their present value, since future benefits are worth less than benefits one has in the present. Likewise, costs that occur in the future have a smaller present value. A simple analogy will show why this is so.

Let an SDA have a sum of $1,000 today. If the SDA spends it all on training services today, it spends exactly $1,000. The SDA, however, could postpone this expenditure and put the funds in a bank. This bank will offer to pay interest on the deposit of, say, ten percent per year. If the SDA leaves the funds in the bank for a year, in one year it will have $1,100. The same $1,000 now is worth $1,100 a year from now. Thus, a dollar now is worth more than a dollar in the future, because present dollars, via interest payments, grow into a larger future sum. Thus, present and future dollars cannot be directly compared unless the future dollars are made to reflect the same value as the present dollars. This is done by discounting a future stream of dollars (whether costs or benefits) to their present value.

The basic equation for estimating a present value is

$$PV \text{ (present value)} = \sum_{t=0}^{T} \frac{B_t}{(1+r)^t}$$

where

B is the benefit (cost) in <u>any</u> time period; say a year;
r is the interest rate used to discount;
t is <u>any</u> year; and,
T is the total time period, e.g. 13 years.

Note that there are three dimensions to the discounting process:

1. What are the benefits or costs in each time period;
2. What is the interest rate to be used in discounting; and,
3. What is the time period over which benefits and costs occur?

Benefits and costs can be directly estimated for those time periods for which complete data exist. If it is assumed that some benefits or costs will occur for time periods beyond that for which there are data, then assumptions have to be made about when these benefits will occur and how large they will be in any time period. Informed judgment is the best guide in this exercise since there is almost no information on the time profile of benefits and costs to subsidized employment and training programs beyond six years after a participant exits such a program.

Choice of an interest rate to be used in discounting is not as easy as it would seem at first blush. There are competing theories as to how to select the correct interest rate. It is not necessary to discuss these theories here except to note that the rate will differ for society as a whole compared to, say, a state government or an SDA. The rate for society as a whole is net of any effect due to inflation, while this is not the case for a local government or agency. The most

practical guidance on the choice of interest rate is to select that rate at which the local government or agency can lend its funds. The accounting office of a local government or agency can provide that figure. Currently, lending rates, are in the neighborhood of ten percent, which is a relatively high number.

The time period over which benefits and costs can occur cannot exceed the average remaining life expectancy of the trained participants. Historically, the average age of trainees has been in the low 30s. This implies a 30 year period or so for which benefits (costs) can occur. This is a very long period. Normally, we would expect the value or productivity of training to depreciate over time, just like the value of any other investment. And, of course, the value of job placement services could depreciate rapidly, depending on the nature of the placement process.

The best advice to offer in these cases it to conduct sensitivity analyses to see how sensitive the estimate of net program benefits are to the size and duration of future costs and benefits and the interest rate used in discounting. In particular, the higher the interest rate used, the less important the other assumptions become. For instance, at a ten percent interest rate a dollar received 25 years from now is worth only 9.2 cents. At a four percent interest rate, however, this discounted dollar is worth 37.5 cents.

Combining Monetary and Non-monetary Benefits and Costs. Other issues arise in the measurement and valuation of benefits and costs. In evaluating a training program, some benefits, such as increased before-tax earnings, result from greater work productivity that can be both measured and assigned a dollar value. However, other benefits, such as psychic benefits cannot easily be assigned a dollar value, if at all. The problem for the analyst is how to include these nonquantifiable and nonmeasurable benefits in the analysis, since, without them, the analysis is incomplete and potentially misleading.

The practice employed to address this problem and to make the analysis as comprehensive as possible is to specify in dollar terms all benefits and costs measurable in dollars; then one should qualify the results with a discussion of measured non-economic outcomes (such as educational performance scores) and nonmeasurable outcomes (such as improved work habits) that would affect the overall benefit-cost calculation. After these nonquantifiable and nonmeasurable effects have been identified, the program's effectiveness would be assessed at a more implicit level using this broader concept of benefits and costs. As with the narrower analysis, if the estimated difference between discounted benefits and costs under this broader concept exceeds zero, the program treatment would be judged to be efficient; if less than zero, it would be judged inefficient or ineffective.

The Focus and Perspectives of Benefit-Cost Analysis. As indicated above, the purpose of a benefit-cost analysis is to judge the social efficiency of a program. It asks the question: "Is the value of the goods and services and other outcomes available to society greater as a result of the program, or would the value have been greater had the resources been put to alternative uses?" As an aggregative measure,

the full societal perspective does not take into account the income distributional effects of a program. To at least partially account for differences in a program's income distributional consequences, a benefit-cost analysis must therefore estimate program effects from several different perspectives.

The modern theory of benefit-cost analysis distinguishes three distinct perspectives for measuring benefits and costs. The first class of benefits and costs consists of those that are received by or borne by program participants. A second class of benefits and costs consists of those received or borne by non-participants, sometimes referred to as "taxpayer" benefits and costs.[3] For an education program, as an example, this class of benefits includes the higher taxes that participants in a program would pay because of increased earnings that result from successful participation in the program. The higher taxes paid by successful program graduates could be used to reduce the taxes of non-participating taxpayers (a direct benefit to them), or more social services could be purchased for everyone (a benefit to non-participants and participants). Costs to taxpayers would include such things as taxes paid to fund the direct cost of operating the program; while certain indirect costs such as the loss of participants' earnings while they are in the program would be borne by the participants and would also represent a social cost. Finally, the most inclusive set of program benefits and costs are those accruing to society at large. This last class is simply the sum of benefits or costs received or borne by participants and non-participants, taken separately. It represents the broadest perspective on the overall performance, or social efficiency, of a program, and it is the perspective taken in the JTPA Act.

A benefit-cost model that enumerates benefits and costs from the perspective of participants, non-participants and society is both sound conceptually and appropriate as an analytic tool. As noted, the major value of the approach is that it formalizes a conceptually sound taxonomy of benefits and costs and it can be used to distinguish broadly between inter-group transfers of benefits and benefits that change (increase) the total goods and services available to society. The transfer of income between program participants and non-participant taxpayers does not represent an increase in resources available to society at large, since the increased benefit to one group is cancelled by the increased cost to the other. For example, if the JTPA treatment reduces the need for AFDC grants payments, this would represent a benefit to taxpayers (who no longer are required to pay as much taxes to support public assistance programs) that is exactly offset by a cost to program participants (who no longer receive as much public assistance). On the other hand, benefits (costs) that accrue to one group that are not offset by corresponding costs (benefits) to the other group, for example, increased output from program participants, represent a real increase in benefits (costs) to society--potentially all members of society.

---

3 Obviously, not all non-participants are taxpayers, and, indeed some participants are also taxpayers.

IDENTIFICATION AND MEASUREMENT OF BENEFITS AND COSTS

The taxonomy of benefits and costs from the perspective of society, the program participant and the non-participant (taxpayer) is outlined in Exhibits 1 and 3. Benefits are discussed first, followed by a discussion of costs. Key measurement issues are discussed when appropriate.

A GENERAL TAXONOMY OF BENEFITS

Benefits to a government subsidized program of employment and training such as the JTPA can occur during and after the program treatment. As with costs, there are benefits to the participants and the non-participants, the sum of which equals social benefits--a full economic accounting of benefits. In addition, within social benefits, a subset of benefits accrues to employers. These are discussed in turn. Exhibit 1 sets them forth schematically.

Increased Output during Program Participation. These before-tax benefits can occur for two reasons. First, if a participant is in an on-the-job (OJT) training program, the value of the output he or she creates during that training is a social benefit. In general, the value of output is measured by the earnings paid to the worker, net of any program subsidy due to the OJT program treatment. The value of the output is defined as the net value added created by the worker. This value added due to labor is the reason why a worker is hired and is the basis upon which the worker is paid. Ideally, the worker receives in earnings the full share of value added created by him or her. Note also that this value added is a benefit to the employer for which the employer is willing to make a payment in return for its creation by the worker. As a practical matter it is very hard to identify that portion of an OJT worker's total wage that is a subsidy to pay for training and that portion which is compensation for the value added created by the worker during the OJT process; a worker will create some valuable output in the process of being trained. No one to our knowledge, has ever successfully unraveled this problem for government subsidized OJT projects.

Second, in a classroom training situation, such as an auto repair class, output is created that can be sold or, if not sold, its value can be imputed. While this may not be a large quantity, it is worth mentioning in order to maintain completeness in the accounting of benefits.

Increased Output after Program Participation. This is by far the largest measurable component to post-program benefits and the one with which we are most familiar. However, there is considerable confusion as to how it is properly measured. In particular, as discussed below, there is a tendency to double-count increases in post-program output. The manner in which we have broken down the components of post-program output is designed to avoid such double-counting.

First, note that the best summary measure of net post-program output is the increase in before-tax earnings--item 2d in Exhibit 1. The measurement of this quantity is central to the analysis of program

net impacts, as discussed in Volume V and VI in this series. Total before-tax earnings is the best summary measure because,

1. If the labor force participation rate increases, other things equal, total earnings will increase;

2. If total hours worked increase (unemployment decreases; time to find a new job decreases), other things equal, total earnings will increase; and,

3. If the hourly wage rate increases (or any other related measure such as monthly salary increases), then total earnings will increase.

In summary, any increase in labor force participation; any reduction in unemployment (increase in hours worked) from whatever source such as reduced time to find a job or reduced job turnover; and, any increase in the hourly wage rate (or its equivalent) will increase total before-tax earnings. While it is sometimes important for policy purposes to decompose the component causes of any earnings increase, it is not absolutely necessary to do so, and often, due to data problems or cost and time constraints, it cannot be done. Thus, total before-tax earnings is the most versatile as well as the key summary measure of increase in post-program output.

Since governments, agencies and taxpayers have an interest in the measure, we have included as a benefit increased income and excise taxes. These tax benefits to non-participants are the result of increased total before-tax earnings and increased expenditures caused by the increase in total after-tax earnings, respectively. Again, these increased taxes are benefits to non-participants and costs to participants. For society as a whole, these two sums net to zero. Society is only interested in the total increase in output. Members of society--both participants and non-participants--can then agree to divide up that increase in output any way they see fit; that is, they can agree to tax each other.

Note that it will very seldom be the case that net tax gains will equal or exceed program costs. Indeed, the correct measure for the return on such an investment is fundamentally the total before-tax increase in output (plus other measures of well-being) relative to total social costs-- all resources used (cost incurred) to operate the program. There is often a temptation to relate social costs to non-participant (taxpayer) benefits. This is an error since it is akin to contrasting apples with oranges. As can be seen from Exhibits 1 and 3, the social and non-participant concepts are defined differently. They cannot be directly compared. Indeed, non-participant benefits and costs are a part of social benefits and costs.

Reduction in Welfare Dependency. There can be a reduction in welfare payments either because one has a lower propensity to qualify for welfare, benefit level held constant, or the benefit level could drop, propensity to use welfare being constant. Changes in total welfare payments represent benefits and costs to participants and non-participants but they net to zero for society since they are transfer payments. The true gain to society comes from the increase in

Exhibit 1

A GENERAL TAXONOMY OF THE BENEFITS OF THE JOB TRAINING PARTNERSHIP ACT: AN ACCOUNTING PERSPECTIVE

	Social[a]	Participant	Non-Participant
1. Increased output during program participation--measured by before-tax earnings paid to the participant, which is the value added by the worker.	+	+	o
2. Increased output after program participation.[b]			
a. Increased probability of being a member of the labor force	+	+	o
b. Increased hours (weeks, months) worked, given one's labor force participation	+	+	o
c. Increased before-tax wage rate per hour worked	+	+	o
d. Increased before-tax earnings (the product of 2b and 2c above)	+	+	o
e. Increased income and excise taxes paid as a result of increased before-tax earnings and consumption from increased after-tax earnings	o	-	+
3. Reduction in welfare dependency			
a. Reduced probability of being on welfare	+	+	+
b. Reduced payments, given one's probability of being on welfare	o	-	+
4. Other benefits			
a. Psychic benefits of being a more productive member of society, e.g. increased self-esteem	+	+	o
b. Increased or improved physical and mental health, including such things as reduced drug abuse, reduced suicide, reduced spouse or child abuse	+	+	o
c. Improved work attitudes and work discipline.	+	+	o
d. Reduced crime against property and persons outside of the family group			
i. property	o	-	+
ii. persons	+	o	+

	Social[a]	Participant	Non-Participant
5. Employer benefits			
a. Increase in value added.[b]	+	+	+
b. Employer benefits not included in increased output after program participation.[b]			
i. Reduced job search costs; this will be due in part to more efficient labor market institutions and information brought about by the JTPA and reduced job turnover due to better worker/training/job matching			
-- A quicker job match, with the average quality of the job match unchanged	+	+	o
-- An improved match, with the average time for a job match held constant	+	+	o
ii. Reduced training costs due to reduced job turnover. (The less often a person switches jobs, the less likely the person will have to retrain, either formally in class or informally on-the-job.)	+	+	o

Notes: [a] The sum of participant and non-participant benefits equals social benefits. Social benefits represent the total increase in economic and non-economic benefits.

[b] Note that the increased total earnings of the participant represent the value of that participant to the employer. This value added is the value of the extra output created by the worker and for which the employer hires and pays him or her.

total output. This increase in total output is what enables one to depend less on welfare or pay others more welfare. Thus, changes in welfare payments are neither social costs nor social benefits.

However, there is an intrinsic value to society, the participant and the non-participant to a reduction in the dependence of society's members on welfare, as long as no one is made worse off as a result. Thus, reducing the degree of welfare dependency, other things equal, is a social benefit.

Other Benefits

Psychic benefits, while not directly measurable, exist in fact and must be noted. They are reflected in such concepts as increased self-esteem. Psychometric scales can be used to provide an index measure of this self-esteem, if desired. Note that total earnings will increase in part due to increased self-esteem, so there is a possibility of double-counting when considering any program benefit that has psychological or physical well being dimensions. Yet, it is clear that there are gains to individuals and to society from the increase in self-esteem alone. Increased self-esteem is valued in its own right.

Improved Physical and Mental Health. It is well known that business cycles induce unemployment and economic distress and that these factors in turn lead to reduced physical and mental health, including increased suicide. Reversing this process can improve mental an physical health. The gains are directly internalized in the person and there is a reduction in social resources required to treat these illnesses induced by economic adversity. Again, the increase in total earnings will pick up some of the improvement in mental and physical health--absenteeism will be less, for instance, and productivity while on the job will be higher. But, as with self-esteem, these benefits to the quality of life add to well-being independent of their economic contribution.

Improved Work Attitudes and Work Discipline. A properly designed and executed employment or training program can change behavioral habits as well as add to human capital. This positive change in habits has intrinsic personal and social value in itself. In addition, of course, such positive behavioral change will also be reflected in increased total before-tax earnings. Getting to work on time, following orders, reducing absenteeism--are all reflected in higher earnings.

Reduced Crime. Ironically, the reduced gains from crime is a cost to the individual participant. There are two aspects to reduced crime. First, crime against property represents an illegal transfer payment--a fancy term for theft. Thus, property crime--independent of losses due to damage--nets to zero for society. However, the legal system and all other avoidance costs represents a drain on society's resources. If crime drops, the use of these resources drops, which is a social benefit.

Crime costs are very hard to directly measure. The best efforts to date have been in the Mathematica Policy Research Inc. (MPR) study of

the Job Corps (Mallar, 1982). For youth, a reduction in crime may be a major component of social gain. The reduction in crime appeared to be about one-half of the total gain to the Job Corps Program. But it may not be a significant factor for prime-age males and females. In any case, this program outcome is extremely costly to measure with any degree of statistical accuracy and is almost never measured on basic program MIS statistics. Thus, states and SDAs may be precluded from making such estimates.

Employer Benefits

The nature of employer benefits is not well understood. The employer gains from a program like JTPA in several ways. These are:

1. A reduction in job search costs both for the initial hire and because of (ideally) reduced job turnover due to more efficient matching of worker, training, job and firm.

2. A reduction in job training costs, both initially and due to reduced job turnover as a result of more efficient matching.

A word must be said here concerning specific and general training. The employer gains from reduced job training costs if the training is firm-specific--the skill for which a participant is trained literally can be used in no other firm except the firm in question. Such training, when provided by JTPA, is a subsidy to the firm, not the worker. Such training, if not subsidized, is a cost to the firm but not to the worker. In principle, the worker will not contribute to the cost of firm-specific training because the worker will get none of the benefits from it. The firm captures all of the benefits to firm-specific training. Historically, some government subsidized employment and training projects have subsidized firm-specific training.

In contrast, if the training is general in nature--the worker gets the benefits--no specific firm will pay for it, because the firm cannot recover its investment costs. This type of training, when provided for by the JTPA, is a subsidy to the worker, not the firm. Most, but not all JTPA training, is general in nature. OJT programs reflect the existence of general training when a worker is paid lower wages during the OJT period. These lower wages are reflections of the cost of training. The worker pays the cost through lower wages. The firm just happens to be the locus of training. When the training ends, wages rise to those of the typical worker in that occupation.

As one might expect, it is often hard to measure whether there is any specific training occurring, since it can occur alongside general training. But, it does exist and any reduction in the cost is a benefit to the firm.

Value Added is the third major benefit to the firm; however, this is the value added that is contributed by the worker to the firm and is the worker's share of the total output created by the firm when it combines land, labor and capital in production. The wage rate is the direct measure of this value added. Total earnings is the total

increase in value added by a worker over a given time period. Of course, this value added, in the form of total before-tax earnings, is also a measure of social benefit. Therefore, while the reduction in job search costs and specific training costs to the firm can be added directly to total social benefit, value added cannot. To do so would be double-counting. Nevertheless, it is important to understand the concept of value added. It is the reason why firms hire workers in the first place. It is what firms pay wages for and it is the reason why such wages can be paid at all.

Double-Counting. The problem of double-counting needs mentioning because it is a common problem area and source of error in benefit-cost analysis. This is a particular problem when effectiveness-cost ratios are used or when they are combined with a benefit-cost measure.

As noted in the beginning of this chapter, there is no fundamental conceptual difference or methodological advantage between benefit-cost analysis and effectiveness-cost analysis. Any given project or program treatment can be jointly evaluated by both approaches. However, as Exhibit 2 shows, many of the most important effectiveness measures are a component of the fundamental program outcome of total before-tax earnings over some period of time, usually a year.

Consider the following: The JTPA is designed to improve the economic condition of the legally-defined target populations. These people are generally poor, either due to temporary shocks to their income (e.g., displaced homemakers) or earnings capacity or due to more permanent problems such as lack of English speaking ability. To relieve this relative poverty, the main problem is to increase their income, the major component of which is earnings. To increase one's earnings, the following sequence of events must occur as a result of a JTPA treatment:

- One must enter the labor market. That is, there must be labor force participation, or nothing else will happen.

- Having entered the labor market, one must become employed. One has to get a job.

- Having become employed, one must work more hours per week, holding weeks worked constant, or one must work more weeks per year, holding hours constant, or both hours and weeks worked can increase.

- One's wage rate must be positive. Ideally, one's wage rate will increase over what it would have been had one not entered the JTPA program. This is a clear-cut measure of program success, but not the only measure.

- Finally, earnings is the product of hourly wage rate times total hours worked.

Clearly, an inspection of Exhibit 2 and a review of the other types of benefits listed in Borus and Tash (1970) show that many, if not most, of the benefits listed are contained in and are a total or partial result of improved earnings. Thus, improved housing or health can occur if earnings increase so one can afford to purchase more of

Exhibit 2

EXAMPLES OF PARTIAL OR TOTAL DOUBLE-COUNTING IN THE MEASURE OF BENEFITS TO EMPLOYMENT AND TRAINING PROGRAMS

Type of Benefit	Contained in a Measure of Total Before-Tax Earnings in Whole or in Part?
• Reduced unemployment	Yes
• Increased employment	Yes
• Increased productivity of the labor force--the ability to earn higher hourly wage rates	Yes
• Improved equity in the distribution of:	
- Income	Yes
- Employment	Yes
• Increased labor force participation	Yes
• Increased self-esteem	Yes
• Increased job satisfaction	Yes
• Increased work week for those desiring more work	Yes
• Reduced unnecessary turnover	Yes
• Increased mobility of the labor force	Yes
• Reduced dependency on government	Yes
• Reduced time to find a job	Yes
• Improved physical health	Yes
• Improved housing	Yes
• Improved skills and education	Yes

Adapted from Borus and Tash (1970). Table 3.

Exhibit 3

A GENERAL TAXONOMY OF THE COSTS OF THE JOB TRAINING PARTNERSHIP ACT: AN ACCOUNTING PERSPECTIVE

		Social[a]	Participant	Non-Participant
1.	Program operating cost			
	a. Overhead and start-up costs (fixed costs)	-	0	-
	b. Variable costs of program operation	-	0	-
2.	Administrative cost above the project level			
	a. Overhead costs	-	0	-
	b. Variable costs	-	0	-
3.	Participant opportunity costs			
	a. Foregone earnings plus fringe payments	-	-	0
	b. Foregone non-market activities	-	-	0
4.	Increased class attendance costs			
	a. Transportation	-	-	-
	b. Child care	-	-	-
	c. Other (except psychic costs)	-	-	-
5.	Psychic costs of participating in training programs	-	-	0
6.	Income maintenance transfers (stipends)	0	+	-

Note: [a] The sum of participant and non-participant costs equal social costs. Social costs represent full economic costs.

these goods and services. As noted above, better mental health or self-esteem is a result of getting any job or getting a better job. In short, care must be taken to understand the logical relationships between different components of benefits so that double-counting does not occur; or, if it occurs, it is clearly identified by the analyst.

A GENERAL TAXONOMY OF COSTS

There are six different cost components included in a benefit-cost model of an employment or training program. These cost categories include: 1) program operating costs; 2) administrative costs at and above the SDA level; 3) opportunity costs to program participants--noted as indirect costs below; 4) other costs associated with a participant's attendance in the program treatment; 5) psychic costs to participants of undertaking the program treatment; and 6) program stipends or their income transfer payments, such as UI benefits received during training.

Operating and Administrative Costs. The cost of operating employment and training programs includes both direct operating costs such as salaries for instructors, materials and staff development costs and the cost of administering the program. None of these costs are counted as costs to program participants, as program participation is free. Both operating and administrative costs do, however, represent costs to non-participants and society. In Exhibit 3 these costs are represented as a zero to program participants and as a negative in the other two columns. (The negative sign is used to denote the presence of a cost item.)

Opportunity Costs to Participants. Persons participating in government-sponsored employment and training programs may forego earnings opportunities they may have had otherwise had they not participated. The wages they would have earned while participating in the program are a cost to them of participating in the program. This opportunity cost to participants is not balanced by corresponding benefits to non-participants, and thus enters the social benefit-cost calculation as a cost to society, as well as to program participants.

Another way to view this cost is that, from society's point of view, a person's decision to enter a subsidized employment training program may reduce the output the person might have produced if the person were employed at that time. This foregone output is a net cost to society; its value is measured as foregone earnings. In Exhibit 3, these losses are characterized by minuses for participants and society and a zero for non-participants, who are unaffected by these opportunity costs.

Since a comparison group exists in the data set to be used to measure post-program effects on total earnings, it is possible to estimate foregone earnings of participants while they are members of the JTPA program. This can be done by identifying the modal time period over which training (or any other program treatment) occurs. Indeed, you would ideally separate program participants by their type of program treatment in order to perform this analysis. Next, since there is a comparison group available, the before-tax earnings of this comparison

group can be estimated during this modal time period. These earnings must then be compared to any before-tax earnings that the program participant group under study may also have earned during the modal time period of program participation. The difference between the two estimates will represent the foregone before-tax earnings that should be counted as a social cost. After netting out any income taxes and social security taxes, the resulting difference in after-tax earnings should be counted as a cost to the average participant in the treatment cohort being studied.

Increased Program Treatment Costs. Participants in employment and training programs may incur costs that are associated with attending classes or receiving other treatments such as job search services. These include such costs as transportation, child care, the extra value of meals bought away from home compared to home meals, any special clothing and equipment and expenses for materials and supplies if these are not supplied free to the participant. Some of these costs, such as the incremental costs of meals purchased away from home, are difficult to measure and may represent only a small share of the total costs of the program treatment to the individual and society at large. As a practical matter due to evaluation cost and time constraints, one would confine the measurement effort to the largest and most easily measured cost items--usually those where a cash outlay is involved and, hence, where they are easily identifiable. In general, these costs are those reported on JTPA Annual Status Report and its underlying documentation.

Psychic Costs of Participating in Training or Job Search Programs. Anyone who has ever sweated out an examination or attempted to puzzle through a mathematical problem or the translation of a complex sentence in a foreign language is familiar with the psychic costs of learning. So, too, is anyone who has had to search for a job. These types of costs accrue to program participants and to society but not to non-participants. They are inherently unmeasurable but are included here for the purpose of completeness. It would be tempting to argue that these psychic costs are balanced at the margin by psychic benefits that derive from a sense of accomplishment. But this is an empirical question and we have no way to measure directly these psychic benefits and psychic costs. The availability of psychometric scales qualifies this statement--changes in self-esteem can be indirectly measured. But it is difficult to conceive of a psychometric scale that would measure the pain-cost and stress of training or job search and be commensurable with any index measure of psychological benefit due to the program. Hence, these costs and benefits are noted as important, but not measured.

Stipends and Other Income Transfer Payments. The possibility of paying stipends to individuals who are participating in the JTPA still exists. Likewise, the Act as well as state laws allow the payments of UI benefits to UI eligible individuals, when properly certified, during participation in the JTPA. These payments represent costs to the non-participant and benefits to the participant. Since they are transfer payments, they net out to a zero cost for society; the cost to the non-participant is cancelled by the benefit to the participant. There is no net additional drain on society's resources as a result of payment of these stipends to the program participant.

THE PRACTICAL PROBLEM OF COST ESTIMATION

The previous section identifies costs based on the individual or group that bears the cost. In addition to using accounting concepts that accurately measure economic cost, these costs should be analyzed with appropriate statistical techniques. Although most of the data currently exist to perform such analysis, economic and econometric cost analysis of current government employment and training programs based on a consistent set of economic assumptions is infrequent. Often, this is because the available data are not collected or published appropriately. For instance, there may be inappropriate aggregation of cost categories; it may not be possible to measure the separate costs of two programs or projects because their costs have been added together at some stage and the original records on each project then made inaccessible. Or, a critical cost or explanatory variable may have been omitted in the data gathering process. Practical problems of this nature in the currently available data are discussed below.

Other problems exist. For instance, the problems imposed on cost analysis due to price imputation of program inputs impart considerable opportunity for measurement ambiguity. Reasonable people can disagree over the appropriate evaluation of non-market inputs such as the capital cost of a public school used in a training program. The appropriate economic methods for price imputation, while clear in general terms, are difficult to apply in practice.[4] In general, it is probably best to avoid this price imputation since it can be very arbitrary indeed.

Also, two or more program outputs are sometimes jointly produced from a single input. This creates a problem when one attempts to measure average or total cost for each of the joint outputs since there is no non-arbitrary way in economic theory to split the joint cost between the two outputs. This has no effect on measuring marginal costs, however.[5] Hence, it is an advantage to estimate total cost functions as we discuss them below because the joint cost component in total cost is handled automatically in the econometric estimation as long as the jointly produced outputs are aggregated. An example of two jointly produced outputs would be the operation of two training programs within a common facility and being managed by a common manager. Some part of the facility costs and managerial costs simultaneously contribute to the operation of each training program.

Of course, if a SDA (or a subcontractor) has a considerable proportion of joint inputs in its program, state reimbursement in cost categories where joint inputs occur will result in transfer payments or subsidies to the SDA since the marginal or additional cost to a jointly produced output is zero. In effect, the SDA may receive payments for which

---

4 See, for instance, Roland N. McKean. "The Use of Shadow Prices," in Samuel B. Chase, Jr. editor. Problems in Public Expenditures Analysis. Washington, D.C.: The Brookings Institution. 1968.

5 The relationship between total, average and marginal costs is discussed below.

little or no reciprocal services have been rendered. There is nothing wrong with subsidies, of course, but sound policy decisions should separate what is a subsidy--an income redistribution--from a true economic cost, wherein additional resources are consumed in the production process.

This leads us directly to consideration of costs versus transfer payments.[6] From a policy standpoint, two benefits will follow from this distinction between cost and transfer payment. First, in future contract negotiations involving employment and training programs, program administrators at the state level will be able to evaluate true economic costs more accurately. Second, previously unidentified subsidies in currently operating programs may be identified. Since no necessary service may be rendered in return for these subsidies, they can be reduced or shifted among programs, if so desired. Unlike a cost reimbursement for the actual use of resources, the elimination of a subsidy or transfer payment need not necessarily negatively affect the objectives or desired output of the program.[7]

AN INTRODUCTION TO ECONOMIC COST THEORY

A major aspect of this paper is to present consistent estimates of the average and marginal costs of employment and training programs based on Federal government expenditures and for Federal government plus any state, SDA or other agency or subcontractor expenditures.

In addition, this paper will indicate how to measure marginal cost, the extra cost of enrolling an additional participant into a program. Marginal costs are extremely important in benefit-cost analysis because, when compared to the additional or marginal benefit received from a JTPA project, they indicate the relative productivity of the project. The decision rule facing a program manager is to continue to devote resources to a particular project or activity as long as the additional or marginal benefit exceeds the additional or marginal cost. When the two are equal, total benefits from the project or activity are at a maximum and cannot be increased by additional investment in that specific project or activity.

Cost Interrelationships. Economists employ various measures to show the relation of cost to output; among these are total, average, and marginal cost. (See Appendix A to this chapter for a graphical

---

6 As indicated below, a transfer to an individual is a payment for which no equal, reciprocal service or benefit is offered. It is a gift, in effect. A transfer payment is sometimes called a subsidy.

7 Consider the example of a surplus public school that has been converted to JTPA training facility. If the surplus school had no other use in the community and could not be sold to someone in the private sector, then its capital cost or rental cost to the JTPA program is zero; it has no alternative use. It is the existence of an alternative use--a foregone benefit--that defines cost. Any additional renovation, conversion, maintenance or operating costs should, of course, be attributed to the program.

depiction of these costs.) Total cost is the total outlay for a given output; average cost is total cost divided by total output; marginal cost is the addition to total cost due to the production of one more unit of output.

The Measurement of Output. Cost measures are expressed in terms of some measure of output. The output variable in this study is usually some variant of the number of persons trained in a productive activity such as work, school or the military, over a given time period; or, it can be the number of persons subject to placement in a job. Briefly, the problem is this: The number of persons trained or subject to placement is only a proxy for the true output measure of an employment and training program. This true output measure is the increase in the human capital of a worker due to the training or placement process. However, as a practical matter, this increase in human capital cannot be measured directly in a physical sense but only indirectly in a value sense--by how much a person's wage rate rises or earnings increase. A rise in wage rate can be due to an increase in a person's output due to a rise in his or her human capital (assuming the price of the output is constant) or the price of the output he or she produces (assuming that output is constant), or both. Thus, the physical measure of output--one's increase in human capital-- becomes confused with the market value of that output. No one has yet successfully solved this problem of identifying the actual change in embodied human capital. Thus, we rely on a proxy, the number of persons trained at some average level of skill or educational proficiency.[8] Such a proxy measure may overstate or understate the true measure of human capital increase.

Specification of Cost Functions. Costs of education and training programs are analyzed using two cost functions: a total cost function and an average cost function. The total cost function permits inferences to be made about marginal cost, i.e., the extra or additional cost of training one more trainee or placing one more participant. Marginal cost is derived by computing the change in total cost as the result of a unit change in the number of persons trained (placed). As noted above, marginal cost estimates are useful in answering questions on whether to expand a given program or project from its current size and in determining the efficient operating level of the program or project. Also, as noted above, using marginal costs avoids the problem of having to allocate joint costs among different types of related outputs, e.g., to different training classes in the same building, assuming classroom training of all types is aggregated into a single measure of output.

---

8 Numerous studies, mainly of educational production functions, use measures of performance on standardized tests as proxies for the increase in human capital. This can also be done in JTPA programs. However, the exact nature of the relationships between these proxies and the change or level of human capital is not well understood. See Samuel Bowles, "Toward an Educational Production Function". in W. Lee Hansen. Editor. Education, Income and Human Capital. Studies in Income and Wealth. No. 35. New York, NY: Columbia University Press.

The average cost function can also be used to obtain information about the optimum scale of operation for a training program. The average cost function, if it has a U-shape as in Appendix Exhibit 4, permits inferences to be made about the output level that will result in a minimum average cost per person trained. One can identify the least-cost point of output. The average cost function is derived by dividing the total cost function by total persons trained--total output. We do not recommend any attempt to directly estimate the average cost function econometrically.

EXAMPLES OF COST ESTIMATION

This section describes the available data on total costs, total output and certain variables that may affect the nature of total costs as total output is produced. It also presents both hand-fitted and econometrically-fitted total cost functions and discusses their interpretation and limitations.

An Empirical Total Cost Function. As Appendix A sets forth in some detail, the total cost function expresses the functional relationship between total cost (TC) and total output (X). The JTPA Annual Status Report is the source of data for estimating the total cost function. These data are limited, but can be used successfully. At this time, possible complex relationships between total cost and total output cannot be investigated because of sample size limitations. Looking ahead, in this case, the loss of information is not great because limited experimentation with the data suggest that for the set of SDAs in Washington State, the total cost function is linear and marginal costs are therefore constant.

We therefore specify a linear total cost function to analyze statewide SDA cost behavior:

$$TC = a + bX + u$$

Where

TC = total federal costs in $100 per quarter;
X = total output, expressed either as total terminees or total terminee/weeks
a = the regression coefficient that estimates fixed costs;
b = the regression coefficient that estimates marginal costs; and,
u = the error term.

The unit of observation, because of the few number of observations available, is the SDA. Cost analysis is therefore done at the state level. Analysis could not be performed for a single SDA because there were too few quarters of data available on which to perform analysis and data were not available on a monthly basis. An SDA, however, can perform econometric cost analysis as its monthly observations on total cost and total output accumulate. Finally, the time period over which the cost and output data are relevant is 7/1/84 through 6/30/85.

Two definitions of total output are discussed: total terminees and total terminee/weeks. A total cost function is fitted manually for total terminees. Experiments with different forms indicated that total

terminee/weeks was the best way to define output for the econometric cost estimation.

The Cost Data. The cost data that are used as examples for analysis in this chapter are based on actual data for SDAs in the State of Washington. The data are structured for reporting to the Federal government and thus are based on the JTPA Annual Status Report ETA 8580 (August 1983). Exhibits 4 and 5 display selected variables from these fiduciary reports.

Strengths

In addition to the main limitation discussed above there are several additional strengths and weaknesses to these data as reported. A main strength of the data is that they are displayed in a clear and unambiguous way. The definitions of the variables are straightforward. Total expenditures (but not total social costs) of Federal funds are shown for a given time period, as are total participants and total terminees for the same time period. These three data items are the primary building blocks of any cost analysis. Total Federal expenditures can be expressed either as a function of total participants or as a function of total terminees. Exhibits 4 and 5 and the following analysis express total costs as a function of total terminees because reporting form ETA 8580 reports on the socio-economic characteristics of terminees. These socio-economic characteristics can affect the nature and extent of costs and therefore can be used to standardize the relationship between total costs and total output. Indeed, individual participant characteristics can be thought of as inputs entering into the employment and training process.

For instance, there are obvious instructional and resource implications to the fact that some of the JTPA participants and terminees have "Limited English Language Proficiency." Before they can be trained, either their English language proficiency must be improved or the instructional materials and instructor must present the training in the participants' native tongue. Either condition will likely increase both fixed and variable costs, shifting the total cost functions up, i.e., raising average and marginal costs. "Average weeks participated" at time of termination is another important variable. Clearly, other things equal, as average participation in weeks increases, total costs are certain to increase.

If weeks in the program are related to some specific training structure, then you can expect a more complex relation between weeks and total costs. Experimentation with the other socio-demographic characteristics that are available may also show some important empirical relationships.

Weaknesses

Although the above analysis would provide considerable useful information on program costs, there are also certain weaknesses to these data. The most important weakness is that the data do not

display the number of terminees by type of program treatment. While there are about two dozen different, but related, treatment approaches listed in the JTPA, as discussed in previous chapters, they can be condensed to a few main types: classroom training, on-the-job training and its variants, and job search or job placement services and their variants.[9] Also, multiple treatments are possible for a given participant.

If total terminees were disaggregated and enumerated by type of treatment, the following would be achieved:

1. One could estimate a much more accurate relationship between total costs and total outputs since the terminees receiving different treatments represent different inputs and, therefore, potentially different cost structures. Incorporating this additional information into the estimation of the total cost function may improve the statistical reliability of the marginal cost estimates.

2. Even though total costs are aggregated into a single number, marginal costs for each type of treatment can be directly estimated by expressing total costs as a function of each of the terminee treatments considered simultaneously. The equation below expresses this possibility:

$$TC = a_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + c_1 T + d_i Z_i + u$$

where
X1 = classroom training only terminees;
X2 = on-the-job training terminees;
X3 = job placement only terminees;
X4 = terminees with multiple treatments
T = Average weeks at termination;
Z_i = a vector of socio-economic variables such as limited English language proficiency, sex or high school grade, or equivalent and above.

a_0, b_1, b_2, b_3, b_4, c_1, and d_i = coefficients to be estimated; and,
u = the error term.

The outcomes of this estimation will be estimates of marginal cost for each type of treatment even if the SDA or state has aggregated these total costs such that they can't be separated in an accounting sense. This is a major advantage of using econometric techniques to estimate program costs.

A final problem with these data is that they do not include any estimates of costs contributed by the state to the operation of these

---

9 Some program operators have been used to considering job placement as a program treatment. More correctly, it is an outcome; one either is placed in a job or one isn't. The costs of job placement however, are inputs and these inputs represent the proper measure of program treatment.

programs. While not necessarily important for the estimation of costs at the SDA level (if we are concerned with only SDA performance), any actual or imputed costs contributed by the state to the operation of JTPA programs should be added to the quarterly and annual program year total costs to measure state level and full economic costs. It appears that these are likely to be the only major potential omission from the (social) cost data, other than foregone before-tax earnings of participants.

Finally, though income maintenance stipends are allowable in JTPA, such payments will not be a large component in reported federal data. However, participants can also receive unemployment insurance payments while in JTPA and these must be identified when they occur in the accounting record. Both types of expenditures must be omitted from any cost analysis at the social level.

ESTIMATING COSTS AT THE SDA LEVEL

As Exhibit 4 shows, cost data collected at the SDA level are in the form of a series of observations on costs and outputs over time--time series data. The data available on SDAs in unpublished form at the state level will be, at best, quarterly data. As Exhibit 4 shows, given the short history of JTPA, this currently does not provide enough observations to do meaningful cost analysis at the SDA level. Exhibit 4 shows how one should take the raw data and organize the data into a display that allows a straightforward description of the cost and output components of the SDA. With these data, average costs by quarter can be estimated by simple division of total costs by total terminations per quarter or by total termination-weeks per quarter. But, with the existing shortage of observations,this is as far as we can go. Total cost functions and marginal costs cannot be estimated at this time (December 1985).

However, the SDAs maintain program data on a monthly basis. Monthly data for SDA "X" above would provide about 15 observations within the data set shown in Exhibit 4--enough for cost analysis of the SDA's program if it is settled down into its long run course of operation and if there are no significant recording errors in the data. Under this arrangement a total cost function and marginal cost can be estimated. A procedure that has been tested before and which works is described below. It was designed to analyze the cost and performance experience of the Downriver Community Conference Economic Readjustment Program (DCC).[10]

Formulation of a Dynamic Total Cost Function for the Typical SDA.[11]

Due to the nature of an SDA's program there is reason to assume that there is a good deal of inertia in the data; that is, adjustment to any new level of activity is not made instantaneously and unused capacity may tend to persist. Concern over "lumpy" start-up costs is an expression of this phenomenon.

---

10 DCC is now an SDA.

11 This model is taken directly from Kulik, et al., 1982.

EXHIBIT 4

TITLE IIA ENROLLMENT AND COST DATA OF SERVICE DELIVERY AREA "X". TIME SERIES DATA.

Variable	Accounting Period				
	10/1/83 3/31/84 (1)	10/1/83 6/30/84 (2)	7/1/84 9/30/84[a] (3)	7/1/84 12/31/84 (4)	7/1/84 3/31/85 (5)
Total Adults					
Total Participants	258	326	271	456	702
Total Terminations	73	253	61	195	385
Male	36	121	30	87	166
Female	37	132	31	108	219
Less than high school graduate	9	46	4	25	60
High school graduate or more	64	207	57	170	325
Single household head with dependent children	19	75	18	54	119
White	55	207	52	174	345
Black	3	6	0	4	9
Other	15	40	9	17	31
Limited English	13	22	4	7	14
Handicapped	7	34	6	23	59
UI claimant	8	46	9	28	55
Unemployed	66	230	48	145	277
Average weeks at termination	10	12	22	16	16
Total Federal Program Costs in $100 (cumulative)	$5,746	$9,130	$2,428	$6,138	$10,083
Difference: (2)-(1); (3)-(2); etc.	---	$3,384	$2,428	$3,710	$3,945
Total Termination - Weeks (total terminations x average weeks at termination)	730	876	1,342	2,768	3,040
Average Cost/Terminee - Week	---	157	$181	173	130
Average Cost/Terminee	---	1,880	$3,980	2,769	2,076

Source: Job Training Partnership Act Annual Status Report; Various quarterly reports as noted by dates above for SDA "X" in the State of Washington.

Note: [a] 7/1/84 is the beginning of the accounting period. It appears that a large quantity of fixed or "start-up" costs were expended during this quarter.

We first posit a simple cost function in which costs are related to output in the relationship below:

(1) $C = C_0 + C_1 (q)$ (This is the same as $TC = a + bx$.)

Where,

C = cost

q = output

This formulation is relevant when the economic activity is at a steady state and the future can be predicted accurately. In our case, the production and delivery of program services entails a learning process by program managers. Our data will be monthly and during the early months of the program the administrators can be assumed to be moving towards some "steady state". Our problem is to formulate a cost function that reflects (a) the learning process in the short run, and (b) the tendency toward a steady state (a stable operating rate) in the long run. In this sense the cost function is dynamic--it reflects adjustment over time to a stable operating rate.

We can write the cost function as:

(2) $C_t = \alpha + \beta q^*_t$ long run cost function

(3) $C_t = \alpha + \beta q^e_t$ short run cost function

Where

q^*_t = expected output in period t

q^e_t = equilibrium output

In the case of the short run cost function, the SDA providing services expects some level of output and it prepares to meet that requirement with its planned staff and other resources. The costs, therefore, are related to the <u>expected</u> level of output, rather than the <u>actual</u> output.

We assume a learning process by the SDA in the form of adaptive expectations. The term below represents the rate of adaptation. The SDA can be modeled as revising its future expectations on the basis of past experience. Specifically:

(4) $q^*_t - q^*_{t-1} = (q_t - q^*_{t-1})$

or

(4') $q^*_t = \lambda q_t + (1- \lambda) q^*_{t-1}$.

Combining (4') and (3) we get:

(5) $C_t = \alpha + \beta\lambda q + \beta(1-\lambda) q^*_{t-1}$

Lagging (3) one period, multiplying through by $(1-\lambda)$, and subtracting the result from (5), we get:

$$(6) \quad C_t = (1-\lambda)C_{t-1} + \lambda\alpha + \beta\lambda q_t$$

This is the autoregressive form of the cost function which we estimate. A note on the estimation of this cost function is in order. Estimation of an autoregressive function by ordinary least squares will result in biased and inconsistent estimates of the parameters. Yet given the small number of observations that are likely available (15 months) and the transitional nature of the program during the first phase of operation, the use of more sophisticated techniques to overcome bias and inconsistency is ruled out at this time. As more observations--months of data--become available, it will be possible to adjust for the problems in the data more effectively if additional variables exist to adjust for the bias.

ESTIMATION OF PROGRAM COSTS AT THE STATE LEVEL

Much more information on costs and outputs exists at the state level since the state can augment its data by drawing information from all of the SDAs in its jurisdiction. It is not limited to the experience over time of only one SDA.

In the State of Washington there are ten SDAs. Data for a full year of operation are available on each SDA as shown in Exhibit 5. These data are known as cross-section data. They are a cross-cut of experience throughout the state at a given point of time. Since the data reflect a period well after JTPA program start-up, one can assume that the data reflect relatively stable operating patterns. In any case, start-up costs are now spread over a year or more. As with the data for a single SDA, one can array the variables of interest in an easy to read table and make comparisons across SDAs.

If you are willing to assume that, in general, the SDAs use similar approaches to managing their programs, e.g., for certain types of training, they all tend to use proprietary vocational schools, while for other types of training, they all tend to use public vocational technical schools, then it is possible to average the experience of the SDAs across the state. (Any aggregation of data makes assumptions of this nature.) Normally, this averaging will be done with a multi-variate regression model as has been discussed above.

Total Cost Based on Total Terminees. However, to explain the principles of the analysis, we have shown an approximate estimate of the total cost function, and the marginal cost and average cost functions based on it, by fitting a total cost function by inspection--that is, estimating the nature of the total cost relation and fitting a linear total cost function with a straight-edge.[12]

---

12 Early efforts at multiple regression and correlation often described this technique. See Ezekial and Fox, 1959.

Exhibit 5

TITLE IIA ENROLLMENT AND COST DATA FOR SELECTED SERVICE DELIVERY AREAS STATEWIDE. CROSS-SECTION DATA. REPORT PERIOD 7/1/84 - 6/30/85.

Variable	SDA 1	SDA 2	SDA 3	SDA 4	SDA 5	SDA 6	SDA 7	SDA 8	SDA 9	SDA 10
Total Adults										
Total Participants	471	807	3,241	932	936	451	750	623	806	448
Total Terminations	413	650	2,232	717	729	332	532	447	725	365
Male	191	329	1,030	301	449	185	209	188	371	177
Female	222	321	1,202	416	280	147	323	259	408	188
Less than high school graduate	85	117	433	104	171	56	93	94	183	79
More than high school graduate	328	533	1,799	613	557	276	439	353	542	295
Single household head with dependent children	142	187	792	239	182	77	228	144	206	115
White	351	586	1,136	356	634	275	428	365	503	340
Black	19	6	446	85	12	12	27	3	26	0
Other	43	58	650	276	83	45	77	79	196	25
Limited English	2	8	355	232	66	7	14	13	18	8
Handicapped	65	27	385	81	50	55	67	24	55	35
UI Claimant	57	111	250	60	113	26	96	73	134	67
Unemployed	307	584	1,922	634	671	308	444	341	700	312
Average weeks at termination	14	17	19	16	12	18	22	18	19	20
Total Federal program costs in $100	$7,765	$12,885	$37,232	$14,684	$12,124	$12,018	$7,844	$10,231	$12,301	$5,854
Total termination - weeks (total terminations x average weeks at termination)	5,782	11,050	42,408	11,472	8,748	5,976	11,704	8,046	13,775	7,300
Average cost/terminee - week	$134	$117	$88	$128	$139	$202	$67	$127	$87	$80
Average cost/terminee	$1,880	$1,982	$1,668	$2,048	$1,663	$3,620	$1,474	$2,289	$1,697	$1,604

Source: JTPA Annual Status Report for the program year ending June 30, 1985. Various SDAs.

Exhibit 6 shows first a scatter plot of the annual data for each SDA. (SDA#3 was not used since it was an extreme measure.) The solid line is an approximation of the total cost function fitted by hand. If this is a good approximation of the total cost function for the JTPA program in the State of Washington for the 1985 program year, then we see two important observations.

First, total costs are linearly related to total output. There are no significant economies or diseconomies of scale of operation within the range of the data we have available. The linear total cost function reveals that marginal costs are constant and they run in the neighborhood of $1,750 per program terminee, given that all forms of program services have been averaged together.[13]

However, it is much more accurate to fit such a cost function by means of a regression model. Given the few number of observations, this can easily be done on a personal computer such as the IBM-PC. The estimated total cost function based on an ordinary least squares (OLS) regression program has the following form:

$$TC = \underset{(284,094)}{\$412,625} + \underset{(501)}{\$1,193}\ X$$

The adjusted coefficient of determination ($\bar{R}^2$) is equal to .37. The data are interpreted as follows.

First, on the average, for the nine SDAs that were included in the analysis [14] fixed costs were $412,625 per SDA per year. This estimate

---

13 It is important to remind ourselves that as the number of data observations increase in quantity it will be possible to disaggregate the terminees by type of program treatment and thereby estimate marginal costs for each type of program treatment.

14 SDA 3 was omitted from both the hand plot and the regression estimate because it was an extreme value--an outlier. There is no hard and fast rule for editing data to remove outliers or other anomalies. Given sufficient observations, SDA 3 could have been left in the model and a variable been included to adjust for that phenomenon that accounted for its large size. With so few degrees of freedom however, it was best to omit SDA 3.

However, a regression model that included SDA 3 gave the following results:

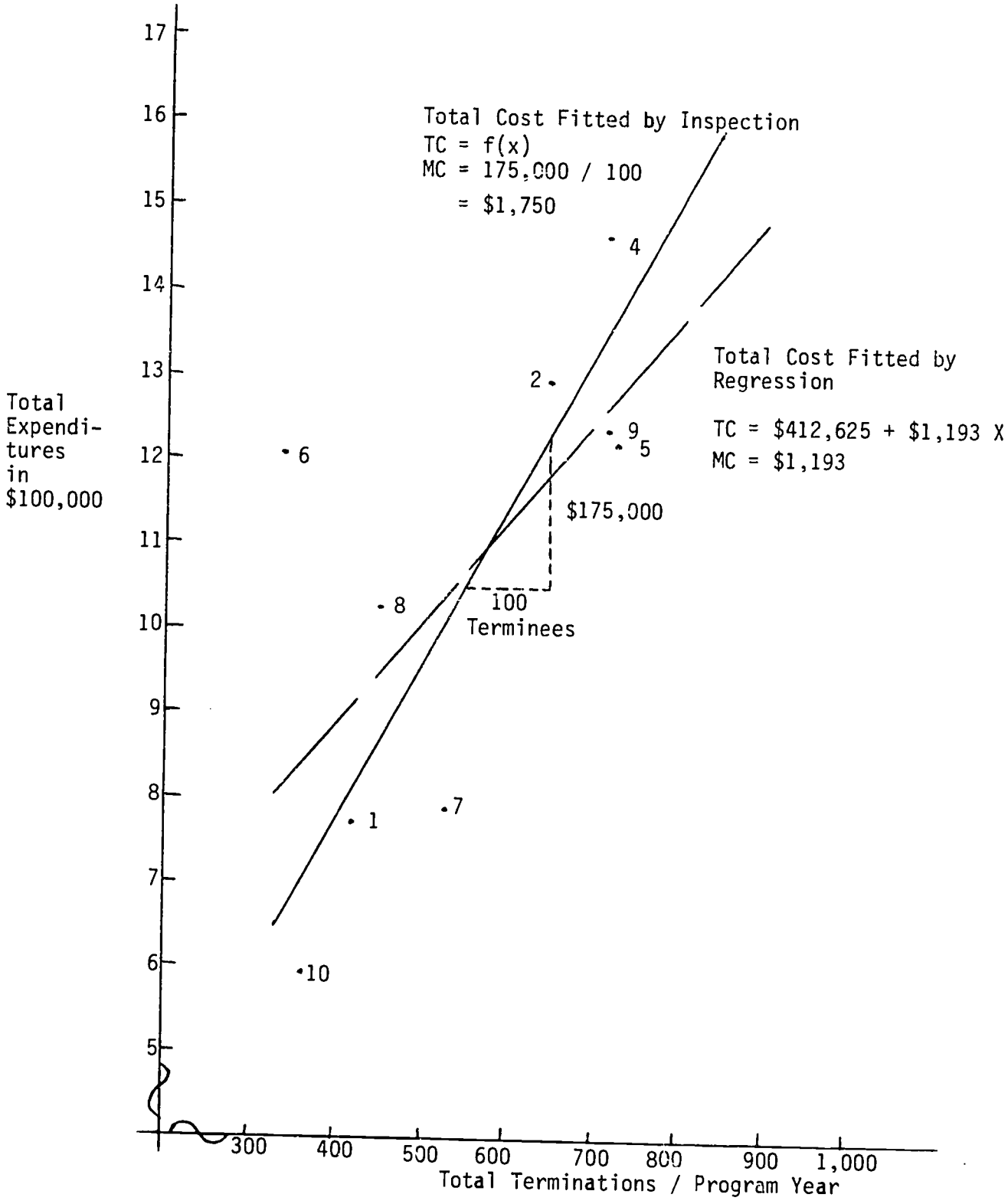
$$TC = \underset{(119,398)}{\$223,986} + \underset{(135)}{\$1,548}$$

with an adjusted $\bar{R}^2$ of .94 or 94 percent.

The high $\bar{R}^2$ suggests a much better fit leaving SDA 3 in the sample. However, another way of working at these results is that SDA 3 dominates the estimation so that the regression is really being fit between two points: SDA 3 and the average of the other nine SDAs. Under this situation the mathematics of regression analysis will give you a biased estimate.

Exhibit 6

COST ESTIMATION AT THE STATE LEVEL: THE SIMPLE RELATIONSHIP BETWEEN TOTAL FEDERAL COSTS (EXPENDITURES) AND TOTAL TERMINEES



is statistically significant at about a 10 percent level of significance. Next, marginal costs equals $1,193; that is, for the program year in question, across all nine SDAs, the additional cost of producing one more program terminee was about $1,200. This estimate is also statistically significant at conventional levels of significance. Finally, we note that total terminees explain 37 percent of the variation in total costs for the program year. Given the few number of observations, this is a reasonable result. With more observations, additional variables could be added that would explain more of the variations in total cost and yield a higher coefficient of determination ($\bar{R}^2$). Breaking this data set into monthly or quarterly observations by SDA, for instance, would improve our ability to estimate the total cost function, though a somewhat different model specification would be required to account for the time series nature of the data. Standard economic procedures exist to do this as are discussed, in part, above, for the dynamic SDA regression model.

BASIC METHODS AND ISSUES IN BENEFIT-COST ANALYSIS

Social Investment Criteria. There are three criteria that can be used to measure the net effectiveness of a social investment: the internal rate of return, the net present value of money benefits and the benefit-cost ratio. Exhibit 7 shows the mathematical expression of these criteria. Two important points should be reviewed concerning the implementation of these criteria. First, as discussed previously, marginal benefits and costs must be employed with these criteria. This is because the incremental benefits of any social investment must be equated with the incremental costs that created them. As long as the incremental benefit exceeds the incremental cost, one should continue to devote resources to the social investment. Once the incremental benefit equals the incremental cost that created it, no further gain in total benefits can be made by further inflow of resources to the social investment program.

Second, recalling Exhibits 1 and 3, one should not mix social concepts of benefit (cost) with participant or non-participant measures of cost (benefit). To do so is the economic equivalent of comparing apples with oranges.

As a final assessment to this discussion, it is necessary to put the nature of any program evaluation in its proper context. Paraphrasing Gramlich (1981), it is sufficient to consider a social investment program a success if the benefits yielded by it represent a net improvement--the present value of benefits exceeds the present value of costs. It is not necessary--or even practically possible--that a program be a panacea. Thus, a job placement strategy that costs an SDA $200 per participant need only to yield $200 in the present value of benefits (at the socially required discount rate) in order to be socially justified. It does not have to guarantee the worker against unemployment for the rest of the worker's life. Of course, if two

complete training approaches cost the same per program participant, but the present value of benefits of one program exceeds that of the other, one should always invest in that treatment that has the largest net gain. Such investment should continue until funds are exhausted or the net benefits equalize, whichever comes first.

Exhibit 7

INVESTMENT CRITERIA

Criterion	Formula
Internal rate of return	$\sum_{t=0}^{T} \frac{B_t}{(1+i)^t} - \sum_{t=0}^{T} \frac{C_t}{(1+r)^t} = 0$
Net present value	$\sum_{t=0}^{T} \frac{B_t}{(1+r)^t} - \sum_{t=0}^{T} \frac{C_t}{(1+r)^t} \geq 0$
Benefit-cost ratio	$\sum_{t=0}^{T} \frac{B_t}{(1+r)^t} \Big/ \sum_{t=0}^{T} \frac{C_t}{(1+r)^t} \geq 1$

B = benefit per unit of time

C = cost per unit of time

t = the time unit

r = the interest rate used in discounting

T = total number of time units

i = the discount rate which gives a net present value of 0

REFERENCES

REFERENCES

Ashenfelter, Orley. "Estimating the Effect of Training Programs on Earnings." *Review of Economics and Statistics*. Vol. 60. February, 1978.

Becker, Gary S. *Human Capital: A Theoretical and Empirical Analysis with Special Reference to Education*. New York, NY: Columbia University Press, 1964.

Bowles, Samuel. "Toward an Educational Production Function." In W. Lee Hansen, editor, *Education, Income and Human Capital*. Studies in Income and Wealth. No. 35. New York, NY: Columbia University Press, 1970.

Ezekiel, Mordecai and Karl A. Fox. *Methods of Correlation and Regression Analysis*. New York, NY: John Wiley and Sons, Inc., 1959.

Borus, Michael E. and William R. Tash. *Measuring the Impact of Manpower Programs*. Ann Arbor, MI: Institute of Labor and Industrial Relations. The University of Michigan-Wayne State University, November, 1970.

Gramlich, Edward M. *Benefit-Cost Analysis of Government Programs*. Englewood Cliffs, NJ: Prentice-Hall, Inc., 1981.

Highlights of JTPA Program Performance for Titles IIA and III during the JTPA Transition Year (October 1983-June 1984). Washington, D.C.: Division of Performance Management and Evaluation. Office of Strategic Planning and Policy Development. U.S. Department of Labor. November, 1984.

JTPA Performance Standards Worksheets for PY (Program Year) 1985. Washington, D.C.: U.S. Department of Labor. Revised January 1, 1985.

Jy, Richard W. "Costs: Theoretical and Methodological Issues." In Gerald G. Somers and W. J. Wood, editors. *Cost-Benefit Analysis of Manpower Policies, Proceedings of a North American Conference*. (May 14-15, 1969). Kingston, Ontario: Queen's University. Industrial Relations Centre, 1969.

Kulik, Jane, with J. W. Frees, Marcia C. Jerrett, Linda Sharpe and Ernst W. Stromsdorfer. *Reemploying Displaced Workers: The Implementation of the Downriver Community Conference Economic Readjustment Program*. Cambridge, MA: Abt Associates Inc., September 28, 1982.

Mallar, Charles, et al. *Evaluation of the Economic Impact of the Job Corps Program: Third Follow-up Report.* Princeton, NJ: Mathematica Policy Research, Inc., September, 1982.

McKean, Roland N. "The Use of Shadow Prices," in Samuel B. Chase, Jr., editor. *Problems in Public Expenditure Analysis.* Washington, D.C.: The Brookings Institution, 1968.

Moayed-Dadkhah, Kamran and Ernst W. Stromsdorfer. *Cost Analysis of Manpower Programs: An Analysis of the State of the Art.* Washington, D.C.: Office of the Assistant Secretary for Policy, Evaluation, and Research. U.S. Department of Labor. November, 1973.

Moayed-Dadkhah, Kamran, Bruno A. Oudet and Ernst W. Stromsdorfer. *An Econometric Analysis of the Costs of Selected Manpower Programs.* Final Report. Washington, D.C.: Office of Policy, Evaluation and Research. Manpower Administration. U.S. Department of Labor. April, 1974.

Sherman, Joel D. and Ernst W. Stromsdorfer. *Model for Benefit-Cost Analysis of Adult Education Program.* Washington, D.C.: Education Analysis Center for Educational Quality and Equality. Pelavin Associates, December, 1984.

Somers, Gerald G. and W. D. Wood. Editors. *Cost-Benefit Analysis of Manpower Policies. Proceedings of a North American Conference.* Kingston, Ontario: Industrial Relations Centre, Queen's University at Kingston. At Wisconsin: Madison, Wisconsin. University of Wisconsin. Center for Studies in Vocational and Technical Education, 1969.

Stromsdorfer, Ernst W., Teh-wei Hu and Maw Lin Lee. "Special Problems in the Economic Analysis of Investments in Human Resources." *Socio-Economic Planning Sciences.* Vol. 5, 1971.

U.S. Congress. *Job Training Partnership Act.* Public Law 97-300. Washington, D.C.: 97th Congress, October 13, 1982.

U.S. General Accounting Office. *Job Training Partnership Act: Initial Implementation of Program for Disadvantaged Youth and Adults.* GAO/HRD- 85-4. Washington, D.C.: U.S. General Accounting Office, March 4, 1985.

Weil, R. L., Jr. "Allocating Joint Costs," *American Economic Review,* December, 1968.

Weisbrod, Burton A. "Benefits of Manpower Programs: Theoretical and Methodological Issues. " in Gerald G. Somers and W. D. Wood. Editors, *Cost-Benefit Analysis of Manpower Policies. Proceedings of a North American Conference.* May 14-15, 1969. Kingston, Ontario: Queen's University, Industrial Relations Centre, 1969.

Data Sources and Data Definitions

Job Training Partnership Act. *Provisions*. Section D: Program Plans and Operations. Revision No.: 3. Effective Date: January 30, 1984. "Provision No. 740 - Program Cost Categories." Olympia, WA: Employment Security Department.

Job Training Partnership Act. *Provisions*. Section E: Management Information Systems. Revision No.: 5. Effective Date: May 30, 1985. "Provision No. 850 - Participant Record Instructions." Olympia, WA: Employment Security Department.

JTPA Annual Status Report. Washington, D.C.: Employment and Training Administration. U.S. Department of Labor.

JTPA Quarterly Fiscal Report. Employment Security Department. Training Program Services Division. JTPA Grants Management. Olympia, WA.

APPENDIX A

EXAMPLES OF TOTAL, AVERAGE AND MARGINAL COST FUNCTIONS

APPENDIX A

This appendix sets forth three common types of relationships between the output of a government-sponsored employment and training program and the inputs or resources used to produce that output.

Economic theory and empirical studies narrow the range of possible equation forms for cost functions. Once a total cost form is specified, the form of the average cost function derived from it is also specified.

The four equations listed below represent the general form of the alternative total cost equations discussed in this appendix.

(1) $TC = a + bX + u1$

(2) $TC = a + bX + cX^2 + u2$

(3) $TC = a + bX + cX^2 + dX^3 + u3$

(4) $\text{Log } TC = \log a + b \log X + \log u4$

Schematic diagrams of the first three equations are provided. The variables in each definition are defined as follows:

TC = total cost;

X = total output;

a, b, c, and d represent partial regression coefficients; and

u1, u2, u3, u4 = error terms to formally complete each model.[1]

Equation (1) is the linear formulation of the total cost equation. The term a represents the amount of fixed costs in the total cost function. The expression bX says that for any change in the number of persons trained, X, total cost will change by some constant multiple relative to the change in persons trained, bX. Therefore, if the number of trainees changes by one trainee--one output unit--then total cost will change by a constant, b, relative to that change of one unit, regardless of the level of total output. Regression analysis is used to

---

[1] As in the previous chapters, the error term encompasses those independent variables which may be relevant to the model but which, for whatever reason, are omitted. It is assumed that the omission of these variables does not bias the estimates of the effects of those variables that are explicitly included in the model.

statistically estimate a linear total cost function. The regression coefficient, i.e., b, for the trainee variable is the estimate of marginal cost in the case of the linear total cost function.

Equation (2) is the quadratic form of the basic total cost relationship. It is quadratic because total cost is hypothesized to be related to the total amount of output plus the square of that total output. In simpler terms, this is just one variation of a non-linear relationship between total output and total input. The marginal cost of the program, i.e., b + 2cX, is again found by using the same method of differentiating. Now, however, in order to determine marginal cost with the quadratic form, once the derivative is found, a specific level of output must be substituted into the equation. In other words, the marginal or additional cost in this equation form is dependent upon the level of output, i.e., marginal cost is not a constant quantity relative to output levels, but changes as the total amount of output changes per unit of time.

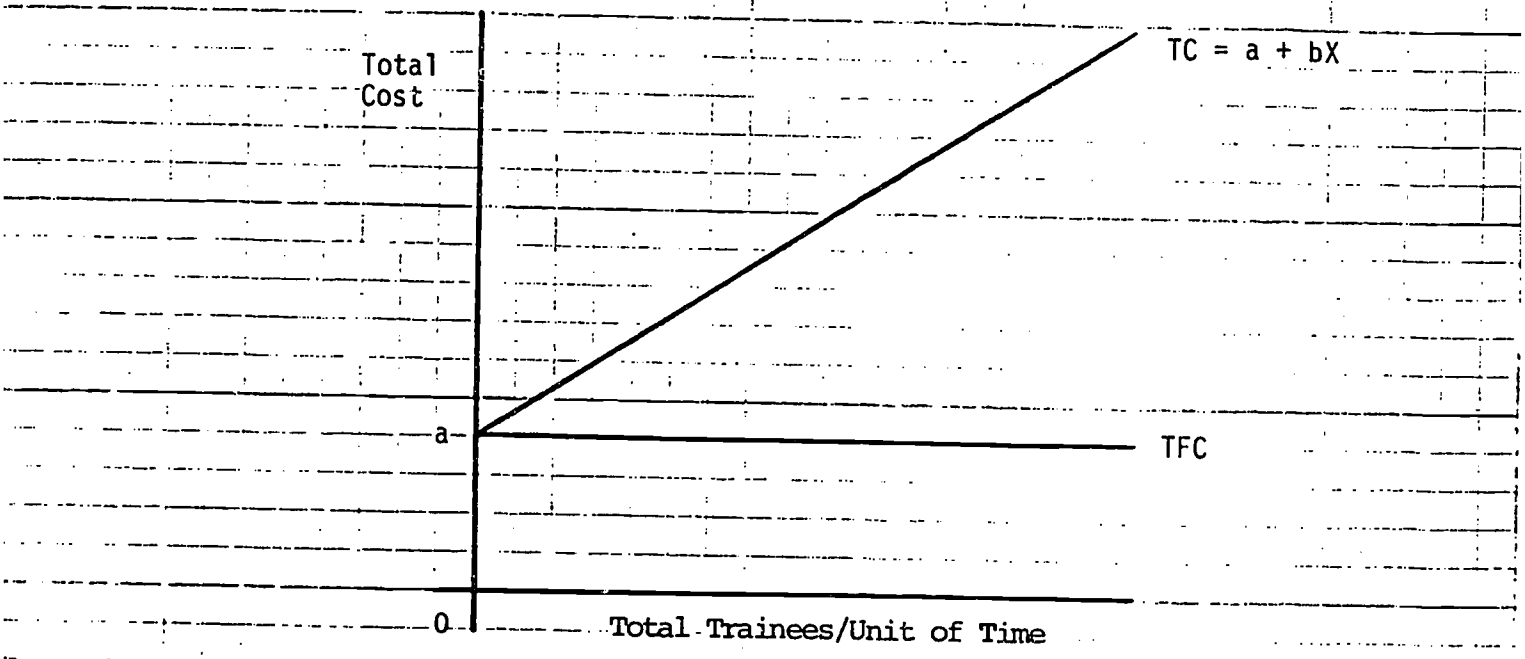
This difference between constant marginal cost and a changing relation for marginal cost reflects two different hypotheses concerning the nature of the relationship between total cost and total output. In the linear case, it is hypothesized that total cost bears some constant relationship to output over a range of output rates. Thus, if output goes up (or down), costs will increase (decrease) by some constant multiple of output. No increasing or diminishing returns to inputs exist as the output expands. Each new additional input unit is just as productive as the last. However, in production processes, including education or training production processes, this may not always be the case. As output is expanded in the short run, the productivity of successive input units may rise or fall. The intuitive reasoning behind this changing productivity, and thus decreasing or increasing marginal cost, is that some inputs are fixed in quantity so that with the addition of more units of one kind of input, either more efficient use of the total set of inputs occurs and productivity increases or crowding and imbalances occur and thus productivity decreases. The quadratic equation form and other non-linear forms specify changing marginal costs because marginal costs vary with output levels.

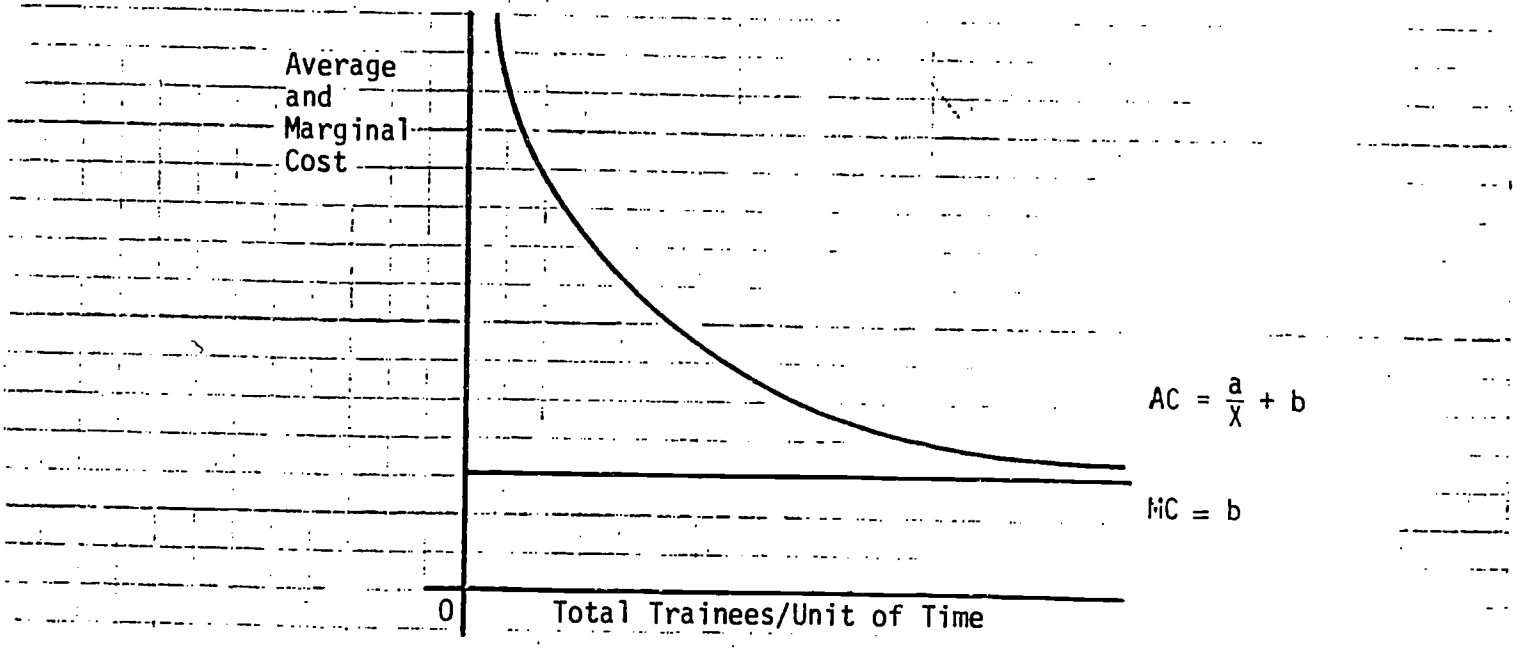
Equation (3) is the cubic form of the total cost function. It is so called because one of the independent variables in equation (3) is a cubic term. Marginal cost calculations can be made using the same method described for equation (2). Equation (3) allows for diminishing or increasing returns to an input used in an employment and training program, as does equation (2).

Equation (4) is the logarithmic formulation of the basic total cost relationship. The total cost equation in logarithmic form, apart from the fact that it may fit a given set of data better than the other three forms, can be used to make direct inferences about the elasticity of costs relative to output. Elasticity in this sense is defined as the percentage change in total cost divided by the percentage change in total output. Elasticity values are expressed in percentage terms and thus enable proportional cost response comparisons to be made among various production processes such as four different employment and training program treatments. The elasticity value is directly

Appendix Exhibit 1

ALTERNATIVE COST CURVES: THE LINEAR FORM





Note: The scales on the axes of the Total and Average and Marginal Costs are not the same.

Appendix Exhibit 2

BASIC DEFINITIONS FOR APPENDIX FIGURES 1, 2, AND 3

- Total Cost (TC) equals Total Fixed Costs (TFC)--Administrative and Start-up Costs--plus Total Variable Costs (TVC)--Operating Costs other than Administrative Costs.

  $TC = TFC + TVC$

- Average Costs (AC) equals Total Costs divided by Total Output (X)

  $AC = \frac{TC}{X}$

- Marginal Costs (MC) equals the change (Δ) in Total Costs as Total Output changes by one additional unit or increment.

  MC = ΔTC with respect to X. It is the slope of the TC function.

  a = TFC

  b, c and d = Coefficients that measure the slope of the TC function.

NOTE: All of the above values are expressed in terms of a common time period, e.g., total costs per week and total trainee completors per week. The length of the time period is not important.

estimated by the regression coefficient for the logarithm of total output--the number of persons trained.

If the elasticity is greater than one, then the percentage change in total cost will be greater than the percentage change in output as output increases. In other words, average costs will rise as a result of an increase in the total number of persons trained. If the elasticity is less than one, then a change in the total number of trainees will result in a smaller percentage change in total costs, and average cost per trainee will fall as output is expanded over some range. If the elasticity value is one, then average cost will remain constant as the result of an increase in total persons trained because the percentage change in total cost equals the percentage change in total output. From cost elasticity measures one can infer the optimum rate of operation, that is, the point where average cost is at a minimum, an important consideration in the efficient management of any enterprise.

Four alternative average cost functions are applicable in this study. The average cost functions are companion forms of analysis to the total cost functions. The average cost functions discussed in this chapter are formulated as follows:

(5) $AC = a/X + b + u5$

(6) $AC = a/X + b + cX + u6$

(7) $AC = a/X + b + cX + dX2 + u7$

(8) $\text{Log } AC = \log a + (b - 1) \log X + \log u8$

The variables used are defined as follows:

AC = total cost divided by total output

X = total output

a, b, c, and d = partial regression coefficients

u5, u6, u7, and u8 = error terms to formally complete each model.

Note that the average cost equations (5), (6), (7), and (8) are derived from total cost equations (1), (2), (3), and (4), respectively, by dividing each of the total cost equations by the total output variable, X. However, equation (8) is not typically estimated because equation (4) directly answers the question of increasing or decreasing total and average costs as output changes.

Technological factors can work to create increasing or decreasing total and average costs because technological levels in employment and training programs may be reflected in training situations that have a discrete output capacity. As a result, training capacity is "lumpy." An SDA may be faced with the situation of choosing a site that can optimally train 100 trainees per month; that is, for this location and size of plant, average cost is lowest at a rate of 100 trainees per

month. However, the projected training center may enroll at a more than optimum rate. Thus, actual average costs will be greater than average cost would be at the optimum rate of output. Average costs thus can be reduced if the rate of output drops back toward 100 per month.

To summarize, for a given capacity, other things equal, average cost is usually higher for a small rate of output because fixed costs are spread over a low rate of output. As the rate of output increases, average cost decreases since average fixed cost per unit of output decreases; a point is usually reached in the rate of output, however, where average cost reaches a minimum. Expansion beyond this point results in a rise in average cost because additional units of variable cost become less efficient as they are forced to work with the fixed inputs.

GRAPHICAL EXAMPLES

A linear total cost function is shown in Appendix Exhibit 1. Total costs change by a constant amount as output increases. Each additional unit of output, e.g., a trainee who completes a training program, costs the same to produce as each previous trainee terminee in the SDA or state. This gives us a flat marginal cost curve because the extra cost of producing an extra trainee terminee remains the same for each additional trainee. Finally, the average cost of producing a trainee terminee declines over the full range of program operation because the fixed cost or start-up costs for running a particular program or program treatment are spread over larger and larger amounts of total output while marginal cost stays constant. Eventually, the average cost effectively equals the marginal cost as a/X approaches zero.

As noted above, one can estimate the average cost curve by dividing total costs by total output, as follows:

$$\frac{TC}{X} = AC = \frac{a}{X} + \frac{bX}{X} = \frac{a}{X} + b$$

Since fixed costs are fixed, they do not change as output changes, marginal cost--the change in total costs as total output changes by one additional unit--is equal to b; or,

MC = b = change in TC with respect to a one unit change in total output.

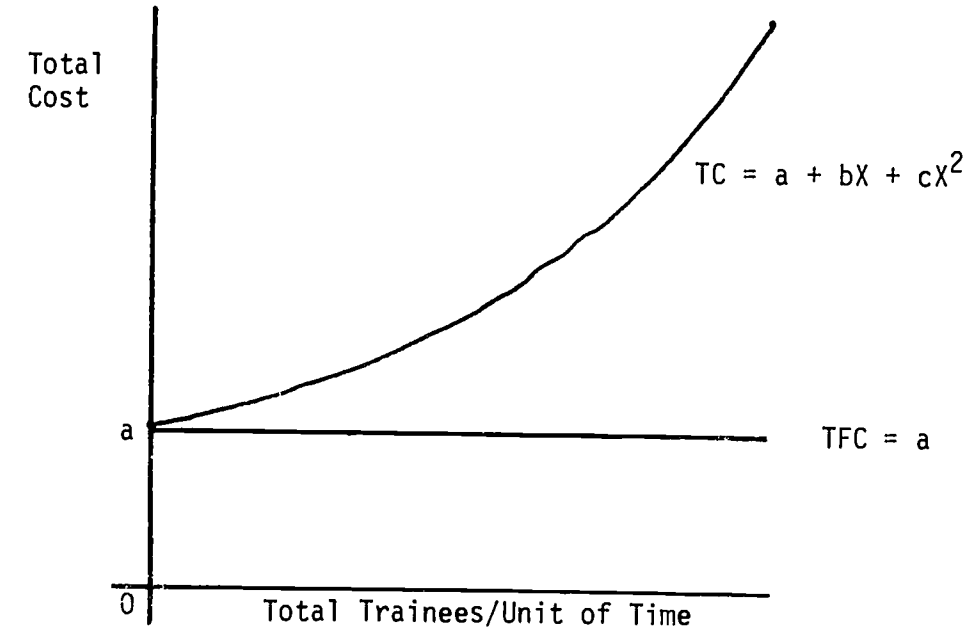
A quadratic cost function is shown in Appendix Exhibit 2. Here, total costs change by an increasing amount as total output increases. The equation for this total cost measure adds an additional variable, X^2, the square of total output, which yields the curvilinear shape of the function. Dividing the total cost function through by total output at any given level of total cost, yields average costs:

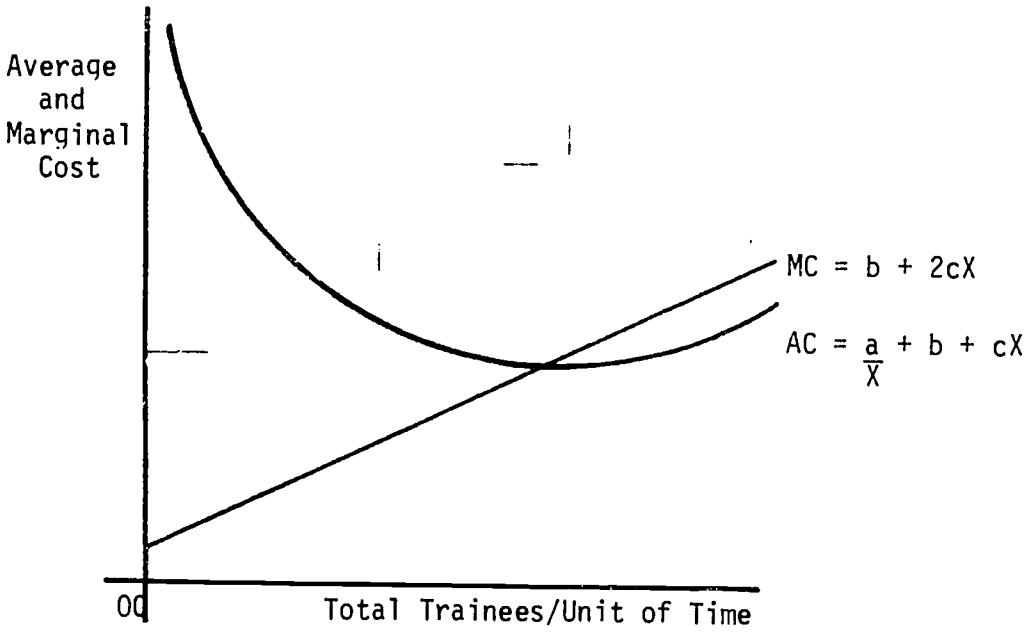
$$\frac{TC}{X} = AC = \frac{a}{X} + \frac{bX}{X} + \frac{cX^2}{X} = \frac{a}{X} + b + cX.$$

Note here that as X increases in size the value of a/X approaches zero,

Appendix Exhibit 3

ALTERNATIVE COST CURVES: THE QUADRATIC FORM





Note: The scales on the axes of the Total and Averages and Marginal Cost diagrams are not the same.

while the value of cX becomes very large. As cX increases in size it approaches the value of 2cX in the marginal cost function. The value of b is constant in both the marginal and the average cost function, so that at high values of output marginal and average costs tend to converge.

Finally, Appendix Exhibit 4 displays the cubic total cost function. When divided by total output, the cubic cost function yields the conventional U-shape to the average cost function.

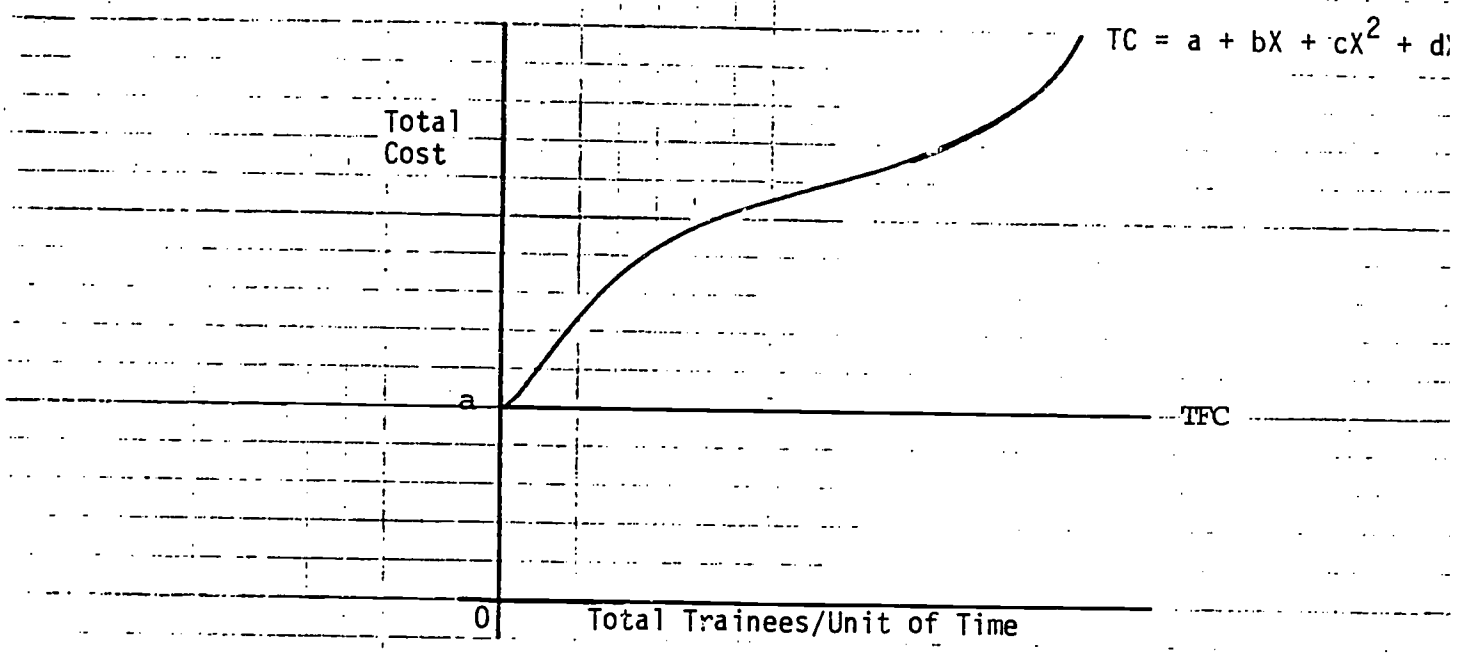
$$\frac{TC}{X} = AC = \frac{a}{X} + \frac{bX}{X} + \frac{cX^2}{X} + \frac{dX^3}{X} = \frac{a}{X} + b + cX + dX^2.$$

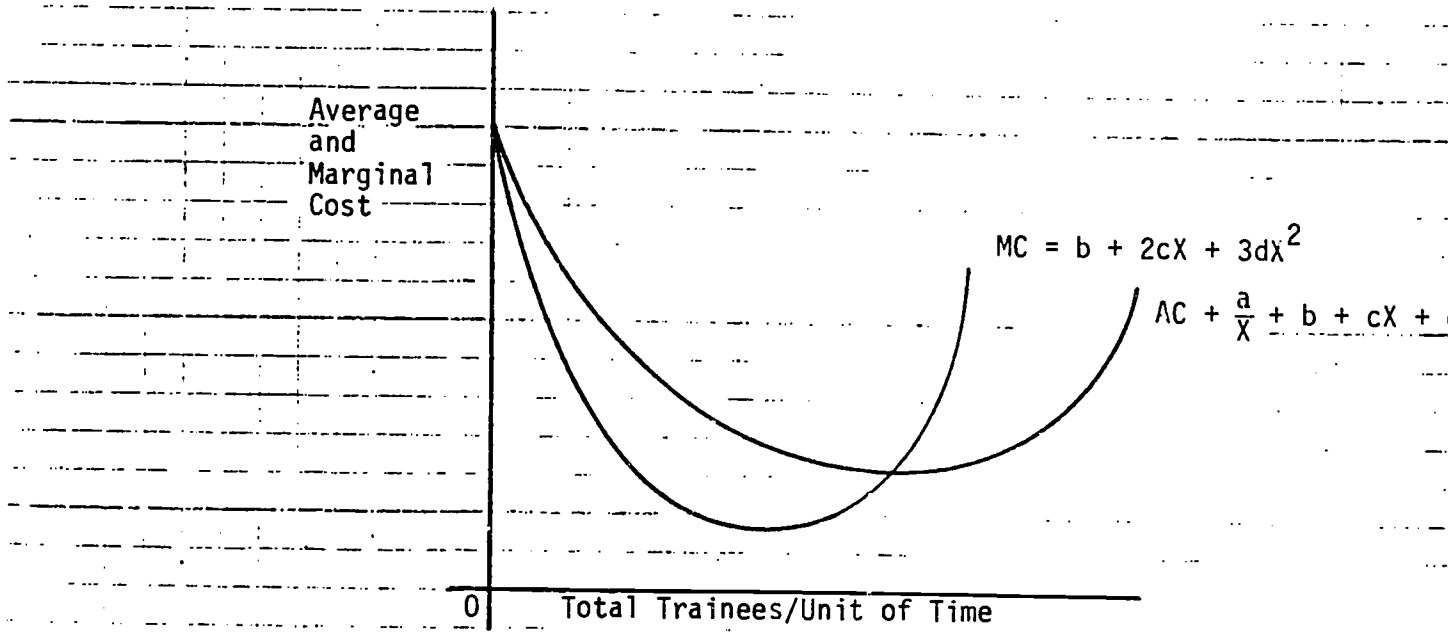
As total output increases, the term a/X causes average cost to decrease. However, at the same time, the terms cX and dX^2 cause average cost to increase. cX and dX^2 dominate a/X as X increases and average cost increases as total output increases. As may be clear at this point, the average cost and the marginal cost curves are related. In this cost relationship, at some point the marginal cost of producing an additional trainee terminee increases with each additional trainee terminee who is added to total cost. This increase is due to the general phenomenon known as the law of diminishing returns. This phenomenon, in turn, is due to the relationship between fixed inputs, such as a single teacher or a classroom of a given size and the number of variable inputs--most importantly the participants to be trained--relative to that teacher or classroom along with any equipment the class may require. As you add more students to the fixed classroom or require a single teacher to instruct more students, as in, say, a vocational-technical class in screw machine operation, the incremental amount of knowledge imparted per unit of teacher input will begin to drop. There are a variety of reasons for this: The room gets more and more crowded, the teacher's span of control over students is attenuated, insufficient attention may be given to each student--all relevant questions may not be answerable, there is insufficient machinery for each student to acquire the optimal practical experience, and so on.

In Appendix Exhibit 4 the law of diminishing returns is fundamentally responsible for the curvilinear shape. However, the relationship between the fixed and variable inputs is such that at first there are increasing returns--incremental output increases as an additional unit of a variable input is added to the set of fixed inputs. This can be due to some complementary relationship between the fixed and variable inputs; for instance, in instructing students in screw machine operation, the most efficient class size may be 20 students, due to the technical relationship between the teacher, the students, the required mix between formal and practical instruction and so on. Thus, the effectiveness of instruction increases up to 20 students, then becomes constant, and, as the class expands further beyond the optimal size of 20 students, instructional effectiveness decreases. Hence, the particular shape of this total cost function--total cost first increases at a decreasing rate, then becomes constant briefly, then switches to increasing at an increasing rate. This is expressed in the U-shape of the marginal cost function and the average cost function. The equations that express the value of marginal and average cost, at any level of total output, are shown on the figure. Their general

Appendix Exhibit 4

ALTERNATIVE COST CURVES: THE CUBIC FORM





Notes: The scales on the axes of the Total and Average and Marginal Cost curves are not the same.

derivation and interpretation is similar to that in Appendix Exhibits 1 and 2.

We should note that an SDA or subcontractor that is operating an employment and training program could conceivably have a total cost function that would take on any one of these three forms (as well as other forms that are not shown here), depending on the nature of the relationship between outputs and inputs in the programs being offered. If one were to break down a program by type of treatment or services delivered, it is possible that each service or treatment could have a different total cost function. In fact, even if the basic form is the same--say, the total cost function is linear for each of several types of training, you could expect the average and marginal costs to differ for each type of training. Empirical analysis of costs has shown this to be the case.

APPENDIX B

SAMPLE OF FEDERAL REPORT FOR OBTAINING PROGRAM DATA

OMB Approval No. 12

U.S. DEPARTMENT OF LABOR
Employment and Training Administration

JTPA
ANNUAL STATUS REPORT

a. STATE/SDA NAME AND ADDRESS

b. REPORT P
From

To

I. PARTICIPATION AND TERMINATION SUMMARY	TOTAL ADULTS (A)	ADULTS (WELFARE) (B)	YOUTH (C)	DISLOCAT WORKER (D)
A. TOTAL PARTICIPANTS				
B. TOTAL TERMINATIONS				
1. Entered Unsubsidized Employment				
a. Entered Registered Apprenticeship Program				
b. Entered Armed Forces				
2. Youth Employability Enhancement Terminations				
a. Entered Non-Title II Training				
b. Returned to Full-Time School				
c. Completed Major Level of Education				
3. All Other Terminations				

II. TERMINEES PERFORMANCE MEASURES INFORMATION		TOTAL ADULTS (A)	ADULTS (WELFARE) (B)	YOUTH (C)	DISLOCAT WORKERS (D)
SEX	Male				
	Female				
AGE	14 - 15				
	16 - 21				
	22 - 54				
	55 and Over				
EDUCATION STATUS	School Dropout				
	Student *(High School or Less)*				
	High School Graduate, or Equivalent, and Above				

SIGNATURE AND TITLE

d. DATE SIGNED

e. TELE. NO.

a. STATE/SDA NAME AND ADDRESS

b. REPORT PERIOD From To

NO.	II. TERMINEES PERFORMANCE MEASURES INFORMATION – *Continued*		TOTAL ADULTS (A)	ADULTS (WELFARE) (B)	YOUTH (C)	DISLOCAT WORKE (D)
0	FAM. STAT.	Single Head of Household with Dependent Children				
1	RACE/ETHNIC GROUP	White *(Not Hispanic)*				
2		Black *(Not Hispanic)*				
3		Hispanic				
4		American Indian or Alaskan Native				
5		Asian or Pacific Islander				
	OTHER BARRIERS TO EMPLOY.	Limited English Language Proficiency				
		Handicapped				
	U.C. STAT.	Unemployment Compensation Claimant				
	LAB. FORCE STAT.	Unemployed				
	Youth Welfare Recipient					
	Average Weeks Participated					
	Average Hourly Wage at Termination					
	Total Program Costs *(Federal Funds)*					

MARKS

B. The Debate Over Experimental vs. Quasi-Experimental Approaches

By Ann Bonar Blalock

June 1986

AUTHOR'S ACKNOWLEDGEMENTS

In the process of carrying out the *JTPA Evaluation Design Project*, John Wallace of the National Commission for Employment Policy requested that a paper be added on the debate among evaluation researchers—and most currently within the U.S. Department of Labor—over the value and feasibility of using field experiments to evaluate programs such as JTPA, rather than sophisticated nonexperimental approaches. He was keenly aware of the significance of this debate for two substantially different but equally important contexts of JTPA evaluation: state and SDA contexts. I appreciate his raising this issue and being responsible for the addition of such a paper to this series of volumes.

I am most indebted to Ernst Stromsdorfer for his careful and thoughtful review and editing of the paper, and for the excellent recommendations for its revision. I am grateful to Terry Johnson for his important suggestions for bringing greater balance and specificity to the discussion. And I want to thank Deborah Feldman for her useful comments on the first draft. My special appreciation goes to Ernest La Palm, who provided the needed support and personal encouragement required to initiate this project and contribute to its products.

Ann Bonar Blalock

PREFACE

The evaluation guide presented in Volume V for studying the impact of JTPA on the recipients of its services utilizes a sophisticated non-experimental research strategy. Before making this choice, an assessment of a range of methodologies was made. Approaches used in evaluating previous employment and training programs were reviewed. JTPA's unique program characteristics and context were considered. The climate in which the program was designed and is being implemented was given attention. The level of interest in the evaluation mandate in the legislation, and the evaluation capability of states and local JTPA administrative organizations were judged.

Despite these careful considerations, the choice not to use an experimental approach--that is, the approach considered to conform most closely to scientific principles and methods--was a complex and difficult one. Why? The following discussion addresses this question by exploring a perpetual methodological debate in policy research.

CONTENTS

INTRODUCTION

A distinguishing characteristic of the human species is intolerance of uncertainty and the unending search for explanations which reduce it. This search has been a persistent feature of all societies irrespective of the nature of their development. The complexity of human needs and abilities supports this pursuit of information about "reality".

The form this search takes may be sacred or secular. The explanations may range from myths to carefully gathered empirical evidence. Policy research, at its best, occupies a position at one end of this continuum, representing a systematic scientific approach for uncovering "the truth" about the effects of social policies and the programs designed to carry them out. But the most competent scientists can only approximate the truth. The best they can do is to state educated probabilities that things "are" as they are able to perceive them through the analysis of unbiased information. And not even the most rigorous scientific methods are perfect in removing all the potential biases which stand between the evaluator and what is true.

This conclusion is important to a discussion of the longstanding debate in policy research between the use of experimental vs. nonexperimental research designs in seeking the truth about social programs. It is often assumed, in this eternal argument, that certainty is possible if only the proper methodology were used. However, the issues are far more complex than that.

This paper probes the nature of this polarized debate to provide a context in which the reader can better understand and appreciate the choices made in the GUIDE FOR NET IMPACT EVALUATIONS, which utilizes a rigorous nonexperimental approach. In studying this controversy, and its implications for the evaluation of JTPA impact at the state and local level, it is important to discuss the following: the background of the debate; the continuum of research approaches typically used in making assessments of social programs and the kinds singled out in the debate; and the scientific, organizational and political tradeoffs involved in making choices.

CONTEXT OF THE DEBATE

The greatest heat in the debate has been generated in skirmishes between policy researchers advocating for the use of experiments, and those arguing that rigorous nonexperimental designs are more appropriate in most policy settings. A lack of consensus among these groups with respect to the definition of concepts central to the debate complicates this argument over which research approach actually brings us closer to the truth, and therefore provides a better basis for improving policies and programs.

But there is nothing new about the controversy. This debate has been raging for at least six decades, beginning with the Thorndike era in psychology in the 1920s. Questions about the use of controlled experiments have continued into the 1980s and are likely to remain. However, there is growing recognition that no one ideal scientific solution is right for all circumstances. Researchers of both persuasions, in the physical as well as the social sciences, find "the truth" almost as elusive now as when the controversy began. What continues to unite these professional peers, however, is the conviction that the application of scientific method to the study of how effective social interventions may be is far superior to more subjective judgments of their value. This common cause is one explanation for the growth of applied research, and its selection for use in the social policy arena in this era of "social accountability".

So what is this debate all about, and why is it a significant issue for states and SDAs in evaluating programs under the Job Training Partnership Act?

Defining the Controversy

The debate turns on the primary goals of scientific inquiry, which influence the choice of issues and methods. These goals involve the endproducts of research: the information acquired and its interpretation. Accuracy and utility are the critical ingredients of this information. More specifically, the important characteristics of scientific information are reliability, validity, and generalizability. In applied reseach an additional characteristic is the usefulness of the information for policy purposes. Beyond this, another goal is the replicability of the research itself, and its conclusions, in other relevant contexts and settings. 1/

To achieve these goals, principles and methods are applied which express logical steps in scientific method. We go to the trouble of conforming to this hard-won historical approach because it brings us closer to "reality" than any other form of assessment. In more technical terms, it helps the evaluator reduce errors and biases associated with the way an evaluation is carried out, and the conclusions arrived at on the basis of the evaluation. In general, the chronology of this evaluation process is as follows:

> Selection of issues which are feasible and useful to study, based on what is already known about these issues and the purpose of the research.

1/ A glossary is provided in Supplement A of this volume which may help clarify the meaning of some of the key terms used here in describing research objectives and methods.

Correct definition and measurement of the factors considered most important in shaping or otherwise affecting these issues.

Development of an appropriate research design guiding the selection of the individuals, groups or organizations to be studied--and the collection, analysis and interpretation of reliable information which can provide valid inferences about the influence of the factors selected for study.

Selection of reliable methods for collecting information about the influence of these factors.

Selection of appropriate methods for analyzing the information gathered.

Cautious interpretation of the results of the analysis, including what insights and conclusions can be drawn, and their generalizability to the intended subjects of the research and to the body of knowledge about the issues studied.

Repetition of the research in alternative contexts and settings, to see if similar inferences can be obtained for the same kinds of research subjects.

The major scientific objective of program evaluation is to guard against error and bias in seeking the truth about the influence of the program's interventions on a particular set of desired outcomes for participants--apart from the influence of other factors which may also influence program outcomes.

It is now well understood that explaining the results of attempts to change people's attitudes and behavior is a monumental task. But it is less well understood that this becomes more difficult if the interventions that are expected to change people are incremental, and other important influences in their lives are not or cannot be measured. Since most social programs involve mainly short-term limited efforts to change those who experience social problems, the evaluator's responsibilties represent a genuine challenge. This challenge is compounded by the increasing public demand for accountability in tax-funded programs, and the growing public appreciation of the technical means now available for judging the effectiveness of such programs. These developments are occurring in the context of severely reduced funding for social programs and their evaluation.

Given this framework within which the assessment of program impact now takes place, the most important scientific problem is developing ways to make valid estimates of the effects of public services and other

forms of social intervention. The debate over what kind of approach should be taken in accomplishing this task must therefore address the problem of error and bias. 2/

Significance of the Debate for States and SDAs

Before discussing further the scientific aspects of the debate, it is important to outline its significance for JTPA. One of the most significant evaluation mandates for states is the emphasis in the legislation on "the return on investment". To comply with this mandate, net impact studies must be carried out. The research designs used to implement such studies affect the accuracy with which the return on investment can be estimated. If JTPA outcomes for the intended population are as the Act directs, and these outcomes can be attributed to the influence of the program's interventions, the benefit warrants the investment in policy terms. However, if the net effect is not positive the policy implications are quite different. The Act essentially places a heavy burden on the evaluator to use those research approaches which best specify the return with a high level of confidence that the estimated impacts are as close to the true effects as possible.

Therefore the continuing debate about what constitutes the best research design affects the credibility of the evaluation choices made at the state and local level. But the state and SDA contexts in which such credibility is judged are significantly different. And these environments are distinctly different from the national context of JTPA evaluation.

States are vulnerable to strong, well organized state-wide public interest group pressures that argue against withholding services from one group of eligible clients while providing them to others in on-going programs, which is a requirement of experimental designs. Also, research designs cannot be allowed to interfere with client, employer or union prerogatives in relation to such programs. Such pressures have a substantial although sometimes indirect effect on the operating decisions of state agencies, including evaluation choices. SDAs are highly dependent on the continuing good will and support of local elected and appointed officials, employer and labor organizations, and social service coalitions, which can influence public, professional and client perceptions of locally managed government programs.

2/ Although the research definitions of "error" and "bias" differ, the term "bias" is used throughout this paper in the nontechnical sense, to refer to both errors and biases in the evaluation process.

Despite pressures placed on the federal agencies by the Congress and well financed and highly organized political, economic and public interest lobbies, the research divisions of these agencies are not usually the primary target of lobbying activities. Although public support for large-scale experimentation at the national level has always been weak, there has been growing tolerance for limited field experiments which affect only a small subset of any state's or local service provider's clients, and occur only in a few selected state and local sites. This is well illustrated by the ability of the U.S. Department of Labor to recommend a small number of JTPA field experiments, while simultaneously analyzing data on a representative sample of all JTPA clients using a nonexperimental approach. The political and organizational compromises at the federal level are based on different tradeoffs.

One must conclude that the research design debate has quite different parameters at the state and local level, where even limited experimentation inevitably involves a significant proportion of the clients of a social program. Those engaging in the controversy often fail to recognize that there are multiple tradeoffs to consider at each level, which require compromise.

SCIENTIFIC TRADEOFFS IN THE DEBATE

The consideration of scientific tradeoffs takes precedence in making evaluation choices, even though political and organizational issues act as a sieve through which high priority research approaches are sifted for their broader feasibility and utility. Therefore major emphasis in this paper is placed on scientific compromises. The following issues are addressed: the issue of error and bias in evaluation research; the continuum of research designs which respond to this issue; common sources of error and bias; major differences between experimental and quasi-experimental designs; and the advantages and disadvantages of both approaches.

The Issue of Bias

It is difficult to avoid subjectivity in determining whether policies and programs work as intended. The sources and varieties of potential bias which scientific method seeks to control are numerous. 3/ These biases can be introduced at any stage in the evaluation process: the conceptualization of issues to be studied; their measurement, the sampling of the individuals to be studied; the selection of a research design for collecting and analyzing information about them; data collection; data analysis; and the interpretation of evaluation results.

3/ An array of examples is provided in Supplement B.

The following are some examples of the kinds of biases that must be dealt with in program evaluations. A primary task of the evaluator is to avoid as many as possible through the way in which the evaluation is planned and implemented, and to try to compensate for those that cannot be controlled by these means.

For example, the outcomes of clients involved in an evaluation may be influenced simply by participating in a study. The additional attention alone may affect their attitudes toward the program and the way they utilize its services. Such attention is an unmeasured non-program intervention which may bias an estimation of the program's own effects. Or the evaluation may focus on the less relevant issues, missing the critical ones. Perhaps the relationship between single head of household and earnings is a key one in understanding the effect of the program on welfare status, but this is not studied. Or it may be that the most important relationships have been properly identified but the variables are poorly measured. For instance, in defining program interventions such as job search assistance, or program outcomes such as employment, too few indices may have been used, or the measures selected may not have been consistently defined across all clients in the study. Evaluation results can also be biased because they are not representative of the program's target group due to inadequate sampling, in which case they may under-represent important groups such as women or youth.

Difficult to control are biases associated with clients' own self-selection into JTPA, and the initial selection of clients by program staff. Not all eligible clients will apply. When applicants exceed program resources staff will naturally enroll those eligible applicants they judge to be most in need, those best able to benefit from the program, or those who will most likely help them comply with performance standards. In some cases these clients may be those most capable of achieving program goals without the program's assistance.

Sometimes clients' outcomes are compared with those of a group of individuals who are not sufficiently similar. If the comparison group's prior labor market experiences are significantly different than the JTPA group, for example, this will bias the results. Or the essential information to be collected may not have been reliably gathered because of the failure of data gatherers to adhere systematically to scientific interviewing principles and procedures. Or the analysis of information may utilize inappropriate methods. Also, differences in program histories, contexts and environments may be ignored in the analysis. Once results are obtained, their interpretation may be biased. Conclusions may be drawn based on the mere existence of a "statistically significant" difference in outcomes, neglecting the strength of that difference. Or the results of one evaluation may be inappropriately compared with those of a substantially different study, such as comparing net impact studies of JTPA with previous net impact studies of CETA.

Minimizing such biases is not easy. The researcher must be astute in identifying those that will impede the discovery of the truth in a particular evaluation, and develop ways to reduce them.

In this respect, the choice of a research design is critical. It is therefore important to recognize the wider spectrum of choice in terms of research designs, to place the debate into better perspective.

A research design is a plan which guides the gathering of reliable, valid information on the appropriate subjects of the evaluation, and directs the analysis of this information to determine causal or other kinds of relationships among the factors considered most important in understanding program implementation or impact. There are a number of alternative designs from which to choose. 4/ The traditional continuum distinguishes between experimental and nonexperimental designs, and nonexperimental designs are frequently differentiated as exploratory, descriptive and quasi-experimental based on their differing research purposes and their degree of rigor in reducing bias.

The usual purpose of exploratory designs is to formulate new ideas and develop and refine research questions for more rigorous study. Descriptive designs are intended to investigate associations or correlations among major explanatory influences and the outcomes of interest, short of attributing cause and effect. Quasi-experimental designs are expected to identify cause and effect relationships between interventions and outcomes, in the absence of experimental conditions and with the help of statistical controls. The purpose of experimental designs is to establish cause and effect, by both randomizing the intervention to the subjects of the evaluation and applying statistical controls.

It would be incorrect to assume that all designs other than the most rigorous are inappropriate or lack utility in policy research. Each type of design makes its own contribution to useful information about a social program. However, in net impact evaluations, the rigorous end of the continuum is involved due to the emphasis on determining whether or not the program "caused" the effects.

Because of the search for causal relationships and the intransigence of bias, the controversy over quasi-experimental and experimental designs has gradually narrowed to those alternatives considered most effective in studying cause-effect relationships in terms of reducing numerous sources and kinds of bias. That is, the debate is now largely confined to experimental designs vs. the most rigorous quasi-experimental designs.

Biases Common To Experimental and Rigorous Quasi-Experimental Designs.

There are some common barriers in making valid estimates of program impacts. Again no step in the research process is exempt. The best developed and implemented experimental and quasi-experimental designs cannot control for all the potential biases that interfere with unveiling the truth.

4/ See Supplement C for a more detailed overview of this continuum.

For example, an evaluator cannot rule out, with either design, those biases due to poor conceptualization of the issues to be studied or their inadequate measurement. If a classic theory useful in studying employment and training programs in the past is inappropriately applied to a contemporary program which is quite different, important factors may remain unmeasured and unstudied. Neither can an evaluator prevent through either kind of design the effects the personality or behavior of the evaluator may have on clients' outcomes.

Also, these designs cannot control for the effects of unanticipated external events which occur to clients, the program or the program's environment during the period of the evaluation, such as unexpected family crises, changes in program leadership or natural disasters. They cannot control for the impact of unknown or unmeasured program implementation policies or practices. Both designs leave uncontrolled those biases which are due to the combined effect of a number of influences on outcomes, such as the combined influence of a client's age, sex and education on postprogram earnings, unless these effects have been anticipated and these characteristics measured.

But it is important to acknowledge that with both designs evaluators can attempt to compensate statistically for common sources of bias by identifying and measuring as many of the factors which may bias results as possible, and adjusting for them in their analyses.

Major Differences Between Experimental and Quasi-Experimental Designs

Despite common problems, there are important differences between the two kinds of designs being debated. The differences are based on the extent to which these designs meet the scientific test in terms of reliability, validity, generalizability and replicability. Within that test, the degree to which they accurately establish cause and effect is at stake. To explain these differences we need to begin with the ideal experiment and work our way toward the realities policy research must confront.

The Ideal Experiment

The goal of the "ideal" experiment is to achieve maximum control over influences which prevent the evaluator from discerning the truth about the phenomenon being studied. The truth in net impact studies is the discovery of which influences contribute to specific program effects in a logical cause and effect series of relationships. In program evaluation a large number of influences may cause the outcomes measured. The evaluator seeks to eliminate factors which compete with the program's treatments in explaining those outcomes.

In the ideal experiment many of the influences other than the service interventions are controlled. That is, they are eliminated as significant influences on program outcomes. Using certain procedures the evaluator can isolate the separate effects of a number of treatments operating simultaneously or sequentially to affect a set of

outcomes. Through the random assignment of individuals to treated and nontreated groups, the experimenter can rule out the effect on outcomes of certain influences that are unknown, or known but unmeasured, thereby reducing bias.

The influences that compete with the program's treatments to explain outcomes are of several kinds: (1) factors which may be confused with or mistaken for the program's interventions--such as a major change in job opportunities in the local labor market; (2) factors which can be controlled by random assignment--such as measured characteristics of clients prior to their receiving services, for example their attitudes toward government assistance; and (3) factors which can be controlled through the use of statistical procedures--such as the selection of clients with particular characteristics to receive certain services, or the duration of the services assigned to them. The objective is to control for as many competing explanations as possible.

In the ideal experiment most competing explanations are controlled through randomization procedures alone, or randomization accompanied by statistical techniques. A well designed experiment permits the evaluator to make a minimum number of appropriate assumptions about such influences, exercise optimum control, and make causal inferences with greater confidence. In the evaluation of program impact, this means that the evaluator can come very close to determining the true effects of the program's interventions.

The key feature of experimental designs is randomization. This process sorts research subjects randomly into an "experimental" group and a "control" group prior to the introduction of the "experimental treatment". Translating this principle for JTPA, it involves randomizing eligible enrollees into a group which receives JTPA services and one from which these services are withheld, before the treated group receives any services.

In the ideal experiment the research design may also involve matching prior to randomization. This process gives the evaluator precise control over particular influences which are known or thought to affect outcomes. Subsequent randomization controls for many of the influences that have not been anticipated.

As an example of matching, assume the experimenter wants to study the differences between the outcomes of male and female JTPA enrollees, and has good reason to believe that age and ethnicity strongly affect those outcomes. The evaluator also believes that attitudes toward work will exert a substantial influence on their employment, earnings and welfare outcomes. Age and ethnicity can be more accurately measured than attitudes, so controlling for their effects poses different problems. Therefore individuals in the two groups can first be matched on the basis of age and ethnicity, which can be assumed to control for these two influences precisely. Subsequent randomization of each of these matched groups into those receiving the program's services and those not, can be assumed to cancel out most of the effects of other factors operating prior to program participation--

such as the attitudes of these individuals (work motivation, for example), and the impact of unanticipated events (such as economic recession) for which controls are not available.

Whereas randomization relies on chance to cancel out the effects of competing influences, matching gives the evaluator more direct control over certain characteristics. Used together, matching and randomization allow the researcher to more closely approximate equivalency between the two groups being compared. Differences in the outcomes of these two groups can then be taken seriously in determining a program's net impact. 5/

An additional element of the ideal experiment is the measurement of the experimental and control groups with respect to critical variables both before and after the introduction of the program intervention to the experimental group. In complying with this principle, the evaluator must assume that the measurement process is not itself affecting outcomes. Given that it is not, preprogram measurement can insure a baseline against which postprogram follow-up information can be compared.

But despite matching, randomization, pre/postprogram measurement and the use of statistical controls, the ideal experiment is not infallible. There are many kinds of interactive effects occurring between the program's interventions and competing explanations, which may jointly affect outcomes. These joint effects are often more complicated than simply adding together the separate effects of each influence. For example, the joint effects of age, sex and the program treatment on outcomes may be different from the separate effects of each of these influences. Even the most rigorous experimental design enables the evaluator to study only a limited number of these interactive effects. Also, the evaluator's assumption that treated and nontreated groups are equivalent may hold true only for the average effects of a program's interventions, or its effects on the "typical" client, and only within the laws of probability--that is, the larger the size of the experimental and control groups, the more confidence one can have in their equivalency.

So, try as we might, certain uncontrolled influences may remain which can obscure or confuse our effort to measure the true effects of program interventions. This simply reveals the natural limitations of the human species and the complexity of reality. Nevertheless, the ideal experiment is the closest we can come to estimating the true impacts of the programs designed to carry out important social policies.

5/ If the experimenter must work with small samples, matching prior to randomization is essential. But this requires the evaluator to accurately identify ahead of time those characteristics of the study group likely to be most strongly related to their outcomes. However, if the evaluator can use large samples, or if a large number of variables can be studied, randomization alone is an effective form of control.

Less Than Ideal Experiments

In actual practice few experiments turn out to be ideal in all respects. As the debate has been narrowed to judgments about rigorous quasi-experimental vs. experimental approaches, there has been a tendency to assume that all quasi-experimental designs are alike, and all experimental approaches the same. In practice, the continuum of designs that fall within each of these two categories can vary considerably in terms of their conformance to or departure from the ideal experiment.

For political, organizational and economic reasons, ideal experimental design principles and procedures cannot always be adhered to consistently in experiments involving ongoing programs. Neither can a meticulously designed quasi-experimental approach always be consistently implemented in real-life program environments. A flawed experiment, just as a seriously compromised quasi-experimental approach, can yield inaccurate information. If there is reason to believe that a field experiment's random assignment procedures will be violated, for example, a well planned and implemented rigorous quasi-experiment is preferable.

Rigorous Quasi-Experimental Designs

It is evident that the evaluation of net impact is more complicated when prior randomization is not feasible or practical. The evaluator must make additional assumptions which may be less plausible, and which may pose a greater potential for bias. The inability to randomize, and therefore to control for most other explanations of outcomes, is the primary problem confronted in using nonexperimental designs. But this problem is most troublesome when the evaluator desires to make causal inferences about a phenomenon which involves multiple interrelated causal variables, and complex relationships between these interventions and their effects. This is the reality faced in studying social programs such as JTPA.

The major task under these circumstances is to identify as many of the competing explanatory factors as possible, measure them and control for them through statistical means. A range of optional designs has been developed which addresses the challenges to reliability, validity and generalizability presented by an absence of randomization, some of which are ingenious adaptations of experimental designs. In these approaches, the statistical frameworks and analysis strategies of experimental designs are used, but not randomization. An example is the design recommended in this set of volumes for studying net impact. It involves pre/post measurement, a carefully constructed comparison group, and the use of statistical methods for identifying and adjusting for biases.

Tradeoffs Between Quasi-Experimental and Experimental Designs

As in the art of politics, research designs involve compromises between the ideal and the possible. These compromises force us to consider tradeoffs with respect to three significant attributes of research: its realism, its scientific rigor, and its representativeness. 6/ Realism is a characteristic of how well the evaluator can identify, select and measure the influences and effects that occur in actual programs in practical situations. Rigor is a characteristic of how well the evaluator can control for bias. Representativeness is a characteristic of how well the evaluator can select the study sample so that the results of the evaluation can be confidently generalized to the intended target population of a program. Severe compromises must often be made among these three qualities.

In general, the primary tradeoff in using well developed and carefully implemented experimental designs is greater rigor. With respect to well planned rigorous quasi-experimental approaches, it is greater realism. Systematic descriptive studies usually afford greater representativeness. Those designs that best combine these positive tradeoffs produce the most replicable evaluations.

Advantages and Disadvantages of Quasi-Experimental Designs

Rigor: As indicated previously, the major concern in using this kind of design to accurately determine the nature of cause and effect relationships is the lack of full control over other explanatory variables which may be confounded with the program's treatments.

Because political, organizational and often ethical constraints discourage the treatment of ongoing programs as experiments, bias estimation and adjustment methods must do double-duty. Even so, competing influences cannot be as efficiently controlled as in a well conceived experiment. Therefore it is imperative that the evaluator be aware of specific variables the research design fails to control. This alerts the researcher both to competing explanations and to potential biases which must be adjusted for through alternative methods.

In constructing a comparison group, the evaluator must use care, caution and imaginativeness in locating the most comparable individuals and the most reliable sources of essential information if bias is to be reduced. Statistical techniques can resolve certain

---

6/ In his article on "Representation, Randomization, and Control", Leslie Kish introduces these terms in discussing design issues. The discussion here reflects some of the positions taken in this article.

control problems by allowing the evaluator to adjust for a large number of competing explanations, although assumptions must be made about how accurate the results of these adjustments may be. But identifying and compensating for all potential influences is beyond the function or capacity of statistical methods. However, it must again be understood that controlling for all intervening variables is a goal which can only be approached in any research.

Realism: One of the advantages of a rigorous quasi-experimental design is realism. The evaluator has greater freedom than with experimental designs to identify, measure and study the broader range of variables and relationships which may more accurately mirror the true complexity of the program. For example, the evaluator can study the complexity of the service treatments as they are assigned by staff and experienced by clients in actual practice. However, such flexibility is limited by sample size.

Representativeness: If quasi-experimental designs are carefully developed there is also the opportunity to achieve greater representativeness than in experimental studies, in terms of revealing the true variation occurring within typical program populations across diverse state and local sites--even though this advantage is again limited by available sample size. At the state and local level where experimentation is less acceptable, for instance, an experiment's study population is frequently limited to that portion of the target group for which an acceptable political or organizational rationale can be provided for randomizing services. The results of such experiments are not likely to adequately represent the target group of the program.

In summary, rigorous quasi-experimental designs are generally weaker on control and stronger on realism and representativesness than experimental designs. They provide more flexibility in resolving conceptual and measurement problems, even though they leave certain research design problems only partially resolved. Because their study populations can better represent a program's intended target group, their findings tend to be more generalizable than those involving an experimental design, even though it may be easier to replicate an experiment because of its economy of scale and scope.

Advantages and Disadvantages of Experimental Designs

Rigor: Experiments are clearly stronger on control, since both randomization and statistical controls are available to reduce bias. Therefore experiments offer a greater likelihood of isolating the effects of a program's interventions.

Realism: Sometimes experiments are poorly conceived in the sense that they may be based on theories which suggest inappropriate explanatory frameworks, given the issues to be studied. The wrong variables and relationships may be emphasized, or the less important rather than the more significant ones. Although this is a problem in all research, experiments are particularly vulnerable because fewer manipulations

be made and fewer variables studied. As a practical requirement, the evaluator must simplify the reality of the program in order to exercise maximum control. Pragmatic and ethical constraints may also place limits on the scope of experimental research. The evaluator must therefore be as correct as possible in identifying and defining the most critical variables and relationships to be studied. Also, because experimental designs must of necessity limit the variety of treatments studied, the variation among the explanatory variables is often sparser than in the real world.

Representativeness: Experiments are most often carried out in a small number of nonrandomly selected sites across the country. These experimental sites tend to involve a population that is less diverse than the overall target population of a program. Such sites frequently have to be limited to those willing to randomize. They often over or underrepresent certain subgroups within the larger program population. Also, for political or organizational reasons experiments may have to confine the study group to a subset of the preferred population, such as to client volunteers, clients provided an incentive for participation, those willing to waive their legal right to privacy, or some other atypical subgroup. This will produce nonrepresentative study samples to varying degrees.

In summary, experimental designs are generally stronger on control and weaker on realism and representativeness than rigorous quasi-experimental designs. They are less flexible in terms of resolving conceptualization and measurement problems. It is difficult to design experiments which represent the larger target population of a program. In addition, contriving the desired realism of a natural program setting is not as feasible in experiments. Therefore, in many cases the findings are less generalizable than those yielded by rigorous quasi-experimental studies.

In conclusion, it is important to recognize that this treatment of some of the tradeoffs between experimental and rigorous non-experimental approaches in evaluating social programs such as JTPA assumes that in practice most experiments can only approximate the "ideal" experiment, and that even the most rigorous quasi-experimental studies will vary in the extent to which they can compensate for the absence of randomization. The main purpose of this brief overview of strengths and weaknesses will be well served if the reader has become increasingly convinced that the debate is more complicated than it is often presented and will continue to lack a definitive winner.

ORGANIZATIONAL AND POLITICAL TRADEOFFS IN THE DEBATE

It is not enough to talk only about the scientific tradeoffs in this lively debate. Organizational and political risks and benefits attach to one or another scientific approach, and can have a major impact on research choices. Just as certain research goals and qualities shape scientific decisions, broad acceptability and immediate utility for

policy, planning and service delivery purposes mold organizational and political decisions about whether to initiate an evaluation and what type of evaluation to perform.

In many program evaluation situations, particularly in ongoing programs as compared with pilot and demonstration projects, experiments are simply not feasible. This is not unique to social programs. It occurs in the physical and biological sciences as well. In organizations, as in politics, practical significance means more than statistical significance. While evaluators have devoted themselves mainly to untangling the scientific tradeoffs, program practitioners and those in politics have preferred to play audience to this Olympic spectacle. But researchers, politicans and program operators alike are now keenly aware of the significant role organizational and political factors play in the debate, and the necessity of considering the characteristics of organizations and the nature of the political arena in tailoring evaluation decisions to pragmatic contexts. The most basic supports for making scientific assessments and using their results depend on this accommodation.

Political Tradeoffs

The larger environment of evaluation is political: the world of elected and appointed officials, special interest groups, and the powerful individuals, groups and community coalitions that support them. Experimental research can be justified to the American public--and therefore to these constituencies--better than before. The public is more sophisticated about evaluation research, and more interested in being given empirical evidence that there is a satisfactory return on their investment. The intrusiveness of experiments, such as dictating a particular form of client selection and service assignment, can be supported as an efficient way to test a new concept or approach before substantial public money is allocated to it. This is especially true if an experiment has a low political profile because only a small proportion of a program's population will be affected, and only a few sites will be involved. This justification may also satisfy the political agendas of state training councils and PICs, who in turn are responsive to pressures generated in the larger political environment.

Despite the public's greater knowledge and approval of scientific options in testing innovations, politicians often feel they cannot rationalize a withholding of services from part of the eligible population of established public programs. Clients themselves play a role here. A high threshold of client resentment about service assignments which occur irrespective of a program's intended appraisal practices, or the impression of client exploitation for research purposes, will jeopardize a program's funding and the sponsoring agency's organizational safety. Also, as a prime political forum, the media are unusually sensitive to inequities. In an open society, inequalities of service provision and possible civil liberties violations are easily identified and exposed. The most visible target

is randomization. Even supporting the use of "those waiting for service" as a comparison group in quasi-experimental studies is not without risk, since this publicizes the existence of incomplete service coverage and invites attention from active client advocacy and legal services lobbies.

These are not the only political problems that must be addressed. There is the issue of the consistency of a program's goals, and its goals vis-a-vis its means for achieving them. Evaluation is perversely affected by program legislation that results from difficult political compromises. Political beliefs define what social needs are to be reduced through the use of available resources. Political rationales therefore underly conceptions of what a social problem is and is not, and direct the choices made among a range of program alternatives for resolving a problem. These rationales and directives are not always internally consistent. If there are competing rationales which must be welded by political compromise, the resulting legislation may be rife with inconsistencies.

Nevertheless, the evaluator is obligated to make judgments about the extent to which programs meet their legislative intent efficiently and effectively. For example, studies of program implementation must directly address the goals and means issue. However, what some political actors may most want in an evaluation is a confirmation of one or another of the competing positions which spawned the compromise. This can define a different purpose for the evaluation, and create disparate public expectations for the findings. In the end, each constituency may perceive different advantages and disadvantages to be associated with any evaluation, or with experimental vs. quasi-experimental studies, depending on what they view as most likely to produce findings favorable to a particular set of interests or political philosophy. When this happens, the luster of the scientific aspects of the debate may give way to the brilliance of its political tradeoffs.

Organizational Tradeoffs

Politics affects organizations. But evaluation decisions pose risks and benefits to organizations as well. These are largely a function of the natural "bureaucratization" of large scale organizations and the sensitivity of government agencies to pressures from their funders and immediate environments. Some of the characteristics of these bureaucracies support and some hinder a full consideration of the debate's scientific tradeoffs. Evaluating programs is a risky as well as potentially beneficial exercise for them. How programs are to be evaluated often increases the organizational threats more than it enhances the informational benefits of research.

All organizations must solve certain basic problems: problems of dividing work, bringing activities together to achieve goals, motivating members to participate through rewards and sanctions, maintaining cohesiveness, and controlling conflict. Resolving these problems stabilizes organizations and clarifies decision-making

authority within them. It enforces accountability and provides checks on the accummulation of power at the top. Many of the so-called "evils" of bureaucratic development, such as formal hierarchies, formal rules and regulations, and the specialization of tasks are intended to insure equity and efficiency. Nevertheless serious constraints may be imposed on policy research by bureaucratic characteristics such as the tendency to replace formal agency missions with organizational goals, to engage in power struggles with "competing" organizations rather than to cooperate, to encourage rigidity and territoriality as the arbiter of the division of labor and decision-making process, and to resist innovation and change.

The last characteristic is most central to the debate. Evaluation implies change. Improvement is its purpose. Yet agency administrators and program operators understandably judge evaluation and evaluation methodologies in terms of bureaucratic risks and benefits--i.e. what is most conducive to the internal maintenance of the organization and to its security outside. These practitioners genuinely appreciate the need to acquire information on a regular basis to judge the value of the programs for which they are responsible. For this purpose they need timely, useful information geared to their administrative and managerial needs and obligations. But information that is likely to suggest that their programs are far less effective than they should be, or possibly ineffective, is information that entails serious risks for an organization's leadership. At the worst, these risks are starkly economic: budgets will be cut. Therefore the decision to evaluate is nearly always a courageous one, and the choice of evaluation methods within that decision must be weighed against many other considerations.

Given that the decision to evaluate has been made, randomization is again the main issue in the debate over the use of quasi-experimental vs. experimental designs. Managers properly view service assignment decisions as their responsibility. Their service delivery staffs derive motivation and reward from the communication of this function downward.

Randomization represents a major organizational intervention. This is particularly true in JTPA, where the careful sorting of clients into treatments appropriate to their individual needs and resources is a legislative mandate.

Historically, randomization procedures have often been subverted for this reason, resulting in unknown and unmeasured biases. Random assignment must be presented such that it is viewed by staff as both justifiable and profitable. The price paid for gaining this kind of acceptance is a reorientation of staff attitudes and behavior But this can reduce the realism and generalizability of the experiment. It would be difficult to apply the findings to programs in which the intended staff identification with selective assignment practices has remained intact.

In summary, organizational and political tradeoffs are a compelling influence in the debate. The intertwining of these considerations with scientific concerns amplifies the difficulties involved in choosing any one "ideal" research design for a particular piece of policy research.

CONCLUSIONS

To a greater extent than we might wish, the beauty of these tradeoffs is in the eyes of the beholder. In terms of scientific exchanges, there are numerous criteria for judgment: reliability, validity, generalizability and replicability--and rigor, realism and representativeness. And certainly more. With respect to organizational and political tradeoffs, feasibility and utility are the significant criteria, and again there are many more. There is no one absolute or universal criterion for determining the best research design among these. One must compromise, choosing the best research strategy to fit resources and circumstances.

Understanding some of the advantages and disadvantages of different approaches should lead to better choices, and each kind of approach can be improved by efforts to compensate for its specific weaknesses. The complexity of choices suggests contributions both evaluators and practitioners in JTPA can make to the debate. Evaluators can become more knowledgeable about the supports, problems, and unanticipated possibilities in the research environment:

> "The social scientist who desires to undertake policy studies, to be an agent of social change and contribute constructively to the solution and amelioration of the myriad of problems in today's communities must recognize that more than technical knowledge and craftsmanship is involved. The way he conducts himself, the types of interpersonal relationships he develops, and the attention he pays to various administrative and organizational tasks are equally crucial to the successful fulfillment of his role." 7/

Practitioners can become more knowledgeable about the scientific tradeoffs in terms of the value of the information they provide:

> "Troubles with experimental controls misled even the great Pavlov into believing temporarily that he had proof of the inheritance of an acquired ability to learn: in an informal statement...Pavlov explained that in checking up on his

---

7/ See Freeman and Sherwood. Social Research and Social Policy, page 142.

> experiments he found that the apparent improvement in the ability to learn on the part of successive generations of mice was really due to an improvement in the ability to teach on the part of the experimenter." 8/

And finally, working together, evaluators and program practitioners can consider the circumstances under which quasi-experimental designs are more appropriate than experimental designs, and vice versa. Anticipating the conclusions of such a joint effort, it is probably safe to assume that in evaluating ongoing programs such as JTPA at the state and local level, the scientific, organizational, and political tradeoffs point to the use of a rigorous quasi-experimental design for determining the net impact of the program. The likely consensus with respect to the use of experimental designs is that, whenever possible, they should be used in evaluating one-time or limited pilot and demonstration projects at the state and local level, which involve a small proportion of the client population of JTPA, and whose specific purpose is to test the effectiveness, feasibility, and utility of various innovations in program implementation to achieve desired changes in client treatments and outcomes: new combinations of services or new service targeting strategies; different program emphases with respect to particular subpopulations of clients; new state or local program policies or guidelines; or other ideas designed to improve state and local training policies and programs. In a very different context, at the national level, it is expected that the tradeoffs will more likely support experimentation.

An important concern, in any case, is to obtain cross-national, cross-state, and cross-SDA comparability with respect to the variables and relationships selected for study, the way these are measured, the manner in which the characteristics of different methodologies are defined, and the extent to which the information obtained can be generalized beyond a single SDA, state, or cluster of experimental national sites.

If such joint efforts, and such information comparability and utility can be adopted as the overriding goals, the debate can pleasantly continue: more as a stress reduction exercise than as a potentially divisive influence in policy research.

8/ See Kish. "Representation, Randomization, and Control", page 269.

REFERENCES

1. Beardsley, P. L. Redefining Rigor: Ideology and Statistics in Political Inquiry. Beverly Hills, CA: Sage Publications, Inc. 1980.

2. Blalock, A. B. and H. M. Blalock, Jr. Introduction to Social Research. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1982.

3. Blalock, H. M. Jr. Causal Inferences in Nonexperimental Research. Chapel Hill, N.C.: University of North Carolina Press. 1964.

4. Blau, P. M. Bureaucracy in Modern Society. New York, N.Y.: Random House. 1956.

5. Campbell, D. T., and J. C. Stanley. Experimental and Quasi-Experimental Designs for Research. Chicago, IL: Rand McNally Company. 1966.

6. Campbell, D. T. "Reforms as Experiments" in American Psychologist. Volume XXIV. No. 4. 1969.

7. Cook, T. D. and Campbell, T. D. Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicago, IL: Rand McNally. 1979.

8. Etzioni, A. A. Modern Organizations. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1965.

9. Freeman, H. W. and C. C. Sherwood. Social Research and Social Policy. Englewood Cliffs, N.J.: Prentice-Hall, Inc. 1970.

10. Kish, L. "Representation, Randomization, and Control" in H. M. Blalock, Jr. (editor) Quantitative Sociology in the Social Sciences. New York, N.Y.: Academic Press. 1975.

11. McLaughlin, M. W. "Implementation Realities and Evaluation Design" in A. L. Shotland and M. M. Mark (editors). Social Science and Social Policy. Beverly Hills, CA: Sage Publications, Inc. 1985.

12. Meyer, J. W. and W. R. Scott. Organizational Environments. Beverly Hills, CA: Sage Publications. 1983.

13. Perrow, C. Complex Organizations. Glenview, IL: Scott, Foresman and Company. 1973.

14. Ross, J. and P. Smith. "Orthodox Experimental Designs" in H. M. and A. B. Blalock (editors). Methodology in Social Research. New York, N.Y.: McGraw-Hill Book Company. 1968.

15. Rossi, P. H. and H. E. Freeman. Evaluation: A Systematic Approach. London, England: Sage Publications. 1985.

SUPPLEMENT A

DEFINITION OF CONCEPTS

CONCEPTS	DEFINITIONS	EXAMPLES
Characteristics of the goals of evaluation:		
RELIABILITY	Reliability is related to the accuracy of measurement and the faithfulness of the data collection effort to the research design. If measures of a given variable do not reliably represent the variable as it was defined, and if these measures are not collected in a consistent manner, the information available for analysis will not be reliable.	In collecting data on JTPA services provided to clients, the evaluator must be assured that information on each desired characteristic of the treatments--the mix, sequence, length, etc.--has been gathered in a consistent way and accurately recorded for each research subject. If data collection guidelines are violated, the information will be of low reliability.
VALIDITY	The interpretation of evaluation information is valid to the extent that the data which were analyzed to provide conclusions accurately represent the definition of the variables studied. Also, the definition of the variables must reflect their true qualities. Validity is also a function of the extent to which other sources of potential error and bias have been reduced. The research framework and design may be well chosen and specified, and the data collected reliably, but the results may still reflect uncontrolled biases that reduce the credibility of the findings.	In studying the gross impact of JTPA, using an appropriate research design and a unique emphasis on a rich array of outcomes, the findings may suggest that women benefit from classroom training more than men, particularly with respect to increased motivation to pursue a career path. "Career path" is difficult to measure in quantitative terms. If measurement error (deviation from the true qualities of a variable) is substantial with respect to outcomes such as these important ones, there will be a lesser degree of validity.
GENERALIZABILITY	An evaluation may be carefully implemented but its results may not be applicable beyond a limited segment of the intended subjects of the research--or beyond the particular kind of context or environment in which the evaluation was carried out.	The use of an experimental design which uses JTPA volunteers as the study group, and which of necessity interferes with the service assignment process intended in the program, will yield results which cannot be generalized beyond an atypical group of JTPA clients in a program implemented significantly differently than intended.
REPLICABILITY	This trait involves the ability to repeat a particular evaluation (its framework and methodology) at a different point in time and in another kind of setting, using similar research subjects--in order to determine the durability of the original findings and to further check on their validity. If replication is not possible, we can never be certain that our results will hold up when the evaluation is repeated elsewhere. Our inferences just might be an artifact of where and under what circumstances we carried out a particular evaluation.	If an evaluation of the net impact of JTPA on minority vs. nonminority enrollees in Chicago has utilized an experimental design, and yielded results that suggest that on-the-job training produces better outcomes for minorities, it will increase the credibility of these findings if a replication of the analysis in Seattle produces similar results. A successful subsequent replication in New York City will further "validate" these conclusions.

CONCEPTS	DEFINITIONS	EXAMPLES
Characteristics of the evaluation approach, or framework:		
EVALUATION ISSUES	Evaluation issues are the research questions to be studied in an evaluation. They convey expectations about the kinds of relationships that exist between certain factors, or variables, which describe these issues. These questions can take the form of hypotheses to be tested, propositions to be investigated, or simply ideas to be explored. In hypothesis testing, and in many investigations of formal propositions, the evaluator focuses on CAUSAL relationships between various influences and effects. In other evaluations, the emphasis may be more on what kinds of ASSOCIATIONS exist between such variables, i.e., how they may be correlated with one another.	In JTPA we are interested in two major sets of evaluation issues: 1) how JTPA services and other formal interventions affect, or are correlated with a number of desired outcomes, and 2) how the unique features of JTPA program implementation affect, or are associated with those outcomes.
MEASUREMENT	Measurement is a process which progressively defines the factors or variables being interrelated in studying evaluation issues. This process begins with what are referred to as "operational" definitions and moves toward one or more specific measures, or indices which "stand for" these variables. The data elements describing variables in MISs are examples of such indices. The accuracy of measurement refers to how well the selected measures represent the influence or effect to be studied.	In JTPA, critical measurement tasks involve precisely defining the characteristics of: 1) the program "treatments", or unique interventions the program provides, and the "outcomes" expected to take place as a result of these interventions; 2) the characteristics of a program's "implementation" which may--in addition to its interventions--influence outcomes; 3) the characteristics of the program "environment" which can effect outcomes, and 4) the characteristics of potential "unintended treatments" which can explain outcomes.
Characteristics of the Evaluation methodology:		
APPROPRIATENESS	This refers to how feasible and efficient the research design is, i.e., the guide the evaluator selected for sampling research subjects, and collecting and analyzing information about them. Feasibility is related to how well the design fits the research situation; efficiency has to do with the degree to which it controls for unknown or unmeasured influences that may explain what is observed just as well or better than the primary ones being studied.	In JTPA the use of an experimental design is inappropriate for studying the influence of program implementation. This is due to a variety of political, organizational and research design problems. For instance, to study the effect of a single characteristic of implementation, one would have to randomly assign that characteristic across a large number of SDAs or project sites. It is obviously not possible to do this.

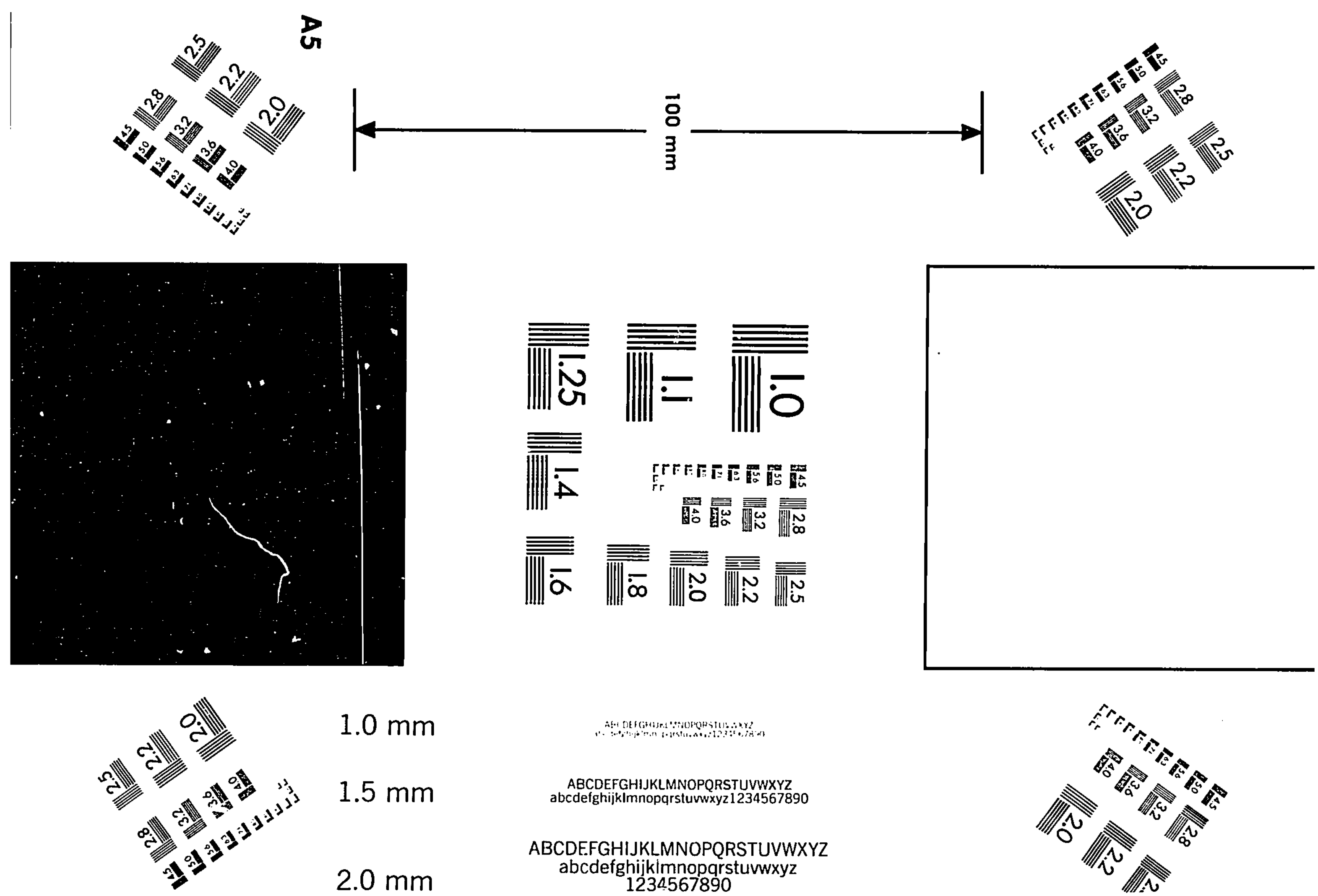
CONCEPTS	DEFINITIONS	EXAMPLES
RIGOR	The rigor of a research design is the extent to which the design yields accurate explanations of the relationships hypothesized, proposed, or expected.	The "ideal" experimental design is considered to be the most rigorous research design. However, rigorous non-experimental designs also involve a high degree of control in research situations where experimental designs are considered inappropriate.
CONTROL	Control involves procedures which reduce the effect of variables that "compete with" program treatments in explaining expected relationships, thereby permitting the evaluator to concentrate on the direct link between the program's intended treatments and the observed effects. Controlling for these extraneous factors, or those whose influence the evaluator would like to eliminate, is a necessary element of an adequate research design. The purpose of controls is to reduce error and bias which can invalidate evaluation findings.	Some of the factors which need to be controlled in evaluating JTPA implementation or impact are the demographic, economic, and labor market characteristics of the program's environment which change over time.

SUPPLEMENT B

SOME SOURCES AND KINDS OF BIAS

SOURCE	KIND	ELABORATION
Those associated with being part of an evaluation.	Client's attitudes and behavior, related to their participation in a special project which affords them a different kind of assistance and attention from that received on the average by the typical program client.	Clients are immediately aware of being "studied", if only from rumors that they are part of an evaluation, or through more frequent or more detailed follow-up surveys than other clients experience. This "test effect" can produce an an impact of its own, which is difficult to separate from the effects of the services clients receive. This is known as the "Hawthorne Effect".
	Staffs' attitudes and behavior, related to being part of an evaluation which involves compliance with policies and procedures that may not be the same as for other staff, or which involves being studied themselves.	Staff tend to behave differently when they know they are being observed or directly studied. They receive greater attention, which may be rewarding, but they are often more closely monitored, which can be frustrating. This influences the very process that is being studied--particularly in evaluations of program implementation. These influences are difficult to measure and hard to isolate.
	Retesting of clients or staff, which involves their being studied more than once over a period of time	If measurements are made at regular points in time over an extended period, and the client or staffperson is aware that data are being collected, the effects of the first data gathering can affect the information collected at subsequent points. The individual becomes familiar with the approach, the data collection interview or questionnaire, and the context of the evalutaion. This affects responses and may bias the information being collected.
	The way the evaluation is conducted, in terms of the evaluator's interaction with staff, clients, employers, and relevant others.	Anything that affects the environment of the evaluation can be confused with the services in explaining client or employer outcomes. Again it is difficult to quantify these nuances, and separate them out from the more direct relationship between services and their effects.
	The evaluator's own personality and behavior.	This may support or constrain "normal" behavior on the part of those being studied, posing a possible bias. If the evaluator's attitudes and general style are consistent, at least the bias is a systematic one, and may be easier to define.
	The influence of interventions introduced by an evaluator's funders.	Some funders are intrusive, requiring frequent on-site monitoring of an evaluation effort as well as continuing consultation. This can introduce biases that attach to the attidues and behavior of the grant officer.

SOURCE	KIND	ELABORATION
Those associated with the "explanatory" framework within which the research issues are to be studied.	Inadequate selection and definition of the key issues, due to the use of an underdeveloped or erroneous knowledge base.	If the evaluator selects the wrong, or the less relevant variables and relationships to study, the evaluation will provide less than useful information. For example, in process evaluations organizational theory will be more appropriate than economic theory in selecting cogent issues to study.
	Use of an explanatory framework which relies insufficiently on the conclusions of previous competent research in that issue area.	If an evaluator carrying out a net impact evaluation of JTPA fails to study the relevant precious research on employment and trining programs over the past two decades, the evaluator may fail to identify the main issues to study. Particularly in the case of experimental or rigorous nonexperimental studies, which usually involve a small set of variables, it is critical that the previous literature be utilized.
	Selection of variables and relationships which are not the most critical ones, given the questions to be answered in the evaluation.	In process evaluations, certain organizational variables are more essential than others, given the distinctive features of JTPA.
Those associated with the measurement of key issues, influences, and effects.	Selection of definitions and measures which do not reliably represent these factors.	If the treatments, or JTPA services, are not defined precisely, the relationships between treatments and outcomes cannot be accurately evaluated.
	The use of too few measures for a variable, or too little "averaging out" of errors and biases in developing indices for variables.	The data elements which are selected to represent the variable may be insufficient in terms of affording multiple measurement of the variable. Multiple measures help to detect and compensate for this type of error, and to describe different aspects of a complex influence or outcome.
	Development of inadequate measuring procedures.	Sometimes the criteria the evaluator uses in developing definitions and indices fail to produce accurate measures.
	Insufficient compliance with adequate measuring procedures.	The evaluator may decide to use a number of indicators of employment as an outcome variable. The criteria for selecting them may be based on an adequate knowledge base and appropriate procedures, but as the evaluation progresses the use of the indicators may be inconsistent from one service contractor to the next.
	An absence of repeated measurements on the same study sample.	Although collecting data on subjects has its own effect, one-time-only measurement can produce biased information. In most impact evaluations, an ideal set of data collection points are: prior to receiving services; during program participation; at termination from the program; and at some point beyond termination.


A5
100 mm
1.0 mm
1.5 mm
2.0 mm
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz1234567890
ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz
1234567890

SOURCE	KIND	ELABORATION
	Lack of reliability among the different measures used to study a variable.	If the evaluator wants to study "work motivation" by collecting information on: 1) satisfaction with the workplace; 2) personal attitudes about the meaning of work quality; and 3) desire to "get ahead", these indices need to be highly correlated with one another. If not, they may be biased indicators.
	Inadequate identification and measurement of variables other than the program's interventions.	Measurement tasks are not limited to the primary variables of interest--i.e., to services and desired outcomes. They also involve the measurement of influences which must be controlled in the evaluation, such as labor market conditions in the community in which the program operates.
Those associated with the definition of the target group to which evaluation findings are to be generalized.	Failure to use a representative sample of the target group as the study population.	An experiment may use only volunteer applicants or enrollees, in order to defuse objections to randomization. However, volunteers are usually not representative of the larger target group of programs, and their different characteristics will bias results. Also, to encourage sufficient volunteers an evaluator may have to provide additional information and incentives, which further sets these volunteers apart from the regular target group and changes the nature of the program treatment.
Those associated with the development of a research design which guides the collection and analysis of information.	Failure to adequately control for the major "competing" explanations of outcomes, such as:	
	1. SELF SELECTION BIAS: due to differences in the characteristics of eligibles who apply for the program vs. those who do not.	Only a subset of eligible individuals apply for JTPA. These individuals may have characteristics that distinguish them from other eligibles, such as level of need, degree of motivation, acceptance of government assistance, etc. These potential differences can bias the results of an evaluation whose findings are intended to be generalizable to the full target population.
	2. STAFF SELECTION BIAS: due to a nonrandom selection of applicants into the program.	Staff attitudes and behavior have a great deal to do with which applicants are actually enrolled. This is an early source of bias in the program. "Creaming" is an example.
	3. FAILURE TO ADHERE TO RANDOMIZATION PROCEDURES IN EXPERIMENTAL STUDIES.	Although randomization of the study sample into treated and nontreated groups is an important aspect of experiments experimental principles are not always applied stringently. If staff are inattentive to randomization procedures, or deliberately resist them, the two groups will not be statistically equivalent on the average, and comparisons between them will be biased.

OURCE	KIND	ELABORATION
	4. INABILITY TO USE A COMPARISON GROUP IN NON-EXPERIMENTAL STUDIES.	Nonexperimental studies are used in situations where randomization is not possible or appropriate. However, extensive control over competing explanations of outcomes can be achieved through the construction of a comparison group which is as "equivalent" as possible to the treated group. The evaluator can use one set of clients as a control for another set in evaluating the differential impact of various services, and can also use statistical controls. However, the use of an adequate comparison group raises the level of control and reduces bias.
	5. INABILITY TO IDENTIFY AND MEASURE A COMPARISON GROUP THAT IS SUFFICIENTLY SIMILAR TO THE GROUP GIVEN PROGRAM SERVICES.	It is sometimes difficult to locate a group of similar individuals who have not received a comparable treatment. It is impossible to be certain that the two groups are, on the average, identical in the statistical sense. And it may be difficult to obtain the kinds of information needed, even when the comparison group of first choice is located.
	6. FAILURE TO TAKE INTO ACCOUNT THE COMPLEX INTERACTIVE EFFECTS THAT USUALLY OCCUR BETWEEN THE PROGRAM'S INTERVENTIONS AND OTHER INFLUENCES AFFECTING OUTCOMES.	Sometimes the evaluator focuses only on the average effects of the program's services, without measuring the extent to which interactions between services and other explanatory factors can jointly affect outcomes.
	7. BIASES DUE TO THE CHARACTERISTICS OF A PROGRAM'S IMPLEMENTATION OR OPERATION.	The evaluator must be able to tease out the effects of service delivery practices from the effects of JTPA services. A poorly managed project can obscure potentially positive program impacts.
	8. BIASES DUE TO THE CHARACTERISTICS OF A PROGRAM'S ENVIRONMENT.	The availability of certain kinds of training slots and jobs may greatly affect client outcomes. Labor market conditions vary across time and projects.
sociated with data on.	Failure to follow systematic data collection principles and procedures.	Data collection efforts need to be standardized to assure reliability.
	Failure to systematically track nonrespondents for follow-up purposes, and retain information on them that explains why they are non-respondents and what their past program statuses are.	Basing results on nonrepresentative follow-up samples introduces serious biases.
sociated with data	Inappropriate analysis methods, given the purpose of the research, the issues to be studied, the kind of design used, and the assumptions made by the evaluator.	Statistical methods vary with the purpose of the evaluation and can range from simple comparisons of means and proportions to complicated multivariate analyses.

SOURCE	KIND	ELABORATION
Those associated with the interpretation of evaluation results.	Limited issues and variables for study.	In both experimental and nonexperimental evaluations, the kinds of issues and influences studied constrain what can be learned. If not enough of the significant factors are studied, the results will have limited utility.
	Drawing conclusions from a comparison of inferences from one evaluation with those from noncomparable evaluations.	If an evaluator treats noncomparable evaluations AS IF they were replications of his/her research, the conclusions will not be valid or useful. For example, comparing evaluations of JTPA with evaluations of CETA are risky, since in some cases the differences in the findings are explainable on the basis of 1) differences in the study samples, 2) the issues addressed, 3) the research designs used, 4) the reliability and validity of the data, and 5) the characteristics of the research sites.
	Using the establishment of statistical significance between program interventions and outcomes as the major single criterion of a successful study, ignoring the STRENGTH of these relationships.	In an evaluation of the effect of job search assistance, the investigator may find that it has a different effect for men than for women which is statistically significant. But this significant difference may be so slight as to be irrelevant for policy purposes.
	Nonrepresentativeness of evaluation sites.	In both experimental and nonexperimental evaluations, the generalizability of the results may be reduced by the limitation of the studies to only selected sites, which are neither randomly selected nor selected with the specific purpose of representing the range of sites within a state or across the country. For example, the selection of sites in experimental studies may favor those in which administrators and program operators are willing to randomize clients, or to those which have successfully submitted unsolicited proposals to funders to study a set of issues of special interest to those particular funders. The use of such selection methods results in evaluations which are difficult to generalize beyond potentially atypical environments.
	"Fine-tuning" the results of nonexperimental studies of a particular program by referring to experimental findings on that program.	Such "calibration" is inapproprate, since experiments and nonexperiments involve quite different research conditions. Their sampling of program target groups differs, as do their data analysis methods. Program implementation and program environments are affected by them differently. Consequently they produce noncomparable information. However, general insights from informal assessments of the differences or similarities in their findings may be useful in identifying evaluation issues that require better or further study.

SUPPLEMENT C

A RESEARCH DESIGN CONTINUUM

TYPE	KIND	PURPOSE	EXAMPLE
Nonexperimental Designs	Exploratory Designs	TO EXPLORE: the main purpose is to formulate new ideas and develop and refine research questions for more rigorous study. Therefore, such designs demand that the evaluator become thoroughly immersed in the object of study and rely heavily on insight and intuition, since dependence on specific hypotheses or a relatively small set of significant variables is neither possible nor feasible. In fact, this is where one develops and refines possible hypotheses for further study. These designs are most appropriate when little information is available from previous research, or the evaluator wants an in-depth look at a complex phenomenon. Evaluations using exploratory designs range from unstructured to highly structured studies. The major advantage is flexibility.	Participant-observation or case studies are very useful in studying program implementation, such as staff attitudes and behavior. Such exploratory designs can contribute important insights about the experiences of clients in JTPA programs.
	Descriptive Designs	TO DESCRIBE: The purpose is to investigate associations among the major explanatory variables short of attributing cause-effect. These designs are well suited to issues for which there is already a body of information, but rigorous quasi-experimental or experimental approaches are not appropriate. These designs permit a range of data collection and analysis procedures, but are not as flexible as exploratory designs. They require careful planning to economize the research effort, protect against bias, and increase the data's reliability and validity. This demands a clear definition of what variables are to be measured and why, and what subjects will be sampled and how they will be selected. Again descriptive studies can vary from simple but well-organized descriptions of phenomena to rigorous systematic surveys. However, they do not control for bias as well as quasi-experimental designs.	Social surveys, such as structured systematic post-program follow-up surveys of JTPA clients and employers, are examples of this kind of design. Descriptive studies of JTPA gross outcomes also illustrate this kind of approach.

TYPE	KIND	PURPOSE	EXAMPLE
	Quasi-experimental Designs	TO ESTABLISH CAUSE-EFFECT: the purpose is to closely approximate the principles and methods of experimental designs, in the absence of randomizing clients into treated and untreated groups. These designs range from more to less rigorous alternatives. The most rigorous are effective in reducing sources of significant bias, such as the effects of major influences other than the treatments. However, this spectrum of designs involves various compromises between feasibility and "ideal" experimental design. In general, these designs require a valid theoretical model of why and how a program works so that cause and effect can be inferred.	Studies of the nature of associations or correlations between JTPA services and outcomes without the use of a comparison group, illustra such designs. More rigorous studies of causal relationships, using a constructed comparison group, are also examples of this approach.
Experimental Designs		TO ESTABLISH CAUSE-EFFECT: the purpose is to make optimal use of scientific method in providing the most accurate and valid information for testing cause-effect explanations of relationships between interventions and effects. These designs accomplish this through a combination of randomization, the use of a control group, and statistical controls for bias. At their best, experimental designs yield the most valid inferences. But experimental designs also range from more to less rigorous, as the evaluator seeks the "ideal".	Field experiments, such as are being recommende for studying JTPA at the national level, are examples. Evaluations of previous national pil and demonstration projects have often utilized such designs. Most evaluations of ongoing programs, however, have not used experimental approaches.