P/PV Briefs

Philanthropy and Outcomes:

Dilemmas in the Quest for Accountability

Gary Walker
Jean Baldwin Grossman

April 1999

Public/Private Ventures is a national nonprofit organization whose mission is to improve the effectiveness of social policies, programs and community initiatives, especially as they affect youth and young adults. In carrying out this mission, P/PV works with philanthropies, the public and business sectors, and nonprofit organizations.

We do our work in four basic ways:

- We develop or identify social policies, strategies and practices that promote individual economic success and citizenship, and stronger families and communities.

- We assess the effectiveness of these promising approaches and distill their critical elements and benchmarks, using rigorous field study and research methods.

- We mine evaluation results and implementation experiences for their policy and practice implications, and communicate the findings to public and private decision-makers, and to community leaders.

- We create and field test the building blocks—model policies, financing approaches, curricula and training materials, communication strategies and learning processes—that are necessary to implement effective approaches more broadly. We then work with leaders of the various sectors to implement these expansion tools, and to improve their usefulness.

P/PV's staff is composed of policy leaders in various fields; evaluators and researchers in disciplines ranging from economics to ethnography; and experienced practitioners from the nonprofit, public, business and philanthropic sectors.

P/PV *Briefs*

Philanthropy and Outcomes:

Dilemmas in the Quest for Accountability

Gary Walker
Jean Baldwin Grossman

April 1999

Philanthropy and Outcomes:
Dilemmas in the Quest for Accountability

"Effectiveness must become the principal criterion for givers of time and money." This clarion declaration is the first of five conclusions of the 1997 report of The National Commission on Philanthropy and Civic Renewal, *Giving Better, Giving Smarter.*[1] The following year, the United Way's Resource Network stated, "In growing numbers, service providers, governments, other funders and the public are calling for clearer evidence that the resources they expend actually produce benefits for people…Many foundations now require programs they fund to measure and report on outcomes."[2] The call for a greater focus on outcomes in philanthropic giving has gained increasing prominence and adherents during the 1990s.

Grantees report that never before have grant negotiations with foundation staffs been so focused on specifying outcomes. Some foundations have employed consultants to work with their staffs so that inputs, operational processes, and intended intermediate and long-term outcomes and impacts are specified and differentiated. A number have added evaluation departments to their organizational structure. Small and medium sized foundations, which have previously given exclusively to direct services, are now asking for and funding evaluations, so that they may know with objectivity and rigor if the projected outcomes are achieved.

The national office of United Way—whose local chapters raised $3.2 billion in 1997—initiated several years ago a major project to both emphasize the importance of specifying outcomes for local giving, and to provide assistance to local chapters on how to go about determining the outcomes of individual grants. This effort, funded by the W.K. Kellogg Foundation and the Ewing Marion Kauffman Foundation, resulted in *Measuring Program Outcomes: A Practical Approach and Focusing on Outcomes.* As of the beginning of 1997, 22,000 manuals have been sold to agencies who are being asked by United Way and others to provide measures of effectiveness in their grant applications.

At first blush this "outcomes movement" seems like an unreservedly good thing. Who can be against more careful specification of what grants are intended to achieve, and more rigorous and objective efforts to assess exactly what they do achieve? This is especially true in a sector that is often criticized for its overemphasis on personal relationships and ideological correctness (in whatever direction) in its giving practices, and for its lack of rigor and public openness in assessing and communicating the effectiveness of its giving. Especially in a country whose dominant sector—the private for-profit sector that creates the wealth that fuels the philanthropic sector—provides on a regular basis an objective, detailed and public accounting of most of its companies' outcomes and performance, and where all of that sector's units prosper or wither by that impersonal accounting.

In that light, the "outcomes movement" is a welcome coming of age for philanthropy, a voluntary descent from its lofty perch on the mountain top to the life of Everyman in the valley below, where the esteem one is held in depends largely on what one achieves, as measured by the cold outcomes of the particular marketplace. You can almost feel the keystone of accountability grind into its long-awaited place.

There is a healthy measure of truth to all of the above. And if the alternative is not caring about outcomes, or not caring about the reliability of how we assess them—not caring about accountability—then the recent emphasis on outcomes is an unreservedly good thing. But, in fact, a more complex reality underlies the current "outcomes movement."

The Context

The overwhelming majority of projects, organizations, initiatives and programs supported by philanthropy are not formally assessed for outcome achievement. Many do not even generate basic descriptive information about the content, quantity and quality of what they do, much less assess what they accomplish. Anecdote, salesmanship, ideology and relationships are as much the basis for many philanthropic funding decisions as is a hard analysis of outcomes. As noted above, it is this feature of philanthropic practice that makes the outcome movement so appealing.

But it is historically inaccurate to see the current "outcomes movement" as a major innovation in the philanthropic sector. The interest in outcomes is as old as the interest in giving. In fact, most philanthropists are quick to declare their intended outcomes, for their very purpose in giving is to *cause change*. The word "outcome" may represent more stylish jargon, but the meaning behind the word—change for the better—has always been there.

Neither is it the case that the "outcomes movement" brings a new focus on measurement, nor on new measurement techniques. Many of the largest and most influential of philanthropic institutions have been deeply engaged in specifying and measuring the outcomes of their giving for several decades now. The Ford Foundation actually created organizations such as Public/Private Ventures and the Manpower Demonstration Research Corporation in the mid-1970s in order to obtain, with the most sophisticated evaluation techniques available, reliable information about the effectiveness of social initiatives in many areas—welfare, employment and training, public housing, transportation, education and youth development, to name but several major "outcome areas." Evaluations of the initiatives cited above have been supported not just by Ford, but by over 100 foundations, including over three-quarters of the country's 25 largest foundations. They have employed random assignment, econometric models, ethnography, qualitative analysis, political science and sociology, among many disciplines and methodologies, both separately and together.

Is then the current "outcomes movement" simply an attempt to extend the practices and knowledge of this substantial group of foundations to all the rest—and to help ensure that recent and projected growth in the number of foundations and in aggregate philanthropic wealth will also include an appropriate emphasis on outcomes specification and measurement, and a sophistication based on previous experience? Certainly that is a worthy goal.

But there is more to the current emphasis on outcomes—and the history behind it—than that some practice it, and many do not, and that the word and the practice need to be spread. In fact, the recent history of philanthropic giving in several important areas of social policy can be characterized as having placed a major emphasis on outcomes assessment—and having concluded that the dominant result of those assessments, is that the intended outcomes were *not* achieved.

For example, in the fields of welfare and employment training there have been, over the past two decades, a significant number of well-done studies, the overwhelming majority of which revealed poor results.[3] Summarizing this major body of work, former Secretary Robert Reich states, "Even successful education and training programs rarely live up to all the expectations placed in them…[They] often cannot lift disadvantaged participants out of poverty." A major evaluation of this kind, the National JTPA

Study (Bloom et al., 1993), was a major impetus for Congress's dramatic rethinking of the nation's employment and training system.[4] These very studies have helped build support for the notion that social interventions "don't work," and that public funding reductions are justified.

Thus, past outcome studies have become an integral part of the politics of social policy—as will products of the current outcomes movement. Failure has produced more than lessons to build on; it has helped produce pessimism about what social policy—the public will applied to social problems—can accomplish.

Thus the current emphasis on outcomes may, for particular foundations, be a new emphasis, and may require the acquisition of new competencies and cause new patterns of giving—but for philanthropy as a whole, the deeper roots of and implications for the current focus on outcomes are in the failure of past initiatives to achieve their specified outcomes.

That failure—and the ample and rigorous documenting of it in fields like welfare and employment training—casts a very different light on the recent emphasis on outcomes, and the actions and priorities it should generate, than does the absence of outcome studies on most grants and at many foundations. Absence would prompt us to action, as quickly as is possible, to fill the void; studies would abound. The light of failure would have the hues of caution and care in proceeding. It would push us to diagnose deeply the strength of the causal relationship between the desired outcome and the initiative to be funded—as well as the likelihood that it will be implemented as conceived. Goodness of motivation and strength of vision would not alone generate studies of outcomes. It would make us examine very closely the practical, as well as conceptual, strength of the evaluation design being adopted, so that the results obtained from the study were not shaped by the manner of collecting information. It would lead us a bit more into the perplexing issues of organizational capacity and implementation, the relationship between deep cultural values and political will, and the interplay between external help and individual change. It might cause us at times to delay beginning an outcomes study until intermediate implementation and capacity goals are met.

In short, the context of the current "outcomes movement" is complex, full of experience and insights that often lead in different directions. That context in no way undercuts the importance of a funder and a grantee being able to articulate what it is they are aiming to achieve, and how they will know whether it *is* achieved—but it does indicate that both the *what* and the *how* may not be as clear cut as they had hoped.

This paper attempts to lay out some of the factors that need to be considered when a philanthropy decides to put a greater emphasis on "outcomes." We divide the factors into three broad categories: Technical (How to Measure); Substantive (What to Evaluate); and Strategic (What to Do).

Technical Issues
(How to Measure)

The first and most fundamental technical issue is that outcomes are *not* the same as impacts. A program or project may specify, measure and achieve its outcomes—and still not have any incremental impact compared to other or no interventions.

How can that be? The National Supported Work Demonstration of the mid to late 1970s provides an excellent example of this phenomenon (Hollister et al., 1984).[5] The program was offered to four groups of unemployed individuals: ex-offenders, out-of-school youth, former drug addicts and AFDC recipients. All the participants were members of a group whose employment prospects were not good, and came to the program in need of jobs. During the follow-up period of two to three years, the employment rates of all the program participants increased significantly—but so did the employment rates of control group members. In fact, except among the AFDC recipients, the employment rates of the comparison group members were higher than those of the program participants. Thus, while the program looked effective when the key outcome variable was examined, the program actually did not improve the situations of the participants any more than what would have happened in the absence of the program.

In the mid to late 1980s, the Summer Training and Education Program (STEP) offered half-time work and half-time remediation to educationally and economically disadvantaged youth (Walker and Vilella-Velez, 1992).[6] The goal of the program was to increase the youth's academic competence. However, over the approximately two months of the summer program, participants' test scores did not increase and actually decreased slightly (Sipe, Grossman and Milliner, 1987).[7] On the face of it, it appeared that the program was ineffective; however, over the same period the test scores of the control group members plummeted nearly a grade level. Thus, rather than being ineffective, STEP was able to dramatically stem the summer learning loss that occurred in these youth.

Does this mean that projecting and measuring outcomes, without assessing impacts by means of a control or comparison group, is without value? No. General knowledge about how similar participants ordinarily do *vis-a-vis* the desired goals, and about the availability of services *vis-a-vis* the number of people who need and want them, can help form a reasoned judgment about a program's value. Detailed knowledge about the quality of each component of a program, why participants stay and why they leave, is also useful. A strong program theory about what should happen to a participant, and detailed knowledge about the actual implementation and course of participation, is even better. These techniques do not provide the certainty of an impact evaluation, but they are clearly useful, and usually better than random impressions.

But it does mean that, in the absence of a sound comparative study, it is often difficult to know to what degree the outcomes achieved are attributable to the initiative funded. This can lead to some situations even more puzzling to the intelligent citizen and voter than is the Supported Work example. For example, the 1983 federal Job Training Partnership Act (JTPA) programs aimed to place poor people with multiple obstacles to employment into jobs; the Act put a major emphasis on quantitative placement rates as a measure of local success in achieving the Act's goals. Local administrators set very high goals for the programs they funded, and offered financial incentives.

As local placement rates around the country began to soar—to over 80 percent in many locations—and were verified as factually accurate, critics speculated that these rates were *too* good, and indicated that most JTPA participants did not have serious obstacles to employment and would have gotten jobs even without JTPA's modest training interventions.

JTPA advocates scoffed. Then an impact study was done. It basically supported the critics. It indicated that JTPA even harmed the labor market prospects of some youthful participants. Well-specified outcomes, careful measurement, incentives and good performance all amounted to very little added value.

So the major "technical" issue for the outcomes movement is that even when outcomes are clearly specified and relevant information is carefully and credibly collected, outcome studies may not lead to accurate conclusions about what the program or initiative actually accomplished. The problem is particularly acute when the outcomes are aimed at producing long-term changes in human behavior (as most important outcomes do).

Are there technical solutions to this problem? Theoretically, yes: one is to use a comparison group so that the counter-factual question—what would have happened in the absence of the program or initiative—can be addressed. But there are numerous technical issues to be addressed in designing and carrying out a credible comparison study. For the purpose of this article, it is useful to highlight three of those issues.

1. **A sound comparison group methodology is not always available.** A good comparison group is one that looks like the program group, at least on key features. Typically, comparison groups are selected so that they match the program group on all the factors that fundamentally determine the key outcomes. Employment program comparison group members are usually matched by age, race, gender and education. The more factors that fundamentally influence the key outcomes, the harder it is to find a group that matches the program group on all of them. For example, it is very difficult to find good comparisons for use in interventions that target entire communities—such as comprehensive community initiatives or enterprise-empowerment zones. Community outcomes—such as economic or social well-being indicators—are influenced by a myriad of factors. Even communities that are quite similar at one point in time generally diverge over the time it takes for the programs to take effect. Therefore, the more factors that influence the key outcomes and the longer an intervention takes to achieve change, the harder it is to find a sound comparison group.

2. **The number of participants is often not large enough.** In order to detect a program's impact by contrasting program and comparison group behavior, one must take into account that the behaviors of even very similar individuals naturally differ. Technically, evaluators get around this complication by not only matching the two groups as closely as possible, but also by comparing the outcomes of large groups of participants and nonparticipants (i.e., comparison group members). The more natural variation there is in an outcome among like individuals, the larger the groups must be. Given the need to average out the natural variation in outcomes, small local programs may find that they do not have enough participants available to statistically detect their program's effects.

3. **The comparison group's activities are often not distinct enough to permit a sound conclusion about the content of the program or initiative's impact.** Most new programs do not offer completely new services, but rather offer higher quality services, a more complete package of services, and/or include critical elements missing in earlier programs. Thus, when testing such programs, it is often the case that comparison group members are able to enroll in fundamentally similar services. For example, many of the comparison group members of Project Redirection—an employment, education and parenting program for teen mothers—found their way into education or training programs, as well as parenting classes, by the end of the study (Polit, Quint and Riccio, 1988).[8] In such cases, when the program group is compared to the comparison group, one is not answering the question, "How did the participants fare compared to what would have happened had they done nothing?" but rather, "How did the participants fare compared to what they would have been able to find on their own?" The answer to the second question is a smaller impact than the first, and thus is harder to detect.[9] Simple participant/comparison group comparisons are often not powerful enough to detect such impacts.

So, the solution to the technical problem generates its own group of technical problems. It is at this point that some who began with great enthusiasm for the outcomes movement lose patience, and wonder if there are not simpler, less expensive and less time-consuming ways to arrive at reasoned conclusions about whether their giving has accomplished its aims.

The answer is "of course." Experienced observation plus basic data plus a sound theory plus some in-depth anecdotes and you can arrive at a reasoned conclusion. The problem—as the Supported Work and JTPA examples highlight—is that the conclusions based on these methods may be incorrect more than a modest percentage of the time. That is a risk worth taking if the funder is not willing to devote the resources and time to supporting a technically sound impact study.

The technical issues discussed above pose substantial challenges to a philanthropy that wants to focus on outcomes. The good news, however, is that there is a substantial body of thought about these technical issues, and that a "soundest approach" can usually be crafted. It just takes resources, and a willingness to use them on this issue. As in many areas of human endeavor, in fact, the progress made on the technical issues surrounding the measurement of outcomes in social policy has been greater and faster than the progress made on how to achieve those outcomes.

Substantive Issues
(What to Evaluate)

A careful look at the many outcome and impact studies conducted over the past two decades fairly quickly leads the observant reader to the conclusion that there must be deep substantive themes connecting many of the initiatives that have been evaluated, some root causes of the weak results they report so consistently.

Examination of the smaller group of studies that report stronger outcomes only strengthens this conclusion, which typically comes in two forms, whatever field of social policy we are looking at: first, that the basic substantive strategy was not, in fact, well implemented; and second, that the basic substantive strategy, even if well-implemented, was not strong enough to reach the desired results. We deal with the first type of conclusion in this section, and the second in the next. These generic conclusions are important for many reasons, not least of which is their implication for the current outcomes movement that neither a greater emphasis on outcomes (the notion that articulating them will improve their chances of occurrence), nor an improved technical approach to measuring outcomes, is likely to increase the probability of generating those outcomes.

During the past 25 years, formal outcome and impact assessments have been instituted in three basic categories of activity: new program initiatives, modified program initiatives and longstanding program initiatives. If one draws back far enough from the details, an interesting pattern is discernible:

1. Evaluations of new program initiatives are dominated by no or negative outcome and impact findings;
2. Evaluations of modified program initiatives are mixed in their findings of no, negative, modest or good outcomes and impacts; and
3. Evaluations of longstanding program initiatives are more likely to have modest or good outcomes and impacts.

This distribution of findings suggests that issues of operational capacity and implementation quality may indeed affect the outcomes and impacts our studies are disclosing. The massive review of employment and training evaluations done by the U.S. Department of labor (DOL) came to a similar conclusion:

> "It often takes time for programs to begin to work. Many of the success stories in training for the disadvantaged have come from programs which were operating for five years or more before they were evaluated." (U.S. Department of Labor, 1995, p.63.)

The logical implication of both organizational development theory and this experience is that funders should probably focus outcome and impact assessment resources only on stable programs and initiatives that have a track record and have refined their substantive strategies based on years of operational experience. It is in examining a mature program that one is likely to provide an accurate assessment of the impact and outcome potential of a particular substantive strategy.

Does this mean that the foundation and policy worlds should abandon pilot or "demonstration" programs? Should no research be done on new programs or demonstrations? No—just not formal outcome or impact assessment until the programs have matured. Since the late 1970s and the negative income tax experiments, policymakers have tried out new social policy strategies in "demonstrations" or pilot programs. Much can be learned operationally from these demonstrations. What are the major hurdles for the program? Does the program model make sense? Is it operationally feasible? What refinements would enable the program to operate better? What types of individuals are attracted to such a program? However, because the operators are just learning how to deliver the desired services, and the kinks in a program will take several years to discover and iron out, dispositive evidence of effectiveness should not be sought from demonstrations of new models. Judgments about the effectiveness of model elements or practices should come from research conducted on established programs.

The strategic rethinking of outcome research we are suggesting has important implications for the outcomes movement in philanthropy. One is that it should stimulate more coordinated efforts among foundations, since the outcomes emphasis should be less on assessing every program and more on (1) finding a stable exemplar to assess, one that represents solid implementation of a particular substantive idea, and (2) developing from that exemplar a set of operational benchmarks by which to assess the progress of newer operational manifestations of that idea.

An example of this approach is the national impact evaluation we carried out on Big Brothers Big Sisters (BBBS) over the 1992-95 period. The mentoring field was (and is) exploding with activity, and yet there were no credible findings about outcomes. BBBS is the brand name in the mentoring business; evaluating it, we felt, would provide a good indication of the utility of the mentoring idea as shaped by years of operational experience (Tierney and Grossman, 1995). Eleven foundations supported this study.

The findings were positive: mentoring had impacts on the initiation of drug use, school behavior and performance, and fighting, among others. Equally important, the study provided the mentoring field with operational benchmarks by which to assess the likelihood that other mentoring programs were achieving similar impacts. It is not necessary to do a new impact study on each and every mentoring program—it is only necessary to assess whether the program meets certain operational quality benchmarks.

Another implication of this strategic rethinking is that it should stimulate foundations to think about a field as a whole, not just particular programs. To follow through on the above example, the challenge for foundations interested in mentoring post-BBBS research is (1) how to expand mentoring, since it has positive effects, and (2) how to ensure that expansion does not dilute implementation quality. These challenges, whether met collectively or by individual foundation efforts, are field-building and maximize positive outcomes—even as they deal with individual mentoring programs.

Strategic Issues

(What to Do)

The second type of conclusion that can be drawn from previous evaluations is that many of the tested substantive strategies were not strong enough to achieve the desired goals. This was the sentiment that former Secretary Reich's earlier quote expressed.

In some cases, the problem was that the goals were simply unrealistic and that what in fact *was* achieved was worthwhile, and needed to be built on. For example, research done on STEP showed that participating youth had powerful and immediate positive educational effects from their summer experiences, compared with a control group of similar youth. But several years later, those effects had disappeared: STEP youth performed no better in school than did their non-STEP peers, having similar dropout and teen pregnancy rates. Because STEP's goal was not simply to provide a halt to learning loss and a quick lift, but rather to "immunize" its participants from the dropout and teen pregnancy bugs, what is remembered about STEP is not the positive effects it did have that needed to be built on during the school year, but that it "didn't work." Many other youth programs suffer from the same diagnoses, even though they do have some positive effects (Grossman and Halpern-Felsher, 1993).[10]

In other cases, the program does not seem to have been strong enough to even build on. For example, there is a consensus that job training as practiced over the past three decades does not work—because it does not help people move out of poverty (U.S. Department of Labor, 1995, p.i). There are a host of well-done studies supported by philanthropy and government, separately and in collaboration, that are pointed to in support of that conclusion. For most people, there are no lesser goals that justify a large public expenditure for job training programs. So, public funding has declined in response, and local entrepreneurs search for new, more effective approaches to job training.

What new insights can we learn from an understanding that many of the programs evaluated were substantially weak, unable to achieve their lofty goals? One important lesson for future initiatives to heed is to set realistic expectations—and then assess whether those realistic expectations are worth the effort. For example, we know that short-term programs rarely lead to long-term change; successful human change usually takes place in steps and generally has a difficult time being sustained in an environment with few other supports. This is especially true for youth, whose life trajectories are just beginning.

But even more important, the primary focus of a philanthropic concern about outcomes in the late 1990s needs to be on improving the strategy, substance and quality of *what* is being funded.

At first blush that statement may not seem so dramatic, and may even offend: after all, who in philanthropy does not care about what is being funded, and who does not want to help improve it? But the very concern about outcomes and accountability —about having a "bottom line"—when combined with our well-developed technical capacity to measure outcomes and impacts, and our increasing reluctance to criticize or shape initiatives that were designed by program operators or community groups, can easily lead funders to rush by the issues of strategy, substance, implementation capacity and quality. They are much more complicated to address, and are not easily amenable to definitive measurement. They are intermediate steps on the way to achieving bottom-line outcomes, and they have not received consistent or strong emphasis over the past several decades. They have not received that emphasis—paradoxically enough from the perspective of the current outcomes movement—in part because we were so anxious to believe and prove that what was funded *did* produce its intended (and often grand) outcomes that we rushed to assess them before we had realistically assessed their strategic, substantive and operational soundness.

This will discourage many philanthropists; it sounds as if we are starting over, just when the "outcomes movement" hints that we are nearing our goals. And the truth is, in many areas of policy, we *are* starting over. But starting over is different than starting from scratch—there are many lessons to build on, both positive and negative, and those lessons make it more likely that newly built policies and approaches can achieve the outcomes we intend. But that *building* process is the first priority—not a rush to outcomes measurement.

For example, this is an excellent time for philanthropy, both as individual institutions and in groups, to support analysis of previous employment training policies and implementation experience with an aim to generating strategies to improve the performance of the entire employment training field, as well as that of particular programs. The recent and prolonged boom in the American economy has hidden our lack of direction regarding *what* (if any) employment training policy and implementation practice America should have or try in the coming years. However, when the private economy inevitably slows down, a new direction will sorely be needed. Such wide-angle work conducted now would not only increase the chances of having more effective strategies, policies and programs—it would also strengthen the resolve and clarity of direction of individual foundations in their dealings with specific grantees and their programs.

But this work is precedent to a focus on the bottom-line outcomes of particular programs. It is the priority if we are to achieve those outcomes.

Some work of this nature has already begun in several substantive areas. For example, the authors are aware of collaborative/philanthropic initiatives aimed at improving the "what" in the areas of youth development and workforce development. Each initiative was developed with an openness to recommendations of substantial change in the policies, funding, substantive and implementation strategies of that field. Many more fields of social policy need the same open reexamination.

An approach that acknowledges the likely need for a major substantive overhaul in a policy field has major implications for thinking about outcomes. The incrementalist approach that has dominated the last several decades assumed that useful changes in policy and practice would come in the form of additions or modest alterations to extant programming—in short, we were just a tinker away from resolving the problem. We thus put more intellectual capital into figuring out how to measure the anticipated long-term outcomes than into figuring out how to dramatically improve substantive strategy and implementation capacity.

The non-incrementalist approach that is necessary now would shift that balance. Rather than focusing on measuring long-term participant outcomes, funders would place greater initial emphasis not only on shaping new substantive approaches, but also on ensuring their quality of implementation and political durability. Adhering to the advice that formal impact assessment is advisable only when a program is stable and mature, the results sought from these new project demonstrations would focus first on issues of implementation soundness and contextual viability, and only when those were satisfied would the focus on impacts begin.

Below are some critical substantive areas that get to the "what" issue, and that need individual and collective philanthropic attention in many social policy domains:

- **New Programmatic Strategies.** As noted above, several key policy areas are in need of serious rethinking regarding their basic substantive strategies. The employment training area is a good example: some analysts believe that there are experiential and theoretical reasons for crafting a very different set of strategies than those utilized in the past several decades. Such crafting is in itself an outcome that needs establishment and assessment.

- **New or Expanded Institutions.** The philanthropic world has, for the most part, devoted its resources to funding "programs" or "initiatives." But often what is most needed to achieve the ultimate outcomes we want in a policy area is *institutions*—institutions that have a reputation for stability and performance. The Ford Foundation's strategy in the arts world during the 1950s-70s is an excellent example.

- **Brand-Name Institutions.** For policy areas that require interaction across the public, nonprofit and commercial sectors, achieving "brand-name" status is often precedent to or at least simultaneous with the possibility of achieving impact. Building brand-name institutions is an outcome in itself.

- **Capacity Building.** In some areas of social policy there are sound programmatic strategies and brand-name institutions—the youth development area, with its Boys & Girls Clubs and Big Brothers Big Sisters, is an example—but there is simply insufficient capacity to reach their intended outcomes at a sufficient scale. Building capacity is an outcome worth defining and measuring.

- **Filling Gaps.** In some areas of social policy, analysis indicates that there are enormous substantive gaps that stand in the way of achieving the outcomes we all want. One example is after-school programming, which has been steadily reduced over the past several decades and which many analysts think is a key both to providing youth with critical development inputs (more adults, more engaging activities, more educational assistance) and to reducing negative behaviors (crime, school dropout, teen parenting).

  This particular "gap-filling" strategy would have as its primary goal the building of a variety of high-quality, well-supervised, engaging after-school activities that attract a broad cross-section of youth. These outcomes might take several years to achieve; only then would it be reasonable to assess whether ultimate outcomes were being achieved. If they were not, the question then would be if the appropriate response were to quit funding after-school activities or to investigate whether other policy areas needed strengthening. Ultimate outcomes cannot always be laid at the accountability doorstep of one programmatic strategy.

As the discussion above suggests, in our judgment the greatest need over the coming years is not simply more declaration of outcomes, envisioning of outcomes and assessment of outcomes; it is figuring out what to do to get those outcomes and, even more important, to get the desired impacts.

We do not recommend a moratorium on all outcome and impact assessments. In some settings, formal impact evaluation is what is called for. What we are arguing for is a strategic rethinking of when to utilize the tools of formal outcome and impact research. The bottom line is that getting to outcomes, measuring them, identifying the benchmarks along the way, and knowing how to influence an entire field are complex issues. It is not always so simple as making the specification of outcomes a priority, or deciding to spend more funds on evaluation.

A commitment to specifying and measuring outcomes is only the beginning step in a rigorous, thoughtful process. It is the commitment to that process that is necessary if the “outcomes movement” is to prove useful to philanthropy, grantees and society.

Endnotes

1 The National Commission on Philanthropy and Civic Renewal, *Giving Better, Giving Smarter,* Washington, D.C., 1997 p.114. The Commission was set up by the Lynde and Harry Bradley Foundation.

2 http://www.unitedway.org/outcomes/why.html, June 1998.

3 The U.S. Department of Labor, *What's Working (and what's not)*, (Office of the Chief Economist, January 1995) summarizes over 100 formal evaluations of employment and training programs.

4 Bloom, Howard, Larry Orr, George Cave, Stephen Bell, and Fred Doolittle, *The National JTPA Study: Title IIA Impacts on Earnings and Employment at 18 Months*, Abt Associates, Inc., January 1993.

5 Hollister, R., P. Kemple, and R. Maynard, (eds.), *The National Supported Work Demonstration*, University of Wisconsin Press, Madison, 1984.

6 Walker, G., and F. Vilella-Velez, *Anatomy of a Demonstration*, Public/Private Ventures, Philadelphia, 1992.

7 Sipe, C., J. Grossman, and J. Milliner, *Summer Training and Education Program (STEP): Report on the 1986 Experience*, Public/Private Ventures, Philadelphia, April 1987.

8 Polit, D., J. Quint, and J. Riccio, *The Challenge of Serving Teenage Mothers: Lessons From Project Redirection,* Manpower Demonstration Research Corporation, New York, 1988.

9 In addition, while the answer to the second is of interest to society, programs themselves want an answer to the former question.

10 Grossman, Jean, and Bonnie Halpern-Felsher, *Research Findings on the Effectiveness of Youth Programming,* Public/Private Ventures, Philadelphia, November 1993.

Board of Directors

P/PV Public/Private Ventures

One Commerce Square
2005 Market Street, Suite 900
Philadelphia, PA 19103

Tel: (215) 557-4400
Fax: (215) 557-4469
URL: http://www.ppv.org