Spring 1996

Reflections on Evaluation of Service-Learning Programs

Maryann Jacobi Gray

Reflections on Evaluation of Service-Learning Programs

by Maryann Jacobi Gray

Somewhere between a thorn in the side and a bloom on the rose of service-learning is evaluation. Whatever one's personal attitude toward evaluation, pressures to demonstrate effectiveness are increasing for practitioners and proponents of service-learning at the postsecondary level. This article describes the factors driving interest in assessing program outcomes and reviews some of the challenges facing evaluators of service-learning programs. Although this discussion focuses on the higher education environment, many of the principles examined are also applicable to high school and middle school programs.

Why Evaluate Collegiate Service-Learning Programs?: To some extent, pressures to evaluate service-learning programs are simply one manifestation of broader concerns about the efficiency and effectiveness of higher education, which have led to renewed interest in assessment of student outcomes (Nettles, 1995). Fiscal constraints have further fueled this interest, because higher education leaders can no longer support educational innovations such as service-learning with new resources but must instead reallocate funds from areas of low priority to areas of higher priority (Benjamin, Carroll, Jacobi, Shires, & Krop, 1993). Service-learning programs compete with other campus-based programs for dollars, space, staff, and equipment. Positive evaluation results would therefore offer a competitive edge.

Given limited institutional resources, many service-learning programs seek external support from such funders as the Corporation for National and Community Service (CNCS), the Fund for the Improvement of Postsecondary Education (FIPSE), and local and national foundations. These organizations routinely require program evaluations. For example, each of the 116 direct recipients of Learn and Serve America, Higher Education grants from CNCS as well as most of the 300-plus subgrant recipients were required to evaluate their programs in 1994–95.

Encouragement to evaluate also comes from within the field as practitioners seek information about service-learning outcomes. NSIEE's 1991 Wingspread meeting began the process of developing a research agenda for service-learning; a second conference in 1993, sponsored by Campus Compact, continued this process. In addition, participants in a 1992 Campus Compact and COOL retreat identified research and evaluation as a high priority for organizational collaboration. The service-learning listserv organized under the auspices of the University of Colorado routinely contains queries about service-learning evaluation and those who have conducted such research are besieged with requests for information and assistance. Giles (1994), Jackson (1993), Liu (1995), and Myers-Lipton (1994) are among those who have pointed out the need for research and evaluation about student outcomes.

Gap between expectations and reality: Over the past three years a number of informative evaluations and research projects have added to our understanding of service-learning outcomes for college students. These include analyses by Astin (1995), Cohen and Kinsey (1994), Dey (1991), Giles and Eyler (1994), Markus, Howard, and King (1993), and Myers-Lipton (1994).

Although these studies offer useful models for the field, those seeking to evaluate their local programs frequently encounter difficulties. The methods and measures used in one setting are often poorly suited to other settings, so local evaluators must design the evaluation plan as well as collect and analyze data.

Furthermore, high hopes and expectations about the benefits of evaluation often give way to disappointment when the results arrive. Ideally, evaluation will provide higher education administrators with a strong reason to support service-learning, but evaluation can also enable administrators to stall or avoid reallocating resources "until all the data are in," and negative or even lukewarm findings can lead to budget cuts or restructuring. Evaluations mandated by external funders are usually intended to help local programs improve their effectiveness, but too often these become tedious exercises with little perceived benefit at the local level.

Disappointment with evaluation is often a result of technical problems that reduce our ability to derive firm conclusions and persuasive recommendations. Many of the technical and methodological challenges faced by local evaluators in the service-learning domain are similar to those confronting evaluations in other domains. For example, because we can rarely assign students randomly to experimental or control groups, most educational evaluations are at best quasi-experiments with corresponding problems in attributing effects to programs. For instance, pre-existing differences between experimental and comparison groups may introduce uncertainty about how to interpret findings. Measurement instruments relevant to service-learning are still under development especially at the postsecondary level, and those that are available may not measure the outcomes of most interest. Thus, the instruments used in evaluations may lack validity. Additionally, relatively large sample sizes may be needed to obtain sufficient statistical power to detect differences between experimental and comparison groups, which can pose serious problems for smaller programs.

These technical problems contribute to a preference among some practitioners and researchers for qualitative approaches, such as ethnographies, portfolios, or naturalistic evaluation (Ostrow, 1994; Patton, 1987; Rossi, 1994; Williams, 1986). Although such methods yield rich data and considerable insight into the service-learning experience for students, they bring new challenges, especially for those who ulti-

v21 n3

mately hope to reach causal conclusions about service-learning.

In addition to these methodological challenges, evaluation funding is generally problematic. Few programs have set aside sufficient funds for evaluation, and the rule of thumb that evaluation should cost about ten percent of the overall costs quickly breaks down for programs with shoestring budgets, characteristic of many service-learning programs. Ten percent of not-much-to-begin-with is rarely sufficient to conduct a rigorous and comprehensive program evaluation, using either quantitative or qualitative methods.

Beyond these challenges, however, service-learning programs pose their own special (although far from unique) problems and issues for program evaluation. Most of these derive from the ambiguous status of service-learning within higher education institutions — often a marginalized, ancillary activity that nonetheless carries widespread implications for how institutions define their responsibilities to their students and to the broader community. For example, service-learning proponents and funders fret about the degree to which service-learning is institutionalized as a core educational enterprise, even as service-learning practitioners value the flexibility and freedom to innovate that smaller, less bureaucratic activities offer (Liu, 1995). Educators disagree about how to best describe the impacts of service on student participants even as they agree that these impacts are highly individualized and diverse, often internal and private, and likely to change over time (Ostrow, 1994, 1995).

Three issues stand out as posing special dilemmas for evaluators of service-learning: unclear and often conflicting goals assigned to service-learning programs; tension between traditional approaches to evaluation and the transformative change experiences sometimes associated with service-learning; and unrealistic expectations about how evaluation findings will be received and used.

A profusion of goals and objectives: By definition, service-learning programs strive to benefit service providers (i.e., students) and service recipients (Kupiec, 1993). In addition, some service-learning programs assign a high priority to goals related to institutional change and development (see the CNCS Learn and Serve America guidelines, for example). Evaluation, then, must either measure outcomes in multiple domains or focus on only some domains recognizing that goal attainment in these domains is a necessary but insufficient indicator of overall program quality. The risks of a limited approach are exacerbated by the potential tension between student and community goals. The activities that promote students' learning are not necessarily those that meet community needs, and the services needed by communities may have little educational value for students.

Further complicating this picture is the multiplicity of *educational* goals attached to service-learning programs. A rough taxonomy of service-learning goals in this domain includes: promoting mastery of skills, including technical (e.g., testing water quality, writing grant proposals), interpersonal (e.g., teamwork, conflict resolution), or communication skills; building knowledge, including both disciplinary (e.g., ecosystems functioning, teaching models) and cross-disciplinary information and understanding (e.g., race and class, the political process); promoting personal and values development, (e.g., development of civic responsibility, career exploration); and learning behaviors (e.g., active learning, integrating habits of volunteer service). Moreover, many of these goals may be best conceptualized as intermediate outcomes presumed to lead to longer-range outcomes such as academic achievement, retention and graduation from college, workforce preparation, and civic participation or leadership.

These educational goals cannot be inferred from observation of students' service activities. For example, programs that involve college students as tutors to K–12 youth may serve such diverse goals as career preparation, teaching students the value of "giving back" to community, improving the tutors' academic knowledge and skills, sensitizing students to issues of race and class in the United States, building tutors' leadership and communication skills, and so forth. Although different goals carry obvious implications for programmatic activities such as reflection, often program managers fail to clarify or prioritize their goals, instead hoping that the service experience will accomplish many different goals based on students' needs, interests, and levels of development.

Even if a single program could effectively promote this array of goals, program evaluation may be unable to measure the full range of possible outcomes. One can imagine the hours of testing that would be required to administer assessment instruments for each relevant goal or outcome dimension. Assuming that different students would show different outcomes, the average change in any single dimension would be difficult to detect and probably very small at the aggregate level. Even if there were some way to identify subgroups of students for whom specific program goals were most salient, most programs are simply too small to enable evaluators to detect program impacts. In the face of multiple or unclear program goals, evaluators have several choices. They can work with program managers and staff to select some goals upon which to focus, based on staff interests, information needs of external audiences (e.g., funders), or even those that can be most easily measured or documented. In so doing, however, evaluators risk seeing their results rejected as either incomplete, irrelevant, or a distorted mirror of program purposes and activities. They can use other evaluation models and methods that reduce the evaluators' need to specify program goals and objectives, such as naturalistic evaluation and other ethnographic or qualitative methods (Lincoln and Guba, 1985; Patton, 1986; Scriven, 1974). Such results may fail to satisfy funders' or policymakers' information needs, however, particularly in the absence of "hard data" on student development or achievement. Another approach, then, is to rely on students' self reports about their service experience using either quantitative ratings (e.g., Likert-type scales for self-perceived change on various dimensions) or qualitative methods (e.g., journals or portfolios). Exclusive reliance on students' self-reports is also problematic, since such ratings are inherently subjective.

Evaluation and transformative change: Traditional experimental and quasi-experimental evaluation designs are based on the study of groups. That is, the determination of a program's impact is based on analysis of the *distribution* of scores within and across groups. These evaluation designs also are based on the assumption that the effects of social and

—continued on page 29

valuable and viable operations — but is it the best use of corporate resources?

- a hardening of attitudes in higher education toward business — educators believing that they have responded to many of the concerns of business, but that corporations simply are not noticing;
- an environment in which higher education is not benefitting from the corporate experience with major organizational change such as downsizing.

What about the Future Relationship between Higher Education and Business? If I had my druthers, I would ask a key group of corporate and higher education leaders to sit down with me.

I would say to corporate leaders...

Look, folks, higher education really is a different kind of business. Instead of making demands of them that are identical to what you have done ("We downsized, why can't they?"), help with intelligent application of successful business strategies to higher education. Offer help, but in a different kind of way.

You really do not do as good a job as you might of identifying your hiring, education, and training needs. This limits higher education's ability to help you. What do you want? General skills? Technical skills? We hear different things from different parts of your organization.

You pay too much attention to credentials and too little to skills. Get beyond the institutional name on the transcript of new hires and determine the skills they have.

Realize that higher education is defined by more than employability. It is also defined by education for life as well as work, education for social and civic responsibility. You tend to ignore this.

I would say to higher education leaders...

Look, folks, you really do have to fundamentally change. You are no longer a unique provider of higher education; your management practices don't fit with the times; you are not adequately focused on results — educational attainment.

As Marion Barry of Washington D.C. said, after he was once again elected mayor, to those who opposed him, "Get over it." Some of the criticisms made by business are valid. Use them to grow and change.

I would say to both of you...

Let's cut through some of the perceptions and the criticisms and let's concentrate on the other topic I have discussed today: work — and especially education for work. In addition to the changes in work I mentioned earlier — knowledge- and information-based with a less stable and permanent work environment, and calling for more general education — business and higher education can acknowledge:

- the need to build education for work on the cutting edge of new technology;
- the impact of a global society on workforce opportunities;
- the need for "out of the box" thinking about education for work.

This is where your energy should be *re*-placed. This is what the country needs from you.

Summary: Rapidly and confusingly, change is a defining characteristic of higher education, work, and business. Higher education is experiencing changes (or demands for change) in funding, role, management, and technology. The kind of work, work environments, and education for work are changing. Business' emphasis on growth, its national role, social commitment, and products are changing. Business and higher education have experienced some dissonance that, ironically, can be used to strengthen their relationship. This circumstance, combined with strengthening the relationship that currently exists between higher education and business, provides a unique opportunity to focus on the changing nature of work and especially redefining education for work, further emphasizing technology, and the impact of globalism, as well as the need for truly creative thinking.

Judith S. Eaton is Chancellor of the Minnesota State Colleges and Universities. Previously, she served as President of the Council for Aid to Education.

Reflections on Evaluation

—continued from page 9

educational programs such as service-learning are relatively constant across people and groups. Thus, effective programs are those that consistently produce significant improvement in targeted domains among groups of participants (i.e., those receiving the intervention).

This definition of effectiveness, however, may not apply to all service-learning programs. Instead of expecting the same gains for most participants within a limited set of domains, perhaps we should expect service-learning to lead to some students benefiting greatly and others not at all or even having negative experiences.

For example, consider a program in which college students tutor high school students in algebra as the service component of a mathematics course. Perhaps most of the college student tutors will show modest gains in their own learning, while some will show no gains, some substantial improvements, and a small number (maybe only one student) will experience profound personal transformations in their values and aspirations. Overall, then, the program may have small or insignificant short term effects on the tutors and service recipients. This aggregate profile, however, does not acknowledge the highly significant impact of the program on some students.

This transformative experience arguably cannot be predicted or determined — one can only provide the setting and the opportunity. To apply deterministic models to measure outcomes that are inherently difficult to predict and erratic is unsatisfying and perhaps inappropriate.

Society's investment in service-learning, from this perspective, is perhaps analogous to our investment in research. The federal government and private foundations distribute millions of dollars annually for research, most of which leads to incremental gains in knowledge and technology. Yet in evaluating the effectiveness of our institutions' research programs, we are less likely to describe the incremental gains that characterize the vast majority of efforts and more likely to describe those few studies that offer dramatic leaps in our ability to cure disease, produce goods, or understand our own culture in relation to others. Because we cannot pre-

—continued on page 30

—continued from page 29

dict which research projects will be most important and useful, research grants seed a wide array of promising projects in the hopes that some will stand out. These transformative findings in essence justify the overall investment in research.

Similarly, service-learning perhaps may best be viewed as a way to extend opportunities to a variety of students and communities with the hope and expectation that most will show small improvement but some will show dramatic gains. While the former is important, it is the latter that rounds out the justification for supporting service-learning programs. Unfortunately, traditional program evaluation methods are poorly suited to identifying and describing these exceptional individuals and outcomes.

This issue may explain some of the differences between those who favor quantitative assessments and those who believe that qualitative methods are best suited to studying service-learning. Tests, surveys, and other quantitative measures typical of most program evaluations cannot fully capture the richness and depth of personal reality. Journals, portfolios, in-depth interviews, and other qualitative methods, however, provide an individual level of analysis that enables identification and description of transformative experiences.

Utilization of Evaluation Findings: Evaluation is often mandated, especially when programs are supported by external grants. Practitioners also, however, evidence considerable interest in evaluation. In many cases, they seek feedback to strengthen their programs. In other cases, they seek justification for additional institutional support, ranging from increased budgets and staffing to recruitment of more faculty to teach service-learning courses.

Those who hope evaluation will aid the quest for legitimacy and institutionalization of service-learning within higher education are likely to be disappointed by administrators' responses to their reports and data tables. Evaluation serves many purposes, but convincing policymakers to change their minds is rarely one of them.

The calls for evaluation of service-learning programs are, in fact, a reflection of service-learning's ancillary status. The same administrators who justify limited expenditures for service-learning due to the lack of evaluation data regularly allocate tens of thousands of dollars to units that may never have been rigorously or systematically evaluated. A lack of evaluation information, from this perspective, enables administrators to delay decisions about reallocating resources. When results become available, they may be used only to legitimize pre-existing decisions (Weiss, 1972).

Why are evaluation results so often neglected in decision making? One reason is that these data are only one form of input to decision making. Tradition, anecdotes, and personal experience, for example, compete with data to influence policy and programmatic directions (Weiss, 1988). Another reason is that evaluations sometimes fail to acknowledge political realities or the trade-offs administrators face in decision making. In the absence of new resources, increased support for service-learning requires administrators to decrease support for other programs, which may also be of high quality or have political constituencies that vigorously defend them. Under such circumstances, evaluation results may be perceived as "nice" but less than compelling. And evaluations that lack rigor or are perceived as advocacy efforts only increase rather than resolve the competition for resources. Also, the incentives within the system pose obstacles to implementation of the recommendations that are generated. For example, even the most positive evaluations of service-learning courses are unlikely to influence faculty behavior until institutions and academic disciplines offer meaningful recognition, rewards, and support to those engaged in service-learning. Finally, the influence of educational values cannot be overstated. When educators are committed to promoting socially responsive knowledge and strong community relations, service-learning is likely to thrive regardless of evaluation results. Conversely, institutions that value research over teaching and foundational or professional knowledge over socially responsive knowledge are unlikely to muster much enthusiasm for service (Altman, 1995).

Evaluations that successfully inform decision making for one audience may not fulfill the information needs of other key audiences. For example, external funders are primarily interested in whether service-learning programs have achieved the goals described in grant proposals. Service-learning practitioners are typically most interested in program influences on students' civic responsibility and participation. Campus administrators, however, are probably more interested in whether service-learning programs have achieved institutional goals such as academic achievement, retention, and community relations, and whether these programs are better or more cost effective than other strategies to achieve these ends. Thus, evaluations that address the goals and interests of program staff or external funders may receive little attention from and exert little influence on campus administrators.

This is not to say that evaluation is useless as a tool for building campus-based support for service-learning. The accumulation of information about service-learning may contribute to both program quality and stability. Evaluation also influences program operations, since knowledge that results will be reported to funders and administrators adds motivation for staff to strive for excellence in service delivery. Also, evaluation may be effective in raising the visibility of service-learning within higher education and promoting discussion and debate about its role within the academy, even if it fails to directly influence decision making (Ewell, 1983).

The evaluation literature indicates that managers and evaluators can influence the degree to which their findings are used in decision making. A key factor is involving stakeholders in the evaluation design and analysis, so that the results are perceived as relevant and legitimate. Other important factors include the presentation and timing of results and developing settings in which results can be discussed (Jacobi, Astin, and Ayala, 1987; Mowbray, 1988). Such steps can, however, increase the cost and time needed to conduct evaluations, and securing the involvement of key stakeholders can prove difficult if a major reason for the evaluation is to capture their attention in the first place.

Conclusion: Three challenges face those interested in evaluating service-learning programs: a profusion of goals and objectives attached to these programs increases the difficulty of conducting comprehensive evaluations and therefore risks studies that are either broad-based but superficial or overly narrow; traditional quasi-experimental and experimental evaluation designs emerge from the assumption that program impacts can be described and predicted at the aggregate level, but service-learning programs might have highly individualized results that are best described through other approaches; and the hope that positive evaluations will increase the status and support accorded service-learning in higher education is questionable, given the subjective and politicized decision making environment within most colleges and universities.

These dilemmas can be resolved, but the solutions may bring new challenges. For example, the use of qualitative methods such as portfolio analyses or ethnographies may enable better description of both the multiple goals of service-learning programs and their transformative effects, but qualitative findings may not convince those seeking concrete evidence of service-learning's impacts on student achievement, cost effectiveness, or benefits in relation to other types of programs or interventions. Conversely, the approach that best responds to the needs and interests of administrators or funders may provide an incomplete portrait of the program that is based on others' perceptions of the program rather than the program's own identify and self-concept. Thus, evaluators and program managers need to consider the trade-offs among different approaches within their institutional context.

That evaluation is a concern for service-learning practitioners is in itself a reflection of the tenuous and ancillary nature of the enterprise within higher education. Often service-learning courses are taught by non-tenure track or untenured faculty, supported through external or discretionary funds, and generally neglected in faculty senate debates about curriculum reform and undergraduate education. Those who would like to see service-learning become more institutionalized must compete with other programs and functions for resources, underscore or build congruence between programmatic and institutional values, and find ways around systemic obstacles and disincentives, particularly related to faculty participation. At best, evaluation can contribute to each of these strategies. Without attending to the issues described here, however, evaluation can only distract practitioners from the real challenges they face.

References:

Altman, I. (1995, August). Remarks in connection with the Award for Distinguished Career Contributions to Education and Training in Psychology, American Psychological Association Convention, New York.

Benjamin, R. S., Carroll, S., Jacobi, M., Krop, C., & Shires, M. (1993). *The redesign of governance in higher education.* Santa Monica: RAND, MR-222-IET.

Cohen, J., & Kinsey, D. (1994, Winter). "Doing good" and scholarship: A service-learning study. *Journalism Educator* 48:4–14.

Dey, E. L. (1991). Community service and critical thinking: An exploratory analysis of collegiate influences. Paper presented at Setting the Agenda for an Effective Research Strategy for Combining Service and Learning in the 1990s, Wingspread Conference Center, Racine, WI.

Ewell, P. T. (1983). *Information on student outcomes: How to get it and how to use it.* Boulder, CO: NCHEMS. ED 246-827.

Giles, D., Jr. (1994). National service-learning research project. *NSEE Quarterly* 19.3: 11.

Giles, D. E. and Eyler, J. S. (1994). The impact of a college community service laboratory on students' personal, social and cognitive outcomes. *Journal of Adolescence*17: 327-339.

Jackson, F. M. (1993). Evaluating service-learning. Ed. T. Y. Kupiec. *Rethinking tradition: integrating service with academic study on college campuses.* Providence, RI: Campus Compact.

Jacobi, M., Astin, A., & Ayala, F. (1987). *College student outcomes assessment: A talent development perspective.* ASHE-ERIC Higher Education Reports 7. Washington, DC: Association for the Study of Higher Education.

Kupiec, T. Y. (1993). *Rethinking tradition: Integrating service with academic study on college campuses.* Providence, RI: Campus Compact.

Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry.* Beverly Hills: Sage.

Liu, G. (1995). National service, government and higher education. Ed. M. Smith. *Service counts: Lessons from the field of service and higher education.* Providence, RI: Campus Compact.

Markus, G. B., Howard, J. P. F., & King, D. C. (1993). Integrating community service and classroom instruction enhances learning: Results from an experiment. *Educational Evaluation and Policy Analysis* 15: 410-419.

Mowbray, C. T. (1988). Getting the system to respond to evaluation findings. Eds. McLaughlin, J. A., Weber, L. J., Covert, R. W., and Ingle, R. B. *Evaluation utilization. New Directions for Program Evaluation 39.* San Francisco: Jossey-Bass.

Myers-Lipton, S. J. (1994). The effects of service-learning on college students' attitudes toward civic responsibility, international understanding, and racial prejudice. Boulder, CO: University of Colorado unpublished doctoral dissertation.

Nettles, M. T. (1995). The emerging national policy agenda on higher education assessment: A wake-up call. *The Review of Higher Education* 18: 293-313.

Ostrow, J. M. (1994). Sticking to our principles in research. *NSEE Quarterly* 19.3: 4-5, 28-29.

—. (1995). Self-consciousness and social position: On college students changing their minds about the homeless. *Qualitative Sociology* 18.3: 357-375.

Patton, M. Q. (1987). *How to Use Qualitative Methods in Evaluation.* Newbury Park, CA: Sage Publications.

Rossi, P. (1994). The war between the quals and the quants: Is a lasting peace possible? Eds. Reichart, C. S. & Rallis, S. F. *The qualitative-quantitative debate: New perspectives. New Directions for Program Evaluation 61.* San Francisco: Jossey-Bass.

Scriven, M. (1974). Pros and cons about goal-free evaluation. Ed. W. J. Popham. *Evaluation in education: Current applications.* Berkeley, CA: McCutchan.

Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice* 9: 5-19.

—. (1972). *Evaluation research: Methods of assessing program effectiveness.* Englewood Cliffs, NJ: Prentice-Hall.

Williams, D. D. (1986). When is naturalistic evaluation appropriate? Ed. Williams, D. D. *Naturalistic evaluation. New Directions for Program Evaluation 30.* San Francisco: Jossey-Bass.

Maryann Jacobi Gray is Associate Social Scientist at RAND in Santa Monica CA.