

11-7-1997

Methodological Problems in Evaluating Service Learning Projects

David A. Payne

Follow this and additional works at: <https://digitalcommons.unomaha.edu/slceeval>

 Part of the [Service Learning Commons](#)

Recommended Citation

Payne, David A., "Methodological Problems in Evaluating Service Learning Projects" (1997). *Evaluation/Reflection*. 50.
<https://digitalcommons.unomaha.edu/slceeval/50>

This Conference Proceeding is brought to you for free and open access by the Barbara A. Holland Collection for Service Learning and Community Engagement (SLCE) at DigitalCommons@UNO. It has been accepted for inclusion in Evaluation/Reflection by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.



ABSTRACT
Methodological Problems in Evaluating
Service Learning Projects

The ever increasing use of "service learning" as an adjunct to the ongoing instructional programs in public schools and higher education has challenged the conduct of both formal and informal evaluations. This article considers threats to internal validity in evaluating "learn and serve" projects with particular attention to data collection design and instrumentation. Mixed-methods designs are likely to be most effective, particularly when the intent is to focus on value-added assessment, and there is considerable variability in the nature and extent of implementation of learn and serve activities.

**National Information Center
for Service Learning**
1954 Buford Ave, Room R200
St. Paul, MN 55108-6107
1-800-808-SERVE

A paper presented at the annual meetings of the American Evaluation Association, San Diego, November 7, 1997.

METHODOLOGICAL PROBLEMS IN EVALUATING SERVICE LEARNING PROJECTS

David A. Payne
University of Georgia

Each school year sees an increase in the number of elementary, middle and high school students who engage community members and their peers in academically linked service activities. These activities are legislatively and operationally referred to as "learn and serve (L&S)." This adjunct to the now nascent but pervasive "volunteerism" movement in the country sings to the enthusiasm of both educational and community leaders. But with increased Federal funding comes accountability, particularly at the local level. Following is a personal narrative (or journal excerpt) of some of the evaluation challenges posed by attempts to evaluate the impact of service learning activities.

The problems discussed are ones encountered by the author as a member of a cadre of evaluators faced with assessing the success of statewide one-time L&S projects. The projects were for the most part modestly funded (e.g., \$3,000 per year), with a corresponding very limited evaluation budget.

It will be obvious to the reader that the problems here discussed are not unique to L&S projects. They are yet additional examples of problems that are encountered everyday by evaluators immersed in the real world. In that sense they reflect the insight of that great baseball evaluator, Yogi Berra, *déjà vu* all over again.

Although emphasized here because of their widespread application, conventional, quantitative designs are not the only designs that can be used. With L&S the by-word is *flexibility*. The variety of L&S treatments coupled with the variety of settings dictates such an approach. A contributory

factor is that the granting agency is primarily interested in value-added evaluations, hoping to demonstrate that engagement in a L&S activity enhances traditional learning outcomes. Some feel that the more traditional methods can best answer those value added questions. Experience suggests that is not always, or perhaps even usually the case. Use of mixed-methods is more likely to yield the data needed to answer the variety of questions that are associated with L&S treatments.

THE NATURE OF SERVICE LEARNING

Millions of public and private educational dollars are spent each year in support of what is referred to as service learning. Although variously defined, service learning embodies both a type of activity and an educational philosophy. As a type of activity service learning focuses on student performances targeted at the community or school. As a philosophy service learning reflects the belief that education must be linked to social responsibility and hands-on experiences for effective learning to take place (Conrad & Hedin, 1991; Giles, Honnet, & Migliore, 1991). The passage in 1990 of the National and Community Service Act has provided contemporary impetus to the historical tenets of John Dewey - we learn and retain most effectively when actively engaged in a meaningful task (Jacoby, 1996).

Service learning activities in today's public schools range from simple recycling efforts (newspapers, glass, plastic, etc.), to more complex and comprehensive environmental activities. Students in East Peoria High School in Peoria Illinois, for example, identified a community partner in the Heartlands Resources Water Council. They began by monitoring the water quality of local streams. As success was experienced, the project expanded to include well water quality and the monitoring of sediment fill in a local lake. The project was so effective that eventually the Environmental Protection Agency began using their results. One science teacher noted: "We have

11 different tests that we perform four times a year at the river's edge. It gives true meaning to chemistry if you know that the measurements that are being taken at this river, which provides drinking water for half the people in Illinois, no one else is regularly testing." At the higher education level, Markus, Howard and King (1993) have reported that students in a large undergraduate political science course where service-learning was a large component were more likely (than a control "no-service" group) to report that they had (a) applied principles from the course to new situations, (b) performed up to their potential and had significantly increased their awareness of societal problems. In this situation service learning took the forms of 20 hours of service opportunities in a homeless shelter, women's crisis center, an ecology center, or tutoring at-risk primary or high school students. Course grades were also statistically significantly higher than those of contrast students, as was participation in community service. Other types of service learning activities seen around the country include beautification and creation of community gardens, development of nature trails, assistance to the homeless and elderly, cross-age tutoring, community-child-care, and peer conflict resolution.

EXPECTED OUTCOMES FOR SERVICE LEARNING

Our public schools make a valiant effort to meet all students needs, cognitive, affective and psychomotor. This is perhaps a virtually impossible task at best, at least with a "traditional" curriculum. Service learning attempts to focus on many developmental needs of adolescents and pre-adolescents. Such needs are peer acceptance, creative self-expression, feelings of self-worth and personal competence, role exploration, capacity for responsible intimate relationships, management skill development and independence. Intellectual, affective and hands-on experiences can help develop and reinforce the satisfaction of these needs.

Each service learning activity generally has a classroom link and service target. Outcomes are project-specific and based on needs analysis data. It is generally expected that service learning projects will enhance (a) the mastery and retention of classroom learning objectives, (b) feelings of civic/school responsibility, (c) ties between school, parents, students, and the community, (d) student self-esteem, (e) student motivation, (f) interpersonal skills, and readiness for the world of work (Shumer, 1994).

The variety of possible outcomes in a single L&S site poses problems for the evaluator faced with the task of presenting evidence of "over-all" impact. The author is working with a L&S project which involves five different teachers at five different grade levels. The projects involve cross-age tutoring, ornamental horticulture, senior citizen support, school service, and environmental protection. To top it all off the majority of the service providers are special needs students. Since generalizability is not of paramount concern it was decided to do five case studies (Merriam, 1988).

PROBLEMS IN IMPLEMENTING SERVICE LEARNING PROJECTS

Evaluation efforts may be complicated by the fact that sometimes different, distant and distinct sites are involved in implementing decentralized service concepts. Diversity of clientele and the multi-site nature of the operation can have both positive and negative implications for evaluation (Turpin & Sinacore, 1991). On the one hand, having different sites suggests that replications of a given approach to service could yield more robust evaluation results. Each site would represent an independent and unique opportunity to see the concept in action. On the other hand, with different stakeholders and administrators involved, uniform program implementation is unlikely. Different service centers are likely to respond to different requests in different ways depending on the nature of resources available to them. There definitely will be program-by-site interaction. Other problem

areas which may be attributed to the multi-site nature of the network include potential lack of standardization in data collection, organization, analysis, and verification. On the positive side of having multiple sites is the sense of ownership of each center by its stakeholders. The site-specific nature of so many projects using the same "treatment" seriously limits the generalizability of the results if external validity were of major concern.

In an effort to meet individual school, student and "community" needs, great latitude is allowed in the selection of the nature and duration of specific learn and serve activities. The end result often is heterogeneous and idiosyncratic treatments. A general service activity may be common to a school or group of schools, e.g., conservation and environmental protection, but have different experiences, e.g., water purity testing or recycling. Each of these experiences in turn might have a different academic link, e.g., science, social studies. Aggregating data to document overall program impact in such cases can be a data collection nightmare.

Just as different treatments may be in force, so might the degree and extent to which the treatment is being implemented in the same or different sites. Several different teachers in the same school and at the same grade levels may exhibit varying degrees of commitment to the "project", e.g., recycling, resulting in a continuum of applications. Again uniform data collection is inhibited and meaningful documentation made difficult. Budget constraints can always inhibit both the implementation of projects and evaluation efforts.

A final implementation problem that has implications for instrumentation and data collection is the focus-unit of the treatment. In some cases emphasis is on an individual student, e.g., gain in reading scores for the recipient of tutoring, in other cases the focus is on a larger group such as a class.

SOURCES OF THREATS TO INTERNAL VALIDITY IN SERVICE LEARNING PROJECT EVALUATIONS

Many factors contribute to evaluation challenges. Among these are limited evaluation and implementation budgets, lack of resources and planning, expertise, and experience and the very nature of the uniqueness and complexity of the service learning experience itself. There is also the tendency to focus on program protection rather than program improvement. Although the threats to internal and external validity are well documented (Campbell & Stanley, 1963) appreciation of them in the context of service learning evaluation has not been addressed in the methodological literature.

Table 1 contains some illustrations of factors which have been shown to contaminate internal validity. These are examples taken over the last several years from the author's observations of projects individually funded statewide with competitors. Some of these factors are more important than others. Mortality, for example, can be an important if sometimes uncontrollable factor. Unless the service activity is non-voluntary e.g., required through a course, lack of motivation to participate and see a project through to conclusion might be an important contributor to the failure to find implementation of the service experience. Selection is another potential source of data contamination. In particular, where self-selection of participants for inclusion or exclusion in service programs is operating the composition of the target population either participants or contrast groups. As is so often the case one of the most challenging problems in conducting service learning evaluations is instrumentation.

Although most evaluators are accustomed to the necessity of having to create or at the very least adapt existing instrumentation, this is almost always the case in L&S projects. This is due in

TABLE 1
EXAMPLES OF TYPES AND SOURCES OF
INTERNAL INVALIDITY IN SERVICE LEARNING EVALUATIONS

<u>Type</u>	<u>Example</u>
History	The ripple effect on the nation of the April 1997 "Summit" on Volunteerism held in Philadelphia
Maturation	Increases in math problem-solving skills of project class are attributed to homeless shelter construction project rather than developmental math instruction they were exposed to together in another class
Instrumentation	Different readers used to assess degree of "altruism" on open-ended instrument which had not been pilot-tested and validated in before and after design
Statistical Regression	Students are selected to participate in a service project at a local nursing home on the basis of their high scores on a social responsibility scale. End of project scores show decline.
Selection	Students are selected to participate in L&S are volunteers who have expressed high social service drive. Their altruism scores are also found to be high. Contrast data collected from convenience class.
Mortality	After experiencing the frustration of tutoring younger students, the less motivated drop out of the project.
Testing	Reactivity/sensitivity of the pretesting with the "Value of Community Service" questionnaire administered to students engaged in beautification project around city buildings. When retested at conclusion of project 86% of scores show increases.
Compensatory Rivalry/ Resentful Demoralization	Within-school students in comparison condition were not invited to participate in service rally at beginning of project year, or in "periodic" celebrations.

large part to the implementation problems previously described. A great deal of effort and time is therefore needed to produce valid and reliable assessments. Usually the kinds of outcome variables addressed in service learning do not lend themselves to "standardized" measurement. A high school class, for example, adopts a local nursing home. One of the expected outcomes of the experience would be an enhanced appreciation for the value of aged populations and the job of care-giving. The measurement of these variables would require a tailor-made device. The assessment of school and classroom learning outcomes always represent measurement challenges. Educational measurement and the resultant interpretations become even more complex when they must be linked to specific service experiences. The use of writing tasks (i.e., essays about service topics and activities) has been found to be an extremely valuable technique in this regard (Neal, Shumer, & Garak, 1994). Scores from norm-referenced batteries are less than optimally relevant to the academic achievement links in most programs. Most instrumentation should be custom-made or at least adapted from existing measures.

It was noted earlier that one of the problems contributing to inefficiency in instrumentation is the fact that different treatment/activity units within the same site have different goals or objectives as they may be starting from different levels. The goals or objectives may be personal development, e.g., to learn how to relate more effectively with adults or peers, or product-based, e.g., develop landscaping or a horticulture greenhouse which will result in plants to be used in beautification projects. An assessment technique useful in collecting data under these conditions is *goal attainment scaling*.

Goal attainment scaling (GAS) historically has been used in a variety of mental health settings where individual patient or client goals need to be addressed. The best single source of information

about the technique is contained in a book edited by Kiresuk, Smith, and Cardillo (1994). In a real sense application of GAS involves standard setting before the implementation of treatment. Goals are negotiated between patient (student) and therapist (teacher/project director). There could be individual or group goals. Two illustrative goals are as follows:

Enhanced knowledge about issues related to water pollution

Value voluntary community service to senior citizens

Once consensus has been reached regarding the goals, indicators must be specified. Indicator data for the first goal were gathered from a 40-item teacher made knowledge test. The second goal was assessed from archival data maintained by the project director. One of the singular advantages of the technique is that expectations are set on an individual basis, so you can adjust different starting positions. Table 2 contains an illustration of how this might be accomplished for our two sample goals. The Level of Attainment scale is fairly standard for GAS applications. Students, project coordinators, or other stakeholders would determine the level of attainment. At the conclusion of the treatment each goal can be evaluated and data aggregated for individuals or groups.

An interesting variation on the GAS methodology would be to use it in a modified form as a retrospective pre-test (Campbell & Stanley, 1963). Used in this fashion the treated group becomes its own control, a particularly useful approach when self-report dependent measures are involved (Howard, et al., 1979).

TABLE 2
ILLUSTRATIONS OF GOAL ATTAINMENT SCALING

Level of Attainment	<u>Scale 1</u> Environmental Knowledge	<u>Scale 2</u> Appreciation of Community Service
Much Less -2 than expected	Less than 4 point gain or loss	Attends 20% of sessions or less
Somewhat less -1 than expected	4-9 point gain	Attends 21-49% of sessions
Expected level 0 of Outcome	10 point gain	Attends 50% of sessions
Somewhat more +1 than expected	11-15 point gain	Attends 51-79% of sessions
Much more +2 than expected	17 or more point gain	Attends 80% or more of sessions

The author has used a modification of the GAS technique in a project focused on assessing school climate across nine schools following the implementation of a site-based management development program. Individual school goals ranged in content from improving student achievement to involvement of teachers in the site-based experience to increasing parent involvement in school programs. Each school generated their own set of objectives. Progress data were contrasted across teachers, parents, and students making a comprehensive picture of project impact. A valuable application of data derived from the use of such a technique can be the assessment of trends over time.

IMPLICATIONS FOR DESIGN

The sum total of the problems just discussed would obviously lead to the conclusion that an eclectic approach to designing L&S evaluations is necessary. Forcing L&S evaluation into traditional Xs and Os configurations doesn't make sense. Neither does relying exclusively on costly qualitative methods. A variety of methods (mixed-methods) to meet a variety of needs would seem to be a reasonable guideline (Greene, Caracelli, & Graham, 1989). This is surely not an earth shaking insight, but when it happens evaluation psyche. The use of case study methodology (Yin, 1994) and aggregated multiple case studies (Miles & Huberman, 1994), in addition perhaps to goal-free and responsive approaches (Stake, 1983) have been profitably used to evaluate L&S projects. But new approaches are always needed. It is hoped that this brief vignette will provoke an exchange of ideas and methods. Perhaps through the National Service Learning Cooperative (a K-12 Clearinghouse on Service-Learning) (serve@maroon.tc.umn.edu). The exciting thing about doing evaluations is the finding a methodology that helps meet an important data-need challenge, and the evaluation of L&S projects surely poses many challenges.

REFERENCES

- Campbell, D. T., & Stanley, J. C. (1963). Experimental and quasi-experimental designs for research. Chicago: Rand McNally.
- Conrad, D., & Hedin, D. (1991). School-based community service: What we know from research and theory. Phi Delta Kappan, 72(10), 743-749.
- Giles, D., Honnet, E. P., & Migliore (Eds.) (1991). Research agenda for combining service and learning in the 1990s. Raleigh, NC: National Society for Internships and Experiential Education.
- Greene, J. C., Caracelli, V. J., & Graham, W. F. (1989). Toward a conceptual framework for mixed-method evaluation designs. Educational Evaluation and Policy Analysis, 11(3), 255-274.
- Howard, G. S., Ralph, K. M., Gulanick, N. A., Maxwell, S. E., Nance, D. E., & Gerber, S. K. (1979). Internal invalidity in pretest-posttest self-report evaluations and a re-evaluation of retrospective pretests. Applied Psychological Measurement, 3, 1-23.
- Jacoby, B. (Ed.). (1996). Service-learning in higher education. San Francisco: Jossey-Bass.
- Kiresuk, T. J., Smith, A., & Cardillo, J. E. (Eds.). (1994). Goal attainment scaling: Applications, theory, and measurement. Hillsdale, NJ: Erlbaum.
- 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(4), 410-419.
- Merriam, S. B. (1988). Case study research in education (A qualitative approach). (2nd ed.). San Francisco: Jossey-Bass.
- Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis. Newbury Park, CA: Sage.
- Neal, M., Shumer, R., & Garak, K. S. (1994). Evaluation: The key to improving service-learning programs. Minneapolis, MN: Center for Experiential Education and Service Learning, University of Minnesota.
- Rossi, P. H., & Freeman, H. E. (1993). Evaluation - A systematic approach. (5th ed.). Newbury Park, CA: Sage.
- Shumer, R. (1994). Community-based learning: Humanizing education. Journal of Adolescence, 17(August), 357-367.
- Stake, R. E. (1983). Program evaluation, particularly responsive evaluation. In G. F. Madaus, M. S. Scriven, & D. L. Stufflebeam (Eds.), Evaluation models (Viewpoints on educational and human service evaluation), 287-310. Boston: Kluwer-Nijhoff.

Turpin, R., & Sinacore, J. M. (Eds.). (1991). Multisite evaluations (New Directions in Program Evaluation). San Francisco: Jossey-Bass.

Yin, R. K. (1994). Case study research: Design and methods. (2nd ed.). Newbury Park, CA: Sage.

