# Measuring the Impact of Youth Voluntary Service Programs

Summary and Conclusions
of the International Experts' Meeting

**THE WORLD BANK**
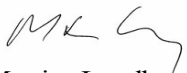
**ICP** INNOVATIONS IN CIVIC PARTICIPATION

**Acknowledgements**

Mattias Lundberg
Senior Economist
Children and Youth unit
Human Development Network
The World Bank

Susan E. Stroud
Executive Director
Innovations in Civic Participation

## Table of Contents

**Executive Summary**

In response to a need for additional evidence on the effectiveness of youth voluntary service programs, the World Bank Children and Youth Unit and Innovations in Civic Participation (ICP) co-hosted a meeting in Washington DC in May 2008, bringing together an international group of youth service practitioners, policymakers, and evaluation experts. The meeting discussed the standards of evidence of impact evaluations, identified available tools and methods for building more rigorous evidence, explored the existing evidence base on youth voluntary service programs, outlined the main challenges in conducting impact evaluations of the programs, and discussed the perspectives and roles of different stakeholders in conducting impact evaluations. The meeting concluded in identifying the next steps in building a research framework and agenda.

*Methods to measure the impact of youth service programs*
Impact evaluations measure whether the reached outcome is being caused by the program rather than by other, external factors. There is not one universal way to conduct impact evaluations. With varying benefits and challenges, experimental, quasi-experimental and non-experimental designs all provide techniques for evaluating the impact of an intervention. Evaluations must be designed individually and it is important to take into account the context, resources, and program design in determining the combination of evaluation approaches to be used. Some general guidelines can be offered on how to conduct rigorous evaluation, however. These include demonstrating the causality between the program and the observed outcome relative to a counterfactual, collecting baseline data against which to measure the perceived changes, using mixed (qualitative and quantitative) methods, and understanding the cost-effectiveness of the program.

*Existing evidence on youth service programs*
There is a lack of formal, publicly available evaluations of youth service programs, especially outside North America and Europe. The available literature is also limited in high-quality data and standardized criteria for program effectiveness. Instead of rigorously evaluating the programs, most existing studies contain case studies describing programs and their outcomes. In addition, very few existing evaluations mention explicit evaluation frameworks, instead providing only retrospective impressions. A review of the existing literature was presented at the meeting as well as six program evaluations demonstrating experimental, quasi-experimental, and non-experimental designs. These program evaluations all provide important examples of existing practice, highlighting the need to take the context into account when designing evaluations, and demonstrating the importance of viewing evaluation as a process, not just as an end result.

*Challenges to evaluating youth service programs*
Throughout the meeting, several challenges for evaluating youth service programs became evident. They can be categorized into conceptual, technical, and operational challenges. The *conceptual challenges* include lack of common definitions, guiding principles and key characteristics of youth service programs, and the enormous contextual variations between the programs that fall under the 'youth service' category. The limited available research on youth service also poses a challenge. The *technical*

*challenges* to rigorous evaluations of youth service programs are related to determining and defining what outcomes to measure (short-term or long-term indicators, direct or indirect outcomes, impact on participants or community, etc.), how to measure them (e.g. issues of lack of tangible impacts, unintended outcomes, and self-reported information), and how to establish causal attribution (logic model, internal & external validity, ceiling effect, fidelity etc.). Finally, *operational challenges* for evaluating youth service programs include limited capacity of youth service organizations, as well as establishing a practice for including young people in different stages of the evaluation.

*Stakeholder perspectives on impact evaluation for youth service*
Not everyone should invest substantial resources to produce evidence of impact through sometimes complex, comprehensive impact evaluations. Evaluation objectives and needs of organizations differ significantly based on their circumstances and program or policy elements. The challenge is to find a combination of approaches that complement each other, and to reach sufficient level of rigor in generating evidence that is at the same time useful to practitioners in enhancing the quality and impact of their programs, convincing and useful to the policymakers, and attractive to funders. Careful monitoring should be part of any organization's core activities, and most small youth service organizations should focus on evaluating their processes, and learning from them. However, even small programs should consider conducting impact evaluations if their program has the intention and capacity to grow.

*Next steps – developing the research framework and agenda*
Meeting participants concluded that there is a need to develop a community of practice in the field of youth service research and evaluation, constructed around three main areas: building a common conceptual framework, agreeing on a research agenda, and refining and adapting evaluation tools and methods. Several remaining key questions were identified, and concrete steps were suggested in all the three areas to address those questions. Meeting participants and others interested in advancing the field of youth voluntary service will be invited to take part in an online discussion forum to continue the exchange of ideas. In addition, working groups will be established to develop work plans for accomplishing the identified, specific tasks, which include i) developing the conceptual framework, typology, overarching logic model, and guiding principles for youth service; ii) developing a research agenda, standards of evidence, and a formalized process for evaluating the programs; and iii) creating a database of tools and instruments and agreeing on a schedule to help youth service programs determine when and how to conduct evaluations.

## 1. Introduction

Youth voluntary service programs exist in many countries around the world, and new programs and policy initiatives are being developed in several others, attracting a great deal of attention and resources. According to the 2000 World Values Survey, one quarter of young people volunteer with religious or secular organizations. International organizations such as UNICEF and UN Volunteers/UNDP, and civil society organizations such as Innovations in Civic Participation (ICP) support the development of these programs. In the United States, president Barack Obama made expanding opportunities for national service "a cause of (his) presidency" during his campaign. The new programs will provide many more young people around the world the opportunity to engage in youth service. However, these programs do not come for free. The *American Prospect* magazine estimated in 2007 that enrolling one million young people in voluntary service programs in the United States by the year 2020 would cost more than US$14 billion per year.

Before more young people are encouraged to participate in the youth voluntary service programs, the stakeholders have to ensure that these programs have a significant positive impact on the youth and societies, and that their benefits exceed their costs. That is, the stakeholders and the evaluation community have a responsibility to develop the evidence of the effectiveness of these programs.

To this end, the World Bank Children and Youth Unit and ICP organized an international experts meeting at the World Bank headquarters in Washington, DC, on May 8-9, 2008 to discuss measuring the impact of youth voluntary service as a development strategy. The researchers, practitioners, policymakers, and donors participating in the meeting discussed the various methods for evaluating the impact of the programs and the standards of

---

**Box 1. Definitions:**

**Impact:** Difference in the changes in outcomes between those with and without the program.

**Impact Evaluation**: "Impact evaluation is intended to determine more broadly whether the program had the desired effects on individuals, households, and institutions and whether those effects are attributable to the program intervention." (Baker 2000)

"An evaluation of the effects – positive or negative, intended or not – on individual households, institutions, and the environment caused by a given development activity such as a program or project" (Baker 2000).

"Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (OECD 2002; Network of Networks on Impact Evaluation, NONIE 2002)

"Counterfactual analysis of the impact of an intervention on final welfare outcomes." (World Bank Independent Evaluation Group, IEG)

**Counterfactual:** Impact evaluation involves the explicit identification of the *counterfactual* – that is, the outcome that would have been observed in the absence of the intervention.

evidence. They also examined the body of evidence on the impact of youth service and discussed strategies for enhancing the data available. This report highlights the broad variety of evaluation methods and tools discussed at the meeting, focusing on how to evaluate impact and how to overcome the obstacles of evaluating youth voluntary service programs. (See Annex A for a list of participants and Annex B for the Agenda of the Meeting.)

This report is divided into six sections, including this introduction. Section Two examines the variety of approaches and methods for conducting impact evaluations, and points out their main benefits and limitations. Section Three outlines the existing evidence base, referring to the evaluations presented at the meeting. Section Four draws attention to the challenges of evaluating youth service programs that were brought to light and discussed at the meeting. Section Five discusses who should conduct impact evaluations and in which situations these types of evaluations should be conducted, drawing on the discussion between stakeholders present at the meeting. Finally, Section Six presents the next steps that were proposed and agreed upon at the conclusion of the meeting.

## 2. Methods to measure the impact of youth service programs[1]

There are very few rigorous impact evaluations conducted on youth voluntary service programs. Youth development programs in general, and particularly those with a voluntary service element, contain several evaluation challenges, which have prevented more evaluations from taking place. There is no one universal way to conduct impact evaluations, but some general guidelines can be offered on how to conduct rigorous evaluation, taking the specific challenges and program characteristics and goals into consideration. However, to be able to move from individual evaluations to building a broader research and evaluation agenda for the field of youth voluntary service, more work needs to be done in building a conceptual youth voluntary service framework.

As stated in the introductory section, one of the main objectives of the experts meeting was to discuss the standards of evidence of impact evaluations used by the various stakeholders at the meeting. The World Bank's *World Development Report 2007: Development and the Next Generation* (WDR07), classified youth service programs as "promising but unproven," suggesting that there was not enough rigorous evidence to support the contention that the programs are *"proven* successful". This raised the question among the practitioners of the requirements by the World Bank and the evaluation community in general of what constitutes rigorous impact evaluation.

This section will outline the different approaches and methods available for evaluating impact. No one single approach can or should be used in all situations. The benefits and challenges of each approach are also explored to imply their applicability in different circumstances. (For additional methodological considerations, see Annex C.)

Impact evaluations measure the effects of the program, making sure that those effects are produced by the program intervention and not by other factors. There is consensus in the evaluation community about this overall definition. (See Box 1 for definitions by different evaluation communities.) Impact evaluation is intended to answer questions that traditional monitoring and process evaluation cannot answer:

- What is the effect of the program on outcomes?
- How much better off are the beneficiaries because of the intervention?
- How would outcomes change under alternative program designs?
- Does the program affect different people in different ways (e.g., by wealth class, ethnicity, region, gender)?

> **Box 2. Definitions:**
>
> **Monitoring:** "Monitoring will help to assess whether a program is being implemented as was planned. A program monitoring system enables continuous feedback on the status of program implementation, identifying specific problems as they arise." (Baker 2000)
>
> **Process Evaluation: "**Process evaluation is concerned with how the program operates and focuses on problems in service delivery." (Baker 2000)

---

[1] This section of the report is mostly based on information from the World Bank Impact Evaluation web-site: http://go.worldbank.org/169GZ6W820, and Baker, Judy L. 2000. "Handbook for Practitioners: Evaluating the Impact of Development Projects on Poverty". The World Bank, Washington, DC.

(from the presentation of Ariel Fiszbein, Lead Economist, Human Development Network, The World Bank)

Monitoring and process evaluation are essential prerequisites for impact evaluation: they assess whether the program was implemented as intended. Monitoring measures the outputs produced by an intervention, and process evaluation assesses whether all the steps of the program are properly executed and whether there are bottlenecks or impediments to successful implementation. (See Box 2) Monitoring and process evaluation should be initiated from the beginning of the program and continued throughout it to enable changes if problems in implementation arise.

## 2.1 Establishing causality and measuring impact

Measuring the impact of a program is an attempt to confirm whether the intervention is likely to have caused program outcomes. This requires demonstrating that the observed change is due to the intervention, and not to other factors. Ideally, one would like to ensure that all other factors that can influence the beneficiary (other programs taking place at the same time, general economic changes in the country, changes to unemployment, new laws, etc.) remain unchanged during the intervention. Since this is not possible outside the laboratory, it is necessary to minimize or control for the impact that these other, unrelated environmental changes might have on the outcome. In this way, evaluators can attempt to understand how the situation of the beneficiaries would have developed in the absence of the intervention. This information can be used to separately identify the developments that can be attributed to the intervention.

Understanding impact usually requires an explicit *counterfactual*, that is, an estimate of what would have happened to the beneficiaries if the intervention had not taken place. This presents a problem: it is not possible to observe the same individual over time both with and without the intervention. The most difficult task for most impact evaluations is to identify a representative comparison or control group: a set of individuals or families who are like the beneficiary group in every way, except that they were not subject to the intervention.

Impact evaluation designs can be broadly classified into three categories: experimental, quasi-experimental and non-experimental. These categories correspond roughly to the methods used to identify a comparison/control group (World Bank Impact Evaluation Web-site).

### 2.1.1 Experimental designs

In experimental designs, the beneficiary and control groups are randomly selected from some well-defined population. Random selection and assignment to beneficiary and control groups is the easiest way to ensure that other factors are not influencing the relationship between the beneficiary and the intervention in a biased way, or that the influence of the other factors is the same across the target population. If the sample size is large enough, there should be no systematic differences between the two groups, other

than the program. In other words, any observed difference in the outcomes between the beneficiary and control groups can be safely attributed to the intervention, and not to underlying differences between the two groups.

The main **benefit** of this technique is the comparative simplicity of constructing the counterfactual and interpreting the results. The impact of the program on the outcome can be measured simply as the difference between the means of the samples of the beneficiary group and the control group. Or, if there are significant differences between the two groups in the initial (pre-treatment) levels of the outcome indicator, the impact of the program can be measured as the *difference in differences* in changes over time between the two groups. Also, since the random selection is done *ex ante* (before the intervention takes place), it ensures that the reasons for participation in the beneficiary or control group are uncorrelated with the outcomes.

There are several **challenges** to this approach, and the meeting participants agreed that not this nor any other approach should be promoted as the "gold standard" for impact evaluations as is sometimes done. Some of the major challenges in this technique include:

- Ethical concerns:  Randomization is considered by some evaluators and practitioners to be unethical because it denies benefits or services to otherwise eligible individuals. It is anathema to governments and service providers, who act in the public interest, to deny benefits to those who need them.  But when budgets do not permit the delivery of interventions or services to all those who are interested and eligible to participate, random assignment ensures that benefits are allocated according to transparent and fair criteria, which give all eligible population equal chances to receiving the treatment.  Alternately, the treatment can be phased in over time, so that all receive the benefit, but those in the control group receive it later than those in the initial beneficiary group.
- Political economy: It may be politically difficult to offer the intervention to certain groups only; or political concerns may favor one group over another. However, experience has shown that populations and governments are more likely to support random assignment into beneficiary and control groups if the selection method is perceived as fair and transparent.
- High cost and time requirements: Experimental designs may be expensive and time consuming, particularly when a significant amount of new data needs to be collected from both beneficiary and control groups.  However, the costs are significantly lower if the mechanisms already exist to collect data for monitoring and process evaluation.
- Unrealistic design: Experimental evaluations generally test a small number of isolated factors. In reality, causality is multifarious – there is a multitude of influences on any given outcome; narrowly-defined evaluations can miss important aspects of context, local conditions and unanticipated events. (American Evaluation Association) Experimental designs need to be accompanied by in-depth qualitative analysis of the changes that occur in the social setting under evaluation, so that unanticipated causal factors can be detected, and are most applicable to evaluating discrete, homogenous interventions.
- Applicability: Many programs are implemented as pilots, and may be too small to permit experimental or quasi-experimental designs which use estimation methods

requiring samples of adequate statistical power. Also, when the evaluation is retrospective, randomization is not possible. (World Bank Independent Evaluation Group)

### 2.1.2  *Quasi-experimental designs*

Where random assignment is not possible or is not the optimal technique, there are other tools available. *Quasi-experimental* techniques generate comparison groups that are not randomly chosen, but are selected so that they closely resemble the beneficiary group, at least in observed characteristics. In these *quasi-experimental designs*, program participants are compared to non-participants using complex statistical methods to account for the differences between the groups and to correct for the *selection bias* that might arise from non-random allocation of benefits.

Two common quasi-experimental methods include **matching techniques and reflexive comparisons**. Matching techniques attempt to select from a larger sample a comparison group that matched the beneficiary group, while reflexive comparisons compare program participants to themselves before and after the intervention. (See a more detailed description in Annex C.)

The main **benefit** of these designs is that they can draw on existing data sources and are thus often quicker and cheaper to implement than randomized designs. They can also be performed after a program has been implemented, provided that sufficient data exists.

The principal **disadvantage** of these techniques is the possibility of selection bias. Moreover, reflexive comparisons cannot say anything about what would have happened to the beneficiaries if they had not received the treatment, nor can it control for unrelated external (non-intervention) changes. The reliability of results also depends on a set of maintained assumptions and on fairly complex statistical methods, requiring considerable expertise.

Selection bias, which is a major **challenge** in these methods, can arise when assignment is not random but purposive, and when there is reason to suspect that some of the characteristics that determine assignment to treatment or control groups might also influence final outcomes.  This can be the case if, for instance, the treated group is initially poorer, or scores higher on some entrance exam, than the treated group.  In these cases, some groups are more likely to receive the benefit than others, and/or they may benefit differently from the program.  Bias can also arise from beneficiaries who behave differently in the program than others – for instance, they are more likely to show up to receive benefits, or they are less likely to drop out of programs that distribute benefits over time.  Participation in youth service programs is usually voluntary; participants are those who demonstrate a desire to serve in the program.  Volunteers are self-selected, and the benefits they receive from participation may not be the same as for someone who is chosen at random from the population.  In fact, the observed differences in outcomes between participants and the control group may be due to those initial differences rather than to participation in the program.  The sources of bias may be observable or

unobservable. In the first case, the bias may be controlled by matching or instrumental variables techniques; in the second case, the bias may be fatal to the comparison.

Finally, a problem with both experimental and quasi-experimental approaches is that they can be conducted as "black box" evaluations, showing the impact of the program on the measured outcomes, without providing any understanding about which particular elements of the program contributed to the impact or paying attention to the process.

### 2.1.3   *Non-experimental designs that permit qualitative and quantitative analysis*

In non-experimental designs, program participants are compared with non-participants using statistical methods to account for the differences between the groups.

The econometric techniques that can be used to compare the two groups and to *correct for the selection bias* include **Regression Discontinuity** and **Instrumental Variables** such as Random Promotion or Encouragement. Regression Discontinuity uses program eligibility criteria to construct a control group: those just within some target cutoff point are assumed to sufficiently resemble those who are just outside it, and it is assumed that differences between the two groups are largely due to the program. Instrumental Variables estimation on the other hand consists of using one or more variables (instruments) that are correlated with participation in the program, but not to outcomes given participation. In principle, this corrects for potential endogeneity of program placement. (See a more detailed description in Annex C.)

### 2.1.4   *Non-experimental designs that permit qualitative analysis*

Some non-experimental evaluation methods do not establish an explicit counterfactual or control/comparison groups but use other sophisticated techniques to attempt to demonstrate causality and program outcomes. There is some debate about the validity of these evaluations, since they may not rigorously establish causality nor explicitly determine the standard against which impact is to be defined and measured. However, they are more capable of capturing the complexity of program outcomes and offer a more flexible and low cost technique that can be applied in situations where experimental or quasi experimental techniques are not appropriate or possible. Examples of these approaches include in-depth case studies, general elimination method, and theory-based evaluations. (See a description of these techniques in Annex C.)

The main **benefit** of these approaches is that they take into account the complexity of the program and the context within which it operates. They provide information about what factors might affect the impact of the program, and thus help to understand the program outcomes.

The primary **disadvantage** of these methods is that the information they provide is complex and difficult to analyze differentially and to generalize from. While many evaluators and practitioners argue that this is not necessarily a disadvantage because programs are complex and the data should reflect this complexity, complex information

that is not generalizable is not very helpful for funders and donors looking to draw lessons about the impact or replicability of a program. Indeed, non-experimental approaches do not, in most cases, permit quantitative analysis or allow for comparisons between different programs. Rather they are most helpful for an in-depth understanding of a specific program in a specific context.

Several other approaches were highlighted at the meeting that can *support* impact evaluations by providing more comprehensive or nuanced picture of impact. These include outcome mapping, success case method, and beneficiary assessment. (See a description of these techniques in Annex C.)

## 2.2  Choosing the right methods

As stated before, there is not one "gold standard" for evaluating impact, but instead "methodological diversity and appropriateness in support of rigor" (Network of Networks on Impact Evaluation, (NONIE)) are of main importance.  Each impact evaluation has to be individually designed, taking into consideration the program's unique characteristics and environment. Each program requires a particular combination of appropriate evaluation approaches, and the evaluator must carefully explore the methodological options and the context in which the evaluation is done in designing the study.

The Network of Networks on Impact Evaluation (NONIE) suggests that for example an analysis of the *contribution* of a program toward achieving its outcomes might be the most appropriate evaluation approach for a program at its early stages, when its design is still changing based on the feedback received from monitoring and process evaluations. It is often not advisable to isolate all the other factors except those from the program, when the program theory design is still undergoing an initial efficacy trial. On the other hand, once the *modus operandi* of a project has been tested and the design is more standardized, it is possible to test its validity and measure the impact produced by the program, through conducting a rigorous causal *attribution* analysis. (See Box 3 for NONIE's definition of the terms)

---

**Box 3. Definitions**

**Causal Attribution** assesses the change in outcome that can be attributed to the program and not to other factors. Many experimental and quasi-experimental designs assess causal attribution.

**Casual Contribution** assesses whether or not the program is one of many causes of the observed change. Causal contribution develops a theory of change with testable hypotheses and predictions telling a credible story regarding how an intervention affects its intended results. (NONIE)

---

Although there is some flexibility in methods, one can establish some general principles of good-practice for impact evaluation.  These include the following:

It is essential that the impact evaluation demonstrate **causality** – that is, a plausible demonstration that the intervention was responsible for the observed outcome in an empirically verifiable and objective manner.  This outcome must be defined relative to a

(preferably explicit) **counterfactual,** to which outcomes of the program can be compared.

The methodologically simplest and most robust way to establish the counterfactual is by using an **experimental approach**, randomly assigning individuals either to receive treatment or to serve in a control group. Randomized control trials should be accompanied by thorough process evaluations and in-depth (qualitative) interviews with beneficiaries in order to avoid becoming "black box" evaluations. Demonstrating causality without random assignment is possible, but considerably more difficult.

All **prospective evaluations** require the collection of **baseline data**, which provides the benchmark against which changes can be measured to discern impact. Baseline data should be collected for both beneficiary and control groups. Retrospective methods, in which one evaluates the beneficiary and control group after the program has been implemented, and then estimates program impact statistically, may also be used to determine impact, but they rely both on more complex assumptions about the populations under study and more complicated analytical tools.

In terms of methods, it is recommended to use more than one method in order to search for regularities in the research data and cross- check the results.Many participants agreed that the best results can be reached by a **mixed methods** approach, which combines well contextualized **qualitative** studies with **quantitative** rigor (World Bank Independent Evaluation Group, IEG). The integration of methods should be carried out during each step of the impact evaluation. Using a mix of quantitative and qualitative data sources (e.g. administrative data and self-reported data), types of data (quantitative vs. qualitative), sampling methods (random and purposeful), measurement instrument (e.g. questionnaire and interview) and analysis methods (e.g. statistics and interpretation) to complement and balance each other, strengthens the quality of evidence and can provide a confirmation of

---

**Box 4. Definitions:**

**Cost-Effectiveness and Cost-Benefit Analysis**

Cost-effectiveness (CE) and cost-benefit (CB) analysis both measure the social returns to investments. CE analysis usually denominates benefits in physical units and CB in money equivalent terms.

Cost-effectiveness analysis estimates the program's total costs (financial costs of the inputs, economic costs of free inputs such as volunteer time, cost of public funds used to provide the benefits, and discounting all of them) and uses it to estimate the volume of outputs per dollar spent on the program. The effects are measured in physical units and they are ideally the intended goals of the program such as number of employed youth. In practice, the measures are often proxies for these goals, often in terms of program outputs (e.g. number of young people who received employment training).

CE denominates the benefits of an intervention in unit terms. CB on the other hand provides information on whether the value of the benefits exceeds the cost of producing them. CB analysis permits the comparison of interventions in unrelated areas, and interventions whose benefits and costs accrue to different people. Determining the value of the benefits delivered to participants is often difficult. Where available, the analysis uses market prices. Where market prices do not exist or are a poor indicator of value, it may be possible to estimate the value by using other methods such as hedonic pricing or contingent valuation.

(Cunningham et. al. 2008)

results obtained through each method independently. Qualitative methods can inform quantitative, and vice versa, and using mixed methods can thus deepen the understanding and interpretation of results, and call attention to factors that might have been missed or misinterpreted by single-method evaluation. (NONIE)

Demonstrating impact is necessary, but it is not sufficient for making decisions on policy. It is also necessary that program evaluations include some understanding of the **cost-effectiveness** of the program, relative to other possible means of achieving similar results, and of the **cost-benefit ratio** of the program, relative to alternative uses of scarce public funds. (See Box 4.)

Generally speaking, the steps to be taken to conduct a good evaluation of youth development intervention can be described as follows:

- A description of the intervention, including its planned activities, methods of delivery, duration and expected outcomes
- Clearly defined goals and indicators that directly reflect these goals (outcome indicators)
- Clearly defined indicators of how well the program is being implemented (process indicators)
- A baseline survey of the population being studied, both beneficiary and comparison (control) groups, which should be conducted before individuals are assigned to each group
- The identification of a comparison group that is as similar as possible to the treated group in all key observable characteristics (random assignment is preferable but is not always possible)
- Periodic follow-up surveys both to monitor process and to evaluate outcomes
- A careful accounting of all the program's costs (this is often overlooked, but is necessary to understand whether the program is successful in economic terms as well as in strictly outcome terms)

(Cunningham et. al. 2008)

### 3. Existing evidence on youth service programs

A literature review conducted by ICP of research on youth service programs reveals that there are very few formal, publically available evaluations of national youth service programs, and almost none outside North America and Europe. The review also illustrates the lack of high-quality data and of standardized criteria for gauging program effectiveness in the evaluations that do exist. The majority of the existing studies take the form of case studies, describing the program rather than rigorously evaluating program impacts. (ICP Framework, Annex D)

Another review of youth voluntary service in Europe conducted by AVSO/proMENTE found the research base and practice to be surprisingly poor, considering the long tradition of youth voluntary service programs in the continent. "In spite of some encouraging results, overall the research conducted to date on the impact of voluntary service has had neither the methodological teeth nor the mandate to really test whether voluntary service works as advertised." (Steve Powell, President and Senior Researcher, proMENTE Social Research) The review included 300 documents of published and unpublished research and evaluation studies and revealed that very few studies mention any explicit evaluation framework. Nearly all use retrospective impressions. Encouragingly, those few studies that did use more objective methods to evaluate youth services revealed some significant benefits, although the measured benefits were much lower than those estimated by retrospective reports.

The AVSO/proMENTE study reinforced the general impression by the meeting participants of the varying quality and generally low comparability of the existing studies. Evaluations differ significantly in type, method and purpose. In addition, the results cannot be applied across cultures due to problems of translation or validation. This renders any formal meta-analysis practically impossible.


### 3.1 Methods used in existing evaluations

To better understand the existing evidence base, ICP collected a selection of the available evaluations to identify the instruments and methods used in each study. Table 1 below categorizes the studies under the commonly used program outcomes for youth service. (Please also see Annex C for ICP's Framework, which includes annotated bibliography of the studies).

Table 1. Existing Youth Service Evaluations

| | Program Goal | Assumptions | Indicator of Success | Instruments | Existing Evidence | Methodology |
|---|---|---|---|---|---|---|
| **Y O U T H  D E V E L O P M E N T** | **Increased participation in civic and political life** | Youth service empowers young people to become more actively engaged in their communities and provides them with a means to do so, increasing the likelihood that they will stay involved in the future | More hours spent volunteering or participating in other forms of civic life; more meaningful/sustained/long-term volunteer commitments; Increase in voter participation among participants in youth service programs; membership to voluntary/community/local associations | - Youth surveys (pre and post, when possible) <br> - Interviews or focus groups with participants | AmeriCorps | Longitudinal study with comparison group |
| | | | | | Unis-Cité | Pre/Post Survey |
| | | | | | Arci Servizio Civile | Longitudinal, Retrospective Surveys |
| | | | | | City Year | Longitudinal Study with control group |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| | **Life Skills Development** | Youth service helps build life skills such as: a sense of agency, independence, self-confidence, leadership, teamwork, emotional and physical well-being, decision-making and communication | Increased sense of agency, independence, self-confidence, leadership, teamwork, emotional and physical well-being, decision-making and communication | - Observational interviews with participants, parents, teachers <br> - Youth surveys (pre and post, when possible) | Katimavik | Retrospective interviews and surveys; Pre/Post surveys |
| | | | | | Green Corps | Retrospective |
| | | | | | Unis-Cité | Pre/Post Surveys |
| | | | | | AmeriCorps | Longitudinal study with comparison group |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| | | | | | Arci Servizio Civil | Retrospective Interviews |
| | | | | | Green Corps | Retrospective |
| | | | | | AmeriCorps | Longitudinal study with comparison group |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **Prevention of Risky Behaviors** | Youth service gives young people a more structured way to spend their out-of-school time; it also helps them to make better choices by equipping them with better decision-making skills. | Decrease in teen pregnancy; decrease in STDs/HIV; decrease in risky sexual behavior or increase in use of condoms; decreased substance abuse; decrease in participation in gangs | - Youth surveys (pre and post, when possible)<br>- Observation<br>- Statistical comparison (local statistics) | Student Partnerships Worldwide | Retrospective, participatory |
| | **Intercultural Competence** | Youth service exposes young people to a more diverse group of people and provides them with training and tools for interacting with different populations | Improvements in cultural competency; Improvements in youth-adult partnerships and intergenerational communication; Improved tolerance and empathy for others | -Pre and Post tests<br>-Interviews with participants<br>- Youth surveys (pre and post) | Katimavik | Retrospective interviews and surveys; Pre/Post surveys |
| | **Improved Education Attainment** | Youth service enhances young people's commitment to education by providing them opportunities to apply what they are learning to real life situations | Increase in grades; increased literacy levels; improved academic knowledge/skills; improved attendance or graduation rate; improved classroom behavior | - Grades<br>- Teacher surveys<br>- Attendance records and graduation rate | Student Partnerships Worldwide | Retrospective, participatory |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| C O M M U N I T Y   D E V E L O P M E N T | **Community Strengthening** | Youth service strengthens linkages between community organizations and young people and improves community morale and cooperation as well as perception of youth in community | Enhanced organizational capacity, stronger civil society, enhanced perception of young people in society | - Focus groups, interviews<br>- Organizational assessments | AmeriCorps | Quasi-experimental |
| | | | | | Learn and Serve America (1999) | Quasi-experimental |
| | | | | | YouthBuild, Inc. | Pre/post |
| | **Peace-building – fostering unity and trust** | Youth service can build/restore trust between participants and communities, particularly when used in post-conflict settings | Reintegration of child soldiers into the community, increased dialogue and cooperation in the community, reduction of inter-group conflict | - Community surveys | | |
| | **Disease Prevention** | Youth service can focus on peer education to reduce risky behaviors, which can reduce STDs or other preventable diseases | Reduced incidence of diseases in the community | - Statistical comparisons<br>- Surveys | | |
| | **Infrastructure Development** | Youth service can mobilize young people to contribute to infrastructure development through | Enhanced community infrastructure, housing, roads, water supply etc. | | AmeriCorps | Quasi-experimental |

| | | | | Un Techo para mi Pais | Experimental |
| | | | | YouthBuild, Inc. | Pre/post |
| | | | | National Youth Service South Africa | Post |
| **Literacy and Education** | Youth service that is targeted toward education and literacy can significantly enhance literacy rates in a country | Increased literacy rates | Statistics, interviews with teachers, students | National Youth Service South Africa | Retrospective |
| **Environmental Restoration** | Environmental service can have positive effects on restoring environmental damages and potentially also on preventing future damages through education and retrofitting | Restoration of environmental damages, such as cleaner waterways, trees planted, etc | | Green Corps | Retrospective |

14

## 3.2 Examples of Existing Evaluations

Six studies listed in the table above were presented in more detail to the meeting. They are briefly described below as examples of the varying evaluation techniques available for different programs. They all provide important information about the extent of existing evaluations and the need to take context into account when determining appropriate evaluation techniques. The different methods used by the programs represented at the meeting also highlighted the importance of seeing evaluation as a process, not just as an end result. While impact evaluation was the focus of the meeting, there was broad recognition that not all programs represented were at a stage where impact evaluation is appropriate. In many instances, the programs were in the process of expanding or conducting formative evaluations to assess program effectiveness rather than impact.

### 3.2.1   *Experimental and Quasi-Experimental Evaluations*

Two of the examples presented attempted to measure the impact of the programs by using experimental or quasi-experimental techniques.

An example of the experimental technique i.e., **randomized control design,** was presented by the non-governmental organization *Un Techo Para mi País (UTPMP)*. This organization is evaluating the impact of home upgrading by youth volunteers in Latin America. The youth-led, volunteer-based organization provides improved and emergency housing for poor households in several Latin American countries. The organization has been evaluating the impact of the program on the well-being of poor families. The baseline survey was

> **Box 5. The Objective of the 'Un Techo Para mi Pais' - Project Evaluation is to Measure the Impact of the Volunteer Experience on:**
>
> **a) Volunteer activities:**
> - Participation in future UTPMP interventions and programs.
> - Participation in other volunteer initiatives.
>
> **b) Perception of social reality in Latin America/social "awareness"**
>
> **c) Labor market decisions**
> - Percentage of ex volunteers who work in public sector.
> - Percentage of ex volunteers who work in NGO sector.
>
> **d) Civic participation**
>
> **e) Educational outcomes and economic behavior**
> - Educational attainment
> - Income
> - Social investment and donations.

collected in 2007-2008. In addition, the evaluation team is currently designing a component to evaluate the impact on the youth volunteers. The baseline study for this evaluation will be scheduled as soon as there is sufficient excess of volunteers to be able to conduct adequate randomization, and the financing for the project has been secured. The core objective of the volunteer component is to expand the social awareness of Latin American youth by offering volunteer opportunities working to improve the lives and livelihoods of the poor. It provides an important opportunity for the youth volunteers - mostly of privileged backgrounds - to interact with different social classes and experience first-hand the realities of living in poverty. There is anecdotal evidence of a "life-

altering" experience for some volunteers; it is hoped that the evaluation will confirm these stories.

The evaluation is designed to be a randomized experiment/control trial. In the recent years, the number of volunteers has on occasion exceeded the slots available, and the program attempts to generate further oversubscription through intensive advertisement/promotion to key target groups such as college students to make the creation of control groups possible. The objective is to have a sample size of 400 eligible volunteers selected for the program, and 200 to 400 for the control groups.

The second example, a recently conducted longitudinal study of the **AmeriCorps** program used a **quasi-experimental design with matched-comparison groups**. AmeriCorps is the largest service program in the United States, with 75,000 participants each year, of whom the majority are between the ages of 18 and 24. The volunteers serve through hundreds of nonprofit and civil society organizations in a variety of service programs. The purpose of the study was to assess the long-term impacts of participation in AmeriCorps on the members who serve.

The research used quasi-experimental design, comparing around 2,000 individuals who served full-time in over 100 programs in 1999-2000 to close to the same number of similar individuals who did not participate in the program. To ensure the close similarity of the beneficiary and comparison group members, the comparison group was composed of people who had indicated knowledge of and interest in AmeriCorps by contacting its toll-free information line and requested information about the program, but who did not enroll during the study period.

Preliminary analysis showed that the beneficiary and comparison groups were similar in age and other characteristics measured at baseline. To minimize the possibility of selection bias and to support causal inferences, the comparison groups were selected using propensity score analysis. (PSA) The study collected information about the background and motivational characteristics that might have affected the individuals' selection into the beneficiary group and the outcomes of interest. Examples of these characteristics include exposure to service during childhood and prior participation in service. This information was used, along with participants' baseline characteristics, to create a measure of each respondent's likelihood to join the program (the propensity score).

Survey data was collected at four time points: baseline survey was administered in 1999-2000, end of service survey in 2000-2001 and 2004, and finally in 2007. The impact of participating on the participants was estimated as *difference in differences*. The study analyzed the program outcomes in terms of the changes between the baseline and post-program values of the same measures, and these changes were compared between program members and comparison group members.

### 3.2.2    *Non-Experimental Evaluations*

In addition to the routine program monitoring and process documentation, some youth service organizations systematically collect baseline and post-project data. This is an important step toward documenting program effectiveness and results.

The evaluation of the *"International Youth Foundation-Nokia Global Youth Development Initiative"*, conducted as an external study by Brandeis University, is an example of a **before-and-after evaluation**. The models of the IYF-Nokia program include youth volunteerism/community service (e.g. the Central European Volunteerism Model); youth leadership; self-expression using arts, media and technology; formal/non-formal education; and workforce development. The program has included 26 countries since 2000.

The study surveyed a total of 3,500 youth (ages 15-25) in 13 of the 26 countries on five continents during 2005-2006 with an objective to measure the impact of the initiative on target youth by documenting life skills changes and other outcomes. Data collected included demographic information, life skills development, schooling, employment, volunteering, aspirations, and opinions about the program. Surveys were conducted before the program (baseline), at the completion of the program, and 3-6 months after the program completion, and once retrospectively.

---

**Box 6. The Outcomes and Indicators of the IYF-Nokia Central European Volunteerism Model Evaluation:**

**Individual level:**
- Increased life skills (team work, communication, conflict management)
- Increased project management skills (planning, implementation, budget management)
- Increased volunteerism/contribution/leadership in communities

**Community level:**
- Tangible benefit of youth projects/activities on the communities
- Youth-adult partnerships formed (parents, local authorities, etc)
- Community perception of young people and value of volunteerism

**Sustainability:**
- Youth projects continue activities with local support
- Integration into formal education systems (as applicable)
- Youth continue volunteering in their communities

---

In addition, in-depth assessments were conducted in the Czech Republic, Hungary, Poland and Russia, which have programs that promote youth volunteerism and civic engagement. In these programs, youth groups (5-10 youths aged 15-25) are asked to submit proposals for projects that address their communities' needs. The groups receive small grants, training and mentoring from an NGO, and the youth themselves lead and implement the projects over a six month period, practicing life skills, project management skills and leadership. Youth are also encouraged to continue service to their communities.

The in-depth assessment included focus groups/interviews with youth, program staff, parents and community members; site visits by evaluators; and review of existing evaluation and other data.

The evaluation of *Student Partnership Worldwide's (SPW) Integrated Approach Youth Programs* in Africa represents an evaluation with strong **youth participation**. SPW engages young people in all levels of programs, including monitoring and evaluation in the highest technical level. SPW is an international development organization mobilizing young people in eight countries across sub-Saharan Africa and South Asia by recruiting and training 18-28 year olds to serve as Volunteer Peer Educators (VPEs). The VPEs live in rural communities for 6-12 months and lead health, environmental, and education programs. SPW currently has almost 1,000 volunteers who are estimated to reach approximately 400,000 youth. The monitoring and evaluation (M&E) efforts aim to capture the intended outcomes of the programs, as well as the unintended outcomes which derive from the intended ones, e.g., life skills component that has proven to have an impact on productivity, health and ambition in addition to the intended increase in life skills. The organization also measures the influence of the programs and the youth on support groups such as teachers and parents, line ministries and departments, opinion and pressure groups, and general community. Data are collected using closed-ended questionnaires, focus group discussions with volunteers, teachers, and other community members, and process monitoring.

SPW believes that the full-time, high-level participation of young people improves the reliability of the measurements. The long-term presence of the young people in the communities permits careful evaluation. In addition, having former volunteers as a resource allows for regular monitoring and feedback, structured surveys among the former volunteers, focus group discussions with

---

**Box 7. The outcomes and indicators used by SPW include:**

**a) Objective 1: Young people will take up significant roles in the policy-making processes that affect their lives**
- Number of young people whose participation in local, national, regional, or international events was facilitated by SPW (disaggregated by VPEs, ex-VPEs and other young people)
- Number of meetings/events organized by other stakeholders (including government, donors, and INGOs/NGOs) where SPW staff/ VPEs were invited to represent youth within the community
- Number of advocacy meetings organized/facilitated by SPW, with relevant government, donors, INGOs and NGOs to advocate the inclusion of young people in the policy formulation process and operations in the country in SPW strategic areas
- % of young people in the National Boards and Trustees of SPW international/country offices
- % of staff under the age of 35

**b) Objective 2: Young people will make responsible decisions regarding their sexual and reproductive health (SRH)**
- % of adults in communities reported supporting a comprehensive and holistic SRH approach for young people (disaggregated by parents, teachers, etc.)
- % of schools where SRH lessons are being delivered to minimum standards by teachers/school staff
- Number of trainings organized for teachers and other government staff for sustainability of interventions
- Number of strategic advocacy meetings carried out with government departments/officials to advocate for adequate attention on SRH needs of young people

**c) Objective 3: Young people will have improved life-skills and livelihood opportunities**
- % of SPW VPEs who are either engaged in higher studies or in economically productive jobs within one year of completion of VPE assignment
- % SPW staff who are former VPEs
- Number of VPEs/Ex-VPEs reported participating in events/activities to build their management skills

support groups such as parents, long term case studies, and evaluations and post-evaluation of the programs.

Finally, the Italian organization **Arci Servizio Civile (ASC)** conducted a **beneficiary assessment** of its national civic service program in 2006 together with the Italian Institute for Social Research (IRS).  The one-year program targets young people between the ages of 18 and 28, who participate in projects implemented by non-profit organizations and the public sector.

The objective of the ongoing monitoring and process evaluation of the program is to measure the quality of the training provided, the achievement of the project's aims, as well as the skills and competencies acquired by the young people. Thirty percent of the volunteers are asked to respond to an online or phone survey within three months from the start of service, and to fill out a monitoring assessment form at 5-6 months of service and at 9-10 months of service. The volunteers are asked to assess the training provided, learning that has taken place, and the project's progress. At 9-10 months they are also asked to compare their values before and after the program.

The purpose of the beneficiary assessment conducted two years after the implementation of the program was to analyze the opinions and attitudes of the participants towards their experience after they finished serving, and to measure the effect of service on the young people in terms of their personal and vocational development and employment possibilities.

The sample of 328 volunteers was drawn initially based on the geographical area of residence. The names of the volunteers were then selected by lottery within each area. Most of the questionnaires were collected through the internet and those who did not have access to the internet were interviewed by phone. The sample was statistically representative of the population of young people who served in ACS in the year in question, and the final number of interviews gathered was sufficient to guarantee that the estimates based on the replies were statistically meaningful. The outcomes measured can be found in Box 8.

> **Box 8. Outcome measures of the IRS Survey of Arci Servizio Civile (ASC), 2006.**
>
> **a) Skills acquired during the civic service period**
> - Ability to deal with different situations
> - Relations with others
> - Ability to work in a group
> - More mature attitude
> - Awareness of territorial problems
> - Vocational skills
> - Civic awareness
> - Time management
> - Use of PC
>
> **b) Social participation by association**
> (e.g. sports, parish, cultural, religious, volunteer work, student, environmental and political associations)
>
> **c) Current employment situation with respect to the situation at the start of the service**
> (student, employed, working, other)

The study measured the variation in participation by calculating the difference between self-reported social participation before and after the program participation, obtaining a

value zero if the behavior of the interviewee did not change, negative value if the ex-volunteer had reduced his/her social activity, and positive value if it had increased.

### 3.2.3   *Cost-benefit Analysis*

There are very few cost-effectiveness or cost-benefit studies conducted on youth voluntary service programs.  Most evaluations that refer to the costs and benefits of the program, mainly estimate the ratio of direct program costs per participant rather than the socio-economic costs and benefits.

One exception is the **YouthBuild cost-benefit evaluation** in the United States. Different YouthBuild programs in the United States have been evaluated at several occasions, using a variety of approaches.  Evaluation of the 'Targeted Intervention Program for Youth Offenders: The YouthBuild USA Offender Project' in 2008 provides an example of a cost and benefits analysis, looking at the monetary value of saving a high risk youth.

In general, the YouthBuild programs target low-income young people ages 16-24, of whom many have a disadvantaged background. Participants spend 6-24 months in a full time program dividing their time between a construction site building affordable housing and developing job skills, and a YouthBuild alternative school working toward a GED or high school diploma. Non-profit organizations sponsor most of the programs. YouthBuild programs include counseling and support services, basic education, leadership and service building community assets, technical employment training, exit opportunities follow-up and support.

The **cost-benefits analysis** of the YouthBuild USA Offender Project collected data on 388 participants in 34 programs over 2 years. The cost of the interventions was US$18.2 million. The socio-economic benefits of preventing crime, heavy drug use, and dropping out of high school were all estimated.

Several other evaluations of different YouthBuild programs have been conducted, and the programs are regularly monitored. The YouthBuild Affiliated Network has specific percentage-wise performance standards to monitor and evaluate the program's success. The performance standards include attendance and retention rates, job placements of completers, average wage for placed completers, educational achievement (GED or diploma), voter registration, and educational testing.  **A formative evaluation** of the YouthBuild Demonstration Project was conducted in 1996, and a **before and after survey** of 900 YouthBuild graduates in 2004 looked at indicators such as having a GED or diploma, using or selling marijuana or hard drugs, arrests, convictions, and being homeless. An evaluation of the YouthBuild Youth Offender Grants was conducted in 2006.

## 4. Fundamental challenges in evaluating youth voluntary service programs

### 4.1 Conceptual challenges

Through the course of the meeting, it became clear that a significant barrier to developing a common framework for evaluating the impact of youth service programs was the absence of common definitions and guiding principles for what constitutes youth service.

In order to understand the evaluation options for youth service programs, it is first important to develop a set of principles to serve as the basis for a common definition of what constitutes youth service, and outline the wide range of program models and goals. For example, programs differ in whether they are voluntary or mandatory, serve advantaged or disadvantaged groups, are school-based, government-run or NGO-led, and paid or unpaid. Programs also differ in the way they address youth participation and youth voice in program leadership and design. And young people can participate in a number of different areas, from environmental service to sexual and reproductive health. A typology would examine all of these different program models and form the basis of an evaluation framework that looks at different program options and goals, and suggests indicators and evaluation tools to measure the quality and impact of these programs.

> **Box 9. Definitions**
>
> **Youth Service:** "An organized period of substantial engagement and contribution to the local, national or world community, recognized and valued by society with minimal or no monetary contribution to the participant." (Global Service Institute)
>
> Common elements of the programs include:
> - a participant's engagement in service is frequent
> - the commitment to service is consistent over time, rather than episodic
> - the service may be remunerated in some way, but at less than market rates
> - the program is structured
> - the program addresses either difficult public problems or needs that have been defined collectively (adapted from Perry, J. L., & Thomson, A. M. (2003). Civic service: What difference does it make? Armonk, NY: M.E. Sharpe.)

It is also important to take into account the enormous variation in cultural, social, and political context, especially in the developing countries. Youth voluntary service takes place around the world in different forms, but is often not called youth service. Thus, when developing a common definition for youth service, it will be important for the field to take into account the different contexts in which these programs operate. For example, some countries in Africa use youth service to promote employability and others use it to promote social cohesion and discipline. These differences inform policy options, funding strategies and institutional arrangements. Like many other social programs, the size and shape of youth service programs depend of public policy choices.

Related to the variety and diversity of programs, it was suggested by some meeting participants that when developing youth service programs and policies it is important to engage the highest levels of decision makers to develop commonly recognized

goals/objectives.  These development goals would then guide the development of the specific program outcomes.

Another challenge is the limited amount of available research on youth service and the need to develop a global research agenda. The field of youth service draws on a number of disciplines, including youth development, youth leadership, resiliency, service-learning, environmental education, community development, values development, global ethics, adolescent development, and other interdisciplinary fields.  Evaluations, too, are conducted from a similarly eclectic range of disciplines. While it is worthwhile to look to other disciplines researching youth service as a component of their larger framework, it is also important to build a field of research focused specifically on youth service.

## 4.2 Technical challenges

Many specific technical challenges to evaluating youth voluntary service programs were discussed in the meeting. The challenges can be divided into those in defining what to measure, those in measuring, and those in defining the attribution or contribution of youth service programs.

### 4.2.1   Choosing and defining what outcomes to measure

A clearer theoretical model is a prerequisite for designing impact evaluations. What is the program trying to achieve, and how does it intend to get there? It is of particular importance in order to develop a clear result chain or logical framework for different types of youth service program interventions, and to define measurable outcome indicators.  The meeting highlighted specific challenges in defining outcomes of youth service programs.

- Most of the currently available evaluations and impact evaluations only measure the **impact on program participants/volunteers**, while the effect of the project is also on the communities that the volunteers serve. Can indicators be developed that capture the **impact on the communities** as well?  Would it be an overly complex task considering that there can be several types of services provided through one project?

- Most currently available evaluations only look at short and/or medium term results. However, some of the project outcomes are expected to be observed in the long term. This raises the question about the most useful and important **short-term and long-term outcome indicators**.

- There is no consensus on the expected outcomes of the volunteer programs. Available programs intend to affect young people in numerous dimensions – health, educational attainment, healthy behaviors, civic values and engagement and so on.  Even within the "civic engagement" or "citizenship" realms, different researchers and programmers have different ideas about the programs' goals and indicators.  What are key **direct and indirect outcomes**?

- There is lack of shared understanding about the **meaning of some commonly stated outcomes** of youth interventions. "Life skills", for example, can mean a wide diversity of skills, including problem-solving ability, communication, cohesion, attending/having a role in community meetings, capacity to make decisions about career, patterns of charitable giving, and ethic of service in the community. The challenge is to define the expected outcomes and systematize their use.

### 4.2.2  *How to measure outcomes*

Another aspect of the technical challenges is related to measurement and data collection tools. The intended outcomes of the youth service programs may not be easily measurable. It must be noted that this may reflect some uncertainty about what the program is trying to achieve.

- Some of the most important impacts expected from the youth voluntary service programs are **not tangible or easily measurable using available tools**. What are good indicators of young peoples' connectedness to the society, sense of identity and sense of belonging, or improved life skills for example? Do more easily measurable behavioral proxies exist for these subtler outcomes (e.g. incidence of volunteering, "taking initiative," exhibiting leadership, joining groups or associations)? Should we look at changes in the community's perception of youth, and should the unit of observation be the community rather than the individual?

- Defining the expected outcomes too narrowly – especially when using quantitative methodologies - can leave out important unexpected outcomes. On the other hand, if the right indicators are used and evaluation is appropriately designed, it can also help capture both positive and negative effects that were not expected. How should the evaluations be designed in order to **capture unintended outcomes**?

- Due to the nature of the programs, all of the evaluations reviewed during the meeting primarily relied on **self-reported information,** either through surveys or interviews, rather than on more objectively-measured outcome measures. Young people may not be able to identify or appreciate changes in themselves. Self-reported data on personal perceptions and beliefs may be biased, or at least incomplete. To address this problem, evaluators compare changes in participants relative to changes in the control group (with assumption that both groups contain similar biases). In addition, the organization can use administrative data such as wages, criminal justice and health data but there are significant administrative barriers to obtaining this type of data. The reliability of self-reported data is affected by ambiguities or incompleteness in the way questions are asked, problems of translation or context. In general, simpler questions yield more accurate information, but may miss important subtle nuances. For example, "Did you attend school this week?" is a very different question from "How do you feel about democracy?"

A positive finding from the review of European impact studies conducted by AVSO/proMENTE was the use of many different sources by some of the studies, and

inclusion of information from volunteers and/or implementing organizations by all of them. However, none of the studies included direct data from the users/beneficiaries of the services. Twenty-five of the studies used qualitative and 30 used quantitative data collection and analysis methods, and some a mixture of the two. (Steve Powell, President and Senior Researcher, proMENTE Social Research)

### 4.2.3 *How to establish attribution*

The most critical technical challenge in evaluating the impact of youth service programs is demonstrating the causal attribution of the program. Trying to understand what would have happened in the absence of the program can be technically challenging, especially if it has to be done *ex post*, after the program has started. Demonstrating the causal attribution or contribution rigorously, without an explicit counterfactual can be technically even more difficult, because it requires a very solid logic model and the ability and tools to test and verify it.

- **Fundamental theory or logic model:** Many programs, and most evaluations, lack an explicit theory of change. When elicited, the underlying and implicit theories are often complex and are not easily expressed in empirically verifiable or evaluable terms. The first step is to clarify, not simplify, the assumed causal links and test these hypothesis.

- **Internal Validity – robustness and reliability.** Without an explicit, objective, and measurable counterfactual, it is difficult to establish that the observed outcomes are really due to the program, and not to something else. In many cases, it is not enough to compare beneficiary outcomes before and after the programs, or beneficiaries to non-participants.

   Because participation in service is often voluntary (self-selecting), those who are more active to start with are more likely to apply; they are also more likely to exhibit the sort of behavior program designers value. More importantly, they are more likely to exhibit this behavior even if they do not take part in youth service. Therefore, attributing differences in outcomes to the program is likely to be misleading. The AmeriCorps study attempted to minimize this potential source of bias by constructing comparison groups using propensity scores to ensure that the groups are on average similar.

   The most robust way to ensure internal validity is by using an experimental design with random assignment. The Un Techo Para mi Pais evaluation team is planning to construct comparison groups through randomly assigning young people who applied to take part in the program to either beneficiary or control groups. However, this method requires that more people apply for the program than can be accommodated. The program need not have the capacity internally to conduct the evaluation, but in that case it must have the resources to contract with an evaluation firm or academic institution to undertake it.

- **External validity – relevance and replicability.** When interpreting the results of summative youth service program impact evaluations, one must keep in mind that the results produced from the evaluation will be most applicable to other groups and situations that are similar to the one from which the randomly or otherwise generated sample was selected. For example, many programs are designed to recruit young people from socially and economically advantaged backgrounds. In that case, the measured outcomes may not be relevant for the general population or for those from more disadvantaged backgrounds, who form the majority of the youth in most developing societies.

- **Ceiling effect.** Given declining marginal benefits of participation, improvements in the outcome indicators will be smaller when the baseline values are already high. Young people from relatively advantaged backgrounds, who participate in the service programs are typically highly civically engaged even before they participate in the programs, leaving little room for growth. The programs might therefore not show strong impact. Conversely, evaluations such as the AmeriCorps study, show higher impacts among minorities and young people from backgrounds that afford them fewer opportunities for growth and development.

- **Fidelity:** The services that volunteers provide are very different – even within programs. Participants in larger youth service programs can work with many different NGOs and CSOs, which offer different services to the community. Thus the "treatment" or experience that the volunteer participants receive also differs. The question was raised in the meeting about the applicability of randomized experiment in evaluating youth service in the presence of non-uniform treatment. The AmeriCorps evaluation offers one answer. Within the AmeriCorps program there are several different ways to serve. In the study, the beneficiary sample included more than 100 programs, which did not share many elements. To address the issue of fidelity, AmeriCorps collected a vast list of possible outcomes, and attempted to evaluate them all in an attempt to identify the commonly shared ones. This, together with a strong logic model and a theoretical framework for why a program leads to certain outcomes, helped to overcome the issue of fidelity.

- Other challenges discussed at the meeting included issues of treatment variation, such as the **intensity and duration** of participation (even if the treatment itself is uniform), and questions of measurement, such as **attenuation** effects (effects wearing off after a while). These are under-explored areas of research, and future evaluations might be designed to identify the prescription (what type of program works) and dose (intensity and duration of the program required to produce desired outcomes). Programs may unintentionally provide benefits to some who are not officially signed up to receive benefits. These **spill-over effects or externalities** can wreak havoc with evaluations, since they blur the distinction between beneficiary and control groups, but they are not necessarily bad for program and service delivery or for the beneficiaries themselves. Finally, individuals who are enrolled in programs may not **take up or complete** the program as intended. Young people are exceptionally mobile. One resulting problem is that young people in the control group may have less of a reason to hang around

simply to get measured, and differential outmigration can cause initially balanced and unbiased control groups to become severely biased.

## 4.3 Operational challenges

- **Limited capacity of youth service organizations:** NGOs operating youth service programs are often small, with limited capacity and resources for conducting experimental impact evaluations. When financial resources are limited and conducting an evaluation would require decreased funding for the project implementation, evaluation often comes near the bottom of priorities for the program managers. Most of the organizations also lack the capacity to design and conduct studies using technically more demanding quasi-experimental methods for example. Capacity to conduct longitudinal studies is particularly limited. This is not an excuse to avoid learning from experience, but it does beg the question of who should be conducting the evaluation (see below).

- **Youth participation:** There is no established practice for youth participation in evaluations, but it can yield important benefits. Discussions at the meeting highlighted the value of youth participation in designing surveys, collecting data, and validating results. Asking the youth beneficiaries how *they* would measure success, and getting the approval of the target group for the survey questions, may lend more credibility to the results. Young people may be more comfortable with young interviewers, leading to greater openness among the respondents.

  Student Partnership Worldwide (SPW) engages young people in all levels of monitoring and evaluation, and credits the success of its M&E efforts to the long-term presence of former program beneficiaries in the communities and their direct participation in the evaluations. SPW provides young people the skills and resources to lead technical M&E efforts. The challenges the organization sees are related to the recognition, financial resources, and commitment at the strategic level to the time and efforts needed to train and build the technical capacity of the youth. AmeriCorps, on the other hand, does not typically engage young people in conducting evaluations. AmeriCorps is a federal program, and its statute requires evaluations to be independent and conducted by individuals who are not directly involved in the administration of the program. To maintain a high degree of independence and credibility, evaluations are conducted by professional evaluators at research organizations. Youth often play a role in evaluations by helping researchers to understand how programs have impacted their lives and their communities (Kevin Cramer, Deputy Director, Research and Policy Development, Corporation for National and Community Service).

  The question of which youth should participate or who should represent the youth was also raised. Currently, in most cases, when young people are included, they are usually university students, and not the targeted or beneficiary youth, or youth from disadvantaged backgrounds.

## 5. Stakeholder perspectives on impact evaluation for youth service

Not everyone should invest substantial resources to producing evidence of impact through complex, comprehensive impact evaluations. Evaluation objectives and needs of organizations differ significantly based on their particular circumstances and program or policy elements. Some of the points made at the experts meeting by the participating stakeholders are outlined below.

- From the **practitioners'/civil society organizations'** perspective, monitoring and process evaluation are important for the project management to understand the day-to-day running of their policy. Evaluations are seen as having value as a useful learning tool, even when they cannot show causal attribution. Evaluation reports can also be used to demonstrate to the management the differential effectiveness of the program or policy for different beneficiary groups and where the gaps are. Some practitioners believe that the use of evaluations is limited, because they are incapable of measuring "transformative change" and other key youth development outcomes that are difficult to define *ex ante*.

  It is important to understand the history and culture of the society where the program is implemented in order to understand how the program operates, and to take into consideration the socio-cultural constraints for collecting certain data. Practitioners also see the evaluations as a tool to prove the programs' value to potential volunteers and their parents, to be able to recruit more young people.

- From the perspective of **larger NGOs** such as IYF and AmeriCorps, impact evaluations with experimental designs are seen as useful for generating evidence for governments and donors rather than deriving from the organization's own needs. The experience particularly in the United States is that the programs that have been evaluated with experimental, random assignment evaluations are gaining increased attention. However, the practitioners' experience also shows that demonstrating good practices and combining them with quantitative data and case studies can also be an effective tool in demonstrating the program's success to funders.

  Some of the practitioners at the meeting voiced the concern that donors and policymakers do not currently reward evaluation. Evaluation may be seen as something that is done only when the program is not perceived as working well, and may in fact be seen as a signal that the program is not working well. The potential down-side risks are not trivial. Ideally, stakeholders should be rewarded for using impact evaluation to close or amend weak programs, rather than feeling threatened as the bearers of bad tidings. In any event, the consequences of negative evaluations for all stakeholders should be made explicit and transparent. Paradoxically, it may be difficult to generate support to conduct (often costly) evaluations of programs that are perceived as doing well. However, governments also recognize the importance of evaluating to enhance program effectiveness and ensure that the money to run the program is being well spent.

This raises the question of who should conduct the evaluation. In-house evaluations can build support and commitment for the program, for the need for evidence and evaluation. On the other hand, independent evaluations may be more credible to external stakeholders. Ensuring the sustainability of the evaluation funding can prove to be a challenge, and there may be reasons to incorporate evaluation into national or international systems for planning and information.

- From **governments' and policymakers'** perspective, evaluation is often seen as something that is done when the program is not perceived as working well. From the point of view of the government representative from South Africa, it is not immediately possible to generate support to conduct (often costly) evaluations of programs that are generally perceived as doing well. However, governments also recognize the value of doing evaluations to enhance program effectiveness and ensure that money is being well spent. Key concerns for the policymakers include: who should conduct the evaluation, at what stage of the program and against which indicators. Lack of "tangible" outcomes, for example, can be a challenge for government representatives attempting to justify funding increases for these programs. (Busani Ngcaweni, Head: Special Projects & Acting Chief of Staff of the Deputy President, The Presidency, Republic of South Africa.) On the question of who should conduct the evaluation, for transparency and success of the evaluation, it should be executed by an independent party and not by the program management staff or policy makers who have a direct interest in the success of the program. It should also be clear how possible negative evaluation results could affect the program managers, policy makers, and other stakeholders. They should be rewarded for using impact evaluation to close or change weak programs, rather than the positions of the policymakers being dependant on the results. Ensuring the sustainability of the evaluation funding can prove to be a challenge, and the evaluation should therefore be part of national or international planning, indicators, and information system. (Dr. Yuanzhu Ding, Sr Research Fellow, Academy of Macroeconomic Research, National Development and Reform Commission and Professor and Director, Center for Volunteering and Welfare, Peking University, China)

- Some *international agencies* such as UNICEF advocate the development of a common, agreed upon set of youth indicators, which could be used to evaluate youth voluntary service, and provide the basis for international examination and synthesis. These commonly-recognized indicators would help to focus the efforts of all the different actors in the field, and particularly within the UN system. As a starting point, the different institutions should look at good practices in process evaluation, and consider building standardized models on that basis.

- For *donors and lending agencies* the questions of most importance include whether the project makes a positive and attributable impact and what is the most effective program among many alternatives. The World Bank and its client countries are increasingly moving toward evidence-based programming and policy-making. When working with country governments that have very limited financial resources and are looking at readjusting their portfolio, the programs and policies that have good evaluations and strong evidence are primary candidates for inclusion in the agenda.

This is necessary to increase the efficiency of expenditure and to make sure that the governments are getting a good return on their investment. Thus, the key question is whether youth service programs have a greater impact on disadvantaged young people than other youth development programs (such as mentoring, internship programs or formal training), or can deliver similar benefits at a lower cost.

To sum up, evaluations may be used to improve existing programs, document best practices and facilitate economic analysis such as cost-benefit ratios, encourage new investment, or convince funders to continue or cease to support a program. The challenge is to find a right combination of approaches that complement each other, and to reach the sufficient level of rigor in generating evidence that is at the same time useful to practitioners in enhancing the quality and impact of their programs, convincing and useful to the policymakers and their technical advisors, and attractive to funders.

Careful monitoring should be part of any organization's core activities, and most small youth service organizations should focus on evaluating their processes and learning from them. Evaluations should be considered even for small programs, always in order to learn from experience, and especially if the program has the intention and capacity to grow. In this case impact evaluation should guide the expansion, and it should be included at an earliest possible stage.

Even if impact evaluation is not conducted, it is beneficial for the program managers and designers to have some idea of the counterfactual from the very beginning. It is important to understand that the program impact is measured relative to something that is explicit and preferably objective and measurable. Program staff can observe near-by communities and their experience in addressing similar problems, or talk to young people who thought about applying to the program, or applied but did not follow through, to try to understand what kind of transformative experiences they have had at the same time that your program has provided stimulus to the beneficiary group. This helps to develop a sense of the causality and the project's effects.

## 6.  Next Steps – Developing the Research Framework and Agenda

Meeting participants concluded that there was a need to develop an international community of practice in the field of youth service, constructed around three main areas of focus: building a common conceptual framework, agreeing on a research agenda, and refining and adapting evaluation tools and methods.

### 6.1 Conceptual Framework

Discussions at the meeting revealed that there is a lack of consensus on what constitutes youth voluntary service and what logic model best captures the theory behind it.  No single model or theory of change has been commonly applied or accepted across programs, and the list of expected outcomes is large and diverse.  In order to develop a common understanding of what constitutes youth service, meeting participants agreed on several key questions still to be answered and agreed upon by the youth service community, including:

- Are there some essential elements or characteristics common to all youth voluntary service programs?
- What are the different categories or activities of youth service? What are their expected outcomes?

It was agreed that before being able to develop an actionable research agenda, this lack of clarity in defining youth service and creating the logic model should be addressed. Several concrete steps were suggested for building the field in this area:

- Develop an agreed-upon conceptual framework
- Create a typology of different program models and activities
- Contribute to ICP's global database of youth service programs
- Examine existing theories of change
- Agree on a set of basic principles for the design for high quality youth service programs

### 6.2 Research Agenda

The conceptual framework for youth service can be informed by ongoing research on youth service from different disciplines. Meeting participants agreed that a number of questions needed to be answered by the community of practice to inform the development of a global research agenda on youth service, including:

- What is the existing research base on the impact of youth service?
- What are the gaps in the knowledge?
- What should be the focus of future research?

Based on these questions, a number of active steps can be undertaken by members of the community of practice and those in related fields:

- Continue to build a comprehensive database of research on youth service, drawing on a number of different fields
- Develop clear principles, standards and norms of practice in research on youth service
- Develop a common research agenda and strategy

**6.3 Evaluation**

While the intent of this meeting was to focus on measuring the impact of youth voluntary service programs, it became clear that efforts must also focus on understanding the principles and purposes of output monitoring and process evaluation. Many organizations do not have solid monitoring and process evaluation systems in place, and are even further from understanding or having the capacity internally to conduct impact evaluations. In order to support more effective evaluations, and enhance the quality of youth service programs, we must better understand:

- What tools or instruments are available?
- Do these tools provide the necessary information, or must new tools be developed?
- How should evaluation methods be chosen?
- For those programs that wish to conduct impact evaluations, now or in the future, what do they need to do to prepare for it?

Based on the answers to these questions, members of the community of practice and/or working group can work to:

- Create a database of tools and instruments, with special focus on what they are used for, what they measure, and when they are most useful
- Agree on a schedule of standards of evidence for evaluation for various purposes
- Develop guidelines to help youth service programs determine when and how to conduct evaluations

**6.4 Follow-up**

Meeting participants and others interested in advancing the field of youth voluntary service will be invited by ICP to take part in an online discussion forum. Working groups will also be established to develop plans to accomplish specific tasks needed in the youth service field based on the conclusions of the meeting. These tasks include: developing the conceptual framework, typology, overarching logic model and guiding principles of youth service through gathering existing logical models and consultation with youth service programs; and developing a research agenda, standards of evidence and a formalized process of evaluation for these programs. These outcomes can then be used for impact evaluation of national youth service programs.

The group has pledged to meet again in two to three years to present the results of evaluations and research that were engendered by this meeting and share the increased knowledge of the impact of youth service - what works, under what conditions and for what cost – to refine the policy advice on the basis of continually-improving evidence.

**ANNEXES**

**Annex A. List of participants**

## Measuring the Impact of Youth Voluntary Service Programs: International Experts Meeting

# LIST OF PARTICIPANTS

| NAME | TITLE | ORGANIZATION |
|---|---|---|
| **Alan Melchior** | Deputy Director and Senior Fellow, Center for Youth and Communities | Brandeis University |
| **Ami Thakkar** | Program Director | International Youth Foundation |
| **Andrew Furco** | Associate Vice President for Public Engagement | University of Minnesota |
| **Arianna Legovini** | Senior Monitoring and Evaluation Specialist | World Bank |
| **Ariel Fiszbein** | Chief Economist | World Bank |
| **Arnie Wickens** | Director CSV Consulting | CSV |
| **Busani Ngcaweni** | Head: Special Projects & Acting Chief of Staff of the Deputy President | The Presidency |
| **Caroline Fawcett** | Advisor | Middle East Youth Initiative |
| **Charmagne Campbell-Patton** | Program Associate | Innovations in Civic Participation |
| **Chris Blattman** | Assistant Professor of Political Science & Economics | Yale University |
| **Constance Flanagan** | Professor of Youth Civic Development, Co-Director, Inter College Minor in Civic and Community Engagement | Penn State |
| **Eva Mysliwiec** | Founder | Youth Star Cambodia |
| **James Cathcart** | Chief Executive | British Youth Council |
| **James Radner** | Executive Director of The Boreal Institute at the Munk Centre for International Studies | University of Toronto |
| **James Youniss** | Professor of Psychology | Catholic University |
| **JoAnn Jastrzab** | Principal Associate | ABT Associates |

| Jon Terry | Consultant | The Search Institute |
|---|---|---|
| **Judith Torney-Purta** | Professor | University of Maryland |
| **Katie Green** | Youth Development Specialist | Peace Corps |
| **Karin Akins** | Program Coordinator, Make a Connection | International Youth Foundation |
| **Kevin Cramer** | Deputy Director for Research and Policy Development | Corporation for National and Community Service |
| **Licio Palazzini** | President | Arci Servizio Civile |
| **Linda Raftree** | Director, Youth Engagement and Action | Plan USA |
| **Lonnie Sherrod** | Professor | Fordham University and the Society for Research in Child Development |
| **Manish Subharwal** | Director of Monitoring and Evaluation | Student Partnerships Worldwide |
| **Mattias Lundberg** | Senior Economist | World Bank |
| **Maxwell Marx** | Director | SPW USA |
| **Michael Baizerman** | Director of Youth Studies Program | University of Minnesota |
| **Michael Patton** | Director | Utilization-Focused Evaluation |
| **Mike McCabe** | Chief of Programming, Inter-America and Pacific Region | Peace Corps |
| **Minna Mattero** | Consultant | World Bank |
| **Naoko Kataoka** | Consultant | World Bank |
| **Nicholas Hartley** | Director of Operations & Programmes | SPW |
| **Peter Levine** | Director | CIRCLE |
| **Ron Kassimir** | Associate Professor of Political Science | New School for Social Research |
| **Ryan Cooper** | Field Coordinator | Un Techo Para mí País |
| **Sally Prouty** | President and CEO | Corps Network |
| **Samantha Constant** | Communications Specialist | Middle East Youth Initiative |
| **Sebastian Martinez** | Economist | World Bank |
| **Shahrokh Fardoust** | Lead Economist | World Bank |
| **Silvia Paruzzolo** | Consultant | World Bank |
| **Steve Powell** | Country Researcher | ProMENTE |
| **Susan Stroud** | Executive Director | Innovations in Civic |

| | | Participation |
|---|---|---|
| **Sylvia Golombek** | Vice President of Programs | Youth Service America |
| **Tim Cross** | President | YouthBuild International |
| **Varun Gauri** | Senior Economist | World Bank |
| **Victor Karunan** | Senior Advisor Adolescent Development and Participation Programme Division | UNICEF |
| **Yuanzhu Ding** | Professor Director - Peking University | Research Center for Volunteering and Welfare |

**Annex B. Agenda of the Meeting**

<div style="border:1px solid black; padding:10px;">

# Measuring the Impact of Youth Voluntary Service Programs: International Experts Meeting

*May 8-9, 2008, World Bank HQ, Room I-200, Washington DC, USA*

</div>

## Background

Youth voluntary service programs exist in dozens of countries around the world, and new programs and policy initiatives are currently being developed in many others, often with the help of international organizations like UNICEF, UNV/UNDP, and International Youth Foundation. This will result in many more young people being offered the opportunity to engage in youth service. Already, fully one quarter of young respondents to the World Values Survey in 2000 report some voluntary action with religious or secular organizations that promote social welfare, conservation, human rights, and many other goals. These programs attract a great deal of attention, and potentially enormous resources. In December 2007, U.S. presidential candidate Barack Obama said that expanding opportunities for national service would be "a cause of (his) presidency". In 2008, the *Time* magazine made the case for National Service in its cover story and a call was issued in the *American Prospect* for a national youth service program, to enroll one million young Americans by the year 2020, at a cost of more than $14 billion per year, in today's dollars. We – the community of researchers, policymakers, and practitioners – have an obligation to make sure that we are doing the right thing in supporting these programs, and that we have the tools necessary for evaluating their impact.

## Overview

The World Bank and ICP are organizing this meeting as an opportunity to bring together, researchers, policymakers and practitioners from the field of youth civic engagement and particularly youth service, to assess the existing research on the impact of youth service programs on young people, explore different evaluation methods, develop a draft evaluation framework, identify gaps in the research, and develop a research agenda to address these gaps. This meeting will initiate a conversation among an international group of researchers, practitioners and policymakers about the evidence base for youth service as a strategy for positive youth development.

**Objectives:**
- Discuss methodological approaches for evaluating the impact of youth voluntary service
- Outline the main challenges in evaluating the impact of youth voluntary service programs
- Review the current research and available evidence on whether and how service contributes to youth development
- Identify the gaps in existing research on the impact of youth voluntary service programs
- Develop a draft framework for evaluating youth service programs, and
- Determine a plan to expand the evidence base in ways that are consistent with the highest standards in quantitative and qualitative research and also useful to

policymakers and practitioners in promoting best practices in youth service; identify the roles of different actors in this process.

**Outputs:**

- Draft framework for evaluating youth voluntary service programs, including discussion of appropriate methods, indicators and outcomes
- Synthesized recommendations, action plan, and endorsed commitments for moving the youth voluntary service research agenda forward
- Final meeting report summarizing presentations, discussions, and research agenda to be published and distributed through WB and ICP website and in print
- Dedicated space on World Bank Children and Youth and ICP's website will be used as a tool for building an international community of interest around the youth voluntary service program evaluation and for distributing knowledge
- Establishing a working group to carry on the work and keep the topic in the agenda of all the organizations that commit to working on it

**Outcomes:**

*Short Term*

- Shared understanding of the current evidence base
- Consensus about what evidence is lacking
- Development of a framework for evaluating youth service programs
- Creation of a roadmap for expanding the evidence base

*Long Term*

- Increased resources for research on the impact of youth service as a youth development strategy
- Rigorous evidence on whether and which programs are most successful

# Agenda

## Day One – May 8, 2008, 8:30am-5:30pm

**8:30 breakfast**

---

**I Welcome & Setting the Stage: background and objectives of the meeting**
**9:00 – 9:45am**

    1. Opening Remarks
    - Wendy Cunningham, Advisor, Children and Youth unit, The World Bank

2. Definition and characteristics of youth service programs along the continuum of volunteering and service. Potential positive impacts of youth service on participants and the importance of measuring these impacts more effectively.
- Susan Stroud, Executive Director, Innovations in Civic Participation

3. Framing the discussion on evaluation of youth civic participation and youth voluntary service programs. The importance and challenges associated with conducting evaluations to produce rigorous evidence of a program's net impact. Why are youth service programs not considered "proven" to be effective in the World Development Report 2007?
- Mattias Lundberg, Senior Economist, Children and Youth unit, The World Bank

---

**II Evaluation Methods and their application to Youth Voluntary Service Programs**
**10:00 – 11:45am**

Chair: Shahrokh Fardoust, Sr Adviser, Independent Evaluation Group, The World Bank

Presentations:
1. Ariel Fiszbein, Chief Economist, HDN, World Bank. Controlled Impact Evaluation.

2. Michael Patton, Director, Utilization-Focused Evaluation, Alternative Methods for measuring the impact of Youth Voluntary Service Programs.

**Topics:** What needs to be proven? Can it be proven and if so, how? What are acceptable standards of evidence?

Discussants:
1. Eva Mysliwiec, Founder, Youth Star Cambodia. Practitioner's perspective
2. Busani Ngcaweni, Head: Special Projects & Acting Chief of Staff of the Deputy President, The Presidency, Republic of South Africa. Policymaker's perspective
3. Victor Karunan, Senior Advisor, Adolescent Development, and Participation (ADAP), UNICEF. International Agency's perspective

Open Discussion

**Topics:** What is the distinction between methods and approaches? What are the needs for evaluation from practitioners', policymakers' and funders' perspectives? What needs to be proven and what are acceptable standards of evidence? What are the challenges?

---

**12 - 1pm Lunch**

**III Existing Evaluations and Challenges**
**1-5:30pm**

**Topics:** Existing evaluations and evidence on youth voluntary service programs. Evaluation challenges and success stories. Presenters will present their program evaluations referring to the main challenges they encountered, including technical issues such as internal and external validity, externalities and spill-over effects, short/long term impact, follow-up & attrition rates, outcomes, indicators & measurement, cost-benefit and cost-effectiveness.

**Session A.**
**1-2:30pm**

Chair: Varun Gauri, Senior Economist, DECRG, The World Bank

Presentations:
1. Kevin Cramer, Deputy Director, Research and Policy Development, Corporation for National and Community Service.
2. Ami Thakkar, Program Director, International Youth Foundation
3. Licio Palazzini, President, Arci Servizio Civile

Presenter-Discussant:
4. Steve Powell, Country Researcher, PROMENTE, Bosnia

Plenary Q&A

**2:30 – 2:45pm Coffee**

**Session B.**
**2:45-4:15pm**

Chair: Ron Kassimir, Associate Provost, The New School

Presentations:
1. Manish Subharwal, Director of Monitoring and Evaluation, and
   Nicholas Hartley, Director of Operations & Programmes, Student Partnership Worldwide
2. Ryan Cooper, Field Coordinator, "Un Techo para mi Pais" project in El Salvador and Peru
3. Tim Cross, President, YouthBuild International/COO YouthBuild USA

Presenter-Discussant:
4. Peter Levine, Director, The Center for Information and Research on Civic Learning and Engagement, CIRCLE

Plenary Q&A

**4:15 – 4:30pm Break**

**Chaired Small Group Discussion**:
**4:30 – 5:30pm**

Discussion on the existing evidence base and challenges of evaluating the impact of volunteer programs. Each group will focus on one of the main research topics identified during the preceding plenary discussion. Total of 6 groups.

**Closing of Day 1**
**5:30pm**

# Day Two – May 9, 2008, 8:30am-5:00pm

**8:30 breakfast**

**I Knowledge and Research Gaps**
**9 – 10:30am**

1. Presentation on the outcomes of the discussions and conclusions of Day 1: Review on the existing evidence base and key research questions remaining

2. Comments and inputs on the presented review, and identifying the main gaps in knowledge and research

Chair: Andrew Furco, Associate Vice President for Public Engagement, University of Minnesota

Discussants:
1. Constance Flanagan, Professor of Youth Civic Development, College of Agricultural Sciences at Penn State University
2. James Youniss, Professor of Psychology, Catholic University
3. Yuanzhu Ding, Director, Research Center for Volunteering and Welfare, Peking University
4. Lonnie Sherrod, Professor of Psychology and Director of Applied Developmental Psychology Program at Fordham University

Facilitated Discussion

Outcomes:  Assessment of the existing evidence base and gaps in it

**10:30 – 11:00 Coffee**

**II  Evaluation Clinic for upcoming youth service projects**

Case studies to be reviewed and commented by a panel of experts

**Clinic I: South-African National Youth Service**
**11am – 12:30pm**

Chair: Arianna Legovini, Lead Monitoring and Evaluation Specialist, AFTRL, World Bank

1. Presentation on the basic features of the project,
   • Busani Ngcaweni, Head: Special Projects & Acting Chief of Staff of the Deputy President, The Presidency, Republic of South Africa

2. Considerations for an Evaluation Design,

- JoAnn Jastrzab, Principal Associate, Abt Associates

3. Open Discussion: floor opened by discussants' comments on the evaluation:
   - Michael Baizerman, Director of Youth Programs, University of Minnesota
   - James Radner, Executive Director, The Boreal Institute for Civil Society, Munk Centre for International Studies, University of Toronto.

**12:30 – 1:30pm Lunch**

**Clinic II: Youth Star Cambodia Program**
**1:30 – 3:00pm**

Chair: Markus Goldstein, Senior Economist, AFTPM, World Bank

1. Presentation on the basic features of the project
   - Eva Mysliwiec, Founder, Youth Star Cambodia (15 min)
2. Considerations for an Evaluation Design
   - Felipe Barrera, Senior Education Economist, World Bank (15 min)
3. Open Discussion (1 hour): floor opened by discussants' comments on the evaluation:
   - Alan Melchior, Deputy Director and Senior Fellow, Center for Youth and Communities, Brandeis University
   - Chris Blattman, Yale University

**3:00 – 3:15pm Coffee**

**III Research & Evaluation Framework and Agenda**
**3:15pm – 5pm**

Evaluation Framework:
1. Presentation on the draft Evaluation Framework for building further evidence
   - Charmagne Campbell-Patton, Program Associate and Susan Stroud, Executive Director, Innovations in Civic Participation (15 min)
2. Facilitated discussion on the Framework; Incorporating suggested Changes (20 min)

Research and Evaluation Agenda:
1. Discussion on the Research and Evaluation Agenda: Roles and Responsibilities and Next Steps (20 min)
2. Small Group Discussions on the Agenda (20 min)
3. Inputs and suggested changes on the Agenda and Commitments to moving the Agenda forward. (30 min)

*Outcomes*: New draft of the Evaluation Framework and Agenda for moving forward

**Closing Remarks**
**5pm**

1. Mattias Lundberg, Senior Economist, Children and Youth unit, The World Bank
2. Susan Stroud, Executive Director, ICP

**Annex C. Specifics on Evaluation Techniques**

Impact evaluation designs can be broadly classified into three categories: experimental, quasi-experimental and non-experimental. These categories correspond roughly to the methods/techniques used to identify a comparison/control group (World Bank impact evaluation web-site).

## *Experimental designs*

In experimental designs, the beneficiary and control groups are randomly selected from some well-defined population.

## *Quasi-experimental designs*

*Quasi-experimental* techniques generate comparison groups that are not randomly chosen, but are selected so that they closely resemble the beneficiary group, at least in observed characteristics.

- Among the quasi-experimental methods, **Matching Techniques** are generally considered the best option. In this method one tries to select from a larger sample, such as national household survey, a comparison group that matches the beneficiary group. The most common type of matching is *propensity score matching*, in which the comparison group is matched to the beneficiary group on the basis of a set of observed characteristics that are believed to influence program outcomes. These characteristics are used to predict participation in the program (the "propensity score"), and the comparison sample is made up of those with similar propensity scores. The method identifies the comparison group from a sample of non-participants closest to a sample of program participants in terms of observable characteristics. The closer the propensity score, the better the match. Matched comparison groups can be selected before project implementation (prospective studies) or afterwards (retrospective studies).

- In **Reflexive Comparisons** on the other hand, program participants are compared to themselves before and after the intervention. That is, the beneficiaries function as their own comparison group. This type of design is particularly useful in evaluations of full-coverage interventions such as nationwide policies and programs in which the entire population participates and there is no scope for a control group.

## *Non-experimental designs that permit qualitative and quantitative analysis*

In non-experimental designs, program participants are compared with non-participants using statistical methods to account for the differences between the groups. The econometric techniques that can be used to compare the two groups and to *correct for the selection bias* include Regression Discontinuity (RD) and Instrumental Variables such as Random Promotion or Encouragement.

- **Regression Discontinuity (RD)** uses program eligibility criteria to construct a control group: those just within some target cutoff point are assumed to sufficiently resemble those who are just outside it, and it is assumed that differences between the two groups are largely due to the program. For example, a program may be targeted to households who fall below the poverty line. In that case, the comparison can be made between those beneficiary households just below the line and those who are just above the line, and therefore not eligible.

- **Instrumental Variables** estimation consists of using one or more variables (instruments) that are correlated with participation in the program, but not to outcomes given participation. In principle, this corrects for potential endogeneity of program placement. The instrumental variables are first used to predict program participation; then the program impact is estimated using the predicted probability of participation rather than the observed participation.

### *Non-experimental designs that permit qualitative analysis*

Some non-experimental evaluation methods do not establish an explicit counterfactual or control and comparison groups but use other sophisticated techniques to attempt to demonstrate causality and program outcomes. Examples of these approaches include in-depth case studies, general elimination method and theory-based evaluations.

- **Theory-based evaluations** can be used for programs and projects that are based on an explicit theory about how and why a program will work. The evaluation is based on assessing each theory and assumptions during the implementation. The underlying theory is presented as many small steps, and data is collected and analyzed to test the underlying assumptions about the chain of causality within each step. If events do not work out as expected, the evaluation can say with a certain confidence where, why and how the failure occurred. The theory-based design provides early indications of program effectiveness, and any breakdowns can be fixed during the project implementation. The evaluation, by following the sequence of stages, can say with a high degree of confidence how the effects were generated; that is, how each of the program components or steps were instrumental in generating program outcomes. (Baker)

- **Cumulative, in-depth case studies** aggregate information from a number of program sites collected over time. Cumulative case studies are often retrospective, collecting information across studies done in the past, but they may be conducted prospectively, structuring a series of investigations for different times in the future. The techniques for ensuring sufficient comparability and quality and for aggregating the information are what constitute the "cumulative" part of the methodology. (Davey 1991) The cumulative case study can capture unintended impacts and ripple effects, and highlight dimensions of outcomes that can be difficult to quantify. Case studies are particularly valuable in situations where individual context differs across programs or where the program is individualized or adapted to a specific situation. (Patton 2002). However, case studies may be biased in favor of programs that demonstrate positive

results, and it can be difficult to verify the quality of the original data and analyses used (Davey 1991).

- **General Elimination Method (GEM)**, sometimes called the Modus Operandi approach, works like crime scene or accident investigation. In these cases, no randomly controlled trial can be conducted to determine the cause of the crime or accident. The General Elimination Method is used to consider alternative causes and weigh the preponderance of evidence leading to determination of the most likely cause. The method tries to establish attribution and causality in cases where evaluation was not integrated into original program design. However, it is difficult to determine rigorously whether it was the program that caused the observed change.

Several other approaches can support impact evaluations by providing more comprehensive or nuanced pictures of impact. These include outcome mapping, success case method and beneficiary assessment.

- **Beneficiary assessments** (BA) are often used in evaluations to gain insights into the perceptions of beneficiaries regarding a project or policy. The overall objective of a BA is to elicit the views of beneficiaries and other local-level stakeholders. BA relies primarily on three data collection techniques: in-depth conversational interviewing around key themes, focus group discussions, and direct and participant observations. (Salmen 2002). BAs are often essential complements to more conventional quantitative impact evaluations.

- The **success case method** involves examination of the experiences of extremes – that is, groups or individuals for whom the program seems to have been unusually effective. This method uses many of the tools of conventional evaluation, such as surveys and statistical analysis as well as interviews and case studies. It does not attempt to provide a comprehensive picture of impact; rather, it offers a simple and fast way to identify and explain the factors that differentiate successful participants from unsuccessful ones. (Kellogg Foundation)

- **Outcomes Mapping** is a technique developed to assess the contributions development programs make to the achievement of outcomes. Outcome Mapping focuses on changes in the behaviour of people, groups and organizations with whom a program works. This methodology acknowledges the complexity of development processes as well as the contexts in which they occur and focuses on examining the logical links between interventions and outcomes, rather than trying to attribute results to any particular intervention. (IDRC)

## Annex D. ICP Framework

| |
|---|
| **Framework on Evaluating the Impact of Youth Service Programs**<br>Innovations in Civic Participation<br>May 2008 |

*Thank you to Alan Melchior from Brandeis University, Michael Patton from Utilization-Focused Evaluation, Manish Subharwal from Student Partnerships Worldwide, Ami Thakkar and Karin Atkins from the International Youth Foundation, Steve Powell from Promente, Minna Mattero, Silvia Paruzzolo, and Naoko Kataoka from the World Bank for their initial feedback and contributions to this document. As this is a working document, we welcome additional feedback. Please send any comments to buhrer@icicp.org.*

## I. Introduction

Innovations in Civic Participation (ICP), in partnership with the World Bank Children and Youth Unit and participants at the International Experts Meeting on Measuring the Impact of Voluntary Youth Service Programs, has developed this document to provide guidance for evaluating youth service programs in different contexts. To prepare this document, ICP staff conducted a literature review of research on youth civic engagement programs, with an emphasis on youth service programs (please see accompanying bibliography).[2]

Review of the research found very few formal, publically available evaluations of national youth service programs outside North America and Europe. The majority of studies of programs in both the US and the developing world take the form of case studies describing the program rather than studying program impacts. Before promoting further investment in youth service programs globally, it is important to ensure that they are having positive effects on young people and on the communities they serve. To do this requires a stronger evidence base for the impact of youth service programs, especially in regions where little research has been conducted to date. It is also important not to make assumptions about the applicability of impact results for programs in the US or Europe for programs in other parts of the world. Context and program outcomes should determine program and evaluation design, and given the difference in contexts, evaluation designs cannot be easily transferred from one setting to another.

The research also demonstrates a lack of high-quality data and of standardized criteria for gauging program effectiveness. Youth service programs have a complex relationship with the community and cannot be isolated from the context in which they are operating or the specific

---

[2] We would like to acknowledge that the bibliography is not complete, and we welcome suggestions for additional studies that might be added. We also acknowledge that the literature search is incomplete. To date we only searched for documents in English and have not carried out searches for documents in other languages. Several other organizations have useful sources of bibliographies and research reports on youth service and volunteering: the National Service Learning Clearinghouse, Corporation for National and Community Service, Institute for Volunteering Research in London, VOSESA in Johannesburg, ABT Associates, International Center for Research on Civic Engagement and Service-Learning, Center for Social Development at the Washington University in St Louis, among others.

outcomes for which programs are designed. However, we also recognize that in order to stimulate more investment in youth service programs, it is important existing programs are rigorously evaluated and demonstrate significant positive impact. We hope that this document will provide practical guidance regarding evaluation of youth service programs that meet the needs of program practitioners and policymakers, as well as potential investors like the World Bank and other donor organizations.

This document lays out a framework for evaluation providing information on key questions, methodological considerations, identification of outcomes and indicators and an annotated bibliography of youth service evaluations.

## II. Definitions

### Youth
For this framework, we use the UN definition of youth to include young people between the ages of 10 and 24 years of age. We recognize, however, that each policy and program may have its own definition of youth. What is important is that the target group is explicit at the outset of programming.

### Assets-Based Approach
This approach views young people as positive actors in their lives and communities. It sees investing in young people as a significant opportunity to harness their potential to improve their communities, while developing skills for future success. It focuses on the contributions young people can make rather than harm they can cause. Within this context, young people need to be supported to engage positively and effectively in their communities' development through youth-friendly education emphasizing the importance of participation, life and livelihood skills development, and access to a variety of structured opportunities for civic engagement appropriate for young people's individual interests, goals, and skill sets. This approach is contrasted with a needs-based approach, which views young people as "at-risk" or in need of help. Programming stemming from a needs-based approach tends to focus on services provided to young people rather than by them.

### Positive Youth Development
According to the National Research Council, "youth development is the transformation of children into competent, confident, connected, and contributing people of character who are fully prepared and fully engaged in their communities."[3] The NRC identifies four areas for youth development: physical, intellectual, psychological and emotional, and social. The Search Institute has also developed a measure of youth development, based on the "40 Developmental Assets" that identifies concrete, common sense, positive experiences and qualities essential to raising successful young people. These assets have the power during critical adolescent years to influence choices young people make and help them become caring, responsible adults.

### Civic Engagement
Defined as individual and collective actions designed to identify and address issues of public concern. Civic engagement means working to make a difference in the civic life ofcommunities and developing the combination of knowledge, skills, values, and motivation to make that difference. Civic engagement can take many forms, from individual voluntarism to organizational involvement to electoral participation. It can include efforts to directly address an issue, work with others in a community to solve a problem or interact with the institutions of representative democracy.

### Youth Service

---

[3] Eccles, Jacquelynne and Jennifer Appleton Gootman (Eds). (2002). *Community Programs to Promote Youth Development.* Washington, DC: National Academies Press.

Although civic engagement can take on many forms, our focus for this project is on voluntary youth service, defined by the Global Service Institute as "an organized period of substantial engagement and contribution to the local, national, or world community, recognized and valued by society with minimal or no monetary contribution to the participant." Common elements of the programs we are focusing on include:

- a participant's engagement in service is frequent
- the young person is viewed as an active participant in the program rather than a beneficiary of services
- the commitment to service is consistent over time, rather than episodic
- the service may be remunerated in some way, but at less than market rates
- the program is structured
- the program addresses either difficult public problems or needs that have been defined collectively[4]

### *Impact*
Impact is the extent and nature of the positive and negative changes resulting from a specific intervention, program or policy. Impacts can be measured at different levels in individuals, communities and at a national level.

### *Impact evaluation*
The following evaluation definition has been adopted by the NONIE (Network of Networks on Impact Evaluation): "Positive and negative, primary and secondary long-term effects produced by a development intervention, directly or indirectly, intended or unintended" (DAC Evaluation Glossary June 2002).

### III. Key Questions[5]

1. What is the purpose of the evaluation?
   Impact (or *summative*) evaluations answer questions such as: *Does the policy (program, intervention) work? How large is the likely effect size?*
   Process (or *formative*) evaluation answers questions such as: *How, why, and under what conditions does the policy (program, intervention) work?*

2. Who is the audience for the evaluation?
   Internal, external

3. Is the evaluation intended to be program-specific or generalizable (i.e., internal or external validity)?

4. Who are the stakeholders? Who needs to be involved in the evaluation?
   Consider whether the evaluation will use participatory methods.

5. What is the perspective of the evaluation?
   Impact on participants, community, institutions?

6. What are the units of analysis?
   Individuals, groups, program components, whole program, organizations, communities, time periods, etc.

7. What will be the sampling strategy?
   Purposeful sampling, probability sampling. Variations in sample size from a single case study to a generalizable sample.

---

[4] Adapted from Perry, J. L., & Thomson, A. M. (2003). *Civic service: What difference does it make?* Armonk, NY: M.E. Sharpe.

[5] These questions were derived from a set of questions in Patton, Michael Quinn. 2002. *Qualitative Research and Evaluation Methods*. Thousand Oaks, CA: Sage

8. What types of data will be collected?
   Qualitative, quantitative, mixed?[6]

9. What type and degree of control will be exercised?
   Experimental, quasi-experimental, no control

10. What analytical approach or approaches will be used?
    Inductive, deductive

11. How will the validity of and confidence in the findings be addressed?
    Triangulation, multiple data sources, multiple methods, multiple perspectives, multiple investigators

12. What resources are available to conduct the evaluation?

13. Who will conduct the evaluation?
    Internal evaluation team, external team comprised of evaluation expert, universities

14. How will the cultural and political contexts be taken into account during the evaluation?

In general, the analysis of program impact requires a common set of preliminary information and steps:

- Identification of the intended outcome, and selection of measurable indicators.
- An empirically verifiable logical frame of the chain of events that explains how the program is expected to affect beneficiaries (individuals, households, institutions, etc.)
- Information about the implementation of the program activities via process evaluation and output monitoring to understand whether the program is properly executed.
- Measurement of the outcomes of the intervention.
- Analysis: do the outcomes change, and are changes due to the intervention?

**IV. Methodological Considerations[7]**

Measuring the impact of a program is an attempt to confirm whether the intervention is likely to have caused program outcomes. Impact evaluation designs attempting to demonstrate if an observed change is caused by the intervention can be broadly classified into three categories: experimental, quasi-experimental and non-experimental.

In **experimental designs**, the beneficiary and control groups are randomly selected from a well-defined population. When designed with a large enough sample size and controlled to prevent or limit influence from outside factors, any observed difference in the outcomes between the beneficiary and control groups can be safely attributed to the intervention and not to underlying differences between the two groups. The main **benefit** of this technique is the comparative simplicity of constructing the counterfactual and interpreting the results. Some **challenges** to this approach include ethical concerns (denying benefits or services to otherwise eligible individuals), difficulty in offering the intervention to only certain groups, a high cost and time requirement, an

---

[6] Qualitative data tends to stem from three kinds of data collection: 1) in depth interviews; 2) direct observation and 3) written documents. Quantitative data tends to stem from surveys, tests, experiments, secondary data. Mixed methods combine qualitative and quantitative data collection. (Patton 2002: 13)

[7] The Methodological Considerations are adapted from the International Experts Meeting on Measuring the Impact of Voluntary Youth Service Programs Report produced by the World Bank and ICP.

unrealistic design for certain programs, and an inapplicability to pilots, small programs and other program models.

Another category of evaluation includes **quasi-experimental** techniques which generate comparison groups that are not randomly chosen, but are selected so that they closely resemble the beneficiary group. In these designs, program participants are compared to non-participants using complex statistical methods to account for the differences between the groups and correct for the selection bias that might arise from non-random allocation of benefits. The main **benefit** of these designs is that they can draw on existing data sources and are thus often quicker and cheaper to implement than randomized designs. They can also be performed after a program has been implemented, provided that sufficient data exists. A major **challenge** to quasi-experimental techniques is selection bias, which can arise with non-random assignment and when it is possible that some of the characteristics determining assignment to treatment or control groups might also influence final outcomes. Moreover, some quasi-experimental designs cannot say anything about what would have happened to the beneficiaries if they had not received the treatment; nor can they control for unrelated external (non-intervention) changes. The reliability of results also depends on a set of maintained assumptions and fairly complex statistical methods.

Furthermore, a problem with both experimental and quasi-experimental approaches is that they can be conducted as "black box" evaluations, showing the impact of the program on the measured outcomes, without providing any understanding about which particular elements of the program contributed to the impact or paying attention to the process.

In **non-experimental designs**, program participants are compared with non-participants using statistical methods to account for the differences between the groups. Some non-experimental evaluation methods do not establish control/comparison groups but use other sophisticated techniques to attempt to demonstrate causality. The main **benefit** of these approaches is that they take into account the complexity of the program and the context within which it operates. These techniques offer a more flexible and low cost technique that can be applied in situations where experimental or quasi experimental techniques are not appropriate or possible. The primary **disadvantage** of these methods is that the information they provide is complex and difficult to analyze differentially and to generalize from. While many evaluators and practitioners argue that this is not necessarily a disadvantage because programs are complex and the data should reflect this complexity, complex information that is not generalizable is not very helpful for funders and donors looking to draw lessons about the impact or replicability of a program. There is also some debate about the validity of these evaluations, since they may not rigorously establish causality nor explicitly determine the standard against which impact is to be defined and measured.

The different design categories are all appropriate methods in different circumstances for different programs. There is not one "gold standard" for evaluating impact, but instead "methodological diversity and appropriateness in support of rigor" (NONIE) are of main importance.  Each impact evaluation has to be individually designed, taking into consideration the program's unique characteristics and environment. The methodologically simplest and most robust way to establish the counterfactual is by using an experimental approach, randomly assigning individuals either to receive treatment or to serve in a control group. However, a mixed methods approach, combining qualitative and quantitative methods, can provide the most comprehensive results as these techniques combines well contextualized studies with quantitative rigor.

### V. Identification of Outcomes and Indicators
ICP has developed a list of possible/likely program outcomes for youth service programs based on our experiences with youth service programming and the research we have conducted up to now.

| | Program Goal | Assumptions | Indicator of Success | Instruments | Existing Evidence | Methodology |
|---|---|---|---|---|---|---|
| **Y O U T H  D E V E L O P M E N T** | **Increased participation in civic and political life** | Youth service empowers young people to become more actively engaged in their communities and provides them with a means to do so, increasing the likelihood that they will stay involved in the future | More hours spent volunteering or participating in other forms of civic life; more meaningful/sustained/long-term volunteer commitments; Increase in voter participation among participants in youth service programs; membership to voluntary/community/local associations | - Youth surveys (pre and post, when possible) - Interviews or focus groups with participants | AmeriCorps | Longitudinal study with comparison group |
| | | | | | Unis-Cité | Pre/Post Survey |
| | | | | | Arci Servizio Civile | Longitudinal, Retrospective Surveys |
| | | | | | City Year | Longitudinal Study with control group |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| | **Life Skills Development** | Youth service helps build life skills such as: a sense of agency, independence, self-confidence, leadership, teamwork, emotional and physical well-being, decision-making and communication | Increased sense of agency, independence, self-confidence, leadership, teamwork, emotional and physical well-being, decision-making and communication | - Observational interviews with participants, parents, teachers - Youth surveys (pre and post, when possible) | Katimavik | Retrospective interviews and surveys; Pre/Post surveys |
| | | | | | Green Corps | Retrospective |
| | | | | | Unis-Cité | Pre/Post Surveys |
| | | | | | AmeriCorps | Longitudinal study with comparison |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| | | | | | Arci Servizio Civil | Retrospective Interviews |
| | | | | | Green Corps | Retrospective |
| | | | | | AmeriCorps | Longitudinal study with comparison group |
| | **Prevention of Risky Behaviors** | Youth service gives young people a more structured way to spend their out-of-school time; it also helps them to make better choices by equipping them with better decision-making skills. | Decrease in teen pregnancy; decrease in STDs/HIV; decrease in risky sexual behavior or increase in use of condoms; decreased substance abuse; decrease in participation in gangs | - Youth surveys (pre and post, when possible) - Observation - Statistical comparison (local statistics) | Student Partnerships Worldwide | Retrospective, participatory |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **Intercultural Competence** | Youth service exposes young people to a more diverse group of people and provides them with training and tools for interacting with different populations | Improvements in cultural competency; Improvements in youth-adult partnerships and intergenerational communication; Improved tolerance and empathy for others | -Pre and Post tests<br>-Interviews with participants<br>- Youth surveys (pre and post) | Katimavik | Retrospective interviews and surveys; Pre/Post surveys |
| | **Improved Education Attainment** | Youth service enhances young people's commitment to education by providing them opportunities to apply what they are learning to real life situations | Increase in grades; increased literacy levels; improved academic knowledge/skills; improved attendance or graduation rate; improved classroom behavior | - Grades<br>- Teacher surveys<br>- Attendance records and graduation rate | Student Partnerships Worldwide | Retrospective, participatory |
| | | | | | Nokia-IYF | Pre/Post and Retrospective Surveys |
| C O M M U N I T Y   D E V E L O P M E N T | **Community Strengthening** | Youth service strengthens linkages between community organizations and young people and improves community morale and cooperation as well as perception of youth in community | Enhanced organizational capacity, stronger civil society, enhanced perception of young people in society | - Focus groups, interviews<br>- Organizational assessments | AmeriCorps | Quasi-experimental |
| | | | | | Learn and Serve America (1999) | Quasi-experimental |
| | | | | | YouthBuild, Inc. | Pre/post |
| | **Peace-building – fostering unity and trust** | Youth service can build/restore trust between participants and communities, particularly when used in post-conflict settings | Reintegration of child soldiers into the community, increased dialogue and cooperation in the community, reduction of inter-group conflict | - Community surveys | | |
| | **Disease Prevention** | Youth service can focus on peer education to reduce risky behaviors, which can reduce STDs or other preventable diseases | Reduced incidence of diseases in the community | - Statistical comparisons<br>- Surveys | | |
| | **Infrastructure Development** | Youth service can mobilize young people to contribute to infrastructure development through building homes, roads or other critical national infrastructure needs | Enhanced community infrastructure, housing, roads, water supply etc. | | AmeriCorps | Quasi-experimental |
| | | | | | Un Techo para mi Pais | Experimental |
| | | | | | YouthBuild, Inc. | Pre/post |
| | | | | | National Youth Service South Africa | Post |

| Literacy and Education | Youth service that is targeted toward education and literacy can significantly enhance literacy rates in a country | Increased literacy rates | Statistics, interviews with teachers, students | National Youth Service South Africa | Retrospective |
|---|---|---|---|---|---|
| Environmental Restoration | Environmental service can have positive effects on restoring environmental damages and potentially also on preventing future damages through education and retrofitting | Restoration of environmental damages, such as cleaner waterways, trees planted, etc | | Green Corps | Retrospective |

### VI. Annotated Bibliography of Youth Service Evaluations

Aguirre International (1999). *Making a difference: Impacts of AmeriCorps\*State/National direct on members and communities 1994-1995 and 1995-1996*. Report prepared for the Corporation for Community and National Service.

> A two-year, three-tier investigation to assess the outcomes and impacts of AmeriCorps State/National Direct programs on beneficiaries, members and communities served. Study used a multi-tiered evaluation design utilizing quantitative and qualitative approaches. Programs studied had meaningful service accomplishments for the populations served, including increased life skills, civic engagement and educational attainment.

Anderson, Leslie M, Laguarda, Katrina G. and Williams, Imeh J.. (2007). *The Effect of the City Year Experience Over Time: Findings from the Longitudinal Study of Alumni*. Prepared for City Year by the Policy Studies Institute.

> Longitudinal study was based on a random sample of 107 corps members who participated in the 2002-03 program year. The study examined participants' levels of civic engagement, civic leadership, and development of social capital over four years, compared with a group of 85 similarly situated young adults. Findings suggest that City Year has generated a greater amount of social capital for its alumni than would have been expected, based on the attitudes, values, and behaviors of these studies' respective comparison groups.

Anderson, Leslie M. and Fabiano, Lara. (2007). *The City Year Experience: Putting Alumni on the Path to Lifelong Civic Engagement.* City Year Inc.

> City Year Alumni Study presents a set of three interlocking studies designed to assess City Year's impact on alumni at various intervals of time after the completion of their City Year community service experience through survey data from 2,189 City Year alumni. Findings show that City Year increased the social capital of alumni across racial/ethnic categories (with the exception of Asian alumni) and all levels of prior education, but also reduced the gap between alumni who came with large initial civic resources (i.e., who are not from minority groups and/or who came with a bachelor's degree) compared with those alumni who came with fewer civic resources (i.e., members of minority groups and/or having only some college).

Bequet, Valérie. (2007) *Service Civil : quel impact sur les jeunes ?* Paris: Unis-Cité.

Evaluation of Unis-Cité national youth service program in France. Study used pre and post questionnaires that included questions about satisfaction with experience and background; participants are from various education levels, primarily from urban areas, more women than men, from different socio-economic backgrounds. Key findings include information about behaviour change and skill development.

Burghardt, J., Schochet, P.Z., McConnell, S., Johnson, T., Gritz, R.M., et al. (2001). *Does Job Corps work? Summary of the National Job Corps Study*. Princeton, NJ: Mathematica Policy Research.

Study based on a national random sample of all eligible applicants to Job Corps in late 1994 and 1995. Findings suggest that Job Corps centers effectively deliver the planned services called for by the program model, provides extensive education, training, and other services, and substantially increases the education and training services that youths receive as well as skills and educational attainment. Other findings include reduced involvement with crime and earning gains.

Corporation for National and Community Service, Office of Research and Policy Development. 2008. *Still Serving: Measuring the Eight-Year Impact of AmeriCorps on Alumni.* Washington, D.C., 2008.

Results from a rigorous evaluation of AmeriCorps show that AmeriCorps has long-term impacts on members' years after they serve. Over the last eight years, since 1999 CNCS has followed more than 2,000 individuals in AmeriCorps State and National and NCCC to look at the effect of service on their future civic engagement and volunteering, employment and careers, and educational attainment. Researchers controlled for factors which may influence study participants' life outcomes such as demographic characteristics, economic status, and prior service and volunteering. Findings reveal that AmeriCorps is a pipeline to careers in public service and creates civic leaders who continue to serve in their communities long after their service has been completed.

Department of Education, Training and Youth Affairs, Commonwealth of Australia. *"It's why you're swinging the mattock." An Evaluation of the Green Corps Programme: Final Report.* May 1999.

Evaluation of Green Corps, an environmental service programme for Australians aged 17 to 20. The main data sources were administrative systems, post programme employment, education and training information collected three months after programme participation, a survey of 149 former participants from fifteen projects selected for environmental assessment, and consultations with stakeholders. A range of benefits for participants were identified including improved employment prospects, the acquisition of vocational and life skills, and increased environmental awareness, and self esteem. Evaluators also conducted an environmental assessment to determine impact of environment from volunteers. Overall the findings of the evaluation were positive.

ECOTEC Research and Consulting. (2001) *Evaluation of the European Volunteering Services Programme Executive Summary.*

Study looks at impacts of EVS programme on participants and finds improved language skills, improved self-confidence, social skills and enhanced cultural awareness. Also identifies benefits to host organizations and communities. Calls for an improved evaluation framework and more developed set of indicators.

Evaluation Services. (2006) *Summative Evaluation of Katimavik.* Department of Canadian Heritage, Corporate Review Branch.

Study of the Canadian National Youth Service Program, Katimavik. Used surveys, focus groups, opinion polls and background research to assess the impact and success of the program.

Family Health International. *Rapid Appraisal Student Partnerships Worldwide Zambia.*

A rapid appraisal was conducted to gauge the relative importance of the five major program elements within specific school contexts from a wide range of perspectives. The appraisal found widespread belief that the program has an impact on knowledge and risk-taking behaviors of Zambian students.

Hahn, Andy, Susan Lanspery and Tom Leavitt. (2006) *Measuring Outcomes in Programs Designed to Help Young People Acquire Life Skills.* The Heller School for Social Policy and Management Center for Youth and Communities, Brandeis University.

A 20-month outcomes study of a global youth development initiative funded by the Nokia Corporation and managed by the International Youth Foundation addressed the knowledge deficit regarding whether life-skills can be strengthened and the effect of life-skills programs in helping young people realize their potential. The outcomes study generally found positive results from participation in the program in the areas of life-skills, greater community engagement, stronger school performance, improved employability and more positive attitudes toward the future and achievement of personal goals.

Hahn, A., Leavitt, T.D., Horvat, E.M., & Davis, J.E. (2004) *Life after YouthBuild: 900 YouthBuild graduates reflect on their lives, dreams and experiences*. Report available at YouthBuild USA.

Results of research study using surveys of 882 YouthBuild graduates from more than 60 sites and in-depth interviews with a cross-section of 57 graduates from eight sites. The study showed significant positive results for the graduates on a number of key measures, and a very positive assessment by them of the value of the program.

Harris, Colette. *Annual Review of Students Partnership Worldwide (SPW) Programme.* Department for International Development, Sierra Leone. (not date)

An evaluation of SPW programme in Sierra Leone. Evaluation consisted of site visits, interviews, and observations. Project indicators show that most elements are on target. Study looks at program impact and included a risk assessment of the programme.

Harris, Lewis and Associates, Ind. (1969). *A Study of Returned Peace Corps Volunteers.*

Study examined past service activity, career goals, level of community involvement, marital and military status, political views and activity. Volunteers were compared with a control group composed of people who had applied to Peace Corps but later declined invitations to join. Major findings indicated that service was valuable to participants, service contributed to changes in their career goals, and relative to control, volunteers were more liberal and maintained those views for a longer period of time.

Jastrzab, J., Giordono, L., Chase, A., Valente, J., Hazlett, A., LaRock, R., & James, D. (2004). *Serving country and community: A longitudinal study of service in Americorps.* Cambridge: Abt Associates.

A longitudinal study of AmeriCorps members. Demonstrates the impacts of national service on members' civic engagement, education, employment, and life skills. Findings of the study reveal that AmeriCorps alumni are more connected to their communities; more knowledgeable about problems facing their communities; more likely to participate in community activities; and more likely to choose public service careers.

Jastrzab, JoAnn, Michele Ciurea,Carolyn Cohen,Carl Hostica,Deborah Small,Anne St. George. (2000) *Evaluation of the Washington Service Corps -- Final Report.* Cambridge: Abt Associates.

Study finds that the program provides important services by making a difference in communities. Alumni are overwhelmingly positive in their assessment of the program experience, reportedly learning valuable skills and earning access to an education award that helped support their continued education. Finally, program benefits outweighed program costs by 1.67 percent, not counting additional benefits which could not be monetized.

Jastrzab, J., Masker, K., Blomquist, J., and Orr, L. (1996). *Impacts of service: Final report on the evaluation of AmeriCorps*. Report prepared for the Corporation for Community and National Service.

An evaluation of the participant and community impacts of the Conservation and Youth Corps. In general, participants found the corps experience worthwhile. The value of program output averaged $13.24 per service hour. The eight sites generated services worth nearly $14 million. The study found that large, mature programs produced a net monetary benefit of $1.04, over and above costs, for each hour of service.

Jastrzab, J., Masker, J., Bloomquist, J. and Orr, L. (1996). *Impacts of service: Final report on the evaluation of American Conservation and Youth Corps.* Cambridge, Mass.: Abt Associates, Inc**.**

An experimental design was used to assess the impact of YC on participants. Forty-one outcome measures in nine categories were used (e.g., civic, social, and personal development; current and planned involvement in other social service; educational aspirations and expectations). Program applicants were randomly assigned to either the treatment group or the control group. A follow-up period that covered about 15 months after enrollment occurred in conjunction with a follow-up telephone interview. The most significant impacts were related to employment and earnings. The treatment group was more likely to have worked for pay and worked more hours (40 percent more than the control group) over the follow-up period. Other results were that program participants were less likely to have been arrested and less likely to have earned a technical certificate or diploma (suggesting that participation in YC may have been a substitute for additional education, in the short run). The study also looked across subgroups (race, gender) for significant differences between the impacts of program participation and non-program participation. The most significant impacts were on African-American males.

**Melchior, A. (1999). *Summary Report: National Evaluation of Learn and Serve America*. Waltham, MA: Center for Human Resources, Brandeis University.**

This evaluation report describes the results of a two-year evaluation conducted by Brandeis University and Abt Associates that examined the impacts of the Learn and Serve America programs in 17 middle schools and high schools across the country, using a variety of quantitative and qualitative methods. The study concluded that at the end of one year of service, the program had a positive impact on participants' civic attitudes, involvement in volunteer service, educational attitudes and school performance.

Mitchell, Maxine V. *et al*. (2003). *Evaluation of the YouthBuild Program.* Chicago, Illinois: Applied Real Estate Analysis, Inc.

Comparison of YouthBuild program to Job Corps, JOBSTART, and the youth components of Supported Work Demonstration (SWD) and Job Training Partnership Act

(JTPA). Finds JobCorps to be most effective. Finds programs are high cost, similar academic achievements across programs, and limited employment achievements.

Perry, J., Thomson, A., Tschirhart, M., Mesch, D., & Lee, G. (1999). *Inside a Swiss army knife: An assessment of AmeriCorps.* Journal of Public Administration Research and Theory, 9(2): 225-250.

Study identifies five AmeriCorps goals: satisfying unmet social needs, developing corps members, enhancing the civic ethic, reinvigorating lethargic bureaucracies, and bridging race and class. The evidence of AmeriCorps' effectiveness is not definitive. Self-reports from recipient programs, selective cost-benefit analyses, and some survey evidence indicate some positive results. More fine-grained survey and field research raise questions about AmeriCorps' overall effects.

Rappaport, Catherine Dun; Anne Hazlett. (2004). *Civic Works at Ten Years: A Program Assessment.* Abt Associates.

Study surveying 2002/2003 corps members and conducted case studies and a cost/benefit analysis of two Civic Works service projects.

Starr, Jerold M. (1994) *Peace Corps Service as a Turning Point.* International Journal of Aging & Human Development, v39 n2 p137-61.

Longitudinal study of 21 U.S. Peace Corps volunteers, first interviewed in 1960s during tour of service in Republic of Philippines and then 20 years later as middle-aged adults. Life events reported after their service and high degree of response agreement in two interviews confirm that Peace Corps experience constituted turning point in their life courses.

Student Partnerships Worldwide. (2007). *Evaluation Report of SPW Nokia School Empowerment Programme, India.*

Evaluation to assess outcomes against outputs and inputs; assess impact of programme and stakeholder perception of programme and make recommendations for improvement. Overall conclusions include: the program is meeting all basic objectives as per logframe and programme design; the programme enjoys strong government, school and community support; the program model is replicable and scalable.

Wolf, Wendy, Sally Leiderman, and Richard Voith. 1987. The California Conservation Corps: An Analysis of Short-Term Impacts on Participants. Philadelphia: Public/Private Ventures.

Four-year assessment of the California Conservation Corps (CCC) that included baseline and follow-up interviews with 943 participants and 1,083 comparison group members. The assessment addressed the economic and non-economic effects of CCC and found that the program pays for itself in the value of the work and the earnings it produces for corps members.

## VII. Evaluation Resources

Berger, Michael A. "Studying Enrollment Decline (and Other Timely Issues) via the Case Survey." Educational Evaluation and Policy Analysis, 5:3 (1983), 307-317.

Bonbright, David. (December 2007) "Measuring Impact, Who Counts?" *Alliance*, Vol. 12, No 4.

Center for Global Development. (2006). *When Will We Ever Learn: Improving Lives through Impact Evaluation*. Report of the Evaluation Gap Working Group.

Chatterji, Madhabi. (September 2007) "Grades of Evidence: Variability in Quality of Findings in Effectiveness Studies of Complex Field Interventions." *American Journal of* Evaluation. Vol. 28, No. 3, pp. 239-255.

Compton, Donald, Micahle Baizerman and Stacey Hueftle Stockdill. (2002) <u>The Art, Craft, and Science of Evaluation Capacity Building</u>. Jossey-Bass, Number 93.

Datta, Lois-Ellin (1990). Case Study Evaluations. Washington, DC: U.S. General Accounting Office, Transfer paper 10.1.9.

Davey, Lynn. 1991. *The Application of Case Study Evaluations.* ERIC Clearinghouse on Tests Measurement and Evaluation, Washington DC. [http://www.ericdigests.org/1992-5/study.htm](http://www.ericdigests.org/1992-5/study.htm)

Earl, Sarah, Fred Carden, and Terry Smutylo. 2001. *Brochure on Outcome Mapping: The Challenges of Assessing Development Impacts*. IDRC. [http://www.idrc.ca/en/ev-64698-201-1-DO_TOPIC.html](http://www.idrc.ca/en/ev-64698-201-1-DO_TOPIC.html)

Flores, Kim Sabo. (2008) <u>Youth Participatory Evaluation</u>. San Francisco: Wiley & Sons.

Institute for Volunteering Research. *Volunteering Impact Assessment Toolkit.*

Miles, Matthew B., and Huberman, A.M. (1984). Qualitative Data Analysis: A Sourcebook of New Methods. Beverly Hills, CA: Sage.

Nichols, Austin. *Impact Evaluation: Regression Discontinuity* The Urban Institute <[http://www.urban.org/toolkit/data-methods/regression.cfm](http://www.urban.org/toolkit/data-methods/regression.cfm)>

Patton, Michael. "Appropriate Methods: The Paradigms Debate and a Utilization-Focused Synthesis." In *Utilization-Focused Evaluation.*

Patton, Michael. (2002). *Qualitative Research and Evaluation Methods*. Thousand Oaks, Ca: Sage

Pawson, Ray & Nick Tilley. *Realist Evaluation*. Development Policy Review Network. Utrecht, The Netherlands. [http://dprn.fss.uu.nl/Realistic%20Evaluation.pdf](http://dprn.fss.uu.nl/Realistic%20Evaluation.pdf)

Robert O. Brinkerhoff, *The Success Case Method* (2003) Berrett Koehler : San Francisco

Salmen, Lawrence F. 2002. *Beneficiary Assessment: An Approach Described*. The World Bank.

Williams, Bob and Iraj Imam (2006). Systems Concepts in Evaluation – An Expert Anthology. AEA Monograph, EdgePress/AEA Point Reyes CA.

WK Kellogg Foundation. *Success Case Study Summary.* [www.wkkf.org](http://www.wkkf.org)

World Bank Independent Evaluation Group. *Impact Evaluation- The Experience of the Independent Evaluation Group of the World Bank.* Independent Evaluation Group

The World Bank (2004). *Monitoring & Evaluation: Some Tools, Methods & Approaches*. Operations Evaluation Department.

The World Bank (2006). *Conducting Quality Impact Evaluations Under Budget, Time and Data Constraints*. Independent Evaluation Group.

The World Bank. *Impact Evaluation - Beneficiary Assessment*. [go.worldbank.org/IF3Z0L5G20](http://go.worldbank.org/IF3Z0L5G20)

Yin, Robert K. (1989). Case Study Research: Design and Methods. Beverly Hills, CA: Sage.

## VIII. Bibliography of Research

**YOUTH SERVICE and VOLUNTEERING**

Angel, W.D. (2003). National *youth policy and national youth service: Towards concerted action*. In H. Perold, S. Stroud, & M. Sherraden (eds), Service Enquiry: Service in the 21<sup>st</sup> century. Johannesburg: Global Service Institute and Volunteer and Service Enquiry Southern Africa (pp. 33-45).

Association of Voluntary Service Organizations. (2003) *Voluntary service: Opening doors to the future, research report. Brussels: AVSO.*

Cabero, Javier. *Programa Nacional de Jóvenes con la Participación Popular*. UNV Bolivia

Cohen, C. (1996). *What service teaches about citizenship and work: The case of AmeriCorps.* Seattle, Washington.

Danzig, Richard and Peter Szanton. (1986) *National Service: What Would It Mean?* DC Health

Deguchi, Masayuki. *Civil Society and Civic Service: Linguapoltical issues on Japan's "Civic Service" programs to the international community.* National Museum of Ethnology and Graduate University for Advanced Studies.

Eberly, Don. (1988) *National Service: A Promise to Keep.* Rochester, New York: John Alden Books.

Eberly, D.J., & Gal, R. (2006). *Service without guns*.

Febbraro, A. (2001). *Encouraging volunteering among Ontario youth*. Retrieved January 31, 2007 from http://www.givingandvolunteering.ca/reports/1997_ontario_youth/page02.asp

Foley, P. (2003). *Youth service for employment: The Umsobomvu Youth Fund initiative in South Africa*. In H. Perold, S. Stroud, & M. Sherraden (eds), Service Enquiry: Service in the 21<sup>st</sup> century. Johannesburg: Global Service Institute and Volunteer and Service Enquiry Southern Africa (pp.159-171).

Gal, Reuven. (2007) *National Youth Service ad a Psycho-Social Process*. Paper.

Gaskin, Katherine. (2004). *Young People, Volunteering and Civic Service.* A report for the Institute for Volunteering Research.

Gouault, Jean-Marie. *Pour Avec, National Service, what are the choices?* UNESCO Coordinating Committee for International Voluntary Service. No Date

Haan, Norma. (1974) "Changes in Young Adults After Peace Corps Experiences: Political-Social Views, Moral Reasoning, and Perceptions of Self and Parents." *Journal of Youth and Adolescence*, 3, 3, 177-94.

Hall, M., Lasby, D., Gumulka, G., Tryon, C. (2006). *Caring Canadians, involved Canadians: Highlights from the 2004 National Survey of Giving, Volunteering, and Participating* (Catalogue No. 71-542-XIE). Ottawa, ON: Statistics Canada.

Hills, Greg and Adeeb Mahmud. (2007) *Volunteering for Impact: Best Practices in International Corporate Volunteering.* FSG Social Impact Advisers.

Institute for Volunteering Research. (2004) *Young People and Volunteering: Preliminary findings and emerging lessons from primary research.*

James Pratt Consulting. (2005) *Volunteering as a Stepping Stone for Marginalized Youth*. Prepared for Volunteer Victoria. Accessed from http://www.volunteervictoria.bc.ca/_pdfs/stepping_stone.pdf

Johnson, L., Benitez, C., McBride, A. M., & Olate, R. (2004). *Youth volunteerism and service in the Latin America and Caribbean region: A potential strategy for social and economic development* (Research Background Paper).  St. Louis: Washington University, Center for Social Development.

Kelly, P. (2007). *National Youth Service – Transforming lives*. Government of Jamaica, Ministry of Education and Youth, Jamaica Information Service. Retrieved on August 23, 2007 from http://www.jis.gov.jm/education/html/20060724t100000-0500_9511_jis_national_youth_service___transforming_lives_.asp

Kuperminc, G.P., Holditch, P.T., & Allen, J.P. (2001.) "Volunteering and community service in adolescence.I *Adolescent Medicine: State of Art Reviews,* 12, 445-457.

McBride, A.M., & Sherraden, M. (2007). Civic Service Worldwide: Impacts and Inquiry. Armonk, NY: M.E. Sharpe.

**McBride, Amanda Moore et al. (January 2003). *The Forms and Nature of Civic Service: A Global Assessment*. Washington University: Global Service Institute, Center for Social Development.**

Metz, Edward, Jeffrey McLellan and James Youniss. (March 2003) "Types of Voluntary Service and Adolescents' Civic Development." Journal of Adolescent Research. Vol 18, No. 2 pages 188-203.

Mysliwiec, Eva. (2005) *Youth, Volunteering and Social Capital in Cambodia. Results of a feasibility study conducted for a Cambodian youth service program.* Youth Star Cambodia.

National Youth Agency. (2007). *Young people's volunteering and skills development: A research project undertaken by the National Youth Agency on behalf of the Department for Education and Skills from October 2006 to March 2007; Executive summary*. Leicester, U.K.: National Youth Agency.

National Youth Agency. (2007). *Young people's volunteering and skills development.* Leicester, U.K.: National Youth Agency. www.dfes.go.uk/research

Obadare, Ebenezer. (2005) *Statism, Youth and Civic Imagination:  A Critical Study of the National Youth Service Corps (NYSC) Programme in Nigeria.* Center for Social Development, Global Service Institute.

Ojo, Folayan. (September 1980) "Youth Employment and the Impact of the National Youth Service Corps on Labor Mobility in Nigeria." *African Studies Review*, Vol. 23, No. 2, pp. 51-62.

Perold, H., Stroud, S., & Sherraden, M., (eds.). (2003). Service Enquiry: Service in the 21st century. Johannesburg: Global Service Institute and Volunteer and Service Enquiry Southern Africa.

Perry, J. L., & Thomson, A. M. (2003). <u>Civic service: What difference does it make?</u> Armonk, NY: M.E. Sharpe.

Powell, Steve and Bratović, E.. (2007) *The impact of long-term youth voluntary service in Europe: a review of published and unpublished research studies*. Brussels: AVSO.

Powell, S. & Čelebičić, I. (2008). Outcome Mapping evaluation of six civil society projects in Bosnia and Herzegovina: summary report. Sida, Sarajevo. [www.promente.org/sida2eng](www.promente.org/sida2eng)

Powell, S. & Bratović, E. (2007b). We can. We volunteer. Pro-social values/behaviour and employability amongst young people in SEE and the impact of volunteer work camps. SEEYN: South-East European Youth Network. [www.promente.org/seeyn5](www.promente.org/seeyn5)

Quarter, J., Mook, L., & Richmond, B. J. (2002). *What volunteers contribute: Calculating and communicating value added.* Toronto: Canadian Centre for Philanthropy.

Reinders, Heinz and James Youniss. (2006) *Community Service and Civic Development in Adolescence: Theoretical Considerations and Empirical Evidence.* In A. Sliwka, M. Diedrich, & M. Hofer (eds.) <u>Citizenship education: theory, research, and practi</u>*ce.* Munster, Germany: Waxmann, pp. 195- 208

Schroer, R. (2003). *Voluntary service: Opening doors to the future; the integration of young people from disadvantaged backgrounds in transnational voluntary service; a cross-national analysis of France, Germany, Greece, Italy and United Kingdom*. Brussels: AVSO.

Shannon, Charlene and Brenda J. Robertson. (2007) *Engaging Youth Ages 8 to 12 as Volunteers: An Opportunity for Youth Development?* Volume 2, Number 2.

Sherraden, Michal. (October 2001) *Civic Service: Issues, Outlook, Institution Building*. Washington University: Global Service Institute, Center for Social Development.

Sherraden, M., & Eberly, D. (1982). *The impact of national service on participants*. In M. Sherraden & D. Eberly (Eds.), <u>National service: Social, economic, and military impacts</u> (pp. 179-187). New York: Pergamon Press.

Sherraden, M., Sherraden, M., & Eberly, D. (1990). *Comparing and understanding non-military service in different nations.* In D. Eberly and M. Sherraden (Eds.), <u>The moral equivalent of war?: A study of non-military service in nine nations</u> (pp. 159-190). Westport, CT: Greenwood Press.

Sutherland, D., Doerkson, M., Hanslip, T., Roberts, J., Stewart, S., Sagnes, E., & Friesen-Storz, T.  (2006). *Youth volunteerism: Measuring the benefits of community service learning programs*. Toronto, ON: Imagine Canada, Knowledge Development Centre

Thoits, P.A., & Hewitt, L.N. (2001). "Volunteer work and well-being." *Journal of Health and Social Behavior*, 42, 115-131.

Thomas, Franklin A. (1984) *National Service: an Aspect of Youth Development* (adapted from remarks given before a group of civic and business leaders in Monte Rio, CA on July 22, 1984)

Thomson, A. M., & Perry, J. L. (1998). "Can AmeriCorps build communities?" *Nonprofit and Voluntary Sector Quarterly*, *27*(4), 399-420.

Tolman, Joel and Karen Pittman. (2001) "Youth Acts, Community Impacts: Stories of Youth Engagement with Real Results." *Community and Youth Development Series*, Volume 7. Takoma Park, MD: The Forum for Youth Investment, International Youth Foundation.

United States Department of Labor. (2007). *Volunteering in the United States*, 2006. Retrieved January 31, 2007, from http://www.bls.gov/news.release/pdf/volun.pd

UNV. *Youth Volunteering for Development: Africa in the 21st Century.* Discussion Paper 4 October 2006.

Vian, Taryn. (2007) *Measuring the impact of international corporate volunteering: Lessons learned from the Global Health Fellows Program of Pfizer Corporation.*

Viva Rio (2005) *A Policy for Preventing and Providing Alternatives to Youth Involvement in Urban Violence in Brazil.* Viva Rio.

VOSESA. (2007) *Research Partnerships Build the Service Field in Africa.* The Social Work Practitioner-Researcher, Journal of Social Development in Africa

Youniss, J., McLellan, J., & Yates, M. (1997). *Community service and social responsibility in youth*. Chicago: University of Chicago Press.

**SERVICE-LEARNING**

Astin, A.W., Vogelgesang, L. J., Ikeda, E. K., & Yee, J. A. (2000). *How service-learning affects students* (Executive summary). Los Angeles: University of California, Los Angeles, Higher Education Research Institute.

Berkas, T. (February, 1997). *Strategic Review of the W. K. Kellogg Foundation's Service-Learning Projects, 1990-1996*. Battle Creek, MI: W. K. Kellogg Foundation.

Billig, S. H. (2000). "Research on K-12 school-based service-learning: The evidence builds." *Phi Delta Kappan, 81*(9), 658-664.

Billig, S. H. (2002). "Adoption, implementation, and sustainability of K-12 service-learning." In A. Furco & S. H. Billig (Eds.), *Service-Learning: The essence of the pedagogy* (pp. 245-267). Greenwich, CT: Information Age Publishing.

Corporation for National and Community Service. *The Impact of Service-Learning: A Review of Current Research.* January 2007.

Eyler, J. (2002). "Reflection: Linking service and learning -- Linking students and communities." *Journal of Social Issues, 58*(3), 517-534.

Eyler, J. & Giles, D. (1997). "The importance of program quality in service-learning." In A. S. Waterman (Ed.), *Service-Learning: Applications from the research* (pp.57-76). Mahwah, NJ: Lawrence Erlbaum Associates.

Eyler, J. & Giles, D. (1999). *Where's the learning in service-learning?* San Francisco: Jossey-Bass Publishers.

Kraft, N. & Wheeler, J. (2002, October). *Service-learning and resilience in disaffected youth: A research study*. Paper presented at the 2nd Annual International Service-Learning Research Conference, Vanderbilt University.

Melchior, Alan and Lawrence Niel Bailis. (2002) *Impact of Service Learning on Civic Attitudes and Behaviors of Middle and High School Youth*. In Billig, Shelly and Andrew Furco, eds. Service-Learning: The Essence of the Pedagogy. Advances in Service Learning Research. Greenwich, CT: Information Age Publishing, Inc.

Muscott, H. S. (2000). "A review and analysis of service-learning programs involving students with emotional/behavioral disorders." *Education and Treatment of Children 23*(3), 346-368.

National Youth Leadership Council. *The National Survey on Service-Learning and Transitioning to Adulthood.* December 2005.

O'Donnell, L., Stueve, A., San Doval, A., Duran, R., Haber, D., Atnafou, R., et al. (1999). "The effectiveness of the Reach for Health Community Youth Service Learning Program in reducing early and unprotected sex among urban middle school students." *American Journal of Public Health, 89*(2), 176-181.

Pritchard, I. A. (2002). Community service and service-learning in America: The state of the art. In A. Furco & S. H. Billig (Eds.), *Service-Learning: The essence of the pedagogy* (pp. 3-21). Greenwich, CT: Information Age Publishing.

Scales, P., Blyth, D., Berkas, T., & Kielsmeier, J. (2000). The effects of service-learning on middle school students' social responsibility and academic success. *Journal of Early Adolescence, 20*(3), 332-358.

Scales, P. & Blyth, D. (1997). Effects of service-learning on youth: What we know and what we need to know. *Generator*, Winter, 6-9.

Shumer, R. (1994). "Community based learning: Humanizing education." *Journal of Adolescence, 17,* 357-367.

Tapia, M.N., & Mallea, M.M. (2003). *Service-learning in Argentina*. In H. Perold, S. Stroud, & M. Sherraden (eds), Service Enquiry: Service in the 21st century. Johannesburg: Global Service Institute and Volunteer and Service Enquiry Southern Africa (pp. 203-215).

**CIVIC ENGAGEMENT**

Baizerman, Michael, Ross VeLure Roholt, R.W. Hildreth. "The Place of Evaluation." In *Deepening the Craft of Youth Civic Engagement.*

Calvert, Matthew, Shepherd Zeldin, and Amy Weisenbach. 2002. *Youth Involvement for Community, Organizational, and Youth Development: Directions for Research, Evaluation, and Practice.* Madison, WI: University of Wisconsin and Innovation center for Community and Youth Development/Tides Center.

Eccles, J.S., & Barber, B.L. (1999). "Student council, volunteering, basketball, or marching band: What kind of extracurricular involvement matters?" *Journal of Adolescent Research*, 14, 10-43.

Flanagan, Constance A., Jennifer Boes, Britta Jonsson, Beno Csapo, and Elena Sheblanova. (1998) "Ties that Bind: Correlates of Adolescents' Civic Commitments in Seven Countries," *Journal of Social Issues*, Vol. 54, No. 3. pp. 457-475.

Flanagan, Constance and Nakesha Faison, "Youth Civic Development: Implications of Research for Social Policy and Programs," *Social Policy Report* XV:1 (2001):

Hildreth, R.W. (2000) *Theorizing Citizenship and Evaluating Public Achievement.* University of Minnesota.

Kahne, J. (2007) *Developing Citizens: A Longitudinal Study of School, Family, and Community Influences on Students' Commitments to Civic Participation* Paper presented at the annual meeting of the American Political Science Association http://www.allacademic.com/meta/p209253_index.html

Keeter, Scott, *et al*. (2002). *The Civic and Political Health of the Nation: A Generational Portrait.* CIRCLE.

Lopes, Mark Hugo and Jason Kolaczkowski. (2002)*Civic Engagement Among Non-College Atending 18-25 Year Olds.* CIRCLE

Michelsen, Erik, Jonathan F. Zaff, and Elizabeth Hair. (May 2002). "Civic Engagement Programs and Youth Development: A Synthesis." *Child Trends.* Edna McConnell Clark Foundation

Reinders, Heinz and James Youniss. (2006). "School-Based Required Community Service and Civic Development in Adolescents." *Applied Developmental Science*. Volume 10, Issue 1, pages 2 – 12.

Torney-Purta, J., Lehmann, R., Oswald, H., & Schulz, W. (2001). *Citizenship and education in twenty-eight countries: Civic knowledge and engagement at age fourteen.* Amsterdam, The Netherlands: International Association for the Evaluation of Educational Achievement.

Yates, Miranda and James Youniss, Ed. (2006). Roots of Civic Identity. Cambridge University Press.


**YOUTH DEVELOMPENT**
Baizerman, Michael and Doug Magnuson (Eds.) (2007). Work with Youth in Divided and Contested Societies. Rotterdam: Sense Publishers.

Cahill, Michelle. (1997) *Youth Development and Community Development: Promises and Challenges of Convergence*. New York and Baltimore: Ford Foundation and International Youth Foundation.

Eccles, Jacquelynne and Jennifer Appleton Gootman (Eds). (2002). *Community Programs to Promote Youth Development.* Washington, DC: National Academies Press.

Jelicic, Helena, Deborah L. Bobeka, Erin Phelps, Richard M. Lerner, and Jacqueline V. Lerner. (2007) "Using positive youth development to predict contribution and risk behaviors in early adolescence: Findings from the first two waves of the 4-H Study of Positive Youth Development." *International Society for the Study of Behavioral Development*. 31 (3), 263–273

Knowles, James C. and Jere R. Behrman. (2003) *Assessing the Economic Returns to Investing in Youth in Developing Countries*. The International Bank for Reconstruction and Development / The World Bank.

Lerner, Richard. *Thriving Youth, Flourishing Civil Society: How positive youth development strengthens democracy and social justice*. Bertelsmann Foundation White Paper.

Ministry of Youth, Employment Opportunities and Sports. (2005) *Life Skills Education Report.* Fiji.

Patterson, Jan. (2001) *The Concept of Youth Development: A review of literature from the United States of America on the concept of youth development.* Occasional Paper: AusYouth.

Perold, H. (2000). *Worldwide workshop on youth involvement as a strategy for social, economic, and democratic developm*ent. New York: The Ford Foundation.

Rappaport, Catherine Dun and JoAnn Jastrzab. (2003) *Promising Practices for Helping Low-Income Youth Obtain and Retain Jobs: A Guide for Practitioners.* Report produced for the National Association of Service and Conservation Corps by Abt Associates with funding from the U.S. Department of Labor.

Roth, J. L., & Brooks-Gunn, J. (2003). "What exactly is a youth development program? Answers from research and practice." *Applied Developmental Science*, *7*, 94-111.

Search Institute. (2003). *40 developmental assets*. Retrieved February 4, 2007, from http://www.search-institute.org/assets/40AssetsMC.pdf

Inter-Agency Working Group on Children's Participation (IAWGCP). (2008). *Children as Active Citizens: Commitments and society's obligations for children's civil rights and civic engagement in East Asia and the Pacific*. Bangkok.

UNICEF. (2007). *Adolescents and Civil Engagement: Social Entrepreneurship and Young People: A summary of the role of social entrepreneurship in the development of young people, communities and the achievement of the MDGs.* Adolescent Development and Participation Unit Programme Division, October 2007. New York.

UNDP. *Civic Engagement.* ESSENTIALS No. 8 October 2002

Wesley, C. (1995-2005). "Children as catalysts of change: Children's participation in rural development in India." *Children's Environments,* 12 (4), 38-48. www.colorado.edu/journals/cye/

*Sources and Recommended Reading:*

American Evaluation Association: http://www.eval.org/

Baker, Judy L. 2000. "Handbook for Practitioners: Evaluating the Impact of Development Projects on Poverty". The World Bank, Washington, DC.

Cunningham, Wendy; Cohan, Lorena M.; Naudeau, Sophie; McGinnis, Linda 2008. "Supporting Youth at Risk: A Policy Toolkit for Middle Income Countries" The World Bank, Washington, DC. "Moving from Wish List to Action #3. The Importance of Evaluating Youth Interventions" by Mattias K.A. Lundberg.

Cunningham, Wendy; Cohan, Lorena M.; Naudeau, Sophie; McGinnis, Linda 2008. "Supporting Youth at Risk: A Policy Toolkit for Middle Income Countries" The World Bank, Washington, DC. "Moving from Wish List to Action #4. Selecting Programs Based on Cost-Effectiveness and Cost-Benefit Analysis" by Mattias K.A. Lundberg.

Development Assistance Committee (DAC) Working Party on Aid Evaluation. June 2002. "Glossary of Key Terms in Evaluation and Results Based Management". OECD. Also available in other languages.

Network of Networks on Impact Evaluation "Impact Evaluation Guidance". NONIE

Network of Networks on Impact Evaluation, NONIE 2008. "NONIE Subgroup 2: Impact Evaluation Guidance". January 09, 2008

Stroud, Susan; Alessi, Brett; McGinnis, Linda; Holland, Peter. December 2005. "Youth Service: A Strategy for Youth and National Development." Also available in French.

World Bank Impact Evaluation web-site: http://go.worldbank.org/169GZ6W820.

World Bank Independent Evaluation Group (IEG). "Impact Evaluation – The Experience of the Independent Evaluation Group of the World Bank". The World Bank, Washington, DC.

*For more recommended reading, please see the ICP "Framework on Evaluating the Impact of Youth Service Programs" (Annex D).*