

Evaluation and Microenterprise Programs

Mark Schreiner

August 2001

Microfinance Risk Management
6070 Chippewa St. #1W, St. Louis, MO 63109-3060, U.S.A.
Telephone: (314) 481-9788, <http://www.microfinance.com>
Center for Social Development
Washington University in St. Louis
Campus Box 1196, One Brookings Drive, St. Louis, MO 63130-4899, U.S.A.

Abstract

Microenterprise programs attempt to help poor people start or strengthen small businesses. Funding and political support has grown rapidly. Is microenterprise a good use of scarce development funds? Unfortunately, most evaluations have been case studies in what not to do. Because benefits and costs cannot be measured completely nor with perfect certainty, rigorous evaluations should support their necessarily subjective judgements with logic and explicit assumptions. The usefulness of an evaluation lies not in its (apparent) incontrovertibility but rather in its clarity of assumptions and in its openness to meaningful review and critique.

Acknowledgments

A later version of this paper is published in the *Journal of Microfinance*, 2002, Vol. 4, No. 2, pp. 67–91. This work was supported by a grant from the Division of Asset Building and Community Development of the Ford Foundation. I thank for help from Dale Adams, Cynthia Sanders, and Michael Sherraden.

Author's Note

Mark Schreiner is Research Director at the Center for Social Development, Washington University in St. Louis. He studies ways to help the poor build assets through savings and loans.

Evaluation and Microenterprise Programs

Abstract

Microenterprise programs attempt to help poor people start or strengthen small businesses. Funding and political support has grown rapidly. Is microenterprise a good use of scarce development funds? Unfortunately, most evaluations have been case studies in what not to do. Because benefits and costs cannot be measured completely nor with perfect certainty, rigorous evaluations should support their necessarily subjective judgements with logic and explicit assumptions. The usefulness of an evaluation lies not in its (apparent) incontrovertibility but rather in its clarity of assumptions and in its openness to meaningful review and critique.

Evaluation and Microenterprise Programs

1. Introduction

This paper reviews evaluations of microenterprise programs that make loans and/or provide training to help the poor, the unemployed, and recipients of public assistance to start or to strengthen small businesses. Success stories of microenterprise abroad have inspired growth in the number of U.S. programs from almost none in the 1980s to more than 340 in 1999 (Langer, Orwick, and Kays, 1999). Federal support has also grown steadily (Else, 2000). The growth of microenterprise may, however, draw resources away from other, possibly better, interventions (Howells, 2000).

Is microenterprise a good use of scarce development funds? Scholars disagree. Raheim (1996a, p. 69) says that evaluations show that programs “create opportunities for clients to engage in productive self-employment which will restore self-respect, facilitate self-reliance, and above all transform a condition of dependency to one of self-sufficiency.” Bates (1997) says that such claims are not yet backed by careful evidence.

Because microenterprise is new in the United States, most evaluations focus on *process*—they ask “What happened?” Process evaluation is important (Woller, Wheeler, and Checketts, 1999), but as the budget for microenterprise grows, evaluations should also measure *impact*—they should ask “What benefits were created, and at what cost?”

Of course, impact evaluation is expensive and difficult. Measures of benefits and costs—especially in the absence of the intervention—are always incomplete and imprecise. All evaluations—whether qualitative or quantitative—necessarily rest partly on subjective judgements. A good evaluation is not one whose conclusions seem incontrovertible but rather one that is open to review because the logic that derives conclusions from measurements and assumptions is clear. The goal of rigorous evaluation is to put everyone “on the same page” so that discussion can focus on differences in methods of measurement, logic, or assumptions.

Evaluation so far has not shown whether microenterprise programs are worthwhile. This paper discusses how to improve evaluation through more rigorous estimation of benefits and costs both with and without the program. The paper highlights examples that avoid common pitfalls and emphasizes that funders create the incentives for careful or careless evaluation. As a simple first step toward improvement, the paper proposes comparing costs with outputs. The insights are applicable not just for microenterprise but also for evaluation in general.

2. Rigor

What is rigor? This section argues that evaluations are inevitably subjective but that rigor constrains excesses. Rigor implies transparency in the evidence, logic, and assumptions that support conclusions. Rigorous evaluations are susceptible to critique and thus more likely to be improved.

2.1 All evaluations are subjective

Evaluation compares net benefits with an intervention versus net benefits without the intervention. Some types of benefits and costs (such as changes in income) can be quantified. Other types (such as changes in feelings of self-worth) are necessarily qualitative, but all types should be considered in the final judgement (Plotnick and Deppman, 1999). Even if all effects could be converted to common units (such as dollars), the conversion would inevitably involve subjective judgements such as “a year of life is worth $\$x$ ” or “people are willing to pay $\$y$ to feel this change in happiness”.

Quantification is still useful, not for its own sake but rather because it helps to make assumptions and judgements explicit. For example, financial benefit-cost analysis must be explicit about the financial costs and benefits included and must either assume away non-financial effects or make an explicit qualitative judgement about them.

2.2. Rigor constrains excess subjectivity

Rigor aims to improve inter-personal reliability. Some subjectivity is inevitable, but excesses occur when judgements rest on unexamined experience, fuzzy logic, or implicit assumptions. Subjectivity is non-transparency; opaque or implicit factors lack inter-personal reliability, and this might let mistakes sneak through (McCloskey, 1998). The heart of the social-scientific method is not experiments but explicitness.

Objectiveness or subjectiveness inheres not in an effect but rather in its measurement. *Qualitative* benefits and costs are unmeasured, unmeasurable, or measured in units with low inter-personal reliability; *quantitative* measures have high inter-personal reliability. Analyses are more rigorous as they depend less on the experience and judgement of a specific analyst.

Rigorous evaluations measure what they can and then point out the subjectivity that remains. Often, simply making explicit the factors that influence a judgement provokes ideas for improvement or spotlights gaps in logic. Rigor whittles away unneeded subjectivity and highlights unresolved subjectivity.

Evaluations are inevitably subjective to some degree, and this is not bad. Good subjective judgements, however, are more than mere opinion; they are derived logically from explicit assumptions. Rigor makes the factors behind a judgement as transparent to others as to the analysts. This forces analysts to check their work, and it provides a basis for the type of reasoned discussion that could lead to improvement.

3. Cost measurement

Evaluations of microenterprise programs often understate costs in that they ignore opportunity costs, costs borne by participants, and costs of displaced non-participants. The issue is less that costs are ignored and more that they are implicitly ignored. Casual readers might mistakenly infer that all costs were measured.

After introducing the major evaluations of U.S. programs, this section discusses three often-overlooked types of costs. The next section discusses the measurement of benefits. Even without knowledge of benefits, knowledge of costs is useful.

3.1 Major evaluations

There are four major evaluations of U.S. microenterprise programs. The first—the Self-Employment Learning Project (SELP)—tracked seven of the oldest and best-known microenterprise programs (Clark *et al.*, 1999; Edgcomb *et al.*, 1996). Second, Himes and Servon (1998) analyzed six programs affiliated with Accion International, the U.S. network with the most clients. Third, the Self-Employment Investment Demonstration (SEID) followed seven programs targeted to recipients of public assistance (Raheim and Alter, 1998; Raheim, 1997, 1996a, and 1996b; Friedman *et al.*, 1995). Fourth, the Unemployment Insurance Self-Employment Demonstration (UISED) tested the effect of access to microenterprise services for the newly unemployed (Benus *et al.*, 1995).

3.2 Opportunity Costs

Resources employed in microenterprise development could have been employed elsewhere and so have an opportunity cost. Most microenterprise evaluations use incomplete measures of opportunity costs, or no measures at all.

Suppose an evaluation covers T years and that the net benefit of a dollar in the best alternative project (the opportunity cost) is r . The standard assumption is that r is 10 percent per year in real terms (U.S. Office of Management and the Budget, 1972).

Let E_0 be the resources (equity) in the program at the start of the time frame, let P_t be the profit (or loss) in year t if the program had paid market prices for its resources, let F_t be the net funds transferred to the program in year t , and let $\delta=1/(1+r)$. Present cost as of the start of the time frame is then:

$$\text{Present Cost} = (1 - \delta^T) \cdot E_0 + \sum_{t=1}^T [(\delta^t - \delta^T) \cdot F_t - \delta^T \cdot P_t]. \quad (1)$$

The first term, $(1 - \delta^T) \cdot E_0$, is the difference between the present worth of start equity E_0 at the start of the time frame and at the end. The second term, $(\delta^t - \delta^T) \cdot F_t$, is the difference between the present worth of net funds F_t in year t and in year T . Net funds F_t include grants in-cash, the market value of grants in-kind, and expenses avoided due to soft liabilities with below-market interest rates. The third term, $\delta^T \cdot P_t$, is the present worth of profit from year t in year T . The measure of profit P_t excludes

grants recorded as revenue, but it includes expenses that would be incurred in the absence of grants in-kind and soft liabilities.

No major U.S. evaluation measures the present value of costs. UISED and SELP count costs as funds spent, but they do not discount, adjust for soft liabilities, or measure grants in-kind. The cost study for SEID is not yet public (Raheim, 1998). The Accion evaluation did not measure costs.

3.3 Costs for participants

Microenterprise participants experience both benefits and costs. No major evaluation measures price costs or non-price transaction costs borne by participants.

3.3.1 Price costs

Prices paid by participants include interest and fees for loans and for training. Evaluations of programs that make loans should measure price costs as the real monthly effective interest rate—that is, as the inflation-adjusted internal rate of return (IRR)—on all cash flows associated with a loan. The IRR accounts for the effects of non-price terms of the loan contract such as the number, frequency, and timing of installments; deduction of fees from disbursements; term to maturity; and compensating balances. As a picture of the true cost of debt for borrowers, the IRR is preferred to the annual percentage rate that lenders must report by law.

3.3.2 Non-price transaction costs

Both borrowers and trainees incur transaction costs. More difficult to measure than price costs, transaction costs are often overlooked, but they may swamp price costs (Adams, 1995). These costs can be seen in terms of money, miles, and minutes: the out-of-pocket expenses incurred to participate but paid to someone other than the program, the cost of transport, and the opportunity cost of time.

Transaction costs do not vary much with loan size, so they impinge most on the poorest because they get the smallest loans. For example, suppose a loan of \$1,000 has a price of zero and is repaid in one installment after one year. The borrower, however, may be required to take classes, attend monthly meetings, complete a business plan, and join a joint-liability group. These requirements can take up to 90 hours (Raheim, 1995; Clark and Huston, 1993; Else and Raheim, 1992). If borrowers do not want or need extra help, the opportunity cost of time can make even a “free” loan quite expensive. For example, if time is worth \$6 an hour, then a \$1,000 “free” loan that consumes 90 hours is equivalent to \$1,000 of credit-card debt with an annual price of $\$6 \cdot 90 / \$1,000 = 54$ percent. Transaction costs matter most for small loans, which is why credit cards, pay-day loans, pawn shops, and the best microenterprise programs abroad can charge high interest rates but must provide streamlined service (Caskey, 1994).

3.4 Displacement costs for non-participants

A microenterprise program may help one small business at the expense of another (Garfinkle *et al.*, 1992). Because the poor often lack skills and wealth, they cluster in sectors with low barriers to entry (Bates, 1997). There, competition is high, profits are low, and one firm's gain may be another's loss. Displacement is very difficult to measure, but it may be as high as half of the net benefits of participants (Bendick and Egan, 1987). No major U.S. evaluation discusses displacement costs.

4. Benefit measurement

The previous section argued that evaluations of microenterprise programs have understated costs. This section argues that they have overstated benefits.

Participants do get positive net benefits from microenterprise programs; if not, they would drop out. For process evaluations, mere participation may imply success. From a social perspective, however, the key question is not whether there are benefits, nor whether net benefits are positive, but rather whether net benefits are so positive that microenterprise is better than other ways to reach the same goals. Because the poor are plenty but the funds are few, evaluations should measure not just the sign of benefits but also the size of both benefits and costs.

Measurement of benefits faces two challenges. First, as discussed in the next section, participants are observed only with participation, not without. Second, as discussed in this section, benefits are not observed, so proxies based on output or outcomes are used. Issues with these proxies include the proper units, aggregation, absolute versus relative measures, and a one-sided focus on positive outcomes.

4.1 Units

4.1.1 Jobs

The output of microenterprise programs is often measured in units of “jobs created”. Not all jobs, however, are created equal. The best convention is to measure full-time equivalents. No major evaluation does this. For example, SEID reports the

total number of self-employed participants, whether high-paid, low-paid, full-time, or part-time. This overstates benefits because most self-employment is low-paid and part-time (Spalter-Roth, Hartmann, and Shaw, 1993).

Some evaluations count as “strengthened” or “saved” all jobs in businesses owned by participants (Clark and Huston, 1993). Without knowing what would have happened without the program, these numbers do not mean much.

4.1.2 Stocks and flows

Some work fails to distinguish between stocks and flows. For example, two censuses of U.S. programs (Langer, Orwick, and Kays, 1999, and Severens and Kays, 1997) report average loan sizes and numbers of borrowers but do not say whether these are loans disbursed or loans outstanding. Failure to distinguish stocks and flows allows programs to report the largest number.

This also matters for training. For example, the average program in 1999 “served” 202 people, most of whom were trainees (Langer, Orwick, and Kays, 1999). The flow of trainees in a year likely exceeds the stock at any point. (Furthermore, the number of trainees “served” likely exceeds the number graduated.) The right unit depends on the question asked, but units should always be reported.

4.1.3 Business starts

Some evaluations report the number of business starts as if firms sprang to life all at once. This is too simple for three reasons. First, participants may join programs

with a business already open. Second, drop-outs may still start businesses. Third, most new ventures (53 percent) fail within four years (Berger and Udell, 1998). As explained by Raheim *et al.* (1996, p. 93), “starting a business is a process rather than an event”, so evaluations should track progress through a series of landmarks such as a business plan, legal establishment, the start of work, and survival through time.

4.1.4 Income

Income measures have three common weaknesses. The first is to report not changes but levels. Levels overstate benefits unless income is zero without the program.

The second common weakness is to report *income* instead of *returns*. Returns account for time worked and capital invested. Bates (1997) and Drury *et al.* (1994) do report returns. Bates computes profit per hour of owner labor and deducts a 10-percent return to capital. Drury *et al.* (1994) report not just average profits of \$1,200 per month but also average returns on owner labor of \$6.70 per hour. This figure is useful because it can be compared with the \$12.41 per hour earned before unemployment and with the \$10.55 per hour earned by participants in parallel job-training programs.

The third common weakness is failure to define *income*. Himes and Servon (1998) state that self-employment income might be seen as revenue, revenue less expenses, owner’s draw, or change in retained earnings. In fact, the correct measure is business profit before taxes and before owner’s draw. Some evaluations report business revenue, perhaps because it is higher than income. For example, Raheim (1996a)

reports an average “income” in SEID of \$21,000 even though, according to Servon and Bates (1998), revenue net of expenses was about \$3,000.

4.2 Aggregates and averages

Some evaluations report cumulative flows rather than annual flows. For example, Langer, Orwick, and Kays (1999) report that the 341 programs in their census served 250,000 people and made loans for \$160 million (average age was 7.5 years).

Cumulative flows mask current performance by mixing it with past performance. A figure of 250,000 cumulative participants may sound better—but conveys less—than a figure of 100 participants per program per year.

Aggregation over participants is also contra-indicated. For example, Raheim *et al.* (1998) report that total sales were \$3.5 million for participants in the programs in Drury *et al.* (1994). More useful, if perhaps less impressive, is that income per participant per year was about \$12,000.

4.3 Non-positive measurements

Some evaluations accent the positive and downplay or omit everything else. In principle, evaluations are not proofs but tests, and evaluators are “agnostic or open-minded about a project [They] should neither justify nor act as hatchets” (Nares, 1990, p. 33). In practice, evaluations “routinely cite impressive-sounding, yet very selective statistics on program success” (Bates, 1996, p. 28).

4.3.1 Half statistics

A common practice is the “half statistic”. For example, Severens and Kays (1997) note that 30 percent of participants were low-income; they do not mention that 70 percent were not. A good example that avoids this is Raheim and Alter (1998), who report the share of participants whose self-esteem grew, fell, or did not change.

4.3.2 Drop outs

Improvement requires feedback on weaknesses as well as on strengths, but data on drop-outs are conspicuous in their absence. After all, drop-outs are the simplest way to evaluate value to participants: they will leave unless they expect positive net benefits. Measurement of drop-outs is standard for microenterprise abroad (Rosenberg, 2001), but Himes and Servon (1998) is the only U.S. evaluation to discuss drop-outs.

4.3.3 Loan repayment

Evaluations of programs that make loans should analyze repayment. A program that does not recoup its loans cannot help many poor people and cannot be sure that it helps the most deserving. Arrears plague many U.S. programs (Bhatt *et al.*, 2001).

The best measure of arrears is *aged portfolio at-risk*. This is the entire balance of all loans with any installments overdue, divided by the whole portfolio, and grouped by the age of the oldest overdue installment. Few evaluations report aged portfolio at-risk. The measures commonly reported understate the risk of loan losses. For example, Servon (1996) reports the number of loans overdue. This serves some purposes, but it

misses the greater risk of larger loans. *Payments overdue* understates risk because it ignores that one missed payment signals greater risk for the entire loan balance.

Edgcomb *et al.* (1996) report *recuperation*, the ratio of cash received to cash due.

Recuperation usefully measures the rate of loan losses in the past, but it may not reflect current risk. A quick test for the quality of lending program is whether it routinely measures repayment, especially aged portfolio at-risk (Moll, 1997).

5. What would have happened

Impact is what happened with a microenterprise program that would not have happened without it. The “with” case did happen, so behavior of treatments (eligibles or participants) can—in principle—be directly measured. The “without” case did not happen, so behavior cannot be directly measured. The central challenge of evaluation is the estimation of the *counterfactual*, behavior in the “without” case (Moffitt, 1991).

UISED estimated the counterfactual carefully and took impact as with-versus-without; other evaluations took impact as before-versus-after. Before-versus-after allows direct measurement of a counterfactual (before the program), but it falsely ascribes all changes to the microenterprise program. With-versus-without is preferred.

5.1 Before-versus-after

Before-versus-after counterfactuals have three problems: they assume that the program caused all changes, they ignore the self-selection of treatments, and they ignore the self-selection of survivors.

5.1.1 Time trends

Microenterprise programs surely affect some outcomes, but they just as surely are not the sole cause of all outcomes. However, the typical evaluation (*e.g.*, SEID and SELP) ascribes all income and all business start-ups and expansions to the program. The (implicit) counterfactual of no microenterprise without the program is unlikely.

With before-versus-after data, the assumed counterfactual—as always—should be explicit. Also, the analyst should ask participants about impact. Servon (1996, p. 47) found that “nearly all the women interviewed had already started or claimed that they would have started their businesses whether or not [the program] existed.”

5.1.2 Self-selection of participants

People who expect high net benefits are more likely to choose to join a microenterprise program than people who expect low or negative net benefits. Thus, the average net benefit for participants probably exceeds the average net benefit for eligibles, had they participated. This is self-selection bias.

One way to mitigate self-selection bias is to control for differences in observed traits (for example, sex, education, or work experience) between eligibles who choose to join and those who do not. Sanders (2000) controls for a few observed traits with data from SELP and from the Panel Study of Income Dynamics. Sanders finds much smaller impacts than Clark *et al.* (1999), who did not control for any observed traits.

Controlling for observed traits is a useful start, but unfortunately self-selection also depends strongly on unobserved traits. Because entrepreneurship requires independent work with high risks, people with unobserved oomph and hustle have the best chance to succeed and thus have higher expected net benefits. Even with observed traits constant, differences before-versus-after a microenterprise program are due partly

to participation and partly to existing spunk and grit. Ignoring the effects of unobserved traits overstates the effects of the program. Again, this is self-selection bias.

UISED is the only major evaluation to account for self-selection on both observed and unobserved traits. Except for Raheim and Alter (1998), no major evaluation controls even for observed traits.

5.1.3 Self-selection of survivors

People with high expected net benefits are more likely to complete a program, start a business, and keep it open. Expected benefits depend not only on the program but also on observed and unobserved traits. Thus, before-and-after analyses that credit programs for all changes in participants who have not dropped out or closed their businesses overstate impact. This is survivor bias, a type of self-selection bias.

Survivor bias may explain part of the positive link between profits and repeat loans found by Himes and Servon (1998) for Accion and by Clark and Huston (1993) for SELP. Servon and Bates (1998) suggest that survivor bias may also explain why profits in SEID (Raheim, 1996) were three times higher than for self-employed women in a broader sample studied by Spalter-Roth, Soto, and Zandniapour (1994).

5.2 Control groups

In before-versus-after studies, outcomes for treatments before treatment provide the counterfactual. In with-versus-without studies, outcomes for non-participants or non-eligibles—control groups—provide the counterfactual.

5.2.1 Principles

The ideal control group has the same joint distribution as the treatment group for all traits—observed and unobserved—that affect the outcome. The closer the joint distributions, the better the control group. If controls match treatments in all ways except for treatment, then all differences in outcomes can be attributed to treatment.

5.2.1.1 Experiments

A good control group is hard to find. To avoid self-selection bias, controls must lack choice about treatment. (Panel data or econometric models may also allow more complex controls for self-selection [Reichart and Mark, 1998; Moffitt, 1991].) The gold standard for exogeneity in treatment status is random assignment, either through design or through nature (Manski, 1995; Meyer, 1995).

In a classic experiment, the evaluator excludes some eligibles (or qualified applicants) at random. This purges self-selection bias because treatment is uncorrelated with individual traits. UISED is the only classic experiment in microenterprise.

In a “natural” experiment, constraints outside the evaluation eliminate choice for some potential treatments. For example, a metro area might straddle state borders, with only one state providing its residents with access to microenterprise programs. Natural experiments typically require more assumptions than classic experiments, for example, that people do not switch states to take advantage of programs.

Even with some type of experiment, impact evaluation is difficult (Heckman and Smith, 1995). Severe threats to validity plague even medical trials, where evaluations can affect thousands of lives and perhaps billions of dollars (Mason and Drummond, 1995). The most common weakness is the simple failure to make assumptions explicit.

5.2.1.2 Comparison groups

Comparison groups match treatments with non-treatments on a small set of observed traits such as sex, location, and/or receipt of public assistance. Although both groups have the same joint distribution for these traits, they may not have the same joint distribution for other traits. Thus, comparison groups are an inexpensive (but imperfect) substitute for experimental control groups. The only use of comparison groups in microenterprise is Raheim and Alter (1998) and Sanders (2000). Comparison groups are useful as long as the assumptions required to derive estimates of impact are explicit. Otherwise, users may remain unaware that impact is likely overstated.

5.2.2 Practices

Himes and Servon (1998, p. 8) write that “the field desperately needs some research that uses a control group.” Four factors help to explain why no major evaluation except UISED has used control groups.

First, control-group evaluations are lengthy and expensive. Unfortunately, donors, governments, and programs often seek quick evaluations, perhaps to support

requests for the next tranche of funds. The high cost of rigorous evaluation means that even the biggest funders can afford only a few evaluations.

Second, some program staff believe that experiments are ethically wrong. After all, programs were started precisely to combat the arbitrary injustice that random assignment seems to perpetrate. This belief fails to consider that resources in microenterprise might help the poor more in some other development intervention.

Third, some impacts elude quantification even with control groups. For example, increased self-esteem may be the most important effect of participation (Raheim and Alter, 1998; Spalter-Roth, Soto, and Zandniapour, 1994). Many effects are indeed too complex, subtle, diffuse, or long-term to be quantified at low cost, but disciplined arguments about qualitative judgements are still possible if based on ruthless logic and explicit assumptions. Bates (1997 and 1996) critiques SEID not because the conclusions are based on qualitative criteria but rather because the criteria are left unstated.

Fourth and most important, most evaluators have been advocates. Although most were probably unaware of the weaknesses of their methods, few were disappointed when impacts turned out to be large and positive.

The dirty secret in microenterprise (and elsewhere [Adams, 1988]) is that few evaluations are really tests. Except perhaps for UISED, evaluations were funded and conducted by people who already believed that microenterprise was worthwhile. Thus, the projects are called “demonstrations”, not “tests”. They sought not to check whether

microenterprise works but rather to convince funders and policymakers that it works. Rigor is secondary to showing large impacts.

These are disturbing claims. Often, evaluators were simply unaware of the weaknesses in their analyses. Sometimes, however, they were aware and chose nonetheless to disseminate figures biased in favor of microenterprise. For example, descriptions of the benefits from SEID are widely published, but the cost study—though completed—is not yet available. Benus *et al.* (1995) were extremely rigorous in that they estimated two sets of impacts for UISED, one including an outlier with income of \$500,000 a year and another excluding it. Subsequent discussion, however, tends to focus on the (more positive) results that include the outlier. Likewise, UISED had two sites, one with positive impacts and one with almost no impact. The policy derived from the report (Vroman, 1997) focuses on the positive impacts. Finally, the SELP data has some very large figures for changes in income and assets, likely due to misplaced decimal points. Sanders (2000) says medians would be more appropriate measures (or elimination or correction of outliers), but Kays *et al.* (1999) use means.

In short, evaluators did not always fix known biases. Some argue that there is a prisoner's dilemma in evaluation; greater rigor would destroy microenterprise's chances with policymakers against competing, non-rigorous proposals. Thus, challenges to the evaluation of microenterprise probably mirror challenges to evaluation in general. Policymakers, however, are not stupid, and they may grow to regard evaluation as little more than social-science fiction.

6. Discussion and recommendations

Is microenterprise a good way to help the poor? Three of the four major evaluations say that it is. For SEID, Friedman *et al.* (1995, p. 16) say that “proliferation of self-employment programs in the United States since (and because of) SEID bodes well for the broad opening of a realistic self-employment option for welfare recipients.” For SELP, Clark *et al.* (1999, p. 68) say that “microenterprise can offer an effective entry point into the mainstream economy for the poor.” For UISED, Benus *et al.* (1995, p. xi) say that “self-employment assistance is a cost-effective approach to promote the rapid re-employment of unemployed workers and should be permanently incorporated into the U.S. employment security and development system.”

This paper argues for more caution. It concurs with Servon and Bates (1998, p. 28) in that “the microenterprise strategy needs better evaluation”. Undoubtedly, impact has been positive, and some participants have turned their lives around. The question for public policy, however, is not whether some people benefit but whether the poor as a whole are better off with microenterprise than with something else. Three recommendations are offered to guide improvement: beware of backlash, reward rigor, and compare outputs with costs.

6.1 Beware of backlash

If microenterprise is worthwhile, then the case should stand up to rigorous analysis. If it is too soon to judge, then evaluations should note this and then argue from explicit assumptions how quickly the required improvement might occur.

In the long term, people will realize that microenterprise is not a panacea, and the backlash may sting. Bates (1996, p. 28) says that “advocates put their cause at risk when they substitute inflated claims and selective statistics for serious analysis.”

Microenterprise has benefits and costs, and experience abroad suggests that failure to account for both sides can harm the poor in the long term (Adams *et al.*, 1984).

Weak analysis can hurt in three ways. First, advocacy distracts effort from the use of feedback to improve performance. Failure is often a better teacher than success. Second, better measurement might improve work-a-day management. If managers do not know how well they perform, then they are unlikely to try to improve. Third, overstated claims may divert funds from projects that could help the poor more. The goal is not more microenterprise but rather improved well-being for the poor.

Benus *et al.* (1995) is an ideal example of rigor. The thoroughness of the report was such that a detailed critique (with different conclusions) was possible (Schreiner, 1999). Another excellent example is the financial benefit-cost analysis of Individual Development Accounts (matched savings accounts for the poor that may be used to capitalize microenterprise) by Clones *et al.* (1995). This analysis is uncommonly

rigorous not because of its accurate measurement (all the numbers are *pro forma*) but because it carefully enumerates different sources of benefits and costs for different groups and then discusses which benefits and costs might be measured, which cannot be measured, and what assumptions are used to arrive at estimates.

6.2 Reward rigor

Government officials and program officers for private funders are rewarded (with promotions, with invitations to speak at conferences or to write prefaces, with feelings of having done good) more for new, successful ideas than for failed experiments. In turn, evaluators suppose that their own rewards (in future contracts) are greatest if a project is deemed successful. Until the rewards for the people who fund and conduct evaluations depend more on the rigor of the process than on the judgement itself, a sanguine bias is inevitable. This does not impugn the motives or morals of funders or evaluators; it merely recognizes explicitly that who pays the piper calls the tune.

The incentive structure should reward learning, not victory. Learning on such a broad scale as a microenterprise movement, however, is extremely difficult to measure. A focus on changes between interim evaluations might provide some incentives in early evaluations to highlight weaknesses that could be resolved before later evaluations.

Funders might also explicitly forbid emotional appeals (for example, photographs of entrepreneurs in reports). Qualitative investigation can convey ideas that nothing else can, but evidence should trump emotion. What matters are not the faces of specific

cases but rather the faceless millions who will get the benefits (or bear the costs) if funds are correctly or incorrectly allocated to microenterprise.

Evaluators might receive fixed contracts for x jobs and then no more. Like lame-duck presidents, they might be more willing to say whether the emperor has no clothes.

Finally, funders could link rewards for program officers to blind reviews of evaluations that seek to identify unacknowledged threats to validity or implicit assumptions. This would reward evaluators who make caveats explicit.

6.3 Compare outputs with costs

Perhaps the simplest way to inform judgements of the performance of microenterprise programs is to compare outputs with costs. Like any shopper, funders should know the price tag.

It is less expensive to compare outputs with costs than to compare benefits with costs. Of course, it is also less informative; the analyst must still judge whether the unmeasured benefits associated with outputs could reasonably exceed costs. Still, “cost calculations can provide a useful ‘reality check.’ . . . Whatever the true size of external benefits, the [funder] must judge that at a minimum the external benefit exceeds this cost for the intervention to be worth undertaking (Devarajan *et al.*, 1997, p. 40).”

Comparing outputs with costs forces analysts to make their judgements explicit. A good example is Edgcomb *et al.* (1996); they argue that their estimated cost of \$6,000

per participant is “close” to that of federal job programs (\$3,500 per participant) because microenterprise participants are poorer.

Likewise, Edgcomb *et al.* (1996) estimate that a dollar-year of debt produced in SELP cost \$1.47. If social gains are to exceed social costs, then borrowers must get at least \$1.47 of surplus per dollar-year of debt.

Schreiner and Morduch (2001) use data from Severens and Kays (1997) and from Langer, Orwick, and Kays (1999) to compute an average cost per participant for U.S. programs of about \$2,000 in 1996 and \$1,300 in 1999. Raheim (1997) also mentions a cost per participant of about \$2,000 for SEID. Of course, cost estimates alone do not reveal whether microenterprise is worthwhile, but evaluators might start to ask whether benefits per participant are likely to exceed \$2,000.

Servon and Doshna (2000) compare costs per job created across several interventions. (It is not clear whether “jobs” are full-time equivalents nor whether they are truly impacts of the interventions.) In three microenterprise programs, costs per job ranged from \$4,000 to \$6,000. In industrial recruitment, costs per job were between \$2,000 and \$10,000 in one study and between \$11,000 to \$50,000 in another study. Costs per job were \$1,500 to \$2,000 in business incubators, \$5,000 in a public-works program, and \$3,000 to \$5,000 in revolving loan funds. Servon and Doshna (p. 191) conclude that “on this measure, microenterprise development is well within the scope of other economic-development strategies.”

References

- Adams, D. W. 1995. Transaction Costs in Decentralized Rural Financial Markets. In *Agriculture in Liberalizing Economies: Changing Roles for Governments*, edited by D. Umali-Deininger and C. Maguire, 249-65. Washington, D.C.: World Bank.
- Adams, D. W. 1988. The Conundrum of Successful Credit Projects in Floundering Rural Financial Markets. *Economic Development and Cultural Change* 36(2):355-68.
- Adams, D. W., D. H. Graham, and J. D. Von Pischke. 1984. *Undermining Rural Development With Cheap Credit*. Boulder: Westview.
- Bates, T. 1997. *Race, Self-employment, and Upward Mobility*. Baltimore: Johns Hopkins University Press.
- Bates, T. 1996. Reply to Raheim. *Journal of Developmental Entrepreneurship* 1(1):27-9.
- Bendick, M., and M. L. Egan. 1987. Transfer Payment Diversion for Small Business Development: British and French Experience. *Industrial and Labor Relations Review* 40(4):528-42.
- Benus, J. M., T. B. Johnson, M. Wood, N. Grover, and T. Shen. 1995. Self-Employment Programs: A New Re-employment Strategy, Final Report on the UI Self-employment Demonstration. Unemployment Insurance Occasional Paper 95-4. U.S. Department of Labor.
- Berger, A. N., and G. F. Udell. 1998. The economics of small-business finance: The roles of private equity and debt markets in the financial growth cycle. *Journal of Banking and Finance* 22:613-73.
- Bhatt, N., S.-Y. Tang, and G. Painter. 2001. Microcredit Programs in the United States: The Challenges of Outreach and Sustainability. In *Replicating Microfinance in the United States*, edited by J. Carr and Z.-Y. Tong. Washington, D.C.: Fannie Mae Foundation.
- Caskey, J. P. 1994. *Fringe Banking*. New York: Russell Sage.

- Clark, P., A. Kays, L. Zandniapour, E. Soto, and K. Doyle. 1999. *Microenterprise and the Poor: Findings from the Self-Employment Learning Project*. Washington, D.C.: Aspen Institute.
- Clark, P., and T. Huston. 1993. *Assisting the Smallest Businesses: Assessing Microenterprise Development as a Strategy for Boosting Poor Communities*. Washington, D.C.: Aspen Institute.
- Clones, D., R. Friedman, B. Grossman, and C. Wilson. 1995. *The Return of the Dream: An Analysis of the Probable Economic Returns on a National Investment in Individual Development Accounts*. Washington, D.C.: Corporation for Enterprise Development.
- Devarajan, S., L. Squire, and S. Suthiwart-Narueput. 1997. Beyond Rate of Return: Reorienting Project Appraisal. *World Bank Research Observer* 12(1):35-46.
- Drury, D., S. Walsh, and M. Strong. 1994. Evaluation of the EDWAA Job Creation Demonstration. Research and Evaluation Report Series 94-G. Washington, D.C.: U.S. Department of Labor.
- Edgcomb, E., J. Klein, and P. Clark. 1996 *The Practice of Microenterprise in the U.S.: Strategies, Costs, and Effectiveness*. Washington, D.C.: Aspen Institute.
- Else, J. F. 2000. An Overview of the Microenterprise Development Field in the U.S. Geneva: International Labour Office.
- Else, J. F., and S. Raheim. 1992. AFDC Clients as Entrepreneurs: Self-Employment Offers an Important Option. *Public Welfare* 50(4):36-41.
- Friedman, R. E., B. Grossman, and P. Sahay. 1995. Building Assets: Self Employment for Welfare Recipients. Overview of the Findings from the Self Employment Investment Demonstration. Washington, D.C.: Corporation for Enterprise Development.
- Garfinkle, I., C. F. Manski, and C. Michalopoulos. 1992. Micro Experiments and Macro Effects. In *Evaluating Welfare and Training Programs*, edited by C. F. Manski and I. Garfinkle, 253-73. Cambridge: Harvard University Press.
- Heckman, J. J., and J.A. Smith. 1995. Assessing the Case for Social Experiments. *Journal of Economic Perspectives* 9(2):85-110.

- Himes, C., and L. J. Servon. 1998. *Measuring Client Success: An Evaluation of Accion's Impact on Microenterprises in the United States*. Boston: Accion Intl.
- Howells, L. A. 2000. The Dimensions of Microenterprise: A Critical Look at Microenterprise as a Tool to Alleviate Poverty. *Journal of Affordable Housing* 9(2):161-82.
- Langer, J. A.; J. A. Orwick, and A. J. Kays. 1999. *1999 Directory of U.S. Microenterprise Programs*. Washington, D.C.: Aspen Institute.
- Manski, C. F. 1995. Learning about Social Programs from Experiments with Random Assignment of Treatments. Institute for Research on Poverty Discussion Paper No. 1061-95. University of Wisconsin-Madison.
- Mason, J., and M. Drummond. 1995. Reporting Guidelines for Economic Studies. *Health Economics* 4:85-94.
- McCloskey, D. N. 1998. *The Rhetoric of Economics, Second Edition*. Madison: University of Wisconsin Press.
- Meyer, B. D. 1995. Natural and Quasi-Experiments in Economics. *Journal of Business and Economic Statistics* 13(2):151-61.
- Moffitt, R. 1991. Program Evaluation With Nonexperimental Data. *Evaluation Review* 15(3):291-314.
- Moll, H. A. J. 1997. How to evaluate credit projects in ten minutes. *Development in Practice* 7(3):300-3.
- Nares, P. 1990. *The Self-employment Strategy, Gaining Momentum, Expanding Opportunities*. Washington, D.C.: Corporation for Enterprise Development.
- Plotnick, R. D., and L. Deppman. 1999. Using Benefit-Cost Analysis to Assess Child-Abuse Prevention and Intervention Programs. *Child Welfare* 78(3):381-407.
- Raheim, S. 1998. Message to the author, Nov. 25.
- Raheim, S. 1997. Problems and Prospects of Self-Employment as an Economic Independence Option for Welfare Recipients. *Social Work* 42(1):44-53.

- Raheim, S. 1996a. Micro-enterprise as an approach for promoting economic development in social work: lessons from the Self-Employment Investment Demonstration. *International Social Work* 39:69-82.
- Raheim, S. 1996b. Rejoinder: Toward a Broader Understanding of the Needs of African-American Entrepreneurs. *Journal of Developmental Entrepreneurship* 1(1):17-26.
- Raheim, S. 1995. Self-Employment Training and Family Development: An Integrated Strategy for Family Empowerment. In *Reinventing Human Services: Community- and Family-Centered Practice*, edited by P. Adams and K. Nelson, 127-43. New York: Aldine de Gruyter.
- Raheim, S., and C. F. Alter. 1998. Self-Employment as a Social and Economic Development Intervention for Recipients of AFDC. *Journal of Community Practice* 5(1/2):41-61.
- Raheim, S., C. F. Alter, and D. Yarbrough. 1996. Evaluating Microenterprise Programs: Issues and Lessons Learned. *Journal of Developmental Entrepreneurship* 1(2):87-103.
- Reichardt, C. S., and M. M. Mark. 1998. Quasi-experimentation. In *Handbook of Applied Social Research Methods*, edited by L. Bickman and D. J. Rog, 192-228. Thousand Oaks: Sage.
- Rosenberg, R. 2001. Measuring Client Retention. *MicroBanking Bulletin*, 6:25-6.
- Sanders, C. K. 2000. Microenterprise versus the Labor Market Among Low-Income Workers. Ph.D. dissertation. Washington University in St. Louis.
- Schreiner, M. 1999. Lessons for Microenterprise Programs from a Fresh Look at the Unemployment Insurance Self-Employment Demonstration. *Evaluation Review* 23(5):503-26.
- Schreiner, M., and J. Morduch. 2001. Replicating Microfinance in the United States: Opportunities and Challenges. In *Replicating Microfinance in the United States*, edited by J. Carr and Z.-Y. Tong. Washington, D.C.: Fannie Mae Foundation, www.microfinance.com.
- Servon, L. J. 1996. Microenterprise Programs and Women: Entrepreneurship as Individual Empowerment. *Journal of Development Entrepreneurship* 1(1):31-55.

- Servon, L. J., and J. P. Doshna. 2000. Microenterprise and the Economic Development Toolkit: A Small Part of the Big Picture. *Journal of Developmental Entrepreneurship* 5(3):183-208.
- Servon, L. J., and T. Bates. 1998. Microenterprise as an Exit Route from Poverty: Recommendations for Programs and Policy Makers. *Journal of Urban Affairs* 20(4):419-41.
- Severens, C. A., and A. K. Kays. 1997. *1996 Directory of U.S. Microenterprise Programs*. Washington, D.C.: Aspen Institute.
- Spalter-Roth, R., E. Soto, and L. Zandniapour. 1994. *Micro-enterprise and Women: The Viability of Self-Employment as a Strategy for Alleviating Poverty*. Washington, D.C.: Institute for Women's Policy Research.
- Spalter-Roth, R., H. I. Hartmann, and L. B. Shaw. 1993. *Exploring the Characteristics of Self-Employment and Part-Time Work Among Women*. Washington, D.C.: Institute for Women's Policy Research.
- U.S. Office of Management and Budget. 1972. Discount rates to be used in evaluating time-distributed costs and benefits. Circular No. A-94 (rev.). Washington, D.C.
- Vroman, W. 1997. *Self-Employment Assistance: Revised Report*. Washington, D.C.: Urban Institute.
- Woller, G. M., G. Wheeler, and N. Checketts. 1999. Evaluation Practices in Microcredit Institutions. *Journal of Developmental Entrepreneurship* 4(1):59-80.