This PDF is a selection from a published volume from the National Bureau of Economic Research

Volume Title: Targeting Investments in Children: Fighting Poverty When Resources are Limited

Volume Author/Editor: Phillip B. Levine and David J. Zimmerman, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-47581-6
ISBN13: 978-0-226-47581-3

Volume URL: http://www.nber.org/books/levi09-1

Conference Date: September 26, 2008

Publication Date: August 2010

Chapter Title: Issues in Implementation

Chapter Authors: Phillip B. Levine, David J. Zimmerman

Chapter URL: http://www.nber.org/chapters/c11721

Chapter pages in book: (13 - 23)

1

Issues in Implementation

Phillip B. Levine and David J. Zimmerman

1.1 Introduction

As we described in the introduction to this volume, the goal of this exercise is to identify the childhood interventions that are most successful at alleviating subsequent poverty. Although this goal is straightforward in theory, its implementation in practice raises a number of difficulties. It raises a number of conceptual issues that need to be resolved to better frame the scope of the exercise. It also raises a number of empirical issues regarding the specific approach we will use to be able to measure interventions along a scale of success. These specific empirical issues are addressed directly in chapter 13; this chapter focuses on those issues that are more conceptual in nature.

We begin by describing the specific questions we will be able to address and which ones we will leave for further, future analysis. We label these "targeting" issues because they define the target that we are shooting at. What is the subpopulation of poor people that we will focus on? What do we consider success? Who are we seeking to benefit, the individual or society more broadly? The first part of the chapter will more fully describe these issues and describe our approach to resolve them in the subsequent analysis.

This chapter will also address an additional set of conceptual issues that will affect the implementation of our analysis. Broadly speaking, we will consider the standards to be used in culling information from the broad array of program evaluations that have been conducted. We call these "empirical evaluation issues," and our resolution of them will define the way that we

Phillip B. Levine is the Class of 1919 Professor and chair of the department of economics at Wellesley College, and a research associate of the National Bureau of Economic Research. David J. Zimmerman is a professor of economics and Orrin Sage Professor of Political Economy at Williams College, and a research associate of the National Bureau of Economic Research.

plan to incorporate the results of previous research. What evidence counts? Does the scale of the evaluation matter? Which outcomes should we consider? The second part of the chapter will focus on these issues and resolve how we will address them subsequently.

Our approach in this chapter will be to specify individual, substantive challenges that we face that require specific decisions to be made in order to organize the remainder of the study. The decisions we have made are by no means the definitively correct ones. It is certainly possible to question any or all of them. We recognize that there are counterarguments to many of those that we make in supporting our decisions; in fact, we engaged in many of those arguments in the process of formulating this volume. But we also recognize that decisions need to be made to narrow the scope of the exercise sufficiently so that we can arrive at a useful and informative final product. We believe we have made the best ones to meet the needs of our particular analysis.

1.2 Targeting Issues

Targeting issues refer to the scope of the exercise that we seek to conduct. We need to limit its scope because addressing all of the potential issues would quickly overload any productive evaluation. This section will identify those challenges and describe the decisions we have made to resolve them in this volume.

1.2.1 Challenge 1: What Is the Relevant Subpopulation?

Our broad goal is to evaluate the effectiveness of policies designed to help improve the future economic success of the poor. The first challenge that we face in implementing that goal is deciding what we mean by "the poor." The poor is a multifarious group; different policy interventions focus on different subsets of that group. Do we focus on all poor people? Perhaps at a more basic level, we need to decide what we mean by poor. Who is poor?

This latter question turns out to be the one that is easier to answer. One way to address the problem of who is poor is to rely on national standards incorporated into the poverty line. If you are below the poverty line, you are poor, and if you are above it, you are not. Yet the poverty line is not necessarily a perfect indicator of who is poor; it has been widely criticized in the past.[1] If we jettison that definition, we are restricted to more ambiguous concepts. We could, for instance, focus on the most destitute (the homeless?) or perhaps the working poor.

In practice, we take a more pragmatic approach to addressing this challenge. The core of our analysis is an examination of policy interventions that have been conducted attempting to reduce poverty through alternative

1. See Blank (2008) for a useful discussion regarding the measurement of poverty.

mechanisms. Each of those interventions struggled with this issue of the target population. Because our analysis simply synthesizes the results of those analyses, we are bound by the earlier decisions made by past analysts. These decisions generally have not led to a focus on the working poor or the most destitute, but to a broader population of low-income households. Throughout the remainder of our discussion, we will use the term "poverty" as a synonym for low income and not according to its more formal definition.

Now that we have "defined" who is poor (where our definition will just rely on those incorporated into past policy interventions), we need to address whether we want to focus on particular subgroups of the poor. The issues and needs of different groups of poor individuals differ, and policy interventions attacking them take very different approaches. Addressing all of them would be quite an undertaking.

One way that we have chosen to cut down the scope of the exercise is to limit the set of policies we consider to those targeted at those who are young. We explore interventions that focus on the early childhood years, the primary school years, the secondary school years, and the very early adult years (into the early twenties). Our rationale for doing so is two-pronged. From a practical perspective, low-income children are the focus of a large number of policy interventions. This group appears to enjoy greater political support, which is needed to have funds moved in their direction. The willingness to spend resources to support low-income adults wanes in the face of their potential to support themselves. The sweeping welfare reform legislation in 1996 is an example of this. Interventions targeted at children may be more popular because it is "not their fault" that they are poor. For whatever reason, one reason to focus on low income children is that they are the target of more interventions. It just makes more sense to look where the light is shining.

We can justify the focus on interventions targeted at low-income children from a broader, conceptual framework as well, imbedding this decision within a human capital framework. From this viewpoint, investments in children make more sense because any persistent benefit from such an intervention will accrue for a longer time. In fact, James Heckman, along with different coauthors (Carneiro and Heckman 2003; Cunha et al. 2006; Cunha and Heckman 2007), has made this a cornerstone of one of his recent research agendas. He takes the human capital perspective to the extreme, arguing that interventions should target the youngest of the poor population. Skill begets skill, from his viewpoint, and early, effective interventions generate the equivalent of compound interest on the investment return. We are sympathetic to this message but prefer to take a less philosophical approach, letting the data tell us exactly where the greatest returns are found. We are willing, however, to use the human capital perspective to justify focusing our attention on interventions directed at the young.

1.2.2 Challenge 2: What Do We Consider Success?

Policy interventions designed to help the poor often have different goals. Some target their health (mental or physical). Others attempt to improve their safety, making them more secure in their environment. Still others attempt to improve their economic conditions. Among those focusing on economic well-being, some interventions identify short-term improvements as the goal, whereas others emphasize long-term gains in economic standing.

All of these goals are important, tackling a different aspect of the ways in which we can improve the well-being of the poor. If the poor do not have access to adequate health care, their health may suffer and reduce their quality of life. Similar arguments can be made regarding personal safety. Quality of life surely suffers if one is victimized by a violent act or even if one is so worried about victimization that daily activities are modified to reduce its likelihood. Improving one's economic circumstances can help improve these aspects of quality of life as well as others. But policies directed at immediate economic remediation may require repeated interventions to alleviate the problem. Other policies that offer longer-term economic gains may do nothing to resolve current deficits but may reduce the need for subsequent interventions.

In our analysis, we focus our attention on the policy impact on longer-term economic success as the relevant outcome measure. In particular, we consider the adult earnings levels of the poor children who are the targets of the policy interventions considered. This is not to say that other outcome measures are not important; clearly, they are. Yet again, we make this decision because we believe that incorporating all forms of outcomes would quickly overwhelm the analysis.

Consider what would be needed to fully evaluate the totality of the impact of an early childhood education program, for example. Suppose that the program is successful in a number of dimensions. It improves educational performance of the students, enhances their "soft skills" through their ability to communicate and deal effectively with others. The students are more likely to go on to college and get better jobs. All of this shows up in the higher wages that the participants subsequently receive. Incorporating all of this into a measure of success has some pitfalls, but is something we hope to accomplish in this volume.

Now consider the other effects. Suppose the intervention enhances the child's nutritional status so that he or she is less likely to be overweight, which reduces the likelihood of juvenile diabetes and its health implications both today and, potentially, for the rest of the child's life. How do we measure that? The child may learn better coping mechanisms for the violence going on outside his or her home, or perhaps the child's success prompts the family to move to a safer neighborhood. The improved safety results in a lower

likelihood of exposure to violence, which improves the child's emotional well-being. How do we measure that? We could continue by providing additional examples, but we believe the point is clear that these other relevant and important outcomes are a lot more difficult to incorporate into our analysis. It is very difficult, if not impossible, to find ways to measure these outcomes in a way that would enable us to compare across interventions.

Instead, we simply choose not to do so and focus on the long-term economic success that these programs may generate. This is not to diminish the value of studying these other outcomes. In fact, in each of the subsequent chapters focusing on specific interventions, authors have included discussions of these other outcomes, where appropriate, in their reviews of the evidence. We simply focus on earnings as an outcome when we move on to synthesize the results later in the volume.

It is important to recognize the implications of our decisions here for interpreting the results of our analysis. What we will be left with in the end is one or more interventions that would appear to provide a greater ability to improve children's economic standing when they become adults. We believe that this is a useful piece of information that should definitely be used in policy discussions.

On the other hand, it would be unwise to base policy judgments purely on this piece of information. Other factors may come into play that are useful as well. Consider two alternative programs. The first one reduces long-term poverty among children, and the second one has no such effect. The second one, however, will reduce the likelihood of asthma attacks and obesity for the remainder of the child's life relative to the first one. Which program should be supported? We believe that we are not even close to being in a position to answer a question like that, so we choose to avoid the question. The question that we are more confident we can answer is the narrow one regarding the impact on subsequent wages, that is, the question on which we focus.

1.2.3 Challenge 3: Whose Benefit Matters?

This challenge is related to the last one, but we believe that a full discussion of the issue would usefully inform the reader regarding the contribution of our analysis, despite some overlap. The issue at hand here is whether our measure of success is one that focuses exclusively on the target population or whether the real beneficiary is society more broadly. As we just discussed, policy interventions can benefit the target population in a number of ways (health, safety, current income, subsequent earnings, etc.). But each of these benefits to the target population bring along external benefits to society more broadly.

Consider, for example, an intervention that reduces the likelihood that a child will commit crime. That program may benefit the child because he or she is more likely to lead a productive life and less likely to endure the difficulties associated with incarceration. But the crimes that would be

avoided as a result of the program would result in benefits to society more broadly. Others would be less likely to be victimized, and expenditures for the police and corrections department could be reduced. These are benefits that are received by society, but not by the individual.

Standard economic terminology uses the terms "private benefits" and "external benefits" to distinguish these concepts. Private benefits are those that are received by the individual participating in the program him- or herself. External benefits are those that are received by others beyond the program participant.

Benefit-cost analyses frequently rely on the external benefits, particularly when the government is attempting to determine whether a program is "working." Job training programs are designed to place workers into jobs that they would not have obtained otherwise. The private benefit is the higher earnings that the worker would receive relative to what they would have received had they not entered the program. The external benefit is the additional tax revenue that the incremental earnings would generate. From a societal perspective, the job training program is "worth it" if it can pay for itself—the tax revenue generated is greater than the cost of the spot in the training program. Any program that can substantially reduce the likelihood of crime is likely to be "worth it" because crime is so costly to society. This general approach is one that is exemplified by Karoly (1998). It makes sense from a government accounting standpoint where spending money on programs that generate more revenue or cost savings than the intervention costs is clearly desirable.

Our perspective is somewhat different. We want to know the most effective ways to help poor children pull themselves out of poverty by adulthood. The government is looking for a return on its investment based on its budget or on social welfare more broadly. But consider an alternative perspective where a certain amount of money is set aside to help poor children, and we want to get the most return on that investment in the form of subsequent poverty reduction. We would focus on the private return to the individual.

That does not mean that the external benefits to society more broadly are not important. They are, and they may play a role in the process of policy determination. Where appropriate, individual chapters will discuss the external benefits that are generated by specific interventions, but that will not be part of the broader comparison across interventions.

One justification for the approach that we are taking here is that it clarifies the issue of who is being helped by the policy. For instance, beneficiaries of interventions that reduce crime may be individuals who are not poor. Distributional issues like these are relevant but are typically not included in program evaluations. Our approach also helps narrow the scope of the exercise so that we are left with an "answerable" question. What types of policies targeted at poor children result in the greatest reduction in subsequent poverty? That is the question we will answer.

1.3 Empirical Evaluation Issues

The types of programs that we will explore in this book have been the subject of a tremendous amount of research. Reviewing the full body of evidence on any one intervention would be a substantial undertaking, let alone doing so for the array of interventions that we examine. In addition, any synthesis across interventions requires that common ground be established regarding their measured effectiveness. In this section of the chapter, we elaborate on the specific issues we face and the decisions we have made to resolve them.

1.3.1 Challenge 4: What Evidence "Counts"?

Previous research examining the types of interventions on which we focus in this analysis has adopted numerous methodologies in determining program impacts. The first type of study simply compares outcomes for program participants before and after the intervention. We call this approach "non-experimental." This technique is seriously flawed because it does nothing to hold constant the fact that participants who volunteer for the program are not necessarily representative of children from lower-income households. In particular, they are likely to be from households that are more motivated to overcome the obstacles they face. These children are likely to be the ones who would have done well anyway. If we see that outcomes improved for these children after they participated in the program, this does not necessarily tell us that the program had any impact.

What we really want to know is how a program would do with a typical child from a low-income household. It is the ability to properly answer this question that is the key to an effective evaluation. An effective methodology needs to be able to take a child who would have had one set of outcomes if he or she did not participate in the program and distinguish whether he or she experienced a different set of outcomes because of the intervention itself. This is what is needed to show that the program had a *causal* impact on the child. This is where a lot of previous research stumbles.

Another methodology that suffers from this problem is the use of comparison groups. This approach attempts to determine what would have happened to program participants had they not participated in the program by selecting a different group of children who do not participate in the program and comparing their outcomes. This is the right idea in that we need to know a counterfactual; this approach attempts to establish a counterfactual. The problem is that it is very difficult, if not impossible, to find perfect matches in a comparison group for those in the program. Outcomes for the two groups would need to have been similar in the absence of the intervention for this approach to work. But if program participation is voluntary, then we already know of one important difference between the two groups—one was willing to participate. If they are different in this important dimension,

they are likely to be different in other important dimensions as well. This makes it very hard to attribute causality to differences in outcomes between program participants and members of a comparison group.

Multiple regression analysis is another approach that researchers sometimes use to compare children whose outcomes would have been otherwise similar. In a multiple regression, researchers specify a dependent variable, which reflects children's outcomes (like test scores). They also specify a relevant independent variable (like program participation), and they want to know whether that variable has an impact on the dependent variable. The key to a multiple regression is that the researcher can also specify any number of other independent variables (demographic factors, geography, etc.) that are held constant in estimating the impact of the relevant independent variable (program participation) on the dependent variable (test scores). In theory, this would be quite an effective technique for identifying a causal impact of an intervention on children's outcomes. If we can hold constant the things that matter other than program participation, then our estimated impact on outcomes would be causal.

The problem with this approach is that it is difficult to observe all of the things that matter. If we leave some of them out, then our estimate of the program effect may, in reality, capture some of the impact of those omitted factors (assuming those omitted factors are related to program participation). Things like differences in a family's motivation level to help their kids overcome obstacles and differences in internal family functioning are very hard to observe and easily could be related to the child's success. This weakness suggests that we are unlikely to obtain causal estimates using a multiple regression approach.

The gold standard in obtaining causal estimates is a controlled experiment. In this approach, children eligible for a particular program are randomized into control and treatment groups. Members of the treatment group are subject to the intervention, and members of the control group are not. The true strength of a controlled experiment is that randomization (if properly conducted) guarantees that the control group and the treatment group are statistically identical. There may be individual differences between members within the two groups, but, on average, the only differences between the two groups are attributable to random variation. As such, we have a metric for determining how members of the treatment group would have fared had they not experienced the intervention. The control group sets that standard. This approach is the strongest for identifying causal effects.

Practical limitations, however, make controlled experiments relatively less common. The difficulty in recruiting candidates for a program that they may not get to participate in, ethical questions of withholding services to the control group, and the administrative difficulty and expense of setting up the experiment and tracking both control and treatment group members are important roadblocks in their widespread use. Nevertheless, they are

periodically conducted to evaluate interventions like those we study here, and we certainly rely on them wherever possible.

Sometimes circumstances in the world line up in such a way that something resembling a controlled experiment just happens to occur. For instance, subsidized child care programs for young children from low-income households may have too many applicants for the number of slots available. One way to allocate those slots is to randomly draw names of children to enter and others to sit on a waitlist. In this example, the children on the waitlist can act like a control group for the treatment group of students who happened to have been chosen to enroll. We call events like this "quasi-experiments."

Other forms of quasi-experiments occur when different geographic entities establish policies that differ from other places. If those policy differences can be plausibly attributed to factors unrelated to underlying differences in anticipated outcomes (like political factors—a newly elected governor implements components of his or her agenda), then we can similarly define treatment and control groups. An important limitation of this sort of analysis would occur if different geographic units adopted different policies in response to differences in outcomes (like passing school reforms because test scores were falling). Quasi-experimental approaches that rely on this sort of policy variation need to be acutely aware of this problem. Nevertheless, if implemented properly, quasi-experiments may provide an effective approach for identifying causal effects of program interventions.

In summary, despite the vast literatures that exist to examine different interventions directed at lower-income children, we discard large amounts of it because they rely on empirical methods that cannot plausibly provide causal estimates of their impact. Throughout the remainder of this volume, we will place heavy emphasis on evidence obtained from controlled experiments and well-designed quasi-experimental studies.

1.3.2 Challenge 5: Does the Scale of the Program Matter?

One of the disadvantages of relying so heavily on experimental evidence is that the logistical and financial constraints associated with large-scale experiments are formidable. This limits their frequency, and when they are conducted, the number of participants is sometimes reasonably small. This creates two potential difficulties. First, smaller sample sizes lead to estimated impacts that are less precise. If exactly the same intervention were conducted again, the results from the second attempt may be considerably different than that obtained the first time. This would reduce the confidence one would place in the estimate.

Our "solution" to this problem is simply to recognize its existence and interpret the results accordingly. Consider two interventions evaluated with experimental evidence where both interventions are estimated to have the same impact on the adult earnings of the children participating. If one intervention included more participants, we would place greater weight on

the results from that experiment. Throughout our review of experimental evidence, we will be clear to state the number of participants in the evaluation to help accomplish this task.

The second potential problem associated with experimental evidence using smaller sample sizes is the ability to "scale-up" an intervention. Programs that are effective when they are introduced to a group of 100 children may not work as well with 10,000 child participants if there are diseconomies of scale. This is the sort of problem that might be experienced if, say, a state or the federal government tried to replicate a successful, small-scale program that was privately funded. One successful charter school may be hard to turn into 100 successful identical charter schools if, for instance, the availability of teachers able to succeed in the new environment were limited. These are the sorts of problems that economists label general equilibrium effects. The small-scale intervention may be too small to disturb the broader marketplace, but ramping up the intervention to a larger scale may do so.

Again, this is a problem that we "solve" simply by recognizing its existence and interpreting the results accordingly. If we see that certain types of interventions are found to have been successful when evaluated, but those evaluations were based on small-scale demonstrations, we need to take note of that fact. If, for instance, private foundations want to spend more money to replicate those interventions on a relatively small scale with a few more sites or in a different location, then it would be appropriate to base that decision on the existing evidence. If similar interventions are considered by, say, the federal government, then this limitation of the existing evidence needs to be taken very seriously. If there are other interventions that appear to be successful and based on evidence from larger experimental studies, then we should prefer those.

1.3.3 Challenge 6: What Outcomes Are Considered?

Interventions designed to improve the lives of children and youths from low-income households have the potential to affect a wide variety of outcomes. School-based interventions, for instance, may alter students' educational performance, but there are a number of ways to assess educational performance, including test scores, grades, attendance, high school graduation, and college attendance. But these interventions may alter less-concrete outcomes as well, including self-esteem, sense of safety, and ability to communicate. Teen drug prevention and pregnancy prevention programs focus on a different array of outcomes. Even within the same types of interventions, different assessments may focus on different outcomes.

The problem that we face is to determine what outcomes we consider. The first thing to keep in mind is that our goal is not the measurement of each of these individual outcomes, but to translate everything into their impact on subsequent earnings levels. We deal exclusively with the issue of how we accomplish this task in chapter 13. The question here is which

outcomes from previous evaluations are we going to use as inputs into that translation?

Again, we will take a pragmatic approach and use whatever outcomes are available to us from the individual studies previously conducted that we have the ability to translate into earnings. Although school-based interventions may affect a large number of outcomes, assessments of their success tend to focus on a relatively narrow range of them. We will use whichever ones are available to us to convert to earnings affects. For the most part, the outcomes typically considered are things that we can convert to subsequent earnings. We describe in detail the process of taking an outcome like test scores and converting it to earnings effects in chapter 13. Outcomes that are more difficult to translate (self-esteem, ability to communicate, etc.) are typically not measured in these assessments anyway, so we have no inputs available to translate even if we had a method to do so. The contribution of "soft-skills" to subsequent economic well-being is something that we unavoidably will miss in our formal analysis. We will discuss the issue further in chapter 13, however, and offer some input as to its contribution.

References

Blank, Rebecca M. 2008. How to improve poverty measurement in the United States. *Journal of Policy Analysis and Management* 23 (2): 233–54.

Carneiro, Pedro, and James J. Heckman. 2003. Human capital policy. In *Inequality in America: What role for human capital policies?,* ed. James J. Heckman and Alan B. Krueger, 77–240. Cambridge, MA: MIT Press.

Cunha, Flavio, and James J. Heckman. 2007. The technology of skill formation. *American Economic Review* 97 (2): 31–47.

Cunha, Flavio, James J. Heckman, Lance Lochner, and Dimitriy V. Masterov. 2006. Interpreting the evidence on life cycle skill formation. In *Handbook of the economics of education,* ed. Eric Hanushek and Finis Welch, 697–812. Amsterdam: North Holland.

Karoly, Lynn A. 1998. *Investing in our children: What we know and don't know about the costs and benefits of early childhood interventions.* Santa Monica, CA: Rand Corp.