



Evaluating social policies: how and why?

Theory and Evidence from Mexico and Colombia

Orazio P. Attanasio

IFS- Centre for the Evaluation of Development
Policies- EDePo & UCL

Bologna 8/6/2007

July 28, 2004

World Bank Challenged: Are the Poor Really Helped?

CELIA DUGGER

The New York Times
ON THE WEB

“Wealthy nations and international organizations, including the World Bank, spend more than \$55 billion annually to better the lot of the world's 2.7 billion poor people.

Yet they have scant evidence that the myriad projects they finance have made any real difference, many economists say.

That important fact has left some critics of the World Bank, the largest financier of antipoverty programs in developing countries, dissatisfied, and they have begun throwing down an essential challenge. It is not enough, they say, just to measure how many miles of roads are built, schools constructed or microcredit loans provided. You must also measure whether those investments actually help poor people live longer, more prosperous lives.”

The Lancet – Editorial , August 28th 2004

“A crucial question, which extends beyond the World Bank, is whether aid of any kind is really better than debt forgiveness. [...]

There are recent signs that the Bank is taking long-needed steps to answer th[is] question, at least in part. This summer it is initiating a series of randomised trials to determine whether its aid projects are doing any good.

Whereas "success" at the Bank has sometimes been calculated by the number of loans made, now more rigorous methods should replace that simplistic measuring stick. Impact evaluations being undertaken by the Bank and its collaborators, [...], are a novelty for the Bank, where, astonishingly, only 2% of the projects it has funded for the last few years have been critically appraised. This is an appalling statistic: such evaluations are public goods, and public accountability surely demands them.

Without evidence, how can one know whether to modify, delete, or expand an existing programme?”

Importance of evaluations

- When designing and implementing a welfare program that uses a substantial fraction of scarce resources, one would like to know its effects.
- Different kinds of policies require different evaluation tools.
- Evaluations are, to a certain extent, a public good:
 - With a good evaluation one can try to use one experience in different contexts (scaling up)
 - Data are created that can be used for several purposes
 - ... and also lead to transparency, accountability, capacity creation

Outline

- What do we mean by 'evaluating' development policies
- The difficulty of obtaining good and credible evaluations
- Treatment- control comparisons and structural models
 - An example: Mexico's *PROGRESA/Oportunidades*
- The importance of measurement
 - An example: The evaluation of Colombia's *Familias en Accion*
- The political economy of evaluations

What do we mean by 'evaluating' development policies

- Given a specific intervention or welfare programme we would like to know what its effects are on:
 - (potential) beneficiaries
 - Non-beneficiaries
 - The functioning of a given market or institution
- One evaluates a policy with the idea of:
 - possibly scrapping the intervention,
 - improving its design,
 - changing its parameters,
 - expanding it to different contexts.

What do we mean by 'evaluating' development policies

- We will be not talking about the actual working of a program (operation).
- Neither will be talking about targeting.
- This is not to say that these aspects are unimportant.
- They are complementary to a good impact evaluation

The difficulty of obtaining good and credible evaluations

- The effect of a welfare program on a specific outcome is defined as the outcome variable when the program is operating minus the outcome we would observe in the absence of the program
- The problem, of course, is that we do not observe the latter if the program is implemented and the former if it is not implemented.
- A possible and (natural) solution is to compare individuals, households, or communities that receive the program to individuals, households or communities that do not.

The difficulty of obtaining good and credible evaluations

- The problem with such a strategy is that participation into the program might not be independent of the outcome of interest
- This is either because the individuals who (choose to) participate into a program are different or because of targeting of the program by the government
- This is referred to as the difficulty of observing counterfactuals
- Additional difficulties for the evaluation:
 - General equilibrium effects

The difficulty of obtaining good and credible evaluations

- In the evaluation literature there is now a strong emphasis on randomization:
 - Random allocation of a program to individuals or to localities
- This is certainly useful, as it creates controlled variation that can be used to obtain credible results
- PROGRESA in Mexico is a prime example of such a strategy

The difficulty of obtaining good and credible evaluations

- Is randomization the panacea to the evaluation problems?
- Probably not:
 - We can evaluate the effect of a program as a whole, but not of its components
 - We cannot extrapolate
 - It might be politically very difficult.
- PROGRESA in Mexico is again a good example

Example 1: PROGRESA in Mexico

- Conditional cash transfers to improve health, nutrition and education
- Model evaluation based on assigning the program randomly to a set of evaluation communities
- Widely perceived to be successful
- It has been exported widely throughout the world and in particular in LA:
 - Nicaragua, Honduras, Brazil, Argentina, Colombia

PROGRESA: the program

- Targeted first at the locality level then at the household level (proxy means testing)
- Nutrition component: cash transfer given to households with children 0-6 on the condition that the households participate into the health component (vaccination, growth and development check ups, courses for mothers).
- Education component: cash transfer given to children attending grades 3 to 9 conditionally on school enrolment and certified attendance; grant increasing with grade
- Grants paid to mothers

PROGRESA: the evaluation

- The program targeted a very large number of localities, so that it went through an expansion process that lasted more than 2 years.
- The administration of the program identified a sample of 506 localities that were drawn into the evaluation sample
- A large data collection exercise was started before the start of the program in 1997.
- 186 randomly chosen localities of the 506 in the sample were 'put at the end of the queue: the program there started at the end of 2000
- From 1998 to 2000, six waves of high quality comprehensive data were collected in these communities
- ... an additional wave was collected in 2003

PROGRESA: the results

- PROGRESA is widely considered a success story.
- Positive results on school enrolment, especially for older children, positive results on nutritional status.
- PROGRESA was exported in many places and, in Mexico, was the first welfare program of its kind to survive a change of administration.
- Moreover, in Mexico, it has recently been expanded to urban areas with a large loan from the IADB

Is the process through which PROGRESA was evaluated perfect?

- Political problems will make such a large scale randomization difficult to replicate
- The evaluation was too concentrated on the total effect of the program: an interesting question that remains unanswered is what would be the effect of a slightly different program
- This is particularly relevant for 'exporting' the program
- The mechanisms through which the program operates are still largely unexplored:
 - What is the role of the conditionalities
 - What is the role of information and education
 - What is the role of the mothers receiving the payments
- Anticipation effects, contaminations etc.

PROGRESA: estimating a structural model of school choice

- One can use the variation induced by the program to estimate a structural model of school choices.
- This model can then be simulated with different versions of the program to establish its effects
- The randomization allows us to estimate particularly rich versions of the model
- This is done in Attanasio, Meghir and Santiago (2004).

PROGRESA: estimating a structural model of school choice

- Children can work for pay or go to school
- The choice is determined by current utility, costs and opportunities (cost of going to school, background, the grant, wages)
- ... and by future opportunities (future choices and eventually return to education in the labour market)

PROGRESA: estimating a structural model of school choice

- The model is then estimated using maximum likelihood
- The fit is remarkably good.

Difference-in-differences estimates of program impact
on educational attendance

age group	eligible 97	ineligible
6-17	0.033 (0.009)	0.016 (0.019)
6-9	0.012 (0.006)	-0.010 (0.011)
10-13	0.024 (0.011)	-0.003 (0.023)
14-17	0.075 (0.025)	0.051 (0.037)

Standard errors in parentheses are clustered at the locality level.

Summary of diff in diff results

- On average the program has a 2.5% increase in enrolment rates
- These effects are much higher for older children and for children with higher education
- (notice that enrolment rates for younger children are already relatively high)

The model

Utility of attending school

$$u_{it}^s = \mu_i + a' z_{it} + b \cdot ed_{it} + I(ep_{it} = 1) \beta^s x_{it}^p + I(es_{it} = 1) \beta^s x_{it}^s + \varepsilon_{it}$$

Utility of working

$$u_{it}^w = (\delta + \theta_i) w_{it}$$

Value of attending school

$$V_{it}^s(ed_{it}, t) = u_{it}^s + \beta [p_{it}^{ed} E \max \{ V_{it+1}^s(ed_{it+1} + 1, t+1), V_{it+1}^w(ed_{it+1} + 1, t+1) \} + (1 - p_{it}^{ed}) E \max \{ V_{it+1}^s(ed_{it+1}, t+1), V_{it+1}^w(ed_{it+1}, t+1) \}]$$

Value of working

$$V_{it}^w(ed_{it}, t) = u_{it}^w + \beta E \max \{ V_{it+1}^s(ed_{it+1}, t+1), V_{it+1}^w(ed_{it+1}, t+1) \}$$

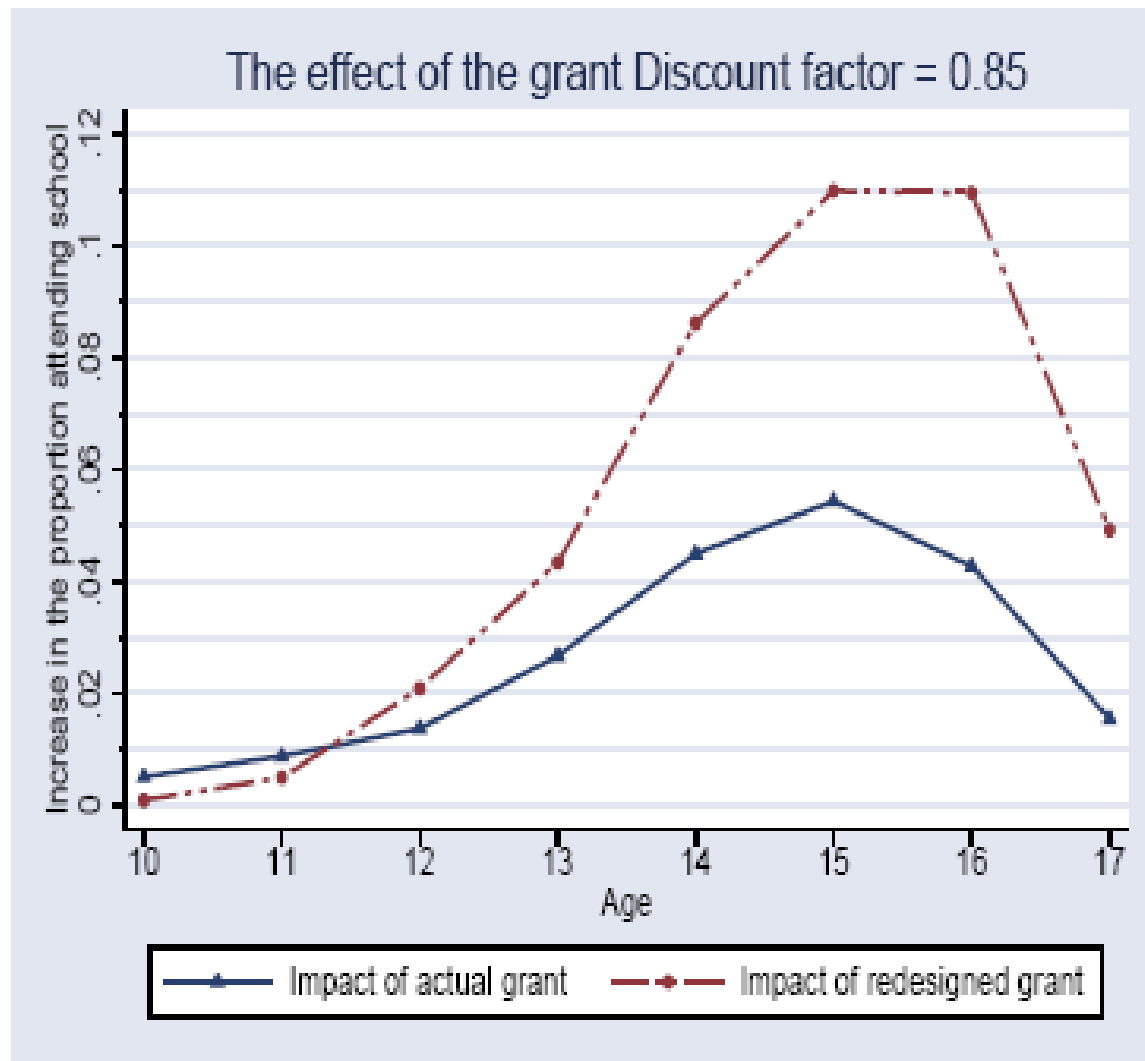


Figure 1: The impact of redesigning the grant

Source: Attanasio, Meghir and Santiago (2005)

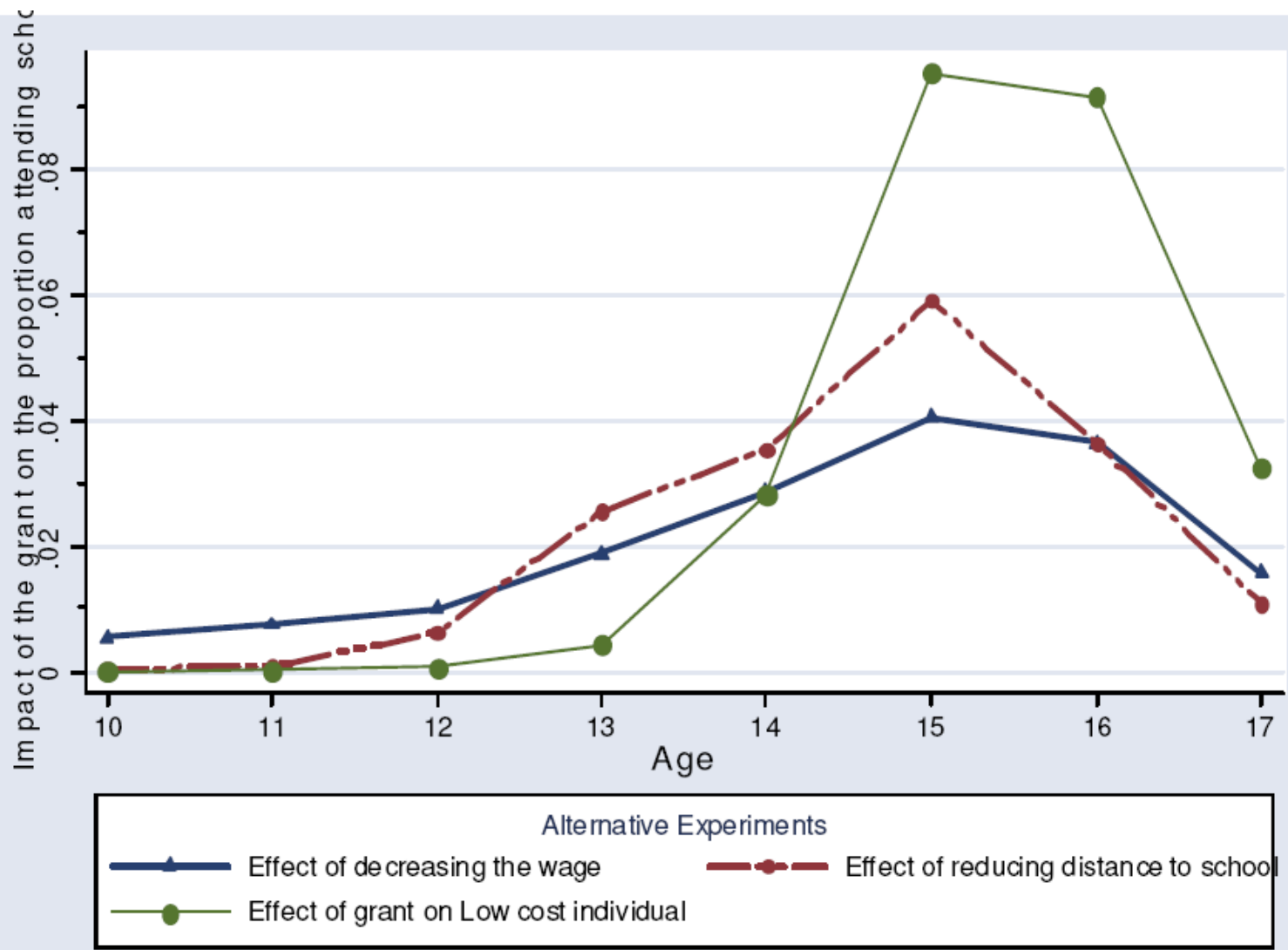


Figure 2:

Measurement

- To estimate a flexible structural model and/or understand the mechanisms behind the estimated effects we need information
- Individual level surveys are essential tools
- In addition to standard variables it is desirable to measure non-standard variables:
 - Expectations
 - Information
 - Biological markers
 - Measurements of cognitive development
 - Attitudes and relative power

Example 2: Familias en Acción in Colombia

- One of the countries where a version of PROGRESA was exported was Colombia.
- The Program Familias en Acción was financed with a loan over a three year period from the IADB and the WB to the Colombian government.
- One of the conditions on the loan was that the program had to be evaluated internationally
- The program (and its evaluation) started in 2002.

Familias en Acción in Colombia

- Familias en Acción is very similar to PROGRESA.
- Targeted to the poorest households of small towns (proxy means testing pre-existed in Colombia)
- Education, nutrition and health components are very similar
- The nutrition component of the program was and is widely perceived as an alternative to Hogares Comunitarios, a pre-existing community nurseries.

Familias en Acción: the evaluation

- The consortium that won the contract started the evaluation work in January 2002.
- The evaluation was contracted out by the Department of National Planning (DNP), while the program is run by Fondo de Inversion para la Paz (FIP) in the president office.
- FIP ruled out randomization of the program early on.
- The main methodology is to compare 'treated town' to 'untreated town' and take into account pre-program differences by having a pre-program measure

The Familias evaluation: how was the experience?

- Baseline data collected July-November 2002
 - 11,500 household interviews in 122 towns (57+65)
- First follow-up July-November 2003 (94% rate)
- Results presented in July 2004 in Bogotá
- Overall positive.
 - Good collaboration with DNP and, after a while, with FIP
 - Important role played by the local institutions (research outfit and data collection firm)
 - Human capital formation
 - Excellent data base created



Examples of new variables

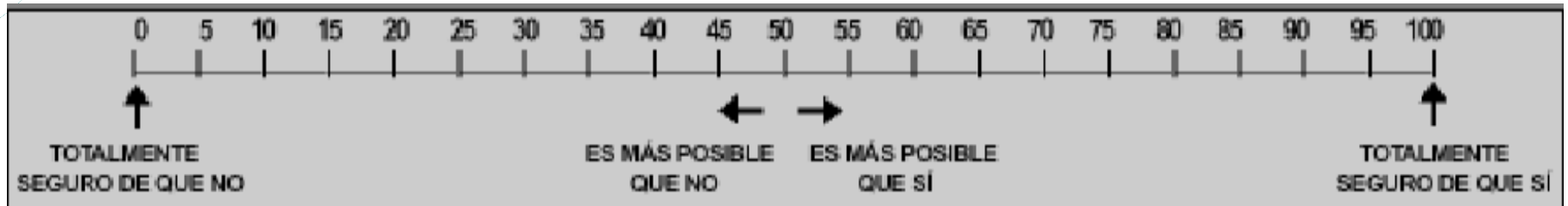
- Income uncertainty
- Social capital

Income uncertainty

- The aim of the question is to elicit the probability distribution of future income
- This is done with a series of questions:
 - Max and min future income
 - The range is divided into subintervals
 - We ask the probability that income will fall within each interval
 - It is crucial to have examples about probability
 - Use a visual device to solicit probability measures

Data: subjective expectations

Explanation of the concept of probabilities

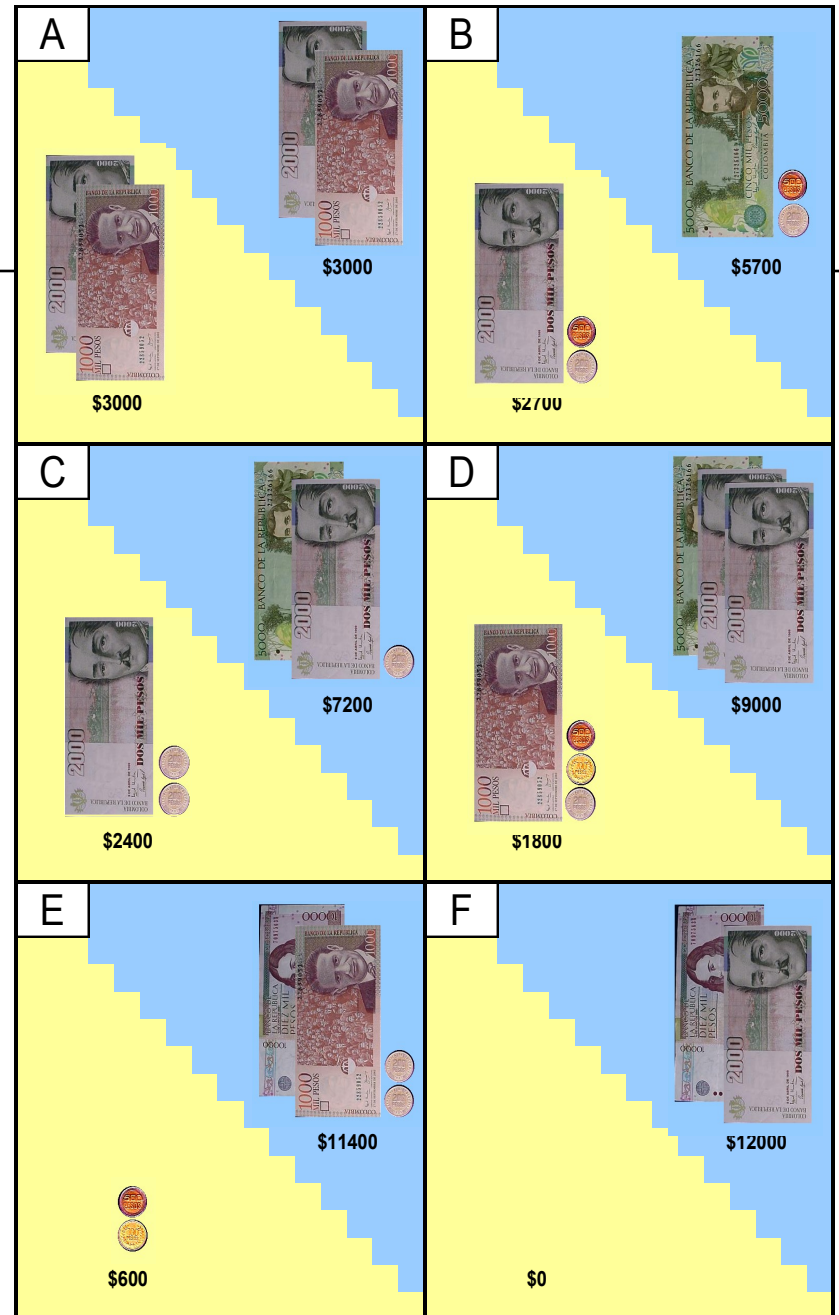


A game that generates data on social capital

- Standard surveys questions on participation provide, in our opinion, only very limited information on social capital.
- A game designed to provide a measure of social capital would be highly complementary to standard questions.
- Data from the game can be analyzed jointly with survey data.
- We looked at several games and eventually chose the risk pooling game (RPG)
- The RPG provides measures of:
 - peoples' willingness to mobilize collectively, to trust one another, and thereby reduce the impact of a shared problem
 - It also provides a measure of their attitudes towards risk - could be used as a control variable in various analyses

The game

- In the first round of the game we introduce people to a problem involving risk
- They have to choose between six gambles
- They can choose a low risk, low expected return gamble like A or B
- or a high risk, high expected return gamble like D, E or F
- The gambles are played out with real money
- Which gamble a person chooses might indicate how risk averse they are



The Risk Pooling Game

- In the second round of the game the players are informed that they can form 'sharing groups'
- Within sharing groups all winnings from the gambles are pooled and shared out equally
- The game is then played individually.
- Individuals then can dropout of the group if they have won. This is done privately.

The Risk Pooling Game

- So, members of groups can afford to take higher risks yielding higher expected returns.
- But you need to trust your friends!
- Whether and to what extent the players form groups in the game provides us with a measure of their willingness to act collectively to solve a shared problem and take advantage of a collective opportunity

The public good game

- Individuals are given a token.
- This can be invested in a private project or a public project.
- The return on the public project is larger.
- If you invest in a private project you get that + a share of what has been invested in the public project
- If you invest in the public project you get only the share of what has been invested in the public project.
- The social optimum is to invest in the public problem.
- The private optimum is to invest in the private project
- The measure of social capital is the share of individuals that invest in the public project.
- The game is played twice, the second time giving the possibility to talk.

The evaluation experience: what would I change?

- Do it earlier, before the program is started
- Try different versions of the program:
 - This is particularly important for these types of programs and in some areas
- Try to use randomization in the design stage
- Understand the mechanisms
- Measure a lot

The political economy of evaluations

- ❑ Politicians do not like evaluations
 - You do not win an election by evaluating
 - The relevant horizon might be too short
 - Randomization is deeply unpopular
 - International organizations and civil servants should and could play an important role in promoting quality evaluations.
 - But this is probably not enough: we need to create a demand for evaluations within the budgetary process.

The political economy of evaluations

- When a program is launched, it immediately creates its own constituency
 - Evaluation should be conducted at a very early stage in the development of a program
 - This has the advantage of trying different versions of the different program...
 - ...possibly randomly allocated across different areas
 - Pilots made by independent agencies
 - Again, international organizations should play an important role in this

The political economy of evaluations

- ❑ How to create the demand for evaluations
 - Evaluations should become important within the budgetary processes as they should be instruments to compete over scarce resources
 - International financial institutions
 - Ministry of finance or National Planning
 - Examples: UK, Australia

Conclusions

- Evaluation of policy interventions is crucial, especially in developing countries.
- We need to learn what programs work, what makes them work.
- Early evaluation is better than late evaluation
- International financial institutions, such as the IADB, have a large role to play
- It is crucial that the demand for evaluation is engrained within the budgetary process