Evaluating Social Programs

Paris, July, 2-4, 2012

J-PAL Europe Advanced Executive Education Course


© Aude Guerrucci

We are very grateful to the Maison des Sciences économiques for its support and its help to organize this course

J-PAL EUROPE
Paris School of Economics - 66bis avenue Jean Moulin, 75014 Paris - T: +33 (0) 171 19 40 70
jpaleurope@povertyactionlab.org - http://www.povertyactionlab.org/europe

Table of Contents

Welcome!

Each year, professors affiliated with J-PAL (Jameel Poverty Action Lab) train dozens of people in the use of Randomized Evaluations. We provide four types of courses:

- **Executive Education** courses are designed for people from a variety of backgrounds: managers and researchers from international development organizations, foundations, governments and non-governmental organizations from around the world. They provide fundamental knowledge for people interested in rigorous impact evaluation and are held in English, French, Spanish and Portuguese.
- **Advanced Executive Education** courses are designed for researchers already familiar with Economics and Econometrics, who want to get a deeper understanding of RCTs: their history, analytical methods, technical issues and practical difficulties.
- **Monitoring & Evaluation** courses are designed for development personnel involved in evaluation or M&E who understand the theory well and need a more hands-on training on good data collection and other M&E practices.
- **Custom Workshops** are imparted in partnership with government departments, development organizations and research institutions. These typically have a sector-specific unifying theme in the context of Randomized Evaluations and teach concepts giving examples from that particular field in order to make them relevant for the trainees.

During these three days, you will attend an ***Advanced Executive Education*** course. The participants in this course come from many different countries and have experience in a variety of different fields: health, education, labor markets, agriculture, microfinance, etc. This diversity will make our course a unique experience for everyone.

You will have the opportunity to meet other practitioners of Randomized Evaluations in different areas and contexts. We hope that you find people with common interests and build new partnerships and projects during these three days.

We hope that this course is a great learning experience for you and imparts useful skills to help you conduct your own Randomized Evaluations. We shall continue to be a resource for you even after the course: please don't hesitate to contact us with questions!

We wish you an interesting and productive course!

Hélène Giacobino
Director
J-PAL Europe

J-PAL: Researchers Against Poverty

Founded in 2003, the Abdul Latif Jameel Poverty Action Lab (J-PAL) is a network of researchers based at the Massachusetts Institute of Technology (MIT) in Cambridge, MA. In 2007, **J-PAL Europe** was officially launched at the Paris School of Economics, and **J-PAL South Asia** was founded at the Institute for Financial Management and Research in Chennai, India. In 2009, **J-PAL Latin America** was launched at the Pontifícia Universidad Católica in Santiago, Chile, and in 2010 **J-PAL Africa** was founded at the University of Cape Town.

J-PAL is a network of 53 researchers from all over the world. Since 2003 they have conducted more than 250 randomized evaluations of anti-poverty programs, of which 100 are already completed.

J-PAL works to improve the lives of the poor by ensuring that development policy is based on scientific evidence, collected through randomized evaluations.

© Aude Guerrucci

What does J-PAL do?

J-PAL has three central objectives:

1. **Rigorous Evaluation of Development Projects:** J-PAL researchers are at the forefront of randomized evaluations, developing methodologies that allow an element of randomization to be introduced into programs in a way that is compatible with the constraints on the ground. As such, J-PAL researchers work on projects ranging from the evaluation of job search support programs in France, microfinance programs in rural Morocco, anti- corruption measures in India, and school feeding in Niger.

2. Capacity Building: Every year, J-PAL runs executive education courses in several locations around the world. These courses have trained hundreds of practitioners in more than 40 countries. Many of these practitioners have gone on to make significant contributions to randomized evaluations, either alone or in conjunction with J-PAL.

3. Diffusion of Results: J-PAL has also been very successful at promoting evidence-based policy through papers, conferences, seminars, and capacity building. Policymakers are increasingly using the evidence generated by J-PAL evaluations to guide their decisions. For example, findings on the impacts of school-based deworming have influenced government policy in Kenya and elsewhere.

J-PAL Funding

In 2005, J-PAL received a substantial gift from Mohammed Abdul Latif Jameel, an MIT alumnus and generous
supporter of poverty alleviation initiatives around the world. The Poverty Action Lab was renamed in honor of his father, Abdul Latif Jameel.

Other donors and financial supporters include the Economic and Social Research Council, Swedish International Development Agency, DFID, The John D. and Catherine T. MacArthur Foundation, the Bill & Melinda Gates Foundation, the Hewlett Foundation, the National Institute of Health, the World Bank, Agence Française de Développement (AFD), the Haut Commissariat aux Solidarités Actives Contre la Pauvreté, the Institute Veolia Environnement, and the National Science Foundation.

Course Schedule

Monday, July 2

08:30 – 09:00	*Participant Registration/ Welcome Coffee & Tea*
09:00 – 09:30	Welcome word and presentation of participants
09:30 – 11:00	Lecture 1: The Why and How of Randomized Experiments: a Historical Perspective, by Luc Behaghel
11:00 – 11:30	*Coffee & Tea Break*
11:30 – 01:00	Lecture 2: Power Calculation for Randomized Experiments, by Marc Gurgand
01:00 – 02:00	*Lunch*
02:00 – 04:00	Group Work
04:00 – 04:30	*Coffee & Tea Break*
04:30 – 06:00	Lecture 3: Dealing with attrition / Lee bounds, by Elise Huillery
08:00 – 10:00	*Dinner*

Tuesday, July 3

9:30 – 11:00	Lecture 4: Conceiving Smart Designs, by Roland Rathelot
11:00 – 11:30	*Coffee & Tea Break*
11:30 – 01:00	Lecture 5: Evaluating from A to Z, by Karen Macours
01:00 – 02:00	*Lunch*
02:00 – 04:00	Group Work
04:00 – 04:30	*Coffee & Tea Break*
04:30 – 06:00	Paper presentation: Discrimination in Hiring and Anonymous CVs in France, by Luc Behaghel

Wednesday, July 4

9:30– 11:00	Group Work
11:00 – 11:30	*Coffee & Tea Break*
11:30 – 01:00	Group Work
01:00 – 02:00	*Lunch*
02:00 – 04:00	Group Presentations
04:00 – 04:30	*Coffee & Tea Break*
04:30 – 05:30	Group Presentations
05:30 – 06:00	*Closing Remarks*

Lectures, Group Presentations the last day, tea breaks and lunches will be held in the conference room on the 6th floor. Group work will be in the rooms 116 or 117 on the 1st floor.

Monday Dinner

On Monday night at 8:00pm, we are pleased to invite you for a dinner all together.

Please see below how to reach the restaurant:

Le Bouche à Oreilles
10 place Paul Verlaine
75013 Paris
Tel.: +33 (0)1 45 89 74 42

Metro: line 5 & 7, station Place d'Italie
Line 6, station Corvisart

Teachers Presentation

Luc BEHAGHEL is an Associate Professor at the Paris School of Economics (INRA) and a CREST Affiliate. His research interests go from labor market and education policies to rural development issues. He is currently conducting several evaluations on education and discrimination in France, and has started orienting his research toward agricultural technology adoption in developing countries.
luc.behaghel@ens.fr

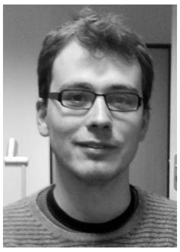
Marc GURGAND is an Associate Professor at the Paris School of Economics. His research focuses on labor market policies, schooling and inequality in both developing and developed countries. He is currently conducting randomized evaluations of counseling schemes focused on the unemployed and welfare recipients. He also has a research program studying inequality in China. Marc Gurgand is a J-PAL Europe affiliate and its Scientific Director.
gurgand@pse.ens.fr

Élise HUILLERY is an Assistant Professor at the Department of Economics of Sciences Po. Her research focuses on policies addressing the lack of human capital (health, education, social capital) in developing countries and in France, with a special interest in understanding the psychological barriers to individual progression. She also has a research program on the colonial history and its long term impact in Africa.
elise.huillery@sciences-po.fr

Karen MACOURS is an Associate Professor at the Paris School of Economics and researcher at INRA. Her current research focuses on conditional cash transfer programs, early childhood development, rural poverty, and agriculture.
karen.macours@parisschoolofeconomics.eu

Roland RATHELOT is a Researcher at the Centre de Recherche en Économie et Statistique (CREST). His areas of interest include labor economics, public economics and economics of immigration, with a particular focus on the spatial dimension. He is currently conducting randomized evaluations of counseling programs dedicated to the youth in France.
roland.rathelot@ensae.fr

Assistants Presentation



Adrien BOUGUEN joined J-PAL Europe in October 2009 and is currently contributing to two randomized evaluations in the school district of Creteil aiming at reducing absenteeism and providing equal opportunities for underprivileged high school students.
abouguen@povertyactionlab.org

Axelle CHARPENTIER joined J-PAL Europe in October 2010 where she is currently contributing to the randomized evaluation of two education programs providing boarding schools for secondary students from disadvantaged backgrounds and reinforcing the link between parents and middle schools in France.
acharpentier@povertyactionlab.org

Victor POULIQUEN joined J-PAL in 2008 and is currently working on three randomized evaluations focusing on education and health in Morocco, Ghana and Kenya. The first one looks at the impact of a conditional cash transfer program in education, the second at the effect of a scholarships program, and the third at the impact of different HIV/AIDS prevention programs. vpouliquen@povertyactionlab.org

Juliette SEBAN joined J-PAL in 2008. She is currently working on two randomized evaluations of programs that aim to increase youth employment and entrepreneurship in France as well as one on the impact of different HIV/AIDS prevention programs in Cameroon.
jseban@povertyactionlab.org

Bram THUYSBAERT is a postdoctoral associate at Yale University and a project director at Innovations for Poverty Action. He is working on series of randomized evaluations in the field of agriculture, livelihood assistance and access to finance in Africa and Latin America.
bthuysbaert@poverty-action.org

Participants List

Name	Organisation	Country of work
ALIMUKHAMEDOVA Nargiza	CERGE-EI, Charles University	CZECH REPUBLIC
ANGERMÜLLER Diane	European Commission	BELGIUM
BENHAMOUCHE Zoubir	Self-employed	FRANCE
CAZZUFFI Chiara	University of Sussex	UNITED KINGDOM
CENDOYA REVENGA Pablo	European Commission	MOROCCO
CHERRIER Cécile	Maastricht Graduate School of Governance	FRANCE
D'ANDON Cécile	J-PAL Europe Morocco – IPA	MOROCCO
DMITRIJEVA Jekaterina	ERUDITE, University of Paris Est	FRANCE
DOAREST Aufa	National Team for the Acceleration of Poverty Reduction	INDONESIA
FREMIGACCI Florent	University of Paris Ouest Nanterre	FRANCE
GASTEEN Maxime	Department for International Development	UNITED KINGDOM
GUILLOT Michel	University of Pennsylvania	UNITED STATES
LAURENT Arthur	MICROSOL	PERU
MALIKI AMADOU Mahamane	World Food Programme	ITALY
MAROUANI Mohamed Ali	University of Paris 1 Panthéon Sorbonne	FRANCE

PASSERARD Françoise	HEC Paris, Marketing Department	FRANCE
PEDEN Alexander	University of Manitoba	CANADA
PHILIPPE Arnaud	J-PAL Europe	FRANCE
PIERNE Guillaume	University of Evry	FRANCE
PRIMA Rizal Adi	Poverty Alleviation Unit - Vice President Office	INDONESIA
ROULAND Bénédicte	University of Maine	FRANCE
SAINT-DENIS Antoine	European Commission	FRANCE
SASU Ionut	European Commission	BELGIUM
SORO G. Amadou	LEMNA	FRANCE
SULTAN Joyce	J-PAL Europe	FRANCE
TRITAH Ahmed	Le Mans University	FRANCE
VAN HAM Christelle	Self-employed	FRANCE
WIESER Simon	Zurich University of Applied Science Institute of Health Economics	SWITZERLAND

Participants Email List:
At the end of the course, a contact list with your email will be sent to all of the participants. If you do not want to give your email address, please let us know and we will remove you from this list.

Help Desk for J-PAL Executive Education Course Alumni:
Last year, J-PAL launched the "RCT Help Online" (RHO). This moderated listserve is aimed at promoting an open discussion amongst participants. Participants will be automatically invited to the list by email after completion of the course.

WiFi Hotspot:
To connect to WiFi, please enter the following codes.
Login: JPAL
Password: jpal2012

Group Presentation

Participants will form 4 or 5 person groups which will work through the design process for a randomized evaluation of a development project. Groups will be aided in this project by both the faculty and teaching assistants with the work culminating in presentations at the end of the week.

The goal of the group presentations is to consolidate and apply the knowledge of the lectures and thereby ensure that participants will leave with the knowledge, experience, and confidence necessary to conduct their own randomized evaluations. We encourage groups to work on projects that are relevant to participants' organisations.

All groups will present on **Wednesday**. The 15-minute presentation is followed by a 15-minute discussion led by J-PAL affiliates and staff. We provide groups with template slides for their presentation (see USB key). While the groups do not need to follow this exactly, the presentation should have no more than 9 slides (including title slide, excluding appendix) and should include the following topics:

- Brief project background
- Theory of change
- Evaluation question
- Outcomes
- Evaluation design
- Data and sample size
- Potential validity threats and how to manage them
- Dissemination strategy of results

Please time yourself and do not exceed the allotted time. We have only a limited amount of time for these presentations, so we will follow a strict timeline to be fair to all groups.

Work groups

Group 1

Assistant: Adrien Bouguen

Diane Angermüller

Pablo Cendoya Revenga

Florent Fremigacci

Mohamed Ali Marouani

Bénédicte Rouland

Joyce Sultan

Group 2

Assistant: Juliette Seban

Jekaterina Dmitrijeva

Guillaume Pierne

Ionut Sasu

Antoine Saint-Denis

Ahmed Tritah

Cécile D'Andon

Group 3

Assistant: Axelle Charpentier

Nargiza Alimukhamedova

Zoubir Benhamouche

Arthur Laurent

Rizal Adi Prima

Amadou G. Soro

Christelle Van Ham

Group 4

Assistant: Victor Pouliquen

Aufa Doarest

Maxime Gasteen

Michel Guillot

Alexander Peden

Simon Wieser

Group 5

Assistant: Bram Thuysbaert

Mahamane Maliki Amadou

Chiara Cazzuffi

Cécile Cherrier

Françoise Passerard

Arnaud Philippe

Bibliography

To begin with…

Duflo, Esther , Rachel Glennerster, Michael Kremer, "Using Randomization in Development Economics Research: A Toolkit" Handbook of Development Economics, Volume 4 (2008). *(a copy is in the USB key)*

Banerjee, Abhijit V. and Esther Duflo, "The Experimental Approach to Development Economics" Annual Review of Economics, (2009).

General Bibliography

Baird, Sarah, Joan Hamory, and Edward Miguel, 2008. "Tracking, Attrition, and Data Quality in the Kenyan Life Panel Survey Round 1" (KLPS-1), mimeo, George Washington University and UC Berkeley.

Banerjee, Abhijit, "Inside the Machine: Toward a New Development Economics," *Boston Review*, (March/April 2007).

Deaton, Angus, "Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development," NBER Working Paper, No. w14690, (January 2009).

Duflo, Esther "Field Experiments in Development Economics," Advances in Economic Theory and Econometrics, Eds. Richard Blundell, Whitney Newey, Torsten.Persson, *Cambridge University Press*, Volume 2(42), see also BREAD Policy Paper No. 002, 2005.

Imbens, Guido "Better LATE than Nothing," NBER Working Paper No. w14896, (April 2009).

Imbens, Guido and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, Vol. 47, No. 1, (March 2009), pp. 5-86.

Kling, Jeffrey R., Jeffrey B. Liebman and Lawrence F. Katz, Experimental Analysis of Neighborhood Effects," *Econometrica*, 75 (January 2007), 83-119.

Examples of RCTs

Banerjee, Abhijit, Esther Duflo, Rachel Glennerster, and Cynthia Kinnan. 2009. "The Miracle of Microfinance? Evidence from a Randomized Evaluation." *MIT Department of Economics.*

Behrman, Jere, Susan Parker, and Petra Todd. 2009. "Medium Term Impacts of the Oportunidades Conditional Cash Transfer Program on Rural Youth in Mexico" with, in Stephan Klasen and Felicitas Nowak-Lehmann, Eds., *Poverty, Inequality and Policy in Latin America, Cambridge* , MA: MIT Press.

Duflo, Esther, Michael Kremer, and Jonathan Robinson, "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya," Poverty Action Lab Working Paper
http://www.povertyactionlab.org/papers/99_Understanding_Technology_Adoption.pdf

Karlan, Dean and Jonathan Zinman, "Observing unobservables: Identifying Information Asymmetries with a Consumer Credit Field Experiment," (forthcoming in *Econometrica)*
http://karlan.yale.edu/p/OU_dec08_v1.pdf

Edward Miguel and Michael Kremer, "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities," *Econometrica* 72(1): 159-217, 2004

We thank Aude Guerrucci for her photographies. www.audeguerrucci.com

Checklist For Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence

Coalition for Evidence-Based Policy

A NONPROFIT, NONPARTISAN ORGANIZATION

Updated February 2010

This publication was produced by the Coalition for Evidence-Based Policy, with funding support from the William T. Grant Foundation, Edna McConnell Clark Foundation, and Jerry Lee Foundation.

We welcome comments and suggestions on this document (jbaron@coalition4evidence.org).

Checklist For Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence

This is a checklist of key items to look for in reading the results of a randomized controlled trial of a social program, project, or strategy ("intervention"), to assess whether it produced valid evidence on the intervention's effectiveness. This checklist closely tracks guidance from both the U.S. Office of Management and Budget (OMB) and the U.S. Education Department's Institute of Education Sciences (IES)[1]; however, the views expressed herein do not necessarily reflect the views of OMB or IES.

This checklist limits itself to key items, and does not try to address all contingencies that may affect the validity of a study's results. It is meant to aid – not substitute for – good judgment, which may be needed for example to gauge whether a deviation from one or more checklist items is serious enough to undermine the study's findings.

A brief appendix addresses *how many* well-conducted randomized controlled trials are needed to produce strong evidence that an intervention is effective.

Checklist for overall study design

Random assignment was conducted at the appropriate level – either groups (e.g., classrooms, housing projects), or individuals (e.g., students, housing tenants), or both.

Random assignment of individuals is usually the most efficient and least expensive approach. However, it may be necessary to randomly assign groups – instead of, or in addition to, individuals – in order to evaluate (i) interventions that may have sizeable "spillover" effects on nonparticipants, and (ii) interventions that are delivered to whole groups such as classrooms, housing projects, or communities. (See reference 2 for additional detail.[2])

The study had an adequate sample size – one large enough to detect meaningful effects of the intervention.

Whether the sample is sufficiently large depends on specific features of the intervention, the sample population, and the study design, as discussed elsewhere.[3] Here are two items that can help you judge whether the study you're reading had an adequate sample size:

- If the study found that the intervention produced *statistically-significant* effects (as discussed later in this checklist), then you can probably assume that the sample was large enough.

- If the study found that the intervention did *not* produce statistically-significant effects, the study report should include an analysis showing that the sample was large enough to detect meaningful effects of the intervention. (Such an analysis is known as a "power" analysis.[4])

Reference 5 contains illustrative examples of sample sizes from well-conducted randomized controlled trials conducted in various areas of social policy.[5]

Checklist to ensure that the intervention and control groups remained equivalent during the study

The study report shows that the intervention and control groups were highly similar in key characteristics prior to the intervention (e.g., demographics, behavior).

If the study asked sample members to consent to study participation, they provided such consent *before* learning whether they were assigned to the intervention versus control group.

If they provided consent afterward, their knowledge of which group they are in could have affected their decision on whether to consent, thus undermining the equivalence of the two groups.

Few or no control group members participated in the intervention, or otherwise benefited from it (i.e., there was minimal "cross-over" or "contamination" of controls).

The study collected outcome data in the same way, and at the same time, from intervention and control group members.

The study obtained outcome data for a high proportion of the sample members originally randomized (i.e., the study had low sample "attrition").

As a general guideline, the studies should obtain outcome data for at least 80 percent of the sample members originally randomized, including members assigned to the intervention group who did not participate in or complete the intervention. Furthermore, the follow-up rate should be approximately the same for the intervention and the control groups.

The study report should include an analysis showing that sample attrition (if any) did not undermine the equivalence of the intervention and control groups.

The study, in estimating the effects of the intervention, kept sample members in the original group to which they were randomly assigned. This even applies to:

- Intervention group members who failed to participate in or complete the intervention (retaining them in the intervention group is consistent with an "intention-to-treat" approach); and

- Control group members who may have participated in or benefited from the intervention (i.e., "cross-overs," or "contaminated" members of the control group).[6]

Checklist for the study's outcome measures

The study used "valid" outcome measures – i.e., outcome measures that are highly correlated with the true outcomes that the intervention seeks to affect. For example:

- Tests that the study used to measure outcomes (e.g., tests of academic achievement or psychological well-being) are ones whose ability to measure true outcomes is well-established.
- If sample members were asked to self-report outcomes (e.g., criminal behavior), their reports were corroborated with independent and/or objective measures if possible (e.g., police records).

- The outcome measures did not favor the intervention group over the control group, or vice-versa.
  For instance, a study of a computerized program to teach mathematics to young students should not measure outcomes using a computerized test, since the intervention group will likely have greater facility with the computer than the control group.[7]

The study measured outcomes that are of policy or practical importance – not just intermediate outcomes that may or may not predict important outcomes.

As illustrative examples: (i) the study of a pregnancy prevention program should measure outcomes such as actual pregnancies, and not just participants' attitudes toward sex; and (ii) the study of a remedial reading program should measure outcomes such as reading comprehension, and not just the ability to sound out words.

Where appropriate, the members of the study team who collected outcome data were "blinded" – i.e., kept unaware of who was in the intervention and control groups.

Blinding is important when the study measures outcomes using interviews, tests, or other instruments that are not fully structured, possibly allowing the person doing the measuring some room for subjective judgment. Blinding protects against the possibility that the measurer's bias (e.g., as a proponent of the intervention) might influence his or her outcome measurements. Blinding would be important, for example, in a study that measures the incidence of hitting on the playground through playground observations, or a study that measures the word identification skills of first graders through individually-administered tests.

Preferably, the study measured whether the intervention's effects lasted long enough to constitute meaningful improvement in participants' lives (e.g., a year, hopefully longer).

This is important because initial intervention effects often diminish over time – for example, as changes in intervention group behavior wane, or as the control group "catches up" on their own.

Checklist for the study's reporting of the intervention's effects

If the study claims that the intervention has an effect on outcomes, it reports (i) the size of the effect, and whether the size is of policy or practical importance; and (ii) tests showing the effect is statistically significant (i.e., unlikely to be due to chance).

These tests for statistical significance should take into account key features of the study design, including:

- Whether individuals (e.g., students) or groups (e.g., classrooms) were randomly assigned;
- Whether the sample was sorted into groups prior to randomization (i.e., "stratified," "blocked," or "paired"); and
- Whether the study intends its estimates of the intervention's effect to apply only to the sites (e.g., housing projects) in the study, or to be generalizable to a larger population.

The study reports the intervention's effects on all the outcomes that the study measured, not just those for which there is a positive effect.

This is so you can gauge whether any positive effects are the exception or the pattern. In addition, if the study found only a limited number of statistically-significant effects among many outcomes measured, it should report tests showing that such effects were unlikely to have occurred by chance.

Appendix: How many randomized controlled trials are needed to produce strong evidence of effectiveness?

To have strong confidence that an intervention would be effective if faithfully replicated, one generally would look for evidence including the following:

- **The intervention has been demonstrated effective, through well-conducted randomized controlled trials, in more than one site of implementation.**

Such a demonstration might consist of two or more trials conducted in different implementation sites, or alternatively one large multi-site trial.

- **The trial(s) evaluated the intervention in the real-world community settings and conditions where it would normally be implemented** (e.g., community drug abuse clinics, public schools, job training program sites).

This is as opposed to tightly-controlled conditions, such as specialized sites that researchers set up at a university for purposes of the study, or settings where the researchers themselves administer the intervention.

- **There is no strong countervailing evidence, such as well-conducted randomized controlled trials of the intervention showing an absence of effects.**

References

[1] U.S. Office of Management and Budget (OMB), What Constitutes Strong Evidence of Program Effectiveness, http://www.whitehouse.gov/omb/part/2004_program_eval.pdf, 2004; U.S. Department of Education's Institute of Education Sciences, Identifying and Implementing Educational Practices Supported By Rigorous Evidence, http://www.ed.gov/rschstat/research/pubs/rigorousevid/index.html, December 2003; What Works Clearinghouse of the U.S. Education Department's Institute of Education Sciences, Key Items To Get Right When Conducting A Randomized Controlled Trial in Education, prepared by the Coalition for Evidence-Based Policy, http://ies.ed.gov/ncee/wwc/pdf/guide_RCT.pdf.

[2] Random assignment of groups rather than, or in addition to, individuals may be necessary in situations such as the following:

> (a) The intervention may have sizeable "spillover" effects on individuals other than those who receive it.
>
> For example, if there is good reason to believe that a drug-abuse prevention program for youth in a public housing project may produce sizeable reductions in drug use not only among program participants, but also among their peers in the same housing project (through peer-influence), it is probably necessary to randomly assign whole housing projects to intervention and control groups to determine the program's effect. A study that only randomizes individual youth within a housing project to intervention versus control groups will underestimate the program's effect to the extent the program reduces drug use among both intervention and control-group students in the project.
>
> (b) The intervention is delivered to groups such as classrooms or schools (e.g., a classroom curriculum or schoolwide reform program), and the study seeks to distinguish the effect of the intervention from the effect of other group characteristics (e.g., quality of the classroom teacher).
>
> For example, in a study of a new classroom curriculum, classrooms in the sample will usually differ in two ways: (i) whether they use the new curriculum or not, and (ii) who is teaching the class. Therefore, if the study (for example) randomly assigns individual students to two classrooms that use the curriculum versus two classrooms that don't, the study will not be able to distinguish the effect of the curriculum from the effect of other classroom characteristics, such as the quality of the teacher. Such a study therefore probably needs to randomly assign whole classrooms and teachers (a sufficient sample of each) to intervention and control groups, to ensure that the two groups are equivalent not only in student characteristics but also in classroom and teacher characteristics.
>
> For similar reasons, a study of a schoolwide reform program will probably need to randomly assign whole schools to intervention and control groups, to ensure that the two groups are equivalent not only in student characteristics but also school characteristics (e.g., teacher quality, average class size).

[3] What Works Clearinghouse of the U.S. Education Department's Institute of Education Sciences, *Key Items To Get*
Right When Conducting A Randomized Controlled Trial in Education, op. cit., no. 1.

[4] Resources that may be helpful in reviewing or conducting power analyses include: the William T. Grant Foundation's free consulting service in the design of group-randomized trials, at http://sitemaker.umich.edu/group- based/consultation_service; Steve Raudenbush et. al., *Optimal Design Software for Group Randomized Trials*, at http://sitemaker.umich.edu/group-based/optimal_design_software; Peter Z. Schochet, *Statistical Power for Random Assignment Evaluations of Education Programs* (http://www.mathematica- mpr.com/publications/PDFs/statisticalpower.pdf),

prepared for the U.S. Education Department's Institute of Education Sciences, June 22, 2005; and Howard S. Bloom, "Randomizing Groups to Evaluate Place-Based Programs," in *Learning More from Social Experiments:*

Evolving Analytical Approaches, edited by Howard S. Bloom. New York: Russell Sage Foundation Publications, 2005, pp. 115-172.

---

[5] Here are illustrative examples of sample sizes from well-conducted randomized controlled trials in various areas of social policy: (i) 4,028 welfare applicants and recipients were randomized in a trial of Portland Oregon's Job Opportunities and Basic Skills Training Program (a welfare-to work program), to evaluate the program's effects on employment and earnings – see http://evidencebasedprograms.org/wordpress/?page_id=140; (ii) between 400 and 800 women were randomized in each of three trials of the Nurse-Family Partnership (a nurse home visitation program for low-income, pregnant women), to evaluate the program's effects on a range of maternal and child outcomes, such as child abuse and neglect, criminal arrests, and welfare dependency – see http://evidencebasedprograms.org/wordpress/?page_id=57; 206 9th graders were randomized in a trial of Check and Connect (a school dropout prevention program for at-risk students), to evaluate the program's effects on dropping out of school – see http://evidencebasedprograms.org/wordpress/?page_id=92; 56 schools containing nearly 6000 students were randomized in a trial of LifeSkills Training (a substance-abuse prevention program), to evaluate the program's effects on students' use of drugs, alcohol, and tobacco – see http://evidencebasedprograms.org/wordpress/?page_id=128.

[6] The study, after obtaining estimates of the intervention's effect with sample members kept in their original groups, can sometimes use a "no-show" adjustment to estimate the effect on intervention group members who actually participated in the intervention (as opposed to no-shows). A variation on this technique can sometimes be used to adjust for "cross-overs." See Larry L. Orr, *Social Experimentation: Evaluating Public Programs With Experimental Methods*, Sage Publications, Inc., 1999, p. 62 and 210; and Howard S. Bloom, "Accounting for No- Shows in Experimental Evaluation Designs," *Evaluation Review*, vol. 8, April 1984, pp. 225-246.

[7] Similarly, a study of a crime prevention program that involves close police supervision of program participants should not use arrest rates as a measure of criminal outcomes, because the supervision itself may lead to more arrests for the intervention group.

Impact Evaluation Glossary
(Sources: 3ie and the World Bank)

Attribution
The extent to which the observed change in outcome is the result of the intervention, having allowed for all other factors which may also affect the outcome(s) of interest.

Attrition
Either the drop out of subjects from the sample during the intervention, or failure to collect data from a subject in subsequent rounds of a data collection. Either form of attrition can result in biased impact estimates.

Baseline
Pre-intervention, ex-ante. The situation prior to an intervention, against which progress can be assessed or comparisons made. Baseline data are collected before a program or policy is implemented to assess the "before" state.

Bias
The extent to which the estimate of impact differs from the true value as a result of problems in the evaluation or sample design.

Cluster
A cluster is a group of subjects that are similar in one way or another. For example, in a sampling of school children, children who attend the same school would belong to a cluster, because they share the same school facilities and teachers and live in the same neighborhood.

Cluster sample
Sample obtained by drawing a random sample of clusters, after which either all subjects in selected clusters constitute the sample or a number of subjects within each selected cluster is randomly drawn.

Comparison group
A group of individuals whose characteristics are similar to those of the treatment groups (or participants) but who do not receive the intervention. Comparison groups are used to approximate the counterfactual. In a randomized evaluation, where the evaluator can ensure that no confounding factors affect the comparison group, it is called a control group.

Confidence level
The level of certainty that the true value of impact (or any other statistical estimate) will fall within a specified range.

Confounding factors
Other variables or determinants that affect the outcome of interest.

Contamination
When members of the control group are affected by either the intervention (see "spillover effects") or another intervention that also affects the outcome of interest. Contamination is a common problem as there are multiple development interventions in most communities.

Cost-effectiveness
An analysis of the cost of achieving a one unit change in the outcome. The advantage compared to cost-benefit analysis, is that the (often controversial) valuation of the outcome is avoided. Can be used to compare the relative efficiency of programs to achieve the outcome of interest.

Counterfactual
The counterfactual is an estimate of what the outcome would have been for a program participant in the absence of the program. By definition, the counterfactual cannot be observed. Therefore it must be estimated using comparison groups.

Dependent variable
A variable believed to be predicted by or caused by one or more other variables (independent variables). The term is commonly used in regression analysis.

Difference-in-differences (also known as double difference or D-in-D)
The difference between the change in the outcome in the treatment group compared to the equivalent change in the control group. This method allows us to take into account any differences between the treatment and comparison groups that are constant over time. The two differences are thus before and after and between the treatment and comparison groups.

Evaluation
Evaluations are periodic, objective assessments of a planned, ongoing or completed project, program, or policy. Evaluations are used to answer specific questions often related to design, implementation and/or results.

***Ex ante* evaluation design**
An impact evaluation design prepared before the intervention takes place. Ex ante designs are stronger than ex post evaluation designs because of the possibility of considering random assignment, and the collection of baseline data from both treatment and control groups. Also called prospective evaluation.

***Ex post* evaluation design**
An impact evaluation design prepared once the intervention has started, and possibly been completed. Unless the program was randomly assigned, a quasi-experimental design has to be used.

External validity
The extent to which the causal impact discovered in the impact evaluation can be generalized to another time, place, or group of people. External validity increases when the evaluation sample is representative of the universe of eligible subjects.

Follow-up survey
Also known as "post-intervention" or "ex-post" survey. A survey that is administered after the program has started, once the beneficiaries have benefited from the program for some time. An evaluation can include several follow-up surveys.

Hawthorne effect
The "Hawthorne effect" occurs when the mere fact that you are observing subjects makes them behave differently.

Hypothesis
A specific statement regarding the relationship between two variables. In an impact evaluation the hypothesis typically relates to the expected impact of the intervention on the outcome.

Impact
The effect of the intervention on the outcome for the beneficiary population.

Impact evaluation
An impact evaluation tries to make a causal link between a program or intervention and a set of outcomes. An impact evaluation tries to answer the question of whether a program is responsible for changes in the outcomes of interest. Contrast with "process evaluation".

Independent variable
A variable believed to cause changes in the dependent variable, usually applied in regression analysis.

Indicator
An indicator is a variable that measures a phenomenon of interest to the evaluator. The phenomenon can be an input, an output, an outcome, or a characteristic.

Inputs
The financial, human, and material resources used for the development intervention.

Intention to treat (ITT) estimate
The average treatment effect calculated across the whole treatment group, regardless of whether they actually participated in the intervention or not. Compare to "treatment on the treated estimate".

Intra-cluster correlation
Intra-cluster correlation is correlation (or similarity) in outcomes or characteristics between subjects that belong to the same cluster. For example, children that attend the same school would typically be similar or correlated in terms of their area of residence or socio-economic background.

Logical model
Describes how a program should work, presenting the causal chain from inputs, through activities and outputs, to outcomes. While logical models present a theory about the expected program outcome, they do not demonstrate whether the program caused the observed outcome. A theory-based approach examines the assumptions underlying the links in the logical model.

John Henry effect

The "John Henry effect" happens when comparison subjects work harder to compensate for not being offered a treatment. When one compares treated units to those "harder-working" comparison units, the estimate of the impact of the program will be biased: we will estimate a smaller impact of the program than the true impact we would find if the comparison units did not make the additional effort.

Minimum desired effect

Minimum change in outcomes that would justify the investment that has been made in an intervention, accounting not only for the cost of the program and the type of benefits that it provides, but also on the opportunity cost of not having invested funds in an alternative intervention. The minimum desired effect is an input for power calculations: evaluation samples need to be large enough to detect at least the minimum desired effects with sufficient power.

Null hypothesis

A null hypothesis is a hypothesis that might be falsified on the basis of observed data. The null hypothesis typically proposes a general or default position. In evaluation, the default position is usually that there is no difference between the treatment and control group, or in other words, that the intervention has no impact on outcomes.

Outcome

A variable that measures the impact of the intervention. Can be intermediate or final, depending on what it measures and when.

Output

The products and services that are produced (supplied) directly by an intervention. Outputs may also include changes that result from the intervention which are relevant to the achievement of outcomes.

Power calculation

A calculation of the sample required for the impact evaluation, which depends on the minimum effect size that we want to be able to detect (see "minimum desired effect") and the required level of confidence.

Pre-post comparison

Also known as a before and after comparison. A pre-post comparison attempts to establish the impact of a program by tracking changes in outcomes for program beneficiaries over time using measures both before and after the program or policy is implemented.

Process evaluation

A process evaluation is an evaluation that tries to establish the level of quality or success of the processes of a program. For example: adequacy of the administrative processes, acceptability of the program benefits, clarity of the information campaign, internal dynamics of implementing organizations, their policy instruments, their service delivery mechanisms, their management practices, and the linkages among these. Contrast with "impact evaluation".

Quasi-experimental design
Impact evaluation designs that create a control group using statistical procedures. The intention is to ensure that the characteristics of the treatment and control groups are identical in all respects, other than the intervention, as would be the case in an experimental design.

Random assignment
An intervention design in which members of the eligible population are assigned at random to either the treatment group (receive the intervention) or the control group (do not receive the intervention). That is, whether someone is in the treatment or control group is solely a matter of chance, and not a function of any of their characteristics (either observed or unobserved).

Random sample
The best way to avoid a biased or unrepresentative sample is to select a random sample. A random sample is a probability sample where each individual in the population being sampled has an equal chance (probability) of being selected.

Randomized evaluation (RE) (also known as randomized controlled trial, or RCT)
An impact evaluation design in which random assignment is used to allocate the intervention among members of the eligible population. Since there should be no correlation between participant characteristics and the outcome, and differences in outcome between the treatment and control can be fully attributed to the intervention, i.e. there is no selection bias. However, REs may be subject to several types of bias and so need follow strict protocols. Also called "experimental design".

Regression analysis
A statistical method which determines the association between the dependent variable and one or more independent variables.

Selection bias
A possible bias introduced into a study by the selection of different types of people into treatment and comparison groups. As a result, the outcome differences may potentially be explained as a result of pre-existing differences between the groups, rather than the treatment itself.

Significance level
The significance level is usually denoted by the Greek symbol, α (alpha). Popular levels of significance are 5% (0.05), 1% (0.01) and 0.1% (0.001). If a test of significance gives a p-value lower than the αlevel, the null hypothesis is rejected. Such results are informally referred to as 'statistically significant'. The lower the significance level, the stronger the evidence required. Choosing level of significance is an arbitrary task, but for many applications, a level of 5% is chosen, for no better reason than that it is conventional.

Spillover effects
When the intervention has an impact (either positive or negative) on units not in the treatment group. Ignoring spillover effects results in a biased impact estimate. If there are spillover effects then the group of beneficiaries is larger than the group of participants.

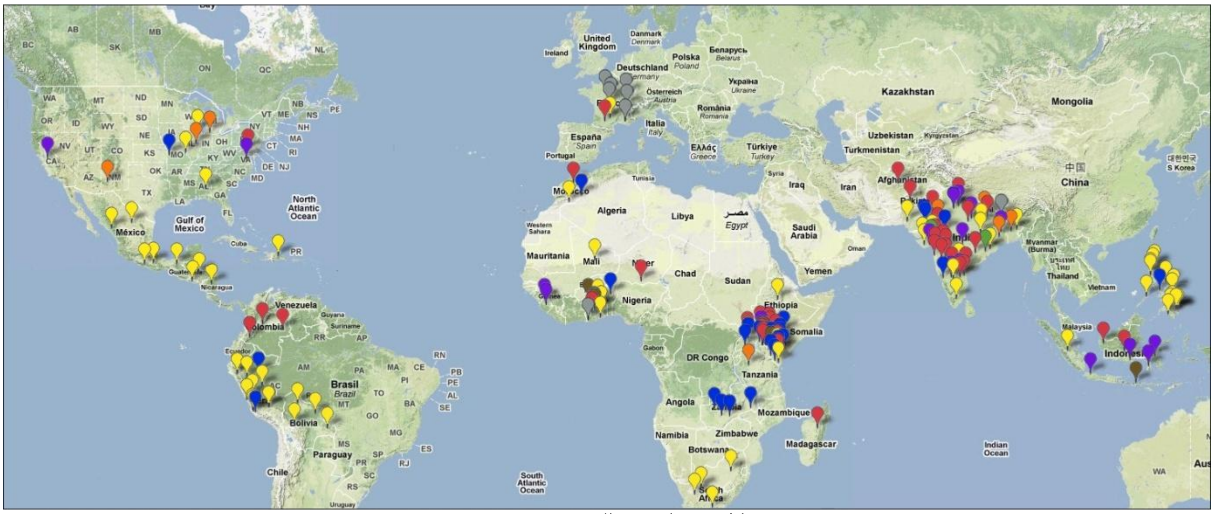
Stratified sample
Obtained by dividing the population of interest (sampling frame) into groups (for example, male and female), then by drawing a random sample within each group. A stratified sample is a probabilistic sample: every unit in each group (or strata) has the same probability of being drawn.

Treatment group
The group of people, firms, facilities or other subjects who receive the intervention. Also called participants.

Treatment on the treated (TOT) estimate
The treatment on the treated estimate is the impact (average treatment effect) only on those who actually received the intervention. Compare to intention to treat.

Unobservables
Characteristics which cannot be observed or measured. The presence of unobservables can cause selection bias in quasi-experimental designs.

J-PAL & IPA Office Contacts



J-PAL projects all over the world

J-PAL Offices

Global Office at MIT
30 Wadsworth St., E53-320
Cambridge, MA 02142 USA
Phone: +1 (617) 324-6566
Email: info@povertyactionlab.org
Website: www.povertyactionlab.org

Europe Office at Paris School of Economics
66bis avenue Jean Moulin
75014 Paris, FRANCE
Phone : +33 (0)1 71 19 40 70
Email : jpaleurope@povertyactionlab.org
Website : www.povertyactionlab.org/europe

South Asia Office at IFMR
Institute for Financial Management and Research
IITM Research Park, A1, 10th Floor
Kanagam Road
Taramani, Chennai 600113, INDIA
Phone: +91 44 3247 50 56
Email : jpalsa@povertyactionlab.org
Website : www.povertyactionlab.org/south-asia

Latin America and Caribbean Office at Pontificia
Universidad Católica de Chile
Instituto de Economía
Av. Vicuna Mackenna 4860
Santiago, CHILE
Phone: +(56-2) 354-1291
Email : jpallac@povertyactionlab.org
Website : www.povertyactionlab.org/LAC

Africa Office at SALDRU
University of Cape Town
Private Bag X3
Rondebosch 7701, SOUTH AFRICA
Phone: +27 21 650 5981
Email : jpalafrica@povertyactionlab.org
Website : www.povertyactionlab.org/africa

IPA

Main Office:
Innovation for Poverty Action
101 Whitney Ave
New Haven CT 06510, USA
Phone: +1 203 772 2216
Email: contact@poverty-action.org
Website: www.poverty-action.org

For the other IPA Offices:
Please refer to www.poverty-action.org/about/contact

Notes

Notes

Notes

Notes

Notes