

IFPRI's evaluation of PROGRESA in Mexico: Norm, Mistake, or Exemplar?

William N. Faulkner

Published in Evaluation, Volume 20, Number 2, April 2014. SAGE Publications.

Abstract:

During the early stages of implementing their new flagship anti-poverty program, Progresá, Mexican officials contracted an evaluation team from the International Food Policy Research Institute (IFPRI). This article critically revisits the narrative of this evaluation's methodology using an approach informed by science studies, finding a number of significant omissions and ambiguities.

By profiling the available information on these facets of the evaluation against, first, the intellectual background of randomised control trials (RCTs) in international development evaluation, and second, the political context of the project, the analysis attempts to extract lessons for today's evaluation community.

These lessons result in two invitations: (1) for the proponents of RCTs in the field of international development evaluation to critically and reflexively consider the problematic framing of the methodology in this case study and what the implications might be for other similar projects; (2) for evaluators critical of how RCTs have been presented in the field to narrow a greater portion of their analyses to specific case studies, known econometric issues with familiar labels, and living institutions with names.

Keywords:

Conditional Cash Transfers; Randomised-Control Trials; Meta-Evaluation.

Glossary of Abbreviations

Abbreviation	Definition
CCT	Conditional Cash Transfer
CONEVAL	Consejo Nacional de Evaluación (National Council on Evaluation)
EBP	Evidence Based Policy
IDB	Inter-American Development Bank
IFPRI	International Food and Policy Research Institute
LGDS	Ley General de Desarrollo Social (General Law of Social Development)
PAN	Partido Acción Nacional (National Action Party)
PRI	Partido Revolucionario Institucional (Institutionalized Revolutionary Party)
PROGRESA (Progresa)	Programa de Educación, Salud y Alimentación (Education, Health, and Nutrition Program)
RCT	Randomised-Control Trial, Randomised Controlled Trial

Introduction

Mexico's Conditional Cash Transfer (CCT) Progresa began in 1997 as an innovative social policy idea and grew to become the nation's flagship anti-poverty program. The program's early survival and potent influence relied upon the integration of an independent impact evaluation by the International Food Policy Research Institute (IFPRI), a prestigious DC-based research non-profit. The IFPRI-Progresa evaluation spearheaded broad changes, not only in Mexican social policy, but also around the world, impelling both the CCT and Evidence-Based Policy (EBP) movements. Also, as one of the first large-scale Randomised-Control Trial (RCT) research projects ever implemented in social program evaluation in a 'developing-country' context, IFPRI-Progresa provided (and continues to provide) an important proof-of-concept, a paragon of the methodology's potential extolled by proponents and held aloft before critics.

At the same time, however, IFPRI-Progresa was not without flaws in both design and practice. This article reviews the project's original documentation, zooming in on features which could serve as bases of productive criticism, specifically the original sampling process, sample attrition and sample contamination. The methodology included a desk review and critical interrogation of official project documents, published articles, books, and statements by participants involved in the evaluation. Although a substantial portion of the text will deal with essentially econometric issues, no original statistical/quantitative work (i.e. sensitivity analysis) will be presented so as not to distract from the central position presented, which is precisely to highlight how numerical results may be shaped, both during and after their production, by socio-political forces. This piece makes no claim as to how well the numbers presented in the IFPRI-Progresa reports represent the impacts of the program on the Mexican rural poor, but instead draws lessons from the process of their production and representation for today's evaluation community.

The analytic approach draws heavily from the field of science studies, particularly the work of Bruno Latour (Latour, 1987; Latour, 1999; Latour, 2004). This approach demonstrates how observing the narrative of scientific production, complete with its instruments, personalities, and institutions, helps to reveal the complex ways in which the results of scientific study engage with their human/societal context. In this view, quantitative results, as they appear in the evaluation documents investigated for this study, are not static facts which represent incontrovertible properties of the social world, but rather palimpsests of that reality produced by a team of trained individuals at a certain moment in a given context employing a particular method.

The mainstream narrative of IFPRI-Progresa smoothes over many of the intricacies of the process of the study. First, Section I revisits three aspects of this evaluation relating to how the samples were selected and maintained, exposing some critical gaps in the mainstream story told in many sources. Second, Sections II and III attempt to fulfill the first-order purpose of this article by enriching the available evidence on these lacunas with background information, specifically the intellectual context of RCTs and the political realities of social policy development in Mexico during the late 1990s. These two backdrops were chosen precisely because the gaps and omissions in the early documentation appear so different when examined against them. The two sections convey two highly divergent impressions of the evaluation, the former highly critical and the latter celebratory.

The final section of this piece reaches for the second-order goal of drawing out the implications of this revised narrative for the field of international development evaluation, extending two invitations:

- (a) For the proponents of RCTs in the evaluation community to critically consider how certain features of the methodology were framed and represented in this case study and what the implications might be for other similar projects.
- (b) For critics of how RCTs have been broadly presented to the evaluation community to focus their analyses on synthesizing evidence and focusing analyses on case studies.

The overall hope is to provide the foundation of a more customized and pointed critique, not of RCTs, but of how one of them unfolded in practice, and what we can learn from this case about how to express their role in evaluation research. This commentary strives to comply with the spirit and ultimate goals of the critic as defined by Bruno Latour:

"The critic is not the one who debunks, but the one who assembles. The critic is not the one who lifts the rugs from under the feet of the naive believers, but the one who offers the participants arenas in which to gather."
(emphases added - Latour, 2004: 246)

Section I: The Process Perspective

Section I cross-examines IFPRI-Progresa from a purposefully narrow perspective devoid of connections to the wider context within which the project emerged. The goal is to describe the basic components of the analysis, before they are overlaid with additional information. The examination of three key issues, the original sampling method, sample attrition, and contamination, exposes some details the

IFPRI evaluation reports tend to omit or simplify, details which imply significant holes in the understandings of how the IFPRI-Progresa study unfolded.

Sampling

The IFPRI-Progresa sampling process, as described in some early sources and a splinter group of later literature, remains somewhat murky, contrary to the clean-and-simple descriptions echoed by the vast majority of work which analyses the IFPRI-Progresa data. Personnel at the Secretaría de Desarrollo Social (Ministry of Social Development, SEDESOL) undertook and directed the actual sampling process in mid-1997, prior to the engagement of most of the IFPRI research team. From this point on, however, at least three stories developed as to how the sampling actually took place.

First, some documentation suggests that the process which selected the treatment and control groups for the IFPRI-Progresa project was quasi-experimental. The treatment and control communities apparently were selected from two distinct universes:

"The design of the impact evaluation of Progresa in communities and households is quasi-experimental... a random sample of communities with 'high' or 'very high' degrees of marginalization were selected which would be incorporated into the program during Phase II (November 1997) and which would serve as the [treatment] communities...Another sample of communities with similar characteristics was randomly designated from those that could have been the object of later selection, and that could function as controls..." (SEDESOL, 1999: 393)

Unfortunately, none of these sources contain information regarding how project managers assigned communities to the different phases of incorporation. The process described in these sources is deceptively similar to a fully experimental design, whereby communities would have been randomised first into treatment and control groups and second assigned to staggered phases of incorporation. The ordering matters enormously, however, because without knowing what process was used to decide the original phasing, it is impossible to know whether communities assigned for earlier or later incorporation might differ systematically.

Second, the vast majority of sources (including the executive synthesis report quoted below) suggest that the design was fully experimental:

"From a set of rural communities in the same geographic region, localities were randomly selected for participation in PROGRESA (treatment localities)

while the rest were introduced into the program at later phases (control localities)." (Skoufias, 2000: 10)

Even within this group of sources, however, there are two subtly different stories. Whereas the quotation above would suggest a clustered randomisation (randomisation at community as opposed to individual level), quotes like the one below portray a clustered matched-pairs design, which uses a statistical matching to align similar pairs and then randomly assigns one of each pair to the treatment and control group (Imai, King and Nall, 2009: 65-66):

"Assignment of communities to treatment or control was based on, first, matching communities on characteristics (levels of infrastructure and economic status measured at the community level) and, second, random assignment of one community in each matched pair to the treatment, the other to the control group." (Rubalcava and Teruel, 2003: 6)

Quasi-experimental evaluations are "still a perfect impact evaluation in theory," (Khandker, Koolwal, and Samad, 2010: 54) and commonly used to evaluate social programs with practical results. The distinction between an experimental and quasi-experimental sampling procedure therefore in no way threatens the validity of the IFPRI-Progresa results. Neither do the conspicuous indications of nonrandomness at the household level uncovered by the report testing the empirical equality between treatment and control groups (Behrman and Todd, 1999) considering that many of the studies employed methods generally brought to bear on quasi-experimental samples. Jointly, however, these issues strongly contest the simplicity of later descriptions of the sampling process.

Attrition

Experiments and semi-experiments use panel data to measure impacts across time. Between survey rounds, some interviewees inevitably drop out of the dataset for a variety of reasons (opting out, moving, emigration, death, etc.). Non-random attrition across time causes sample population characteristics to shift, and sometimes these shifts are uneven between treatment and control groups (selective attrition). What results is "analytically similar" (Heckman et al., 1998) to selection bias in that "attrition lead[s] to selective [read: systematically different] samples" (Alderman et al., 2001). The most common method to mitigate attrition bias is to follow up with non-respondents (e.g. tracking down migrants).

"The IFPRI evaluation did not follow up migrants as part of the evaluation surveys," (Parker, 2003: 21 footnote 9) and yet:

"...close to 35% (45%) of households (individuals) were not found in one more waves; close to 17% (29%) of households (individuals) were interviewed in 1997 and again in November 2000, even though they could have missing information for any wave in between). Attrition differs significantly between treatment and control groups, even after controlling for household characteristics and the eligibility criteria." (Rubalcava and Teruel, 2003: 7)

The presence of nonrandom attrition indicates that even if the samples were experimentally selected and statistically identical to begin with, by the end of the experiment period they were significantly unequal. As explained in the following section, attrition was only tangentially analysed in the original IFPRI analyses.

Contamination

The term 'contamination' may refer to several different concepts. In this piece it will refer to the possibility that "families or individuals from control localities or other localities [can] immigrate to treatment group localities in order to receive program services" (Behrman and Todd, 1999: 3). Over the course of the data collection period, contamination "can appear and intensify...and so affect estimates of program impact" (King and Behrman, 2008: 4). As with attrition bias, the decision to not track down sample out-migrants makes exact quantification of this bias impossible, but the outlook is worrisome:

"The rapid expansion of the program...meant that control communities often literally were surrounded by other communities that were already receiving [benefits]..." (Fiszbein and Schady, 2009: 311)

"The growth in beneficiaries likely did not go unrecognized by the control communities." (Parker and Teruel, 2005: 211)

It is important to note that there are many studies which examine the other concepts of 'contamination' and/or 'spillover,' using the same data (for example: Attanasio, Meghir, and Santiago, 2010; Behrman and Hoddinott, 2005; Bobonis and Finan, 2002; Bobonis, 2004; Buddelmeyer and Skoufias, 2004). None of these sources addresses the potential impact of contamination as defined above. Once again, discussion of contamination of the treatment communities by migrating members of the control-community is all but absent from later descriptions of the IFPRI-Progresa sample.

A hazy sampling process, nonrandom attrition and probable contamination of the treatment communities do not necessarily discredit the IFPRI-Progresa team's early results as representative of what happens when a CCT comes into contact with the Mexican rural poor. They do reveal, however, that what in most

published accounts is described as the airtight execution of an RCT, in reality appears to have had several leaky seams.

Section II: The Critical Scholar's Perspective

The decisions which guided the IFPRI-Progresa research project did not emerge in a vacuum, but instead in a vibrant, shifting landscape of interacting agents, information, theory, and belief. This section will use this background to take a second look at the same features examined by Section I, but this time couching IFPRI-Progresa within the worldwide discussion surrounding 'best practices' evaluation methodologies. The goal of this section is to show how the previously explored features of the IFPRI-Progresa evaluation process transform when seen as emerging within the larger system of evaluation-interested academics and researchers. The claim is that this context is likely to have helped produce what is seen in practice, that is, a selective emphasis on the problems where the experimental design is strong in comparison to others, such as the mitigation of selection bias, and a diversion of attention away from potential weak points.

Three tandem trends harmonized during the past two decades to bring the experimental design to the forefront of evaluation methodologies, and also to the center of widespread debate. First, since the 19th century, economics as a discipline has gradually gained prestige relative to the other social sciences (Beinhocker 2006; Ferraro and Pfeffer, 2010; Holton, 1992). Second, during the post-war decades, econometrics (the application of economic theory to real-world data using statistical models) rose to prominence within the field (Swann, 2006). Finally, over the past two decades a core group of applied micro-economists have forwarded the use of experimental designs as tools for evaluating programs in international development contexts (Ravallion, 2009). The ascension of the experiment has not come without controversy, however. Some authors claim that the emergent methodological hierarchy, with experiments as the "gold standard," (Card, Ibararan, and Villa, 2011: 7) engenders "exceedingly narrow" (Cartwright, 2007: 2) conceptions of evidential relevance, and indeed applies "brakes on the acquisition of knowledge" (Davidson, 2006). This dissidence has instigated

"...an emerging duality of evaluative practice: 'on the one hand, the methodologies of the economist with RCTs and counterfactual analysis...and, on the other hand, a group of alternative methodologies.'" (Cummings – quoted in Guijt, 2010: 3)

The debate stemming from calls for more experiments in social program evaluation, with its critiques, rebuttals, and counter-rebuttals, is now widespread (some key sources from various positions on the spectrum: Duflo and Kremer,

2003; Davidson, 2006 Stirling et al., 2007; Ravallion, 2009; Angrist and Pischke, 2010; Barrett and Carter, 2010; Deaton, 2010; Guijt et al., 2010)

The superposition of the economic, econometric, and experimental trends produced a wave through policy evaluation, upon whose crest rode IFPRI-Progresa. This venture thus occurred at a truly crucial and propitious moment for those actors interested in expanding the esteem of experiments in social policy evaluation and catalyzing their usage worldwide. IFPRI-Progresa both resulted from the aforementioned rise of experiments, and helped to propel them to prominence, enhancing their regard with widespread international support. The concern, therefore, is that the difficulty and novelty of using an experimental approach could have motivated IFPRI-Progresa authors to shape the presentation of the study design as 'free-of-caveats' as possible.

Sampling

Ambiguity and apparent oversimplification plague the documentation delineating the sampling procedure. Particularly in the original methodological documents, it is difficult to imagine an excuse for not accurately and completely recording the procedure employed. As stated in one source, "the lack of documentation by government officials may reflect their perception of the controversial nature of carrying out an evaluation with an experimental design" (Parker, Rubalcava, and Teruel, 2008: 3980, footnote 12). Still, from a scholarly perspective, having literally no unified record of the methodology of such a large and influential study seems unjustifiable. To a lesser degree, the unquestioning and uncritical stance taken by the IFPRI-Progresa authors (the source cited above is the only one to mention the lack of documentation) also seems somewhat blameworthy from a purely academic standpoint.

As previously mentioned, however, a small, splinter group of sources, mostly grey literature, appendices, and footnotes, provides hints of the potential intricacy of the sampling procedure. With the exception of a tangential and contradictory allusion to these details in Skoufias (2005: xi, 27), none of the IFPRI-associated team nor the later authors citing their work mention the 'intermediate' selection universes or any matching procedure. Instead, this body of literature tells the simple, clean story that "the initial phase of Progresa was implemented as a randomised social experiment in which 506 rural villages were randomly assigned to either participate in the program or serve as controls" (Behrman, Sengupta, Todd: 238).

Attrition

As opposed to selection bias, which features prominently in the IFPRI-Progresa documentation, only two final reports mention attrition bias (which, recall, creates essentially the same problem from an analytic point of view). Schultz (2000) carries out analyses of enrollment using both a "panel sample" (households with data across all five survey rounds) and a "pooled sample" (households with data in at least one survey round). While admitting that "it is not possible to implement a satisfactory sample-selection-correction model," (p. 2) the word 'attrition' appears only once. The rest of the report frames the use of both samples as a form of robustness testing. This framing turns attention away from attrition bias, instead highlighting the dual-sample testing as a strength of the analysis. Behrman and Todd (1999) discuss attrition, but only hypothetically, never making any connection to the actual data collected. Their piece states: "The problem of attrition will be examined in detail in a future report" (p. 3). Several documents and publications highlight the extent of attrition and its potential effect on the estimated impacts in the original samples (for instance: Behrman and Hodinott, 2005; Bobonis 2011; Parker, Rubalcava, and Teruel 2008; Parker and Teruel, 2005; Rivera et al. 2004; Teruel and Rubalcava, 2003). Other sources (e.g. Parker and Todd 2010) address attrition in different samples and time periods. Still, it seems unlikely that in the original report Behrman and Todd were referencing these documents which appeared years later and under the auspices of separate projects and institutions. No other IFPRI report mentions attrition whatsoever and "none of the [original IFPRI] analysis considered the possible biasing effects of attrition/migration on estimated program impacts" (Parker, Rubalcava, and Teruel, 2008: 3991).

Contamination

In much the same manner, contamination bias is mentioned in just one of the IFPRI final reports. Behrman and Todd (1999) once again discuss the subject hypothetically, noting that "it would be most useful if individuals could be followed if they leave the locality, so that migration patterns could be understood and taken into account" (p. 3 footnote 6). This report, dated March 26, 1999, came out before the completion of survey rounds in November, 1999. Still, the authors would likely have known that even in the final round migrants would not be tracked down and surveyed, and that "pressure applied by control communities through local and state government officials...in combination with the fast growth of Progresa, [was contributing] to the decision...to end the experiment earlier than planned (e.g., at the end of 1999 rather than in the year 2000...)" (Parker and Teruel, 2005: 211). Why then, does the discussion remain purely hypothetical, if not as a sort of strategic omission? By transforming clearly-

present but unquantifiable biases into hypotheticals, the IFPRI authors shift attention away from this shortcoming.

In general, the intellectual perspective shows that the IFPRI-Progresa narrative appears to have evolved in a way well-described in the science studies literature:

"Metaphorically speaking, statements, according to the first principle, are much like genes that cannot survive if they do not manage to pass themselves on to later bodies." (Latour, 1987: 38)

The move towards the clear-cut characterization does not seem so surprising when contextualized within an economics-heavy environment laden with strong selective pressure for cleanly-derived experimental results. Many thinkers have depicted the ubiquity of these "drift[s]," "detours," or "free moves" within academic writing (Latour, 1987, Pickering and Guzik, 2008). The conflicting descriptions of IFPRI-Progresa's sampling, for instance, did not necessarily have to originate with an individual or group purposefully abridging their narratives. Instead, the large, loose network of researchers laden with responsibilities, time pressures, and power relationships, may simply have gravitated towards the predominant trends within their environment.

Social policy evaluation and experimental designs sit positioned within complex socio-historical circumstances which have brought them to the fore over the course of the last two decades. At the outset of this meteoric rise, garnering support for IFPRI-Progresa and the RCT methodology was "one of the most serious difficulties the program encountered" (author's translation from Spanish, Paz-López, 2007: 851). For the IFPRI-Progresa team, the struggle to implement the evaluation as an experiment thus symbolized economists' larger battle for the methodology's legitimacy in the social policy arena worldwide. The implications of their victory would ripple far beyond Progresa and the rural Mexican poor, deeply influencing the field of international development evaluation.

Section III: Validity – The Political Perspective

Section III views the same three facets of IFPRI-Progresa through the lens of politics, policy, and institutions, resolving yet another image of the project, as concocted and executed within a dynamic political environment, already ripe with various predispositions about evidence, evaluation designs, social policy, and CCTs. Most notably, three hulking political trends cast their shadows over the evaluation project: on the global stage, the ascension-to-prominence of EBP and the growing acceptance of CCTs as viable poverty reduction tools, and in Mexico, the slow destabilization of the Partido Revolucionario Institucional

(Institutionalized Revolutionary Party, PRI). Against this backdrop, IFPRI-Progresa appears in a more positive light, more worthy of the eulogies it has received.

By the time Progresa began to form in the heads of Mexican social policy designers, the controversy surrounding the long rule of the PRI and its social policies was already entrenched. Under 'New Federalism,' president Ernesto Zedillo (1994-2000) attempted to de-politicized social programs. One of the paramount objectives at the core of Progresa's development, therefore, was to establish the program's apolitical status by sustaining it through the presidential switch in 2000. Vicente Fox Quesada, leader of the Partido Acción Nacional (National Action Party, PAN), won the presidential elections on July 2nd, 2000, and took office on December 1st. The IFPRI final reports were released primarily during this interval. Program supporters used the transition period to convince skeptical parties in the new administration of Progresa's effectiveness through personal contact, meetings, and media reports (Lustig, 2011: 9).

Due to the IFPRI evaluations, Progresa gained political hardiness and funding sources, while evaluation became a codified part of Mexican social programming. As if surviving the presidential election cycle, a feat never before accomplished by a Mexican social program, were not hard enough, Progresa endured through the first changeover of the national ruling party in seventy-one years (Behrman, 2007: 6-8; Rocha-Menocal, 2005: 353). President Fox not only continued the program, but expanded it significantly into urban areas (Levy, 2006: 112-113). A loan for the amount of US\$1B from the Inter-American Development Bank (IDB) for the expansion of the program was negotiated during 2001, and disbursed beginning in 2002. At the time, this was the largest loan ever granted by the IDB (Parker, 2003: 28-29). In 2003, the Ley General de Desarrollo Social (General Law of Social Development, LGDS) explicitly codified the requirement for independent evaluation for all federally-run Mexican social programs (Greenspun, 2011: 23) and for this purpose established the Consejo Nacional de Evaluación (National Council on Evaluation, CONEVAL) (Oxman et al., 2010: 428).

The excitement for the new program spread not only through Mexico, but through the IDB and World Bank (Levy, 2006: 112). Both of these organizations extensively employed the example of IFPRI-Progresa to forward both CCTs and EBP (via 'rigorous' evaluation of social programs). The World Bank "featured the program as a model" in the 2004 Conference on Poverty in Shanghai (Rocha Menocal, 2005: 353), and in 2006 the former President James D. Wolfensohn commented "Progresa's rigorous emphasis on evaluation has set a standard for poverty reduction programs in the developing world" (Levy, 2006: viii). The evaluation remains one of the programs' most celebrated features, having "inspired many others in Latin America and other regions" (Székely, 2011: 20).

As mentioned previously, "Progresa had already done the randomisation prior to IFPRI's involvement," (Todd quoted in Behrman, 2007: 80) apparently under the aegis of Daniel Hernández Franco, who also carried out the baseline ENCEL survey in March 1998 "without much input from anyone" (Gertler quoted in Behrman, 2007: 75). When IFPRI became involved, "Progresa was already committed to the program design and evaluation design...and IFPRI's role was to support and reinforce this commitment" (David Coady quoted in Behrman, 2007: 86). From this perspective, the tendency to overlook the ambiguity and incompleteness of the sampling documentation appears quite plausible. To question the sampling procedure too deeply would have been (1) a demonstration of mistrust of the very people so vocally committed to working towards improving social policy through evaluation (not to mention that they belonged to the entity which had contracted IFPRI), and (2) ultimately counterproductive. Who would have benefited from an exposé revealing the project as anything less than an ideal design implementation? Certainly not the well-intentioned representatives of the Mexican government who had dedicated themselves to making strides towards de-politicization and long-term results. Certainly not the well-intentioned IFPRI researchers. Certainly (it seemed) not the Mexican rural poor.

With so much riding on the project, including reputations of so many deserving and well-intentioned individuals, not to mention the growing momentum of an attitude of accountability, transparency, and unconditional commitment to alleviating poverty, what rational individual would risk igniting controversy, derailing the entire project, erasing all the progress, and setting social policy in Mexico back at least five years? Such a controversy would also have condemned future evaluation projects with similarly honorable objectives to even more insurmountable political barriers.

The international political outcomes depended on the perceived external validity of IFPRI-Progresa's results. From the standpoint of the Mexican government, the decision to continue, cut, scale up, or scale down the project depended on the notion that the results not only held up on their own, but also potentially applied to populations outside of the original sample. In this light, complicating the narrative of how IFPRI-Progresa's findings and fully explicating their potential weaknesses (e.g. sample attrition and contamination) would likely have paralyzed both operators and funders, leaving them no justified course of action. Any apparent weakness of the representativeness of findings could be used to ensnare the emergence of the EBP strategy while still in its infancy. Such nitpicking would have flagged the momentum made towards spreading the CCT strategy and tarnished the image of a project which actively and seriously promoted governmental transparency.

Contextualizing the evaluation project within this ambience of excitement regarding CCTs, the use of the social sciences to inform policy, and the momentous transformation of Mexican federal government illuminates the extent of the power placed in the hands of the IFPRI-Progresa team. If the IFPRI-Progresa results were not seen as applicable to a larger context then researchers, lobbyists, policy-makers, and the development-focused public worldwide would cast their money and their attention elsewhere. Moreover, the money designated for Progresa, most of it flowing as cash into the hands of poor women, might end up back in tortilla subsidies or government penchants. If, however, the team could project a clean, apolitical, and positive image of the program, there was enormous opportunity to put wind in the sails of CCTs, bolster the credibility of EBP worldwide, and sustain the direct flow of cash to poor Mexicans.

Conclusions & Pathways

For all of the facts, ideas, and relationships uncovered, there seems to be no coherent, unifying narrative, no seamless truth about IFPRI-Progresa. Sections I, II, and III presented three meticulous narratives which all seem relatively coherent, but end up in very different places. Rather than some single past event, IFPRI-Progresa turns out to have consisted of a plethora of heterogeneous elements within a huge system interacting in only semi-predictable fashion. This opacity, however, is precisely the environment with which evaluators, academics, and politicians must come to grips in social policy evaluation.

The strong reflex is to worry that embracing multiplicity implies falling into the relativist trap and being left paralyzed without a course of action. As elaborated in Section III, this risk was very real for the IFPRI-Progresa team on the relatively short timescale of domestic political decision-making, and perhaps even some years after the presidential/party turnover as the program continued to gain political permanence. In certain situations like that of Mexico's federal government in the final years of the 1990s, there is a strong case that the systematic simplification of the evaluation process, whether active or passive, was a politically reasonable course of action. As argued in Section III, it was probably only because of the myopia of IFPRI-Progresa's early narrative that the evaluation team was able to jolt Mexican social policy out of the viscous cycles of political patronage and constant overturn.

In 2013, more than a decade later, the relevance of this argument has faded dramatically. Reexamining and rearticulating the detailed narrative of IFPRI-Progresa's undertaking now furnishes important lessons for the evaluation community as a whole. This reinterpretation does not provide incontrovertible truths, or singular pathways forwards, but within the arena of international

development evaluation, filling in the story of IFPRI-Progresa with its full degree of socio-historical complexity could help to bring RCT proponents and their critics closer together.

In 2000, as one of the first experiments ever undertaken in a developing country context, IFPRI-Progresa's ostensible success in implementing the design served to legitimate RCTs in this new context. Over the course of the following decade, however, the clean-and-simple IFPRI-Progresa narrative has been repeatedly employed to legitimate an agenda to place RCTs at the peak of a questionable and controversial methodological hierarchy. During the same period, there has been a complete absence of any critical revisit to the process of presenting and disseminating IFPRI-Progresa's results in late 2000. The partiality of IFPRI-Progresa's narrative continues to have two principle effects:

(a) The perpetuation of the forged image that the quantitative results of IFPRI-Progresa were unaffected by their political and intellectual context, an image which aids in preserving a stalemate between the randomistas and their critics.

(b) The starvation of the evaluation community of valuable lessons about negotiating conflicting needs and the intricacies of strategically disseminating evaluation results drawn from direct experience.

The assertion of this article, therefore, is that while perhaps a beautiful example of evaluation's power to change politics and induce more rational social policy-making, the longer-term results of IFPRI-Progresa which emerged within the field of evaluation have significantly aided the installation of a false imbalance between methodologies and an accompanying, unnecessary debate. Perhaps reanimating the IFPRI-Progresa narrative with its messy details, a feat only accomplishable with the help of those who lived through the project itself, would be the gesture necessary to open some version of Latour's gathering space, a dissonant yet respectful assemblage of opinions, where RCT proponents could begin to reshape the presentation of quantitative results so as to publicly and seriously acknowledge their historical, linguistic, and time-dependent facets.

Similarly, this piece has attempted to demonstrate ample room remains for the critics of how RCTs and similar methods are represented within international development evaluation to narrow and specify their analyses. Orienting critiques at case studies, at known econometric issues with familiar labels, at people with names and living institutions requires a deep personal commitment and reputational investment. Direct negotiation of 'ground-level' issues implies more frequent, less formal communication, exactly the type of engagement which

allows both members to iteratively rearticulate their points and reach a mutual understanding.

Fully understanding IFPRI-Progresa means letting go of that ever-present drive to arrive at a singular thumbs-up or thumbs-down conclusion. It is precisely because IFPRI-Progresa's narrative has such easily-identified strong and weak points that it provides such profound lessons for the evaluation community. This article is predicated on the hope that exposing the more detailed narrative of IFPRI-Progresa may engage RCT proponents in a reflexive discussion about how the methodology has permeated the socio-political world of evaluation, and simultaneously encourage the critics of how RCTs are represented in evaluation to delve deeper into specific narratives as a way of inciting discussion and finding common ground.

Bibliography

Alderman H, Behrman J, Watkins S, et al. (2001) Attrition in Longitudinal Household Survey Data. *Demographic Research*, 5, 79–124, Available from: <http://www.demographic-research.org/volumes/vol5/4/> (accessed 1 September 2011).

Angrist JD and Pischke J (2010) The Credibility Revolution in Economics: How Better Research Design is Taking the Con out of Econometrics. 24(2), 3–30.

Attanasio O, Meghir C and Santiago A (2010) Education choices in Mexico: Using a structural model and a randomised experiment to evaluate Progresa. IFS working papers, London.

Barrett CB and Carter MR (2010) The Power and Pitfalls of Experiments in Development Economics: Some Non-random Reflections. *Applied Economic Perspectives and Policy*, 32(4), 515–548, Available from: <http://aepp.oxfordjournals.org/lookup/doi/10.1093/aepp/ppq023> (accessed 25 August 2011).

Behrman JR (2007) Policy-Oriented Research Impact Assessment (PORIA) Case Study on the International Food Policy Research Institute (IFPRI) and the Mexican PROGRESA Anti-Poverty and Human Resource Investment Conditional Cash Transfer Program. Assessment, Washington, DC.

Behrman JR and Hoddinott J (2005) Programme Evaluation with Unobserved Heterogeneity and Selective Implementation: The Mexican PROGRESA Impact on Child Nutrition. *Oxford Bulletin of Economics and Statistics*, 67(4), 547–569, Available from: <http://doi.wiley.com/10.1111/j.1468-0084.2005.00131.x>.

Behrman JR and Todd PE (1999) Randomness in the Experimental Samples of PROGRESA. *Food Policy*, Washington, DC.

Beinhocker ED (2006) *The Origin of Wealth: Evolution, Complexity and the Radical Remaking of Economics*. Harvard Business School Press, Available from: <http://www.amazon.com/Origin-Wealth-Evolution-Complexity-Economics/dp/157851777X>.

Bobonis GJ (2004) Income Transfers , Marital Dissolution and Intra-Household Resource Allocation : Evidence from Rural Mexico.

Bobonis GJ (2011) The impact of conditional cash transfers on marriage and divorce. *Economic development and cultural change*, 59(2), 281–312, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21174881>.

Bobonis GJ and Finan FS (2002) Transfers to the Poor Increase the Schooling of the Non-Poor: The Case of Mexico's PROGRESA Program. University of California at Berkeley.

Buddelmeyer H and Skoufias E (2004) An Evaluation of the Performance of Regression Discontinuity Design on PROGRESA. Washington, DC.

Card D, Ibarra P and Villa JM (2011) Building in an Evaluation Component for Active Labor Market Programs: a Practitioner's Guide. Development, Washington, DC.

Cartwright N and Efstathiou S (2007) Evidence-based policy and its ranking schemes: So, where's ethnography? In: Twente, First Biennial conference of the Philosophy of Science in Practice, Twente University, pp. 1–9.

Davidson EJ (2006) The RCTs-Only Doctrine: Brakes on the Acquisition of Knowledge? *Journal of MultiDisciplinary Evaluation*, 6, ii–v.

Deaton A (2010) Instruments, Randomisation, and Learning about Development. *Journal of Economic Literature*, 48(June), 424–455.

Diaz JJ and Handa S (2006) An Assessment of Propensity Score Matching as a Nonexperimental Impact Estimator Evidence from Mexico's PROGRESA Program. *The Journal of Human Resources*, XLI(2), 319–345.

Duflo E and Kremer M (2003) Use of Randomisation in the Evaluation of Development Effectiveness. Development, Washington, DC.

Faulkner WN (2012) Negotiated Perspectives: A Complexity Approach to IFPR-Progressa in Mexico. Tulane University, Available from: http://wnfaulkner.files.wordpress.com/2012/08/faulknerw_negotiatedperspectives.pdf.

Ferraro F, Pfeffer J and Sutton RI (2005) Economics Language and Assumptions: how theories can become self-fulfilling. *Academy of Management Review*, 30(1), 8–24.

Fiszbein A and Schady N (2009) Conditional Cash Transfers: Reducing Present and Future Poverty. *Journal of Banking & Finance*, Washington, DC, Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0378426603001249>.

Gertler PJ, Martinez S and Rubio-Codina M (2012) Investing Chash Transfers to Rais Long-Term Living Standards. *American economic journal*, (1), 164–192.

Greenspun S (2011) Assessing Mexico's Anti-Poverty Program "Oportunidades": Combining and Comparing Objective and Interview-Based Indicators. Tulane University.

Guijt I, Brouwers J, Kusters C, et al. (2010) Evaluation Revisited Improving the Quality of Evaluative Practice by Embracing Complexity. In: Conference Report.

Heckman J, Smith HJ and Todd P (1998) Characterizing Selection Bias Using Experimental Data. *Econometrica*, 66(5), 1017–1098.

Holton G (1992) Ernst Mach and the Fortunes of Positivism in America. *Isis*, 83(1), 27–60.

Imai K, King G and Nall C (2009) Rejoinder: Matched Pairs and the Future of Cluster-Randomised Experiments. *Statistical Science*, 24(1), 65–72.

Khandker SR, Koolwal GB and Samad HA (2010) Handbook on Impact Evaluation: Quantitative Methods and Practices. Learning, Washington, DC, World Bank.

King EM and Behrman JR (2008) Timing and Duration of Exposure in Evaluations of Social Programs. World, Washington, DC.

Latour B (1999) Pandora's Hope: Essays on the Reality of Science Studies. Cambridge, MA, Harvard University Press.

Latour B (1987) Science in Action. Science in Action, Cambridge, MA, Harvard University Press.

Latour B (2004) Why Has Critique Run out of Steam? From Matters of Fact to Matters of Concern. *Critical Inquiry*, 30(2), 225–248, Available from: <http://www.jstor.org/stable/10.1086/421123>.

Levy S (2006) Progress Against Poverty. Washington, DC, The Brookings Institution, Available from: <http://www.getcited.org/pub/101595616> (accessed 7 September 2011).

Lustig N (2011) Scholars Who Became Practitioners: The Influence of Research on the Design, Evaluation, and Political Survival of Mexico's Antipoverty Program Progreso/. *Wall Street Journal*, (June), 1–20, Available from: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1910074 (accessed 28 August 2012).

Menocal AR (2005) Less political and more pro-poor? The evolution of social welfare spending in Mexico in a context of democratisation and decentralisation. *Nord-Süd Aktuell*, 3(4), 346–359.

Orazio AA, Costas M and Santiago (2010) Education choices in Mexico: Using a structural model and a randomised experiment to evaluate ProgresA.

Oxman AD, Bjørndal A, Becerra-Posada F, et al. (2010) A framework for mandatory impact evaluation to ensure well informed public policy decisions. *The Lancet*, Elsevier Ltd, 375(9712), 427–31, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/20113827> (accessed 13 September 2011).

Parker SW (2003) Case Study: The Oportunidades program in Mexico. In: Shanghai Poverty Conference - Scaling Up Poverty Reduction, Shanghai, China.

Parker SW (2004) Mexico's Oportunidades Program. In: Shanghai Poverty Conference, Shanghai, China, World Bank, pp. 1–3, Available from: <http://info.worldbank.org/etools/docs/reducingpoverty/case/119/summary/Mexico-Oportunidades Summary.pdf>.

Parker SW, Rubalcava L and Teruel G (2008) Evaluating Conditional Schooling and Health Programs. In: Schultz TP and Strauss J (eds), *Handbook of Development Economics*, Amsterdam, Elsevier, pp. 3963–4035.

Parker SW and Teruel G (2005) Randomisation and Social Program Evaluation: The Case of ProgresA. *The ANNALS of the American Academy of Political and Social Science*, 599(1), 199–219, Available from: <http://ann.sagepub.com/cgi/doi/10.1177/0002716205274515> (accessed 5 July 2011).

Paz Lopez M de la (2007) Reseña de “Procesos domésticos y vulnerabilidad: perspectivas antropológicas de los hogares con Oportunidades” de Mercedes González de la Rocha (coord.). *Estudios Sociológicos*, XXV(3), 848–857.

Ravallion M (2009) Should the Randomistas Rule? *The Economists' Voice*, 6(2), 1–5.

Rivera JA, Sotres-alvarez D, Shamah T, et al. (2013) Impact of the Mexican Program for Education, Health, and Nutrition (ProgresA) on Rates of Growth and Anemia in Infants and Young Children: A Randomised Effectiveness Study. 291(21), 2563–2570.

Rubalcava L and Teruel G (2003) *The Effect of PROGRESA on Demographic Composition*. Mexico City.

Schultz TP (2000) *Final Report on the Impact of PROGRESA on School Enrollment*. Evaluation, Washington, DC.

SEDESOL (1999) *Mas Oportunidades para las Familias Pobres - Primeros Avances*.

Skoufias E (2000) Is PROGRESA Working? Summary of the Results of an Evaluation by IFPRI. Washington, DC.

Skoufias E (2005) PROGRESA and Its Impacts on the Welfare of Rural Households in Mexico. IFPRI.

Stecklov G, Winters P, Stampini M, et al. (2005) Do conditional cash transfers influence migration? A study using experimental data from the Mexican PROGRESA program. *Demography*, 42(4), 769–90, Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16463921>.

Stirling A, Leach M, Mehta L, et al. (2007) Empowering Designs: towards more progressive appraisal of sustainability. STEPS WORKING PAPERS, 3.

Swann GMP (2006) Putting Econometrics in its Place. Cheltenham, Edward Elgar.

Székely M (2011) Toward Results-Based Social Policy Design and Implementation. CGDEV Working Paper 249.