Norm, Mistake, or Exemplar?

A Complexity Approach to IFPRI-PROGRESA in Mexico

William N. Faulkner

Abstract:

During the early stages of implementing their new flagship anti-poverty program, Progresa, Mexican officials contracted an evaluation team from the International Food Policy Research Institute (IFPRI). This paper examines the validity of the IFPRI-Progresa process using an approach informed by 'complexity,' and exemplifies how the elaboration of multiple narratives has potential to move the evaluation community beyond the seemingly inimical to-randomize-or-not-to-randomize debates.

Keywords:
Conditional Cash Transfers; Randomized-Controlled Trials; Policy; Meta-Evaluation; Complexity Theory.
## Glossary of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>CCT</td>
<td>Conditional Cash Transfer</td>
</tr>
<tr>
<td>CONEVAL</td>
<td>Consejo Nacional de Evaluación (National Council on Evaluation)</td>
</tr>
<tr>
<td>EBP</td>
<td>Evidence Based Policy</td>
</tr>
<tr>
<td>IDB</td>
<td>Inter-American Development Bank</td>
</tr>
<tr>
<td>IFPRI</td>
<td>International Food and Policy Research Institute</td>
</tr>
<tr>
<td>LGDS</td>
<td>Ley General de Desarrollo Social (General Law of Social Development)</td>
</tr>
<tr>
<td>PAN</td>
<td>Partido Acción Nacional (National Action Party)</td>
</tr>
<tr>
<td>PRI</td>
<td>Partido Revolucionario Institucional (Institutionalized Revolutionary Party)</td>
</tr>
<tr>
<td>PROGRESA</td>
<td>Programa de Educación, Salud y Alimentación (Education, Health, and Nutrition Program)</td>
</tr>
<tr>
<td>RCT</td>
<td>Randomized-Control Trial, Randomized Controlled Trial</td>
</tr>
</tbody>
</table>
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 Randomized-Control Trial (RCT) research projects ever implemented in social program evaluation, IFPRI-Progresa provided (and continues to provide) an important proof-of-concept, 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. As part of a Master's thesis project, the author reviewed the project's original documentation, zooming in on features which could serve as bases of productive criticism. Very simply, this paper asks:

 Were the results of IFPRI-Progresa valid? If so/not, what are the implications for their wider applicability?

The analytic approach draws heavily from the fabric of ideas under the umbrella of 'complexity,' as well as the field of science studies. The methodology included a desk review and critical interrogation of official project documents, published articles, books, and statements by participants involved in the evaluation. The subsequent sections overlay the available evidence with different sets of background information, so as to (by way of metaphor) cook the same ingredients into three different dishes. From each perspective, the analysis highlights distinct facets of this massively complicated project, draws different conclusions, and conveys a different overall impression.

The takeaway message is that there is no single story which cleanly clinches the answer to either of the above questions. Instead, this research exposes how validity and external applicability are not discrete, binary properties, but instead variegated qualities which emanate from many sources and disperse unevenly through the networks connecting policy and social science research. The complexity of this process is but one of many intricacies which later endorsements of IFPRI-Progresa tend to ignore. Simultaneously, the case study displays how the complexity approach can frame and engage with these discordant narratives, pluralizing the potential pathways of action and providing the foundations of customized, appropriate, and constructive critiques, more in tune 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, p. 246)

---

i. Programa de Educación, Salud y Alimentación (Education, Health, and Nutrition Program, PROGRESA); now 'Oportunidades.'
iii. These questions are equivalent to questioning the project's 'internal' and 'external' validity.
There are some issues which definitely call into question whether the results of IFPRI-Progresa are valid even just for the sample population. I’m confused and concerned, and I can’t give a firm opinion on the validity of the results.

Boy, this evaluation doesn’t look as airtight as later authors make it sound! But I guess no evaluation is perfect. They did a pretty good job with a huge project in a difficult context.

The narrative constructed by the IFPRI-Progresa authors plays up the strong points of the evaluation and downplays the weak spots. It had to because they wanted to make RCTs look viable as policy evaluation tools.

Experiments in economics stand on shaky theoretical ground to begin with, and these guys seem to just confirm the worries in practice. The validity of the design and its implementation are still up in the air—just doesn’t seem like it in the reports! This is politics disguised as science!

If the results don’t generalize, they’re not much good to anyone. Without selling them on scalability neither Zedillo nor Fox administrations would have cared about IFPRI-Progresa or evaluation! Without translatability, nobody at the World Bank or IDB would have taken interest. This way, CCTs, EBP, and evaluation are gaining traction around the world!

Debating theory and methodology didn’t put cash into the hands of poor Mexican women. A well-timed, well-sold evaluation did. IFPRI-Progresa did a stellar job, succeeding against all odds!
Section I: Validity – 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 answer the first question regarding the internal validity of the results by describing what the raw ingredients look like before being cooked, manipulated, and overlaid with additional information. The examination of two key issues, sample attrition and contamination, demonstrates serious cause for concern about the internal validity of the findings, but with the caveat that without further evidence, their true influence cannot be known. From this perspective, no final, definitive verdict emerges.

Attrition

RCTs use panel data to measure impacts across time. Between survey rounds, some interviewees inevitably drop out of the dataset for any number 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, p. 21 footnote 9) and yet by November 2000, 17% of households and 29% of individuals from the original November 1997 sample were no longer in the survey. Moreover "attrition differ[ed] significantly between treatment and control groups, even after controlling for household characteristics and the eligibility criteria" (Rubalcava and Teruel, 2003, p. 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.

Worse, post data collection, no statistical method can effectively deal with attrition bias. Selective attrition bias is therefore an unmeasured and acutely significant factor which directly threatens the validity of the results of the IFPRI-Progresa evaluations.

Contamination

Contamination refers 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, p. 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, p. 4). As with attrition bias, the decision to not track down sample out-migrants makes any 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, p. 311)

iv. Whether the samples were similar to begin with is questionable. For further discussion see: Faulkner (2012), Section 2.2.3, p. 24.

v. The author acknowledges that the term ‘contamination’ can refer to several distinct concepts, but in this piece the word will refer to the concept defined above.
"The growth in beneficiaries likely did not go unrecognized by the control communities." (Parker and Teruel, 2005, p. 211)

Nonrandom attrition and all-but-certain contamination of the treatment communities both cast a shadow over the validity of IFPRI's results as representative of what happens when a CCT comes into contact with the Mexican rural poor. What might seem an airtight methodology has leaky seams. Nevertheless, without more information and analysis, none of these Achilles heels is sufficient to entirely undermine the validity of the results. In addition, most of the weaknesses would have been difficult if not impossible to correct in the actual implementation of the project. It may be that the IFPRI-Progressa results were simply as solid as possible given the circumstances.

Section II: Validity – The Critical Scholar’s Perspective

The decisions which guided the IFPRI-Progressa 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 a new recipe to cook (take a second look at) the base ingredients examined by Section I, but this time couching IFPRI-Progressa within the worldwide discussion surrounding 'best practices' evaluation methodologies. The goal of this section is to show how the weakened internal validity of the IFPRI-Progressa results transforms when seen as emerging within the larger system of evaluation-interested academics and researchers.

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." 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, p. 3). 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." The superposition of the economic, econometric, and experimental trends produced a wave through policy evaluation, upon whose crest rode IFPRI-Progressa. During the late 1990s, the novelty of using an experimental approach could have motivated IFPRI-Progressa authors to shape the presentation of the study design as 'free-of-caveats' as possible. RCTs are also uniquely "labor-intensive and costly," (Duflo and Kremer, 2003, p. 19) which incentivizes the extension of their results to broader contexts. The IFPRI-Progressa results had to at least appear internally valid so as to have made their production worthwhile. Indeed, in practice, advocates of the experimental design selectively place importance on the problems where it is strong in comparison to others, such as the mitigation of selection bias, while diverting the attention from potential weak points.

vi. This assertion fleshed out in more detail in Faulkner (2012), Section 3.1, p. 34.
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). The author found two citations, one simply called "Teruel and Rubalcava, 2003" and another of a mimeo titled "Attrition in PROGRESA" from 2007, but it seems unlikely that Behrman and Todd (1999) were referencing these documents which appeared years later and under the auspices of separate projects and institutions. No other IFPRI reports mention attrition whatsoever and "none of the analysis considered the possible biasing effects of attrition/migration on estimated program impacts" (Parker, Rubalcava, and Teruel, 2008, p. 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 taking 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, p. 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.

So: The interpretation of the IFPRI-Progresa results as internally valid also depends on an image highly sensitive to trends in the wider academic context. 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, p. 851). For the IFPRI-Progresa team, the struggle to implement the evaluation as an experiment thus

viii. Although there is a parenthetical reference in the text, a full citation of the document does not appear in the references of Rubalcava and Teruel (2003), p. 7-8.
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.

Section III: Validity – The Political Perspective

Section III views 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. Zooming out from the evaluation process to the entirety of the Progresa allows us to see the shadows of three hulking trends cast 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 much more positive light, and despite acute internal weakness, appears much 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) was attempting 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' effectiveness through personal contact, meetings, and media reports (Lustig, 2011, p. 9).

Due to the IFPRI evaluations, Progresa gained political hardiness and funding sources, while evaluation became a codified part of Mexican social programming. Fox not only continued the program, but expanded it significantly into urban areas (Levy, 2006, p. 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.\textsuperscript{x} 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, p. 23) and for this purpose established the Consejo Nacional de Evaluación (National Council on Evaluation, CONEVAL) (Oxman et al., 2010, p. 428).

The excitement for the new program spread not only through Mexico, but through the IDB and World Bank (Levy, 2006, p. 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, p. 353), and in 2006 the former President James D. Wolfensohn commented

\textsuperscript{x}. 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. Sources: Behrman (2007), p. 6-8; Rocha-Menocal (2005).

\textsuperscript{xi}. At the time, this was the largest loan ever granted by the IDB. Source: Parker, 2003, p. 28-29.
"Progresa’s rigorous emphasis on evaluation has set a standard for poverty reduction programs in the developing world" (Levy, 2006, p. viii). The evaluation remains one of the programs’ most celebrated features, having "inspired many others in Latin America and other regions" (Székely, 2011, p. 20).

All of these 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 applied to populations outside of the original sample. Admission that the IFPRI-Progresa findings could only vaguely inform what would happen if the program expanded to new areas would paralyze 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.

Section IV: Pathways

For all of the facts, ideas, and relationships uncovered, there seems to be no coherent, unifying narrative, no seamless truth about IFPRI-Progresa, and no single way to decisively judge the validity or applicability of the project as a whole. Sections 2, 3, and 4 have presented three 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, together producing myriads of patterns at a diffuse range scales. This opacity, however, is precisely the environment with which evaluators, academics, and politicians must come to grips in social policy evaluation. Plunging into the fog using a complexity-based approach does not provide incontrovertible truths, quick answers, or clear paths of action, but instead a plurality of information types, perspectives, and potential trajectories. The series of loops and caveats emphasized by the complexity approach is mirrored by a series of opportunities and paths of action.

- **Dialogue with evaluators on the ground:** The to-RCT-or-not-to-RCT disputes often occur at a purely theoretical level, when in reality the controversies have deep ties with local situations and individual projects. Evaluators deeply understand the origins of these arguments and viscerally experience their results.
There is no 'Experiment': Clearly, the individuality of each evaluation project makes talking about 'randomized evaluations' difficult. These entities are neither packaged nor predictable. As with IFPRI-Progresa, agency matters. Context matters. Details matter.

Think about the Results of the Results: IFPRI-Progresa did more than simply tell the world about Progresa's effects on the rural Mexican poor. The project is a historic superstar, a best-case-scenario for evaluation and a coup d'état for EBP in Mexico. Certainly the impeccable image commonly forwarded stands on fragile foundations, but perhaps it was precisely the myopia of this image which managed to jolt Mexican social policy out of the viscous cycles of political patronage and constant overturn.

Much More to Do: Over the years, experiments, particularly as a basis for EBP, have shown remarkable adaptability to different contexts across continents, and at a diversity of scales. Nevertheless, they provide no panacea and no final answer to poverty. When writing about Progresa, Levy (2006) might as well have been writing about experiments, or even social policy evaluation in general:

"Although Progresa may be an essential component of the solution, a single program cannot solve a problem that has multiple causes." (p. 20)

Conclusion

Rather than rushing towards a single, assured conclusion, this piece has attempted to express the value of shifting positions, swimming among contradictory opinions, and questioning tacit beliefs. The result, hopefully, is a gathering space, a dissonant yet respectful assemblage of opinions. Section I presented an acceptable evaluation which turned out well enough given the abilities and constraints of those involved. Section II described an evaluation with rather serious deficiencies that could undermine the legitimacy of the findings. Section III, meanwhile, sketched a more positive image of an evaluation which, despite some problems, had an amazingly positive influence both globally and locally. The study of how to construct and act upon multiple narratives must replace the oppressive focus on arriving at a single thumbs-up or thumbs-down conclusion.

The problem is that the pathway between the current state of affairs in social policy evaluation and a better one, much like the pathway between now and a Mexico without poverty and deprivation, is long, complex, and incomprehensible to any single person or from any single viewpoint. Again, just as this complexity disrupts simple, universal judgments, it also reveals that change can emerge from a variety of quarters. Individuals must foment change from many angles (policymakers, researchers, funders, students). For all of these groups, taking time to think deeply and introspectively about social policy evaluation processes (as well as methodologies) should become a basic part of the evaluation itself. There is no step-by-step formula for success. Critical thinking, impassivity, and artful articulation must lead the way forwards.
Bibliography


Greenspun, Samantha. “Assessing Mexico’s Anti-Poverty Program ‘Oportunidades’: Combining and Comparing Objective and Interview-Based Indicators”. M.A. Thesis. Tulane University, 2011.


Acknowledgments

The author gratefully acknowledges the contributions of Dr. Laura Murphy, Dr. Ludovico Feoli, and Dr. Alan Barreca who advised the thesis from which this paper was extracted. Dr. Nora Lustig also deserves recognition for providing access to IFPRI, Progresa/Oportunidades, and their communities of researchers, many of whom are her personal friends. Many thanks as well to Samantha Greenspun, whose work and ideas seeded the entire project. Finally, the author would like to express gratitude to the Stone Center for Latin American Studies, and praise the atmosphere of intellectual diversity which it represents.

Author Biography

Having recently received his M.A., William N. Faulkner is actively seeking employment in the evaluation sector. His interests include complexity theory/systems thinking, science studies, research methods, monitoring and evaluation, poverty and inequality alleviation, and international development. William endeavors, through both study and practice, to shape the globe's powerful networks of knowledge creation and diffusion into a form more beneficial to the marginalized and disenfranchised members of world society. For more information see: wnfaulkner.com. The author may be reached at wnfaulkner@gmail.com.