Workshop on Measurement of Women's Economic Empowerment

By Sonia Laszlo, Kate Grantham and Avril Rios
(McGill University)

ISID
INSTITUTE FOR THE STUDY OF INTERNATIONAL DEVELOPMENT
INSTITUT D'ÉTUDE DU DÉVELOPPEMENT INTERNATIONAL

Produced with support from McGill University and the International Development Research Centre (IDRC). Working papers are in draft form and distributed for purposes of comments and discussion only. The observations and views expressed in this working paper are the sole responsibility if the author(s).

INSTITUTE FOR THE STUDY OF INTERNATIONAL DEVELOPMENT
2018 ANNUAL CONFERENCE MEASUREMENT WORKSHOP REPORT

March 15-16, McGill University Faculty Club

By Sonia Laszlo, Kate Grantham and Avril Rios (McGill University)
Rapporteurs: Stephanie Ferrao (IDRC) and Justine Tupe (McGill University)

We are grateful for the generous support of:

Global Affairs Canada

ISID INSTITUTE FOR THE STUDY OF INTERNATIONAL DEVELOPMENT INSTITUT D'ÉTUDE DU DÉVELOPPEMENT INTERNATIONAL

McGill

IDRC | CRDI International Development Research Centre Centre de recherches pour le développement international

WILLIAM + FLORA Hewlett Foundation

UKaid from the British people

WORKSHOP ON MEASUREMENT OF WOMEN'S ECONOMIC EMPOWERMENT

Description

The Workshop on Measurement of Women's Economic Empowerment (WEE), moderated by McGill's Paola Perez-Aleman, facilitated an open discussion on the challenges of measuring WEE faced by researchers and policy makers specializing in WEE. Measuring WEE is becoming widely important given increased focus on women's and girls' empowerment in international development policy, as evidenced by Canada's Feminist International Assistance Policy. Organizations such as IDRC and Global Affairs Canada presented the interests and concerns of Canadian policy makers regarding scalability and economic efficiency. Their concerns were met with the researchers' in-depth understanding of contextual issues faced on the field. The researchers ("expert panelists") in the roundtable discussion included:

- Agnes Quisumbing from the International Food Policy Research Institute,
- Arjan de Haan from the International Development Research Centre,
- Lotus McDougal from University of California San Diego School of Medicine,
- Markus Goldstein from the World Bank
- Naila Kabeer of London School of Economics and Political Science, and,
- Susana Martinez Restrepo of Fedesarrollo and COREWOMAN

ISID Director, Sonia Laszlo, opened the workshop by giving thanks to participants and event sponsors. She then handed before things over to moderator Paola Perez-Aleman. After a round of introductions, each expert panelist spoke for five minutes on their relevant work and addressed the question: *Why have you chosen the approach that you adopt to measure WEE? What are the blind spots of this approach?* This was followed by a round of three discussion questions, chosen with input from stakeholders at McGill, IDRC and Global Affairs Canada:

1. What is the best approach to measure medium- and long-term transformative change towards WEE for development policy and practice? What are the trade-offs (effectiveness, cost-efficiency, etc.) between choosing one approach over the other?
2. Given the importance of the development context in which an intervention takes place, is it better for monitoring, accountability, and learning purposes to design a standard measurement methodology that would be valid in only a restricted subset of sectors (i.e. agriculture, private sector development, etc.), or to have a more universally applicable methodology, at the expense of being able to compare project results?
3. What would make the biggest difference for our ability to measure WEE programs and policies?

Approach and Blind Spots

Recognizing that no instrument exists to measure WEE across all countries, projects and contexts, expert panelists discussed what guides their own approach to measuring WEE and what blind spots are associated with these measures.

Agnes Quisumbing discussed the Women's Empowerment in Agriculture Index (WEAI). WEAI was developed collaboratively by IFPRI, OPHI and USAID and building on IFPRI's considerable expertise in this area. While it was inspired by Naila Kabeer's conceptualization of WEE, its focus on five domains was partly driven by USAID's project needs. Belonging to the Alkire-Foster class of measures, it has several very useful features as it is both additive and decomposable. It is especially well suited for rural or agricultural populations, and captures decision making around production, access to and control of resources, control over income, leadership in the community and time use. The blind spots of this approach are well known to its designers and users. The main problem is that it was designed for population based surveys, and focuses on agriculture and production. Moreover, it does not consider important areas of such as employment, other parts of the value chain, gender-based violence or health. IFPRI and its partners are currently developing a "Project WEAI" (pro-WEAI) which would incorporate new dimensions such as domestic violence, mobility, and intrahousehold relations. A validation exercise using psychometric measures is being undertaken in collaboration with Emory University. While the WEAI was designed to be applicable in many settings (and indeed has been), its multiple thresholds make calibrating challenging when dealing with different countries and contexts.

Naila Kabeer worked backwards from other people's attempts and drawing from Sen's capability approach. Noting that achievements are quite trivial, we need to really work on measuring agency to capture changes in power relations. One main blind spot is that current measures tend to be very focused on individuals, and yet we need to move from the individual to the collective and to the structural. For instance: What forms of collective agency? How do you measure changing norms? Kabeer notes that some existing measures do capture status in the community, but this does not necessarily have anything to do with WEE. Similarly, a second limitation is how to interpret the data? Many women (e.g. single moms) *have to be autonomous* but that's not necessarily what agency or empowerment is really about.

Lotus McDougal discussed the EMERGE group's scoping exercise. Their approach is not to propose a new measure but rather to put existing measures to scientific testing recognizing that WEE funding and programs are growing but that the science of measurement is seriously lagging. Specifically, they assess existing measures from a psychometric lens looking at multiple dimensions. Psychological, psychometric, social and health empowerment have been well researched, but economic empowerment measures or time use instruments have not been rigorously scientifically tested. The major blind spot according to her: instrument reliability (especially test/re-test) is lacking.

Susana Martinez Restrepo really homed in on the difficulties of transporting WEE instruments from one context to another and illustrated this using the concept of *Supermadres*

(Supermoms) in Latin America. In fact, picking up on an earlier observation by Naila Kabeer, many women (e.g. single moms) are already very autonomous and very "empowered" in a traditional sense within the household. They are already autonomous according to many instruments – they control the household budget, are the main decision-maker concerning children's well-being, and are important breadwinners. A large amount of empowerment programs could in fact be disempowering because women are very time poor. Lack of time and sleep have negative impacts on health and wellbeing so interventions aimed at empowering women must reflect these important time constraints. The main blind spot here lies in the difficulties in measuring subjective measures especially for women with low educational outcomes: abstract concepts are going to be especially hard to measure. This has important implications for the ability to apply a common questionnaire to different cultural contexts: to make the instruments meaningful in measuring WEE, they need to be understood by respondents which may mean that they must be framed in non-abstract terms that they can relate to in their daily life, and that, inevitably, will be context specific. Another blind spot is the researchers' own positionality.

Markus Goldstein described the World Bank's Africa Gender Innovation Lab and the work they are doing in conducting rigorous impact evaluation. He pointed out that projects are vastly different, so they must all have very different measures. The Africa Gender Innovation Lab looks across these to draw a core set of measures (about 11 topics, 50 questions total). Echoing the blind spot identified by Lotus McDougal, Markus Goldstein stressed the need to improve on the quality of the instruments and measures.

Arjan de Haan focused his remarks on the measurement of WEE in IDRC's GrOW projects. Citing the work led by Sonia Laszlo and Kate Grantham at McGill in the GrOW Research Series, 40 different papers and projects led to an equivalent number of approaches to measuring WEE. One major limitation of the work in the GrOW projects is that it wasn't clear from papers why they chose the measure they chose. One important blind spot identified in this exercise was how to handle labour force participation – echoing Naila Kabeer's and Susana Martinez Restrepo's remarks earlier, it is not clear whether these measures of achievement are adequate proxies for WEE. He also brought up the question of how WEE measures integrate with SDG measurements.

Question 1: What is the best approach to measure medium and long-term transformative change towards WEE for development policy and practice? What are the trade-offs (effectiveness, cost-efficiency, etc.) between choosing one approach over the other?

The answer to this question is complex and differs depending on what the objectives of measurement really are.

Quantitative versus qualitative approaches

The large majority of existing measures are of a quantitative nature. While expert panelists generally agreed on the importance of quantitative measurement and the need to get that measurement right, they all pointed out the need to promote the use of qualitative approaches. Context matters. Unfortunately, much of the context is lost in a great deal of quantitative instruments. Life histories are especially important in this context. Getting at changes in norms, for instance, is more easily done via qualitative than quantitative approaches. The need to ensure that voice and choice are captured may also be best handled using qualitative methods. Panelists agreed that a mixed-methods approach to measurement of WEE is ideal, and that quantitative should be paired with qualitative work.

Multi-disciplinarity and local partnerships

Panelists discussed the merits of multidisciplinary work. Collaborating in multidisciplinary teams and with numerous partners is viewed as a generally positive thing to advance the measurement of WEE, though challenges certainly exist. For instance, there are tensions between consensus and comparability in establishing a common research agenda. Collaboration requires asking common questions across different contexts, and this is very difficult because everyone has their own set of priorities. Academics and practitioners also have very diverse sets of incentives driving the work they do, and this is seen by the tension between publishing in top economics journals, which often do not do justice to qualitative work. Similarly, there is also a desire and need to collaborate with local organizations and research centres in the Global South, but these often have very different skill levels and true collaboration may require a deliberate commitment to capacity-building of Southern partners.

Scope versus scale

Interventions and projects require some degree of a rigorous evidence base, making impact evaluations indispensable. Because these must consider a range of outcomes, both intended and unintended, to address possible spillover effects, impact evaluations require surveys with large sample sizes and a large control group. These are expensive to run, and good criteria to select interventions for evaluation is when they have potential for high impact or widespread interventions for which there is limited rigorous evidence (e.g. microfinance).

Time horizon

Most evaluations tend to measure impact over a very short time horizon. The typical impact evaluation in development covers 3 to 5 years from baseline to follow-up. More often than not, this is because of constraints imposed by funding agencies. Few donors or funders are in it “for the long game”. Yet the long-term effects of intervention may be just as, if not more important that the short-term impacts. This is especially relevant for those interventions that may have important spillover effects. Technology (such as using mobile phones to conduct simple surveys) however may help facilitate lower cost follow-ups on narrowly defined outcomes.

National versus international

There is a clear sense emerging from the panel that measurement needs will be very different for national or project purposes than for international comparisons. National and project needs are likely dependent on local context and so will need to be specific to that context. Impact evaluations combined with qualitative methods tend to be preferred in these contexts. Meanwhile, for international comparisons (such as might be relevant for tracking progress towards the SDGs) to be meaningful then survey data using a relatively small core set of questions would be preferable.

Question 2: Given the importance of the development context in which an intervention takes place, is it better for monitoring, accountability, and learning purposes to design a standard measurement methodology that would be valid in only a restricted subset of sectors (i.e. agriculture, private sector development, etc.), or to have a more universally applicable methodology, at the expense of being able to compare project results?

Local context matters so how do we move from the national to the international level?

Picking up on the last point in the previous section, we should discourage using a heavy core set of questions to go across different projects. The problem is that getting a meaningful "slim core" set of measures is difficult to get, and it may even be a "myth" to think that it is possible. Questionnaires are too long, leading some to be concerned about respondent fatigue and hence quality of the data collected. Once you trim the questionnaire to get to such a 'slim core', there will always be tensions to add a module here or there. Also, with WEAI's class of measures, by trimming the questions to make it parsimonious, you lose what is really interesting in the project. For projects that are addressing very different aspects or domains of WEE, you really do want different sets of indicators. With the pro-WEAI, IFPRI is piloting a relatively slim core set of measures, but with optional, validated modules as well to address the local context or project specific needs.

Question 3: What would make the biggest difference for our ability to measure WEE programs and policies?

There is a general agreement that we need to improve measurement across the board, but that there are also different types of measurement to address different policy needs:

1. Diagnostic needs: Measures like WEAI are good diagnostic tools because they cast a wide net, but contextual barriers need to be addressed in order to meet policy needs.
2. Monitoring needs: To catch the different points in logic framework, measures and outcomes will be very specific to interventions and planned objectives.

3. Learning needs: Impact evaluations are key here for capturing the "out of the box" effects of projects as they pertain to women's local and subjective experiences of empowerment. These measures need to be high quality, but not necessarily standardized in the same way as diagnostic or monitoring tools.

Key Lessons

- Ultimately, no "silver bullet" or single approach or instrument exists to measure WEE effectively across all countries, projects and contexts – indeed, some measures make sense in certain contexts but can produce drastically different, even misleading results in other contexts (e.g. *Supermadres* (Supermoms) in Latin America).

- An additional challenge is that empowerment in one area can lead to disempowerment in another and it is difficult for measures to account for this (e.g. increased access to jobs can lead to time poverty for women due to child care burdens and familial expectations).

- Trade-offs inevitably arise when choosing one measurement approach over another (e.g. scope vs. scale; effectiveness vs. cost-efficiency; data specificity (context) vs. comparability; individual- vs. community- or national-level analysis)

- Recognizing these challenges and tradeoffs, several recommendations emerged from the workshop with regard to measurement of WEE:

    - In developing WEE measurement frameworks, a clear definition and program objectives should be defined early on to help with defining good, relevant indicators.

    - Mixed-methods approaches (combining quantitative and qualitative work) to measure WEE are useful for balancing tradeoffs in comparability of data versus attention to local context, and in subjective versus objective data collection.

    - One strategy to measure WEE across multiple projects, countries and contexts is to develop a set of several "core measures" (ideally based on existing, validated indices) that are used across the board, and then supplement these with more context-specific measures identified on a project-to-project basis.
        - Some examples of core measures might include: household decision-making power, attitudes about violence again women; freedom of mobility; legal rights to property and inheritance; access to employment opportunities).

    - Adding to the previous point, respondent fatigue is a real problem, and in order to ensure high quality data collection a "slim" set of core measures (no more than 20-25) is ideal. For longer surveys, researchers should start with core measures and subjective questions (about women's own experiences of empowerment) and end with supplementary measures and more standardized sets of questions.

    - Measurement frameworks should focus on providing broad diagnostic and learning tools as well as additional frameworks to support monitoring processes.

- Moving forward, it was suggested that a cross-disciplinary review of common WEE measures used across contexts would provide a useful mapping exercise to see if some standard or comparability exists. For an example of this type of mapping exercise undertaken for IDRC's GrOW program, see Laszlo et al. (2017).

Expert Panelists

Agnes Quisumbing, International Food Policy Research Institute

Markus Goldstein, World Bank

Arjan de Haan, International Development Research Centre

Naila Kabeer, London School of Economics and Political Science

Lotus McDougal, University of California San Diego School of Medicine

Paola Perez-Aleman, McGill University (Moderator)

Susana Martinez-Restrepo Fedesarrollo

APPENDIX A: WORKSHOP PARTICIPANTS (ALPHABETICAL)

Name	Affiliation
Aylward, Michael	Global Affairs Canada
Brandt, Nicole	Global Affairs Canada
Diaz-Martin, Lucia	The Abdul Latif Jameel Poverty Action Lab
De Haan, Arjan	International Development Research Centre
Dowie, Gillian	International Development Research Centre
Dubé, François-Philippe	Global Affairs Canada
Ferrao, Stephanie	International Development Research Centre
Goldstein, Markus	World Bank
Grantham, Kate	McGill University
Housseini, Bouba	International Development Research Centre
Kabeer, Naila	London School of Economics and Political Science
Laszlo, Sonia	McGill University
Martínez-Restrepo, Susana	Fedesarrollo
Max, Kerry	Global Affairs Canada
McDougal, Lotus	University of California San Diego School of Medicine
Perez-Aleman, Paola	McGill University
Rios Torres, Avril	McGill University
Quisumbing, Agnes	International Food Policy Research Institute

APPENDIX B: ANALYTICS

A live one-way feed of the workshop proceedings was made available to selected partners and stakeholders at the following link: https://youtu.be/ls1Ct44YlRM. As of the end of June 2018, this livestream recording has been viewed more than 95 times. ISID is now in the process of editing this recording to include participants' titles and to improve overall production quality. Once this process is complete a new recording of the workshop will be uploaded onto ISID's YouTube page and made publically accessible.