Introduction

Gender mainstreaming entered the lexicon of the international development community at a time of high optimism. Support for this strategy to advance gender equality gained ground during the decade between the UN World Conferences on Women in Nairobi in 1985 and in Beijing in 1995. The term was formally coined and embraced in the Beijing Platform for Action (UN Women 1995, pp. 79–229). It is a transnational strategy, not rooted within a national or country framework (Walby 2005, pp. 453–54), and the aim is to advance gender equality by amending all mainstream policy arenas.

Optimism in world development—especially among the Western countries that saw the victory of their system over that of the Soviet Union—was combined with a high level of ambition. The strategy to redress women’s unequal position in 12 critical policy areas was proposed in a uniform way in every context (Charlesworth 2005), bringing a gender perspective into any planned action so that the effects for women and men are known in advance and can be addressed accordingly.

Twenty years later, the tone is quite different. “Overall ‘fatigue’ in gender mainstreaming” calls for meetings to reverse the tide (UN Women 2013). There is an “elephant in the room, and dragons at the gate”—and a revision of the strategy to deal with gender discrimination is urgently needed (Sandler and Rao 2012).

In this article I analyze and compare the findings of evaluation studies on gender equality over the past 20 years. Not one study has reported positively about the gender mainstreaming policy. The majority of the evaluations have been preoccupied with corporate policies, institutional processes, and procedures, while results for women have been scantily addressed. I will argue that this is both

It is time to say goodbye to gender mainstreaming, and to move to focused actions toward equality for women, supported by evaluations that matter.
due to the framing of the strategy of gender mainstreaming and to the failure to properly apply the instrument of evaluation. It is time to say goodbye to gender mainstreaming, and to move to focused actions toward equality for women, supported by evaluations that matter.

**Evaluation as a tool in gender mainstreaming**

After Beijing, gender mainstreaming was adopted and applied by nearly every multilateral and bilateral donor agency, government, and nongovernmental organization. Their policy documents on gender equality quote or reflect the definition of gender mainstreaming adopted by the United Nations Economic and Social Council (ECOSOC) high-level panel discussion in 1997:

Mainstreaming a gender perspective is the process of assessing the implications for women and men of any planned action, including legislation, policies or programs, in all areas and at all levels. It is a strategy for making women’s as well as men’s concerns and experiences an integral dimension of the design, implementation, monitoring and evaluation of all policies and programs in all political, economic and societal spheres so that women and men benefit equally and inequality is not perpetuated. The ultimate goal is to achieve gender equality (UN General Assembly 1997).

By presenting monitoring and evaluation as a key aspect of the interventions, the ECOSOC definition acknowledges the potential contribution of evaluation studies “to achieve gender equality.” Over the years, evaluations have been forthcoming, and the reports offer material to research what their contribution has been.

**Gender mainstreaming in evaluations around 2000**

Early 2000s. I coauthored the Organization for Economic Cooperation and Development/Development Assistance Committee (OECD/DAC)¹ Review of Gender and Evaluation (Hunt and Brouwers 2003), a meta-study undertaken to find out how the gender perspective and results for women are addressed in evaluations of development policy and practice. It is the third review conducted by the OECD/DAC; earlier studies were done in 1994 and 1999.

A main conclusion of the 2003 review was that “overall progress in incorporating gender perspectives into general evaluations² of development assistance has been slow and uneven” (Hunt and Brouwers 2003, p. 10). The major findings of the thematic-type evaluations were that: (i) gender mainstreaming is treated as a goal rather than as a means to the long-term objective of gender equality; (ii) the assumption that gender mainstreaming leads to gender equality outcomes prevails, but is untested; (iii) the word “gender” is generally applied as a synonym for “women,” with little or no focus on gender relations or on the impact of development activities on gender equality; and iv) most of the gender benefits identified deal with women’s practical needs, not their strategic interests.

The quality of both types of evaluations is disappointing. Lack of sex-disaggregated baseline data, limited monitoring information, and lack of gender-sensitive indicators have hindered the task of proper research. Many evaluations are conducted with little attention to and involvement of stakeholders in the programs and projects. Two-thirds of the general evaluations lack the required gender expertise on the evaluation team. Compared with the studies of 1994 and 1999, there is not much progress in the depth

---

¹ The reviews had been initiated by the OECD/DAC Working Party on Aid Evaluation in collaboration with the Working Party on Gender.

² The OECD/DAC review of 2003 distinguishes between i) general evaluations, evaluations of regular development interventions, which include some gender analysis or references to gender issues but are not specifically designed to evaluate gender equality policy or initiatives; ii) thematic evaluations, evaluations of gender- specific policy or interventions, designed to evaluate gender equality, gender mainstreaming and women’s empowerment efforts.
The design of the studies

The 21 evaluation studies hinge on gender mainstreaming, the policy adopted by all agencies in line with the agreements of the Beijing Conference. All studies have been conducted by or under the auspices of the agency’s official evaluation department, and they essentially apply the conventional approach of documentary study and interviews. Their focus is on the relevance of the policy, the commitment by the agency’s management and staff, and the implementation of the strategy. The last refers to the extent to which gender has become part and parcel of the mainstream policy, programs, and procedures. Although most evaluations include country case studies, these deal primarily with the question of whether gender issues are reflected in strategies and project design.

The most common interpretation of the gender mainstreaming strategy is the twin-track approach or dual strategy, meaning, on the one hand, that a gender focus should be mainstreamed in all aspects of the organization: themes, sectors, activities, and institutions (ECOSOC definition of 1997), and, on the other hand, that women should be supported directly by stand-alone projects. Not one of the evaluations was designed to differentiate between the results of interventions on these two very different tracks.

With a few exceptions, the evaluations have not investigated money flows. Money is frequently mentioned in terms of “lack of,” “decreased,” or “woefully inadequate,” but it is seldom studied. The few evaluations that deal properly with finances

3 For the identification of evaluations, I consulted the Websites of the OECD/DAC and of the donor agencies, and compared my list of studies with the list of the 2011 meta-analysis of gender evaluations by the African Development Bank. While I may have missed one or two evaluations, I’m confident to have found the most important ones.

4 The bilateral studies are by Norad (three studies), DFID, BMZ, CIDA, SDC, Irish Aid, Sida, ADC; the multilateral studies are by UNDP, UNICEF, WFP, ADB (two studies), WorldBank; the studies of the funds are from Unifem, Danida; AusAid produced the monitoring report; the meta studies are from Norad and AfDB. For a full overview see ISS Working Paper 556 (Brouwers 2013)
provide the percentages of total official development assistance (ODA) that can be attributed to gender equality programs, noting that the figures are not very reliable. Their basis is usually the gender marker system developed by the OECD/DAC.\(^5\) Any failure or inconsistency in financial recording by the responsible units makes the ultimate report unreliable. The UK’s Department for International Development (DFID) evaluation (2006), the most elaborate on finances of all the evaluations in the sample, explains the problem: financial figures concern the whole project in which gender will be mainstreamed, the figures are registered at the design stage, but how much of the budget is actually spent for the promotion of gender equality is never separated out. Gender equality activities in smaller projects may not be included at all. This clarifies the wide range reported—that overall spending increased “consistently over the period from 5 percent in 1995/1996 to 32 percent in 2004/2005” (DFID 2006, annex 4).

A major methodological flaw is that results go unreported. The evaluations offer no insight into what has changed, how women and girls have benefited, or if gender relations have become more equal. Reports may announce a focus on program results, but hardly go beyond assessing the integration of gender dimensions in country programs, project design, implementation, monitoring, and evaluation. Various studies acknowledge this shortcoming and explain the nature of the problems encountered: absence of data, unprocessed data, or meaningless data.\(^6\)

Overall, the evaluations fall short in their design: the focus is confined to policy, strategies, and procedures, no differentiation is made between the two tracks of the mainstreaming policy, there is limited interest in money flows, and there are few efforts to investigate results and benefits.

**What do the evaluations tell us?**

The emerging picture can be summarized by a quote from the most recent evaluation in the sample: “The concept of gender mainstreaming may not require lobbying anymore, walking the talk does” (ADC 2012, p. 4). A main conclusion of every report is the gap between policy and implementation. The policy is found to be inconsistent, ambiguous, and confusing; invisible and unclear; not implemented systematically. Not one of the agencies observed the ECOSOC definition of gender mainstreaming in “all policies and programs, all areas and levels.” Gender performance is best in the areas of education, health, nutrition, water supply and sanitation, population, agriculture, and rural development, and worst in the sectors of infrastructure, transport, energy, program loans, and private sector development.

The required institutional reforms have not taken place, there are no accountability mechanisms, there is lack of authority behind the policy, and staff has little appreciation for the policy, do not know it, or do not know how to apply it. In the mainstreaming philosophy, gender is everybody’s task, but in practice, it has become nobody’s responsibility.

---

5 An activity can either target gender equality as its “principal objective” or as a “significant objective.” A “principal” score (2) is assigned if gender equality was an explicit objective of the activity and fundamental to its design—that is, the activity would not have been undertaken without this objective. A “significant” score (1) is assigned if gender equality was an important, but secondary, objective of the activity—meaning it was not the principal reason for undertaking the activity. A “not targeted” score (0) is assigned if an activity is not found to target gender equality. See OECD/DAC (2012), Aid in Support of Gender Equality and Women’s Empowerment, Statistical Overview, www.oecd.org/dac/stats/gender

6 For instance, the only programmatic result on gender equality reported in relation to the roads program in the 2006 Mid-Term Review of the Sida Country Strategy (Kenya) was that the quota for women’s participation had been mainstreamed and incorporated into the program’s Logical Framework.” (Sida 2010:33)
On a more positive note, the gender analysis needed for an informed practice has increased in country strategies. However, due to low quality or lack of focus, an analysis at the country level seldom affects programmatic choices. The experiences at the project level are better. Due to a proper analysis and incorporation of gender issues, two-thirds of the World Bank’s gender-integrated projects deliver substantial or high outcomes, against 22 percent of the nonintegrated projects (IEG 2010).

A unanimous finding concerns the weakness of documentation, monitoring, and evaluation. There are no baseline data and few sex-disaggregated data, monitoring forms may be filled out but not used, and the quality of the files is low. Less than half of the completion reports deal meaningfully with gender outcomes, so the benefits remain largely unknown. Most of the good gender results encountered are of an anecdotal nature. The better studies about project achievements are the evaluations of DFID, the World Bank, the Asian Development Bank, and the Swedish International Development Cooperation Agency (Sida), but these also leave the progress toward gender equality largely untouched.

For a summary of the findings of the current study, we can simply go back to those of the 2003 OECD/DAC review, as summarized above: i) gender mainstreaming is still treated as a goal rather than as a means to the long-term objective of gender equality; ii) the assumption that gender mainstreaming leads to gender equality outcomes still prevails and remains unproven; iii) the word “gender” continues to be applied as a synonym for “women,” and there is little or no focus on gender relations or on the impact of development activities on gender equality; and iv) most of the gender benefits identified deal with women’s practical needs, not their strategic interests.

The double failure: the mainstreaming strategy and its evaluation

With not one example of success in the evaluations produced over the past two decades, the unequivocal message is that gender mainstreaming has failed.

The strategy remains a promise of how gender discrimination can be ended, but it exists in theory only. “Too much theory to practice,” notes the AfDB in its evaluation synthesis (AfDB 2011, p. 72). In a comparison between the Women in Development theory and the more encompassing Gender and Development theory, which calls for detailed social analysis and fundamental changes in power structures, the AfDB study concludes that donor organizations have exacerbated the situation. The theory of Gender and Development, notably that once power changes become entrenched in the thinking and actions of all stakeholders, the change of social, political, and economic processes would follow in due course, is void, since “the theory has not been widely adopted in the development field” (AfDB 2011, p. 74).

Ironically, the paradigms of gender and gender mainstreaming itself have hindered its implementation. The factors of failure include: the mystifying language of gender, often not understood in other languages; the inherent ambition that a proper application of the approach will transform fundamental power relations between the sexes; and the all-encompassing definition of gender mainstreaming to cover “all areas and at all levels.”

My study confirms the AfDB’s conclusion that evaluation has also failed. Not only has evaluation been remiss in the failure to evaluate and the failure to respond to evaluation findings (AfDB 2011, p. 76), but the preoccupation with corporate policy and strategy of all evaluations has prevented the generation of knowledge about what is happening in the real world. That is where the failure of evaluation hits hardest. For two decades, international development agencies have conducted the same kind of evaluation studies, resulting in identical conclusions, and leading to identical recommendations. How
much repetition is needed? By failing to address the strategy in practice, the evaluation studies have not contributed to knowledge about what works and what doesn’t work. The failure of the evaluation studies and the failure of gender mainstreaming are two sides of the same coin: the result of the corporate approach to gender equality. The failure of evaluations is inherent to the concept of gender mainstreaming.

**What next?**

A contested strategy that has not been able to evoke results for women’s equality and that has reached a state of “overall fatigue” should be waved goodbye. UN Women clearly states the case: “The danger going forward is that the window of opportunity to achieve gender equality is closing as a result of mainstreaming fatigue (UN Women 2013).

Ideas of gender mainstreaming are the product of the 1990s. The concept reflects not only the optimism of the time, but also the political, economic, and social forces of that era. More than a decade into the twenty-first century, conditions have changed. Some changes have been for the better, with progress in female education, life expectancy, labor force participation, and formal rights (World Bank 2012), and with the idea of equality for women a more widely accepted concept. But other changes have been for the worse, like the persistence of high levels of violence against women, the alarming levels of maternal mortality in Sub-Saharan Africa and South Asia, and the increase in fundamentalism denying women fundamental human rights. Elephants and dragons abound.

The world has also changed with respect to institutions—their centrality is fading, and they matter less to people. Nowadays, people are making their way around institutions that have so often let them down, be they financial (the banks), political (politicians and parties), economic (corporations), or religious (churches). This mood will have its impact on institutions for international development and implies that the potential of the institutions to change people’s lives (as through gender mainstreaming) is not the same as it was two decades ago.

Future international development efforts for women’s equality can improve by:

- A change in terminology: The very term gender mainstreaming implies an institutional focus, which has failed and is outdated. Why not replace it with a term that points to what is to be attained—such as equal rights for women—and that is less difficult to translate into other languages than gender mainstreaming, and better understood worldwide.

- A change in focus: support direct action in context. Local involvement and locally formulated priorities are crucial. The evidence in the review that the best progress is made in countries with sympathy for women’s equality and that the “type of aid matters,” with best results in sectors such as education, calls on donors to be selective in their choice of countries and sectors.

- Priority to areas that matter most: areas that are key for women’s empowerment such as education, income, and voice, or that affect women specifically, such as reproductive health and rights, violence against women, and women in conflict areas. The areas selected for the Goal 2 (Empower girls and women and achieve gender equality) of the post-2015 agenda are promising: a) prevent and eliminate all forms of violence against girls and women, b) end child marriage, c) ensure the equal right of women to own and inherit property, and d) eliminate discrimination against women in political, economic, and public life (UN 2013—The report does not once mention the term gender mainstreaming!).

- Results-oriented evaluations: based on concrete data, applying theories of change, seeking evidence. The 21 evaluations should be the last ones that complain about lack of data. This is an archaic statement in a digital age, where most of our personal data and doings are registered and, if so
desired, controlled. The current data revolution can, if properly applied, be of immense support to future evaluations (UN Global Pulse 2012). For a subject such as women’s equality, which evokes such strong emotions, it is all the more important to have good data about performance, progress, and failure. The challenge has been issued to collect these data and to find methods that go beyond the measurements of efforts, activities, and outputs, and that substantiate what has been achieved and what is working. It is time to go ahead with evaluations for women’s equality that deal with real life issues instead of corporate misunderstandings.

7 See, for example, UN Global Pulse, Big Data for Development Challenges and Opportunities, May 2012. The term big data refers to digitally generated data of various sources that may be turned into actionable information, and maybe relevant to supplement limited and often unreliable traditional data like official statistics and survey data.

References

A complete list of the 21 studies reviewed by the author can be found in Table 1 of Working paper 556, April 2013.


EVALUATION Matters


Profile of the Author

Ria Brouwers has been involved in the Netherlands’ policy for women’s rights and gender equality in international development since the policy started in the 1980s. She worked as researcher, advisor and evaluator for the Netherlands government and for Dutch private development organizations; and assisted other bilateral donors, the OECD/DAC and various multilateral agencies. As senior lecturer, International Development Policy at the International Institute of Social Studies (ISS) in The Hague, she combined such external assignments with teaching, making the practical experiences beneficial for teaching in evaluation of development policy and programmes. Besides women’s rights and gender equality, her areas of interest are international development policy, research in development, civil society and theory and practice of evaluation. Ria retired from the ISS in 2012 but pursues here interest in these fields through her firm Bravo.