Journal of MultiDisciplinary Evaluation
Volume 16, Issue 34, 2020

JMDE
Journal of MultiDisciplinary Evaluation

ISSN 1556-8180
http://www.jmde.com

Equity Implications Evaluating Development Aid: The Italian Case

Marina Forestieri
University of Messina
Laboratorio di Valutazione democratica-Roma

Background: In the field of development aid, social equity is an emerging issue that concerns the evaluation community in its theoretical and practical dimensions.

Purpose: A widely held belief is that evaluators do not apply theory. In this paper, we intend to verify this statement about equity in the field of cooperation projects.

Setting: Not applicable.

Intervention: Not applicable.

Research Design: We considered equity-focused approaches and found three common factors: stakeholder participation, attention to context, and focus on marginalized groups. These elements operate as screening criteria in identifying equity issues in a case study.

Data Collection and Analysis: The paper examines a practical experience of Italian cooperation. This involved a review of evaluations reports completed between 2013 and 2014. The reports are analyzed according to the three screening criteria.

Findings: The use of the three criteria has proved its worth in grasping the issues of equity neglected and often not recognized in reports. Once again a gap emerges between theory and practice. The availability of theoretical approaches is not sufficient. The paper, therefore, proposes a reflection on the responsibility of evaluation towards social justice.

Keywords: *social equity; development aid; theory and practice gap.*

Introduction

In the area of development aid, the demand for evaluation has been oriented by the debate on the effectiveness of aid (Solheim, 2012). Evaluation has been increasingly required to provide rigorous evidence that expenditure in aid is warranted, focusing in particular on impact (CGD, 2006). The impact to be evaluated has always referred to poverty reduction, which was the main objective of interventions. However, efforts to reduce poverty were not successful if they did not also address inequalities (UNRISD, 2010; UNICEF et al. 2013). In the broad debate on the development agenda, the lack of attention to issues of equality and the scarce commitment to the most marginalized and vulnerable groups have been recognized as being a limitation of the Millennium Development Goals (MDGs) (Kabeer, 2010; UN Task Team, 2012). A broader interpretation of poverty has been provided in terms of economic resources and access to services and resources. This multidimensional view of poverty is therefore consistent with the Human Rights-Based Approach (HRBA) (OHCHR, 2004). The focus has been shifted from economic growth to social equity, recognizing its function in development. It has been observed that greater equity offers more opportunities for worst-off groups, furthermore promoting poverty reduction (World Bank, 2006).

This new understanding of social equity gained through the various international initiatives and the review of experiences (UNDP et al., 2011) has obviously affected the evaluation process as well: it was necessary to look beyond simple economic and quantitative measurements (Poverty Analysis Discussion Group, 2012). A decisive contribution to assessing and considering Human Rights and Gender Equality (HR & GE) has developed within the United Nations, where, in addition to the various initiatives and publications by UNICEF (Bamberger et al., 2011; Segone, 2012a), a specific organization, the United Nations Evaluation Group (UNEG), has been considered for this purpose.

The focus on effectiveness has shifted from "aid effectiveness" to "development effectiveness" (Stern, 2008; De Toma, 2010). The desired results are the improvements (not only the economic, but also social) that can increase human development. It is important to know whether money has been properly spent, not only from the donors' point of view, but also from that of the final recipients, and to verify to what extent results contributed to reducing inequities (Peersman, 2016, p. 38). This has become a specific task of evaluation.

Purpose

A widely-held belief is that evaluators do not apply theory. In this paper, we intend to verify this statement about equity. Given the increasing importance of the demand for evaluation in the field of development aid, where social equity is an emerging issue, and considering that there are several appropriate theoretical approaches, many evaluations should prove to be sensitive to equity.

The European Union (EU) is the world's largest development aid donor. It is committed to eradicating poverty and promotes good governance for human and economic development in partner countries. The principles of EU development cooperation are embedded in the 2005 European Consensus on Development. In the context of poverty eradication, the EU aims to prevent social exclusion and to combat discrimination against all groups, while improving people's lives in line with the MDGs and promoting gender equality and equity (Joint Statement, 2005). That is why every European cooperation project that aims to improve a given aspect, albeit marginal, of a country's reality, also needs to make reference—even if only implicitly—to the improvement of the conditions of well-being. Therefore, the evaluation of a project must examine the results in terms of equity, because the effects are inevitably different for those with different situations.

As indicated by Ofir and Kumar, each program must be evaluated by examining its contribution to improving the quality of life, promoting equity, and reducing discrimination (Ofir et al., 2013, p.14). The research hypothesis proposed here is that all evaluations of cooperation projects have implications for equity and that it is possible and useful to find them.

The first part of this paper discusses the theory and considers evaluation approaches particularly appropriate for equity, and the indications are provided to formulate an analysis framework for research. The second part addresses the practical aspects and examines evaluation reports based on three criteria: stakeholder participation, context, and marginalized groups. The work ends by stressing the role that the evaluation community, theorists and practitioners, can play in promoting equity in the development process.

Theory

Evaluation Approaches

There are many approaches sensitive to equity issues. Some refer to particular categories of discrimination, such as gender (Seigart, 2005; Podems, 2010), ethnicity (La France et al. 2010; Bowman et al. 2015) and race (Dean-Coffey, 2018; CEI et al., 2017). Others, though having a comprehensive approach to all possible forms of discrimination, such as the human rights approach or the equity-focused evaluation, do not have a specific evaluation framework (Segone, 2012b, p. 10; D'Hollander et al., 2014, p. 79).

Other types of evaluation, namely democratic, empowerment, responsive, transformative, and participatory, have a specific framework attributable to one or more theorists, and, at the same time, address all situations of inequity. These approaches are the most notable for the study at hand for two reasons. First, they are particularly used for development aid because their roots are often in developing countries, so much so that the theory has been developed alongside experience in the field. The second and more relevant reason for this choice is that all the approaches are particularly geared toward equity.

The democratic approach as a whole is consistent with the need to contrast situations of inequity. Moreover, the first theorist of this approach, Barry MacDonald, affirmed the value of an evaluation that distinguishes itself from autocratic and bureaucratic models. An evaluation qualifies as democratic because it serves the whole community (MacDonald, 1974, 1978). House, another democratic theorist, later asserted that this choice is an application of Rawls' conception of social justice (House, 2004, p. 223). As regards to empowerment evaluation, evaluation has an instrumental role in solving problems from a social justice perspective (Fetterman, 1996a, p. 383). Participatory approaches have qualified as processes aimed at the issues of political and social justice (Cornwall et al., 2001). Explicit indications to overcome inequalities regarding social structures, situations of privilege, and power have also been provided through transformative (Mertens, 2007) and responsive evaluations (Hood, 2001, 2004; Stake, 2004a; Abma, 2006).

Aspects Common to the Approaches

The aforementioned approaches share three relevant elements for equitable development:

- stakeholder participation
- context
- marginalized groups

Stakeholder participation. Stakeholder participation has become a central theme in development aid since the 1970s when the attention shifted from donors to other stakeholder groups, as a reaction to the "top-down" approach to development aid that neglected local needs (Mansuri et al., 2013). This shift has spread participatory approaches (Cullen et al., 2011a) that are alternative and people-centered. These methods, applied in most countries in the world (Chambers, 1995), have allowed ordinary people to determine their priorities and improved efficiency and effectiveness. For these reasons, aid agencies and development organizations have included the participatory approach in their evaluation policies (OECD/DAC, 1997; Estrella, 2000). This empowers local people and ensures that evaluations address equity (WHO, 2013, p.5). In practice, these methods share the principle of stakeholder involvement (Cousins et al., 1998) by continuously reviewing who is benefiting from a project and who is not, thus enhancing the equity of outcomes (Hilhorst et al., 2006, p. 20).

More recently, the Most Significant Change and Outcome Mapping have also been advancing. The former is divided into various phases in which stakeholders play a crucial role in the process ranging from the presentation to the selection of the more meaningful stories for change and then discuss the reasons, choices, and criteria (Davies et al., 2005). The latter goes beyond the traditional cause-effect logic and enhances the dialectical input of stakeholders in identifying the factors of change (Earl et al., 2001). In addition, in these approaches, there is a profile characterizing the participatory methods, that is, change is identified through the experiences and indications of the stakeholders, based on the assumption that the beneficiaries, as agents of their own development, also play an active role in the evaluation (Rogers, 2012a, p. 146). Stakeholder involvement has become an acknowledged evaluation practice (Bryson et al., 2011, p. 3). Broadly speaking, engagement should be seen widely ensured in two aspects: the evaluation phases and the categories of stakeholders. Responsive and transformative evaluations are in line with this formulation.

In 1975, responsive evaluation overturned the traditional vision that first examined the objectives, placing the points of view of all the various stakeholders at the forefront (Stake, 1975). For the democratic evaluation, "the involvement of all those who have an interest" is not only one of the characterizing aspects, but is “the fundamental principle” (House et al, 2000a); therefore, the commitment of evaluators is to ensure that the interests of all the stakeholder groups involved are represented (House et al. 2000b, p. 5). Finally, with the empowerment approach, all the stakeholders share and build the evaluation tools with the evaluator and acquire independence in decision-making (Fetterman, 1996b, p. 4). This approach is useful in marginalized communities, where involving recipients is essential.

Context. Each evaluation unfolds within an environment comprising multiple dimensions, namely, social, economic, and cultural, and this can influence the intervention. All the approaches considered have highlighted the importance of context. The environment in which the project is realized is not regarded as a separate element, but as the arena in which the change is pursued; this, according to Greene, is even more evident in the participatory and democratic approaches (Greene, 2005, p. 83). The participatory approaches are based on the principles of inclusion, emancipation, and democratization of power (Cousins et al., 1998), favor the contribution of the different stakeholders and enhance the local community, ensuring a better understanding of context. The democratic approaches also consider the various aspects that influence the context: the elements of time and space, the socio-political structures, and the value system; to pursue its ideal in any case, an evaluation must consider all these situations that bind it inextricably in practical application (House et al., 2000b, p. 3). This indication also applies to the transformative perspective that stresses the need to capture “the complexity of the context” (Mertens, 2010, p. 7) and identify, through careful analysis, cultural differences and asymmetries of power (Mertens, 2016, p. 105).

In the remaining approaches, however, the focus on the context is not specifically aimed at favoring social justice, but rather at obtaining a thorough understanding of needs and structures. This is precisely the path of responsive evaluation, whose first theorist, Stake, has repeatedly stressed the need to root the evaluation in the local reality because only through the context is it possible to interpret the results and understand their actual meaning (Stake, 2004b, p. 173). Evaluators should immerse themselves in the context to grasp its various aspects, to understand the interests and even the language of its interlocutors (Stake, 1975). By paying attention to the cultural aspects, the perspectives of the participants are better understood; therefore, the evaluation can more easily grasp the needs of the environment in which it operates (Hood, 2005, p. 98). The cultural profiles of context have subsequently characterized this approach, which requires constant and specific attention to cover the entire evaluation process (Frierson et al., 2010). Finally, in the vision of empowerment, the commitment to be in tune with the community's cultural and political context is considered to be the element that

differentiates this approach from traditional evaluations (Fetterman, 1996b, p. 32).

Marginalized groups. Another element common to the approaches considered is the attention paid to the poorest and most marginalized people. This is consistent with their use in development programs because the approaches are designed to reduce inequalities and intervene in situations where beneficiaries are regarded as worst-off groups. These include the disadvantaged, vulnerable and the marginalized groups (Schwandt, 2015, p. 52).

Already in the 1990s, the democratic approach—observing that evaluations often neglected the interests of minority groups—recommended holding them in "special consideration" (House, 1993, p. 157). Similar attention is given by the empowerment approach, which aims to increase people's capacity and prioritizes disenfranchised people: minorities, people with disabilities, and women (Fetterman, 1994, p. 12). In the transformative perspective, the excluded groups share situations that limit their access in terms of social justice (Mertens, 2012, p. 27). This approach is concerned precisely with those stakeholders whose voice would be neglected in the evaluation, interpretation and use of the results. It aims to include them by respecting their cultural positions (Mertens, 2007, 2016; Cram et al., 2016).

The attention to the worst-off groups is therefore current in the aforementioned approaches. However, the attention appears to be particularly evident in those expressly characterized as participatory. There is a firm belief that participatory approaches are those that best involve those who have less power and "voice" in the program (Mansuri et al., 2013, p. 121). Participatory approaches take an active role in managing change among primary stakeholders, especially those at risk of exclusion (Ottier, 2005).

Practice

The elements common to the theories examined—stakeholder participation, context, and marginalized groups—are chosen as screening criteria to identify equity issues in a practical experience. Our research consisted in a review of all evaluation reports carried out in the two-year period 2013–2014, on behalf of the Italian Ministry of Foreign Affairs (MAE) and published on its website (Valutazioni, 2015). The case of Italian cooperation is considered to be exemplary of the current situation because the attention to the issues of social justice is constant and very detailed and it can be observed in all documents, both general and specific. The sectorial guidelines and programmatic indications contain ample references to equity issues: human rights, inclusion, poverty reduction, and gender equality (MAE, 1999, 2009, 2012a, 2012b, 2012c, 2014). These topics have long been affirmed and repeatedly reiterated in official documents and in the guidelines expressed by international organizations and the most recent meetings in Paris, Accra, and Busan. Express references to these documents are found in Italian and European cooperation policies.

Moreover, the period examined justifies the choice because it marks a phase of renewed attention to evaluation that was previously uncommon in the Italian administration (Stame, 1998). In 2010 the Italian Ministry of Foreign Affairs published several guidelines (MAE, 2010a, 2010b, 2010c, 2010d, 2010e) and set up an office to establish the strategy and quality of the evaluations, which also considered the recommendations of the DAC Peer Review (OECD, 2009).

Table 1
Acronyms Used to Indicate Evaluation Reports Referenced

Acronym	Linking Questions
Strada AFG	Riabilitazione della strada MaidanShar – Bamiyan in Afghanistan
Budget MOZ	Valutazione indipendente del budget support in Mozambico
Studio ETT, PS	Studio sulle diverse modalità di accesso al credito di aiuto e allo sviluppo dell'impresa informale in Egitto e in Palestina
Debito ETT	Egitto. Programma di conversione del debito – II Fase
Rom MKD	Macedonia, città di Stip e Prilep: miglioramento della qualità della vita dei Rom e avvio dell'integrazione
Ong ETH	Progetti gestiti da Ong in Etiopia
Emergenza AFG, RLB	Programmi di Emergenza in Afghanistan e Libano
Art Gold ALB	Art Gold Balcani, Art Gold 2 in Albania: appoggio alle reti territoriali e tematiche di cooperazione allo sviluppo umano
Sanità MOZ	Iniziative nel settore sanitario in Mozambico
Rientro AFG	Sostegno alla reintegrazione di famiglie afghane rientrate da lungo esilio nelle province di Balkh e Sar-I-Pul, Nord Afghanistan
Educazione MKD	Pilot activities for education and culture in Macedonia
Crediti TNS, ETT, PS	Crediti d'aiuto in Tunisia, Egitto e Territori Palestinesi
Ong RAG	Progetti gestiti da Ong in Argentina
Debito PER	II Accordo di Conversione del Debito tra Italia e Perù
CHYAO WAL	Trust Fund alla Banca Mondiale "Children& Youth in Africa" – CHYAO in Sierra Leone

Criterion 1: Stakeholder Participation

An equity-oriented evaluation is characterized by the participation of the stakeholders involved in the project (UNEG, 2008)

Stakeholder participation has long been affirmed (Cartland et al., 2008; Greene, 2012); so much so that it has become a basic concept (Mathison, 2005, p. xxxiii). The emphasized advantages of stakeholder participation are mainly due to different yet closely linked factors: utility, credibility and equity. The first reason is the most obvious for the immediate operational implications. Evaluation is oriented toward action; it does not end in judging an intervention, as it must help to make decisions to change it, continue it, or propose it again. Experience has shown that the evaluation results are better understood and then used if stakeholders were included during the evaluation process and could express their ideas (Patton, 1997; CIDA, 2004; Cullen et al., 2011b). This situation also increases the possibility of reaching a consensus, overcoming initially conflicting positions (Weiss, 1983; Morra et al., 2009). The other factor, namely credibility, is ensured when the process of involving the various stakeholders occurs through actual consultations, with continuous information on the objectives and methods of evaluation. These are aspects of openness and transparency that build trust in the results of the evaluation (UN, 2011a). The last yet most important reason is equity, which in development aid draws attention to the most

vulnerable and marginalized groups (UNEG, 2008).

In focusing on human rights and gender equality, it is necessary to find means to include members of marginalized communities. Although the process is not always easy, it becomes indispensable in cases where these groups are the target of the project, as understanding their perspective is necessary. Notably, the attention to stakeholder involvement has grown since the attention of aid evaluation has shifted away from aid delivery towards the impact on beneficiaries, and the understanding is that it cannot be measured without involving them (Cracknell, 2000, p. 318).

It is not useful to involve all stakeholders, neither in the same manner nor at all stages of the evaluation process. To strike a balance between the needs for inclusion and efficiency, an informal selection is often made even though there is a specific tool for it: the stakeholder analysis. The stakeholder analysis in the reports examined, though conducted in the planning phase, is not considered in the evaluation phase. This is also true of the only report that expressly refers to it as one of the three evaluation tools used: deviation from planning analysis, stakeholder analysis, and institutional analysis (Sanità Moz). The report, however, focused only on planning analysis, thus missing the opportunity to highlight those significant changes to social justice which could have been captured through the other two tools.

The Evaluation Guidelines recommend involving "as widely as possible the local counterparts and beneficiaries" (MAE, 2010a). This broad involvement can be understood by considering two aspects: which and how many stakeholders to identify and what opportunities and modalities to offer for their contribution. These two aspects match two of the characteristics identified in a well-known work by Cousins and Whitmore: defining which groups will be involved and the degree of their involvement (Cousins et al., 1998, p. 10). For the first point, notably, interest can be direct as in the case of donors, institutional bodies, or recipients, but in a broader sense, even a community as a whole. Among the possible distinctions between the various categories of stakeholders, we simply distinguish the key stakeholders and final beneficiaries. In the reports examined, all Terms of Reference (ToR) propose a very detailed list of key stakeholders to be consulted, whereas the indications for the final beneficiaries are more general. Sometimes, the beneficiaries are identified through their representatives, and more often, at the end of the list of key stakeholders, the wording "and the most important beneficiaries" is added. This can be interpreted either by looking only at the recipients of funding or the final recipients of the intervention. The evaluation reports follow the indications of the ToR for the key stakeholders, whereas, for beneficiaries, they broaden the range: individual beneficiaries, representative figures, families and the local community. Therefore, regarding the characteristics of the groups involved, these aspects, according to Cousins et al. (1998), are a precise signal of participatory evaluation. But this signal should be linked to the other characteristic, namely the degree of stakeholder involvement (Cousins et al., 1998). There are different modes of participation considering the various phases of evaluation and the type of commitment made by the stakeholders. In general, for key stakeholders, the reports identify areas of involvement at all stages, whereas for the final beneficiaries, only in the implementation phase.

Based on the experiences of development aid, it is widely believed that participation must concern the entire evaluation process; effective involvement consists in sharing the responsibilities from the initial phase to the dissemination of the results. All these aspects are present in three reports: in one case for the key stakeholders only (Budget Moz), and the other two cases, also for the final beneficiaries (Debito ETT, Credito TNS, ETT, PS). The next step in understanding the role that stakeholders have in the evaluation is to determine the actual space they have had along a continuum, ranging from providing information to sharing decisions. In an equity perspective, stakeholders have the right not only to be consulted but also to share decisions on what and how to evaluate. Therefore, "measuring" participation throughout the process becomes crucial (UNEG, 2014, p. 32). An interesting

perspective is therefore observed in the widespread request of the ToR to consider "the level of stakeholder participation in the evaluation." Although all the reports address the perspective, none of them respond precisely to the request. A curious observation is that the only report that engages in measurement, providing indices of expected and achieved participation for stakeholder categories, refers to the project activities and not to the evaluation process (Educazione MKD). Arguably, there is more experience, and therefore awareness, in participation during the phases of design and implementation, that many other reports have highlighted.

The proposal to measure could have served as an incentive to determine means of more comprehensive participation in the evaluation. But the answer of the reports appears to be absolutely vague. Even those reports that declare using a participatory approach (ART Gold ALB; Ong RAG) do not differ from the others. By specifically considering the methods of communication, it should be noted that those with one-way communication prevail over dialogical ones and are limited to providing information. The typical situations that may hinder participation (obstacles related to time, place, and the availability of means of communications, and security) (UNEG, 2014, p. 46) have also occurred in the experiences reported. Among the difficulties encountered in including beneficiaries, the most common were: the time gap between the evaluation and the end of the project, the impossibility of reaching places, the lack of updated lists of beneficiaries, and change of staff. In the general modest application of participation, some evaluation reports stand out by underlining the broad composition of the representatives, specifying tools and methods, and identifying aspects of a fruitful collaboration and possible empowerment (Emergenza AFG, RLB; Educazione MKD).

In conclusion, while the key stakeholders are involved throughout the evaluation process, the final beneficiaries, instead, are left in the background; clearly, the final beneficiaries are generally the mere recipients of the intervention, and this situation is also reflected in the evaluation. Even in the case of a report that defines the key stakeholders as "actors whose contribution is essential," no elements support this statement (Crediti TNS, ETT, PS).

Stakeholder participation allows linking the impact of change on human rights and gender equality, only if involvement occurs throughout the complete evaluation process. As this phenomenon is not found in the documents examined, it can be assumed that the evaluation was aimed more at accountability than learning. Notably, the cases examined have not created a space to share ideas, which would have ensured the attention for issues of social justice. The prevalence of accountability, found in these experiences, is a common situation that has been debated for some time now (Guijt, 2010). Also in the context of EU cooperation, it has been recommended to promote and stimulate learning (Bossuyt et al., 2014, p. 46).

Criterion 2: Context

To properly evaluate the context, the United Nations Development Programme (UNDP) states that “understanding the political, cultural and institutional framework of evaluation can give important indications on how to conduct it to ensure impartiality, usefulness, and credibility” (UNDP, 2009, p. 167).

The evaluation, to determine the results and express a judgment on what has worked, must consider the overall situation in which the intervention occurred; therefore, it needs to identify the elements and modalities that have played a role. Experience has shown how the context as a whole, through the different factors of the environment in which a project is realized, affects the results (IFAD, 2015, p. 29). In particular, the factors most relevant to the type of intervention should be highlighted. That is why Quality Standards require a detailed description of the context (OECD/DAC, 2010, p. 12). For these reasons, each evaluation report usually has a description of the institutional and political framework, with the most significant data on income, poverty, and social problems, and an analysis of the specific sector in which the action occurs. The clearer the relationship between the two parties is, the better the understanding of the project’s logic. In the evaluations examined, this connection

between the context of the country and context of the sector has been well explained and is accurate, especially in cases where the intervention is carried out in a single country and directed at a specific sector (Strada AFG; Sanità MOZ; Debito PER). More often, however, in the same situation, only the part relating to the sector is present, because there is a widespread tendency, also common in these reports, to reserve the most consistent and detailed analysis for the part specifically concerned with the type of intervention. A situation of greater complexity is that in which the projects, even if directed toward a single sector, are conducted in different countries. Two evaluations focused on this type of experience: in the emergency sector (Emergenza AFG, RLB) and the credit sector for SMEs (Crediti TNS, ETT, PS). The purpose of a moment for a comprehensive evaluation is to draw lessons valid for the thematic area concerned and to apply them to future programs. This is precisely what is found in an emergency report offering useful indications and outlining a series of recommendations. When the aim is more ambitious and attempts to make generalizations from the results through comparison, the task becomes more difficult. Each context has its own characteristics that interact with the project in ways that are just as specific and that can make comparisons difficult. These difficulties have undermined the commitment of the evaluation team on credit. The notable success observed in one country, Tunisia, and an equally remarkable failure in another, Palestine, are not explained. A subsequent detailed study was required, which underlines the impossibility of making a comparison (Studio ETT, PS). Yet, in the same Palestinian territories, in the same area, at the same time, the interventions of the French cooperation had satisfactory results (Trésor, 2014). Perhaps the comparison with this experience would have been more fruitful even without the commitment to a joint evaluation, for which, however, the appropriate conditions were in place, according to the "Evaluation Guidelines" (MAE, 2010a, p. 6).

How reports examine the context. The logical framework (LFA) and the theory-based approach (TB) are among the most widely used approaches to explain the results and both use causal logic and require satisfactory knowledge of the context. In the documentation examined, the ToR always require at least one of the two approaches, sometimes both. There is an explicit reference to the context in the request to examine the LFA and an implicit reference in the request to judge the "validity of project design". The latter can be understood as TB, because the theory of the program is based on the logic of the project (Vogel, 2012), and it is necessary to know the conditions of the context that were favorable or unfavorable to the change pursued.

The evaluation reports present a range of answers, which, on the whole, accurately respond to the ToR indications and consider the context. The logical framework is always at the forefront, but the attention to the hypothesis of change is evident and often reconstructed based on the LFA. Those that used theory-based approaches to analyze the context have specified the links and outlined the evaluation matrix (Strada AFG; Ong ETH; Sanità MOZ; Educazione MKD). The group of theory-based evaluations also includes a well-known type, realistic evaluation (Pawson et al., 1997). It explains the results (O) that consider the conditions of that particular context (C) and the mechanisms that generate them (M) according to the formula (C + M = O). It starts from an initial hypothesis of this configuration, which, through the analysis of the different combinations and observations of the results, whether positive or negative, is also redefined several times.

This method has been adopted in only one report (Budget MOZ) to evaluate one of the most relevant aid modalities in recent years: budget support (E.C., 2008). This tool, considered appropriate to ensure aid effectiveness, was proposed by the European Commission as a "vector of change," which refers to the "promotion of human rights and democratic values" as key issues to be addressed for development to occur (E.C., 2012, p. 12). Budget support is configured as a set of interventions, and in addition to the multiplicity of donors, there is a plurality of objectives that include economic growth, poverty reduction, and improved governance (Dijkstra et al., 2015). The evaluation thus becomes particularly complex and cannot be

based on a simple chain of linear causality. Therefore, after several experiments, a specific method has been developed, the Comprehensive Evaluation Framework (CEF). The CEF is based on the choice of a program theory to hypothesize a sequence of effects through the various levels interacting with the context (OECD, 2012).

The evaluation of budget support in Mozambique complies with the CEF and indicates the realistic evaluation, where for the first step the report must specify the program theory (Budget MOZ, p. 45). The report does so without declaring the initial hypothesis and therefore not even its subsequent redefinition. The report identifies the overall budget support as a mechanism and presents the main changes in the context. However, it is unclear how many configurations (C + M) have generated the different effects (O), making understanding the application of the realistic evaluation difficult, because what has happened in between is missing, namely the most interesting part on how the program achieved those results. Therefore, neither the entire budget support nor the other interventions defined as "mechanisms" in the report can be referred to a realistic evaluation. It is an odd limitation in a report in which everything else is very accurate: the analysis of the context is done comprehensively and follows the other CEF steps.

Another tool, the SWOT (Strengths, Weaknesses, Opportunities, Threats) analysis, was proposed by the ToR, and has been used in many reports. The SWOT framework considers, through a simple diagram, the main factors that can influence the results (Evalsed, 2003; Start et al. 2007). The SWOT analysis, initially used for the ex-ante evaluation, is also adopted during and after the action; in the reported experiences, this analysis is used in the ex-post phase, but only partially, because it considers only the internal factors—strengths and weaknesses—and applies them to the traditional DAC criteria. This simplification neglects the other two items of the matrix: the external factors—opportunities and threats—, which would be very useful instead for understanding the context.

The context in different phases. Many observations in the reports highlight an inadequate understanding of the context in the initial phase that in some cases concerns the entire logical framework and in others its articulations. The limits are identified in the absence of clear objectives, precise results, or well-defined indicators. However, in the difficult situations in which these interventions occur, the lack of information limits the application of the logical framework. There is often a lack of clearly defined information, data and indicators, and land, health, and school registers, which are necessary to identify the recipients of the services. This inadequacy of data, which does not allow for an evaluation of the changes resulting from interventions, has even been identified as "the main project weakness" (Rom MKD). This difficulty is encountered frequently enough, and many recommendations suggest a more in-depth context analysis.

By examining the common insistence on the lack of data, one observation is that the data seldom refers to what would be useful for understanding the situation regarding human rights and gender equality. This attention is found instead in those reports that do not attribute the limits to the absence of data. These reports offer an analysis of the broad context with precise references to the social aspects of development and inequality, thus showing sensitivity to issues of social justice. They have been able to use experiences, studies, and evaluations that refer to broad programs, with more donors and more sectors, and have a perspective of action extended to the international context (Budget MOZ; Art Gold ALB; Sanità MOZ; CHYAO WAL). Among these is the assessment relative to Albania that expresses an original approach that is not only sensitive to issues of fairness but is also able to dynamically relate to the context, involving more actors and more sectors.

Attention to the context is important especially in the implementation stage; the major problems are felt because of continuous changes. In developing countries, the speed of political changes, institutional fragility, and weakness in the administrative system make understanding the context and adapting to its changes more difficult (Ofir et al., 2013). The documentation contains many remarks about these aspects. There have been different responses to the change of context. The first group includes several projects that have not

had any difficulty finding means to respond to change, that is, from simple adjustments to a complete review of the intervention logic (Ong RAG; Ong ETH). In these experiences, a favorable element was sharing with the local actors. Sometimes the change, though consisting of an improvement, requires a revision because the change can alter the project's setting. Some examples are: the increase in liquidity available in the case of SMEs (Crediti TNS, ETT, PS) and the rapid economic growth in Mozambique. The report on Mozambique was very exhaustive in presenting the numerous adjustments made by the program in response to changes, sometimes problematic, in the context (Budget MOZ). Positive adaptation can even comprise rejecting a change that at first sight was thought to be an improvement. A notable case is the refusal to modernize by choosing a more qualitative concept of growth, based not on economic parameters but on those of equity (Ong RAG).

In the second group, numerous interesting activities have succeeded in significantly modifying the context towards social justice. We observed evident results supported by data and others, which, apparently more modest, have nevertheless opened up new paths. These initiatives were aimed at reintegration into the community of origin or families; or promotion of opportunities for dialogue and collaboration between ethnic groups, or between public institutions and civil society. We can also include other initiatives that, though not producing results with great visibility, have certainly had a positive impact on the context, favoring social inclusion. When the transition from normative affirmation to the actual exercise of rights is complex, several phases may be necessary for effective implementation. This step was facilitated by actions aimed at overcoming resistance and changing cultural bias towards worst-off groups and by those that fostered awareness and the exercise of social and economic rights (Debito ETT; Debito PER; Emergenza AFG, RLB).

Finally, the third group includes those cases, unfortunately many, that highlighted those limitations of context that have had a particular weight on fairness. It is largely an essential category, namely that of access to the services offered: school, training, and laboratory. The reasons given were different, ranging from the geomorphological and infrastructural characteristics of the environment (climate, terrain, and road distances) to those of an institutional and political nature (regulations, procedures, government guidelines). Other reasons, though present, were highlighted less, and were often difficulties that are deeply rooted in an often neglected dimension of context, namely culture.

Cultural aspects of context. Every context is widely permeated by cultural elements and a context cannot be correctly assessed without taking these elements into account. Culture is not homogeneous and intersects with various factors, namely, institutional, economic, and social, which are expressed through habits, language, values, and ways of thinking and communicating (McBride, 2011; Kirkhart, 2010).

In 2004, it was acknowledged that there was still a long way to go to include the cultural aspects of context into evaluation practice (Sengupta et al., 2004, p. 11). The acquisition of this new requirement has grown so rapidly, that not many years later it was possible to state that the time had come to conduct evaluation in a culturally responsive manner (Frierson et al., 2010, p. 93). The awareness of these aspects had paved the way: the affirmation of the concept of “cultural diversity” at the international level (UNESCO, 2001), the evaluators ethical responsibility to respect differences in culture (UNEG, 2008, p. 7; DAC/OECD, 2010, p. 6), the need to include vulnerable groups in the evaluation (Samuels et al., 2011, p. 184), and the complexity of contexts with multiethnic and multicultural realities (Chouinard et al., 2009, p. 458). Numerous articles and monographic publications have been devoted to culture in evaluation journals (Thompson-Robinson et al., 2004; Woolcock, 2014; CJPE, 2015). The American Association has dedicated a specific document to it (AEA, 2011), and the European Council, recognizing the various dimensions of development, has stated that for each of them culture constitutes an "essential component" (Council, 2015, p. 6). Recently, significant importance has been attached to the revision of the standards for evaluation of the United Nations. Cultural aspects must be considered

at all stages; it is necessary to be sensitive to the customs, beliefs, and habits of the social and cultural environment. It is advisable to include local experts when considering the different perspectives to facilitate acceptance by the local community (UNEG, 2016).

However, precisely in the area of development cooperation, culturally sensitive evaluations are still scarce; therefore, awareness of the cultural importance of the context cannot be considered a given in planning approaches based on local culture (Chouinard et al., 2015). Yet the culture of the context should also be emphasized in evaluating appropriateness, a relevant category that consists in the cultural acceptance and feasibility of the activities or methods (UNDP, 2009, p. 168). Notably this aspect, in the reports considered, appears to be neglected. While the cultural dimension, whether in whole or in part, is ignored, it may happen that a needs analysis, though correct, chooses a project, which, while responding to the need, will be difficult to achieve. It is not enough to offer education for girls or vocational training for women if this contrasts with the conception of the female role. It is, therefore, possible to explain that the same difficulties are encountered in different projects and countries that have the identical gender view, so that girls in Addis Ababa and women in Peru did not receive training, and in Afghanistan girls neither had access to the new road nor women to the market. When looking at culture it is possible to understand the difficulties of those initiatives, which, directed to change the context, were only partially successful and bound to finish with the end of the project. These actions had an impact on culture, particularly on the division of roles, the articulation of power, and the social structure: the girls' clubs in Ethiopia, the Multicultural Centers in Macedonia, and the new responsibilities in Sierra Leone and Macedonia. The same applies to topics such as democracy and governance, which have been relevant in development policy guidelines in recent years and referred to in the majority of reports as objectives to be pursued. In this perspective, mechanisms of participation and integration between ethnic groups have been launched, opportunities for meetings and collaboration have been organized between the central and local level or between the public and private sector, but implementation was partial or unsatisfactory; therefore:

- good governance records a less positive impact (Budget MOZ)
- participation is not achieved (Studio ETT, PS) or has no roots in the mentality of the country (Art Gold ALB)
- integration between ethnic groups is penalized by mutual distrust (Rom MKD)
- the public–private partnership has difficulty in establishing relations (Debito ETT) in the initial stages (Educazione MKD) or subsequent definition (Art Gold ALB)
- the collaboration between the center and the periphery is struggling to work (Educazione MKD)

The assessments are limited to taking note of these failures without asking too many questions, even wondering whether:

> "Despite all the conferences on women's rights... the condition of women is still precarious" (Rientro AFG).

> The project, paradoxically, despite having achieved the planned activities….did not reach any results (Crediti TNS, ETT, PS)

The assessments that stop at the surface consider that they have not achieved the objectives because they do not imagine that the causes lie elsewhere, in the cultural dimension of the context. The evaluator should take these aspects into account, but does not do so, neither in these reports nor in general, as there are substantial and inevitable limits in their cultural competence. Hood theorized the culturally responsive evaluation (CRE) because understanding aspects of a culture other than one's own experience is almost impossible (Hood, 2005, p. 97). It should be considered that every evaluator has a training and experience that characterize their methods, from their hypotheses to the results. The problems are more evident when an evaluator's culture is Western, or Eurocentric, and the context is that of a Southern country, where, furthermore, there are communities with different characteristics within the same

context. Therefore, awareness of differences is necessary, even among the different groups, so concepts such as development, participation, and empowerment can be interpreted differently (Ofir et al., 2013, p. 22). This can also occur for the seemingly simpler terms of everyday life. For instance, even the definition of home can be different in different cultures (Murdie, cit. in Lee, 2004).

The same applies to the methods: the broad use of interviews, considered useful to understanding the ideas of recipients, may be unwelcome or incongruous in some environments (Fitzpatrick, 2012). It is no coincidence that the main exponent of empowerment, working with the most marginalized populations, has underscored the risks of the rationale and procedures of Western thinking, which not only prevent full participation but exclude different ways of knowing (Fetterman, 1996b, p. 24). All this explains the increasing distrust in evaluations marked by the culture of the agencies or donor countries (Stern, 2006), and there has been talk of decolonizing evaluation (Hopson et al., 2012; Johnston-Goodstar, 2012).

Evaluation must include a broader view to examine the local situation without the preclusions of a Western perspective. If an evaluation uses, instead, the "development lens," the differences can be recognized and not favor the stances of the majority (Kirkhart, 2011). The methods must be flexible to adapt to the needs and reflect the values of the context, and not only those of donors (Carden, 2013), and to respond to questions that may also come from the different perspectives of the various stakeholders (Samuels et al., 2011; Hopson et al., 2012; Ofir et al., 2013). According to the useful distinction expressed by Carden and Alkin, the evaluation can therefore reject the "adopted" methodologies imported from the culture of Western countries and choose those "adapted" that are coherent with the culture of the context (Carden et al., 2012). Evaluation can become functional not only for use, credibility, and validity (Kirkhart, 2010; UNEG, 2016), but even for social justice (Newman et al., 1996, p. 147).

Criterion 3: Marginalized Groups

> The evaluation should examine whether the process of development has been fair, and the benefits of growth have reached the most marginalized groups in society (Donaldson et al., 2013 p. 13).

Development aid has always had the general objective of reducing poverty, but this is insufficient to reduce inequalities. Although there has been much progress, it has been found that despite the improvements in income growth inequalities persist; in some cases, these inequalities have worsened, and in others, the differences within the same country have increased (Ferreira et al. 2008; Melamed, 2012; UN HLP, 2013; UNDP, 2014; Fuentes-Nieva et al., 2014). Sometimes that economic growth is concentrated in the most advanced areas, therefore, the inequalities with peripheral areas or with excluded groups increase. This problem has also been observed in the experiences examined:

> "The poor have benefited less from growth than the non-poor" (Budget MOZ).

> "While poverty has been substantially reduced, inequalities have decreased little and social conditions outside the urban areas are worse" (Debito PER).

Poverty is the most common cause of exclusion but is often matched by further discrimination based on a social status considered subordinate. The most common exclusionary elements are gender, ethnicity, race, religion, and disability (World Bank, 2013), and result in disparities in access to services, the control of goods, and access to property (UNDESA, 2013), which, by preventing these categories from benefiting from different opportunities, restrict human rights (UNDP, 2013).

There is an implicit assumption that cooperation is committed to helping the most disadvantaged situations, and this has been affirmed in numerous international as well as Italian documents (MAE, 2009; L. n. 125/ 2014; MAE, 2012b). The call to equity is also evident in the ToR, where it is expressed as the involvement of marginalized categories sometimes in the implementation, and

sometimes when verifying the impact. In both cases reference is made to poverty, which, however, being a multidimensional phenomenon, can include other situations of exclusion.

The attention to marginalized groups is present in all the projects and therefore in the respective evaluation reports. This attention is more explicit when they constitute the main objective, in the case of projects conducted by NGOs related to a specific group, and when, as in emergency projects, the spectrum of vulnerable groups is very broad: refugees, returnees, prisoners, the mentally disabled, minors, young people, and women. This attention, though less evident, is present also as a secondary objective, for example, for the territorial position, such as some provinces in Mozambique or the Hazara ethnic group in Afghanistan.

Finally, as development cooperation is being examined, it comes as no surprise that significant elements for the worst-off groups can be found even in those projects aimed at a specialized sector such as debt conversion. In the case of Egypt, we examine the improvement of human development indicators, and in Peru, the expected results include the participation of women and the most vulnerable groups in the exercise of rights.

Human Development

Vulnerability is caused by the limitations people encounter in the material resources available and possibilities of choices essential to their development (UNDP, 2014, p. 23). A key to understanding vulnerability is offered by the human development approach (HD), which has strong implications for evaluation (Sen, 2005). The human being is the cornerstone on which to measure development, so one looks not so much at economic growth, but at expanding people's capacity. Attention is paid to inequalities, and active participation in community life is valued. Human development has long entered the international debate, and hints of it have been included in various reports, though limited to the use of specific indicators (Educazione MKD; Rientro AFG; Sanità MOZ; Chyao Wal). Therefore, a more interesting endeavor would be to investigate how this profile was treated in the two evaluations that followed this approach.

The first is the evaluation concerning the conversion of Egyptian debt (Debito ETT). In this case human development is one of the areas indicated in the Joint Declaration, opportunities for social inclusion are provided for, and initiatives for poverty reduction are directed specifically to vulnerable groups and women's empowerment. Furthermore, the consistency of the program regarding the orientations of the international organizations is emphasized, and the country must commit itself to strengthening the capacities of all citizens. This way, they can realize their potential and enjoy the expanded possibilities of choice they all have (Debito ETT, p. 53). This reference corresponds exactly to Sen's (1999, 2005) capacity approach. For the evaluation of the overall program to consider the macro and micro levels, we considered using mixed quantitative and qualitative methods for human development, but the signs that appear are too synthetic and do not allow an assessment of these aspects as well. This is the case, for example, of the Social Contract Advisory, Monitoring and Coordination Center aimed at promoting the participation of civil society in decision-making processes to rebuild trust between citizens and institutions through democratic governance. It is defined as a specific tool for the achievement of human development and the report presents its results as satisfactory. However, no elements can support the conclusions and, furthermore, it is stated that governance, for the immaterial dimension, is difficult to measure and there was no proper monitoring. There is only one other explicit reference to human development in the project for training in mechanics. Many other projects could be evaluated according to this perspective, including the network promoted for the recognition of the rights of small producers, or the project for the inclusion of people with mental disabilities. Therefore, the human development issue is, though present, little appreciated, and not evaluated. In the great profusion of economic data, no space has been found for those significant for this approach. Even if there are "actions aimed at improving the human and social development indicators"

in the Upper Egypt area, there is no reference to supporting data nor the survey periods.

A similar situation can be found in the other evaluation, which, moreover, concerns the only project that explicitly stated the use of the human development approach (Art Gold ALB). Among the project partners was the United Nations Development Programme (UNDP), an agency that played a strong promotional role, including through publications and annual reports (Human Development Reports). This explains the great attention that the ToR placed on the issue by asking to: "evaluate according to the UNDP evaluation policy" and "verify the level of achievement of the UNDP's commitment to the human development approach and whether the aspects related to the parity and gender are present in the implementation of the program." The UNDP evaluation policy is guided by the principles of human development and human rights and considers the values of fairness, justice, gender equality, and respect for diversity (UN, 2011a, p. 3), but these principles are not reflected in the report. The evaluation, despite being under way, has expressed a positive opinion on implementation, but the development on which it has focused is regional and never refers to the human development paradigm or specific indicators. Human development has appeared only as a content of training courses. In the perspective of this approach, one aspect of interest for equity, namely marginalized groups, has been neglected in the report, which admitted: "The activities aimed at the most disadvantaged sectors of the population have not been identified as priorities." In truth, all the indications at international level consider them as priorities; moreover, the project was a local operational tool of the strategy that stated that national priorities would be concentrated on vulnerable groups to ensure better access to services and reduce obstacles to participation (UN, 2011b, p. 11). The program included the participation of vulnerable groups in the areas of health, education, economy, and land management, the improved access to basic services and job placement of people with disabilities (UNDP/UNOPS, 2009, pp. 12, 14).

The analysis of these two evaluations showed that the evaluation approach was not directed to human development, perhaps because of the difficulty in grasping the effects for skills and opportunities in all dimensions (Chiappero-Martinetti, 2010).

Gender

The lack of attention to gender issues in evaluations reveals an attitude present in many cases. Gender equality, as a cross-cutting issue, should, in any case, be considered; although when it receives attention it is often as an extra, offered only thanks to the sensitivity of the project managers. For example:

> "Above all, it should be considered as a result of added value, the introduction in the training activities of the issue on gender aspects, normally not addressed when not directly specified by the projects" (Ong RAG).

> "Compared to the indicator of job creation the relevance is very high, although it should be noted that the issue of female employment was not considered a priority" (Crediti TNS, ETT, PS).

> "The activities targeting gender equality and women empowerment were not clearly planned.....however, particular attention was given to support the women's leadership" (Art Gold ALB).

In this latter case, an evaluation emerges which takes into account neither the programs nor the indications of ToR. Yet "One UN" asserted that the program's goal concerning gender was to ensure active and recognized participation in the democratic governance of Albania and that for good governance, action had to be taken on the causes that limit women's participation in public life, and that at the local level these factors were most available (UN 2011b, p. 23). All these indications were consistent with the program that aimed to promote the active participation of women and stated that gender mainstreaming would be pursued at all phases of the projects (UNDP/UNOPS, 2009, p. 4). To the key question envisaged in the ToR, "To what extent has AGA 2 promoted gender equality in its activities?" no real answer is given, citing the absence of specific documents and referring to projects that are only

potentially in strategic plans. The question, in any case, asks to find the parity in the initiatives of Aga 2, and not in the plans of other actors. It appears that neither the working group nor the evaluators were interested in gender equality.

Recent analyses of European projects in developing countries found similar difficulties: despite the extensive investments in the sector, gender inequality persists (E.C., 2015a). Yet the importance of gender equality for inclusive and sustainable development is widely recognized, but gender is not sufficiently integrated into the objectives of the countries or in the evaluation systems, due to poor context analysis and limited use of indicators. In most countries, European programs suffer from a superficial and scarcely documented understanding of the context in which issues of equality and empowerment arise (E.C., 2015b). Consistent with these analyses, the Commission in a subsequent document calls for a rigorous gender analysis for all external initiatives, adapting the approach to the specific characteristics of the partner country's context and the use of disaggregated data (E.C., 2015c).

A more immediate and simpler way might be to involve experts from the local context as evaluators (Rogers, 2012b; UN Women, 2014; IEO, 2015). This indication is reflected in a report that analyzes three projects conducted by NGOs in Ethiopia (Ong ETH). The expert's contribution proved to be crucial in the project for the promotion of cultural heritage, where he promoted the involvement of women in all activities and obtained good results. In the same country, however, in the other two projects, the evaluation noted that the gender issues were neglected, limiting the success of projects. That could demonstrate that the expert's support was the factor that contributed to making the difference.

These evaluations do not suggest solutions to problems, which are certainly, as they define them, “difficult”, nor any explanations beyond the usual absence of gender indicators are given, whereas a contextualized gender analysis would be more useful. A more promising and more complete approach is the so-called "twin-track”, also adopted by the EU, and considered in the Agreement between Italy and Peru. The report of Peru evaluates the corresponding program (Debito PER). The first track considers gender a cross-cutting theme (gender mainstreaming) for all projects. Article 3.3 of the Agreement adheres to it, calling for projects to "improve the condition of women". The second track, which comprises implementing specific interventions for equal opportunities, is set out in Article 3.4, which establishes "the priority" for the projects concerning gender issues. The evaluation focuses more on gender mainstreaming and about second track refers to excellent results, determined through a specific study, not reported, not even in part. The evaluation nevertheless describes several promising initiatives to change the system.

Measurement and Interpretation Questions

Some evaluation problems emerging in these reports refer to issues that are quite widespread in the international debate, and they are difficulties related to limitations in the use or choice of indicators that are reflected in the evaluation design and the interpretation of results. The change cannot be established only with quantitative data, which give information only on the effectiveness of expenditure, while it would also require qualitative data and long-term measurements (ICHRP, 2012). It may indeed happen that, by choosing a set of a few simple indicators, for example, the lack of schooling, a simple causal link can be established, assuming that the construction of school buildings is the right solution. This causal link, where the relationship between intervention and outcomes are directly connected, does not work in practice (Schwandt, 2015, p. 23). An example is an initiative where basic education was to have an even more ambitious purpose, i.e., social development, but it was determined that:

> “Strengthening of infrastructure has not had a direct effect on the increase in school registrations (Ong ETH)”

With these approaches, these evaluations are forced to discover that the expected results were not achieved, whereas the interpretation would be facilitated by a careful reading of the

hypotheses underlying the theory of change. On the other hand, oversimplification is often the case. Therefore, a hypothesis is that by intervening in one dimension the desired result can be achieved, but this dimension can be characterized as "necessary but not sufficient" because other factors interfere (UNEG, 2013, p. 19). The construction of the new road in Afghanistan appeared "necessary" to free the Hazara ethnic group from isolation, but was not "sufficient" because security from Taliban attacks was not guaranteed; the same road was “necessary” for the girls to access school, but not “sufficient” for the cultural limitations preventing use by the girls (Strada AFG).

Many remarks of this type are in the evaluations examined. Two exemplary cases concern training. In Afghanistan, training in weaving did not automatically increase the empowerment of women, because they could not directly sell their products. In Mozambique, the training of doctors coming from remote areas did not guarantee that they would stay to work there, as there was also the issue of the shortage of housing. From this perspective, it is easy to understand the reasons why some results have been considered unsatisfactory or are uncertain about sustainability. This case was also true for services offered in vain although consistent with the needs: credit for SMEs, sanitary and laundry services for the Roma, cataloging computer programs in Macedonia, and the pilot plant for marble processing in Peru.

If an evaluation is oriented toward fairness, difficulties increase, and indicators must be sought to measure progress in terms of rights or improvement of social justice, when even the most common concept of development has numerous meanings (Fuhr, et al., 2018) and is not easily measurable. We found that interventions often lack clear and specific indicators on human rights and gender equality. To reflect situations of exclusion and discrimination, they should be specific. However burdensome and complex this process can be, disaggregation of data would be very useful for an evaluation that aims to "leave no one behind” (UN HLP, 2013, p.7).

Findings

Concerning the three analysis criteria, it can be noted that the elements for stakeholder participation appear to be incomplete in terms of involvement with regard both to the categories and to the phases. Clarification is also lacking with regard to the procedures adopted and the role played by the participants; that is, the function limited to consultation was not able to facilitate the inclusion of the needs and ideas of those with less voice. Those recommendations which propose involving vulnerable groups, e.g., indigenous peoples, civil society, business representatives, and final beneficiaries, suggest involvement limited to the planning phase.

By examining the context, it can be observed that the description is always broad and accurate but tends to remain detached from the rest of the report. Therefore, the characteristics of the context are not linked to the results. This phenomenon occurs especially, as observed, due to the difficulties of a cultural nature that prevent interventions from being fully effective and are completely beyond the evaluative analysis. What is more difficult to understand is the absence of remarks and interpretations of these links in the many positive cases in which the activities have affected the context by changing toward greater social equity. In the recommendations, the context is highlighted in a few cases, where careful analysis is recommended only, as in the case of participation, for planning. Planning, however, will always be limited, given that there is no indication to consider cultural aspects and the root causes of inequity.

Finally, for the third criterion, the marginalized groups, we note that all evaluations specify the commitment to poverty reduction, whereas the reference to the reduction of inequalities is not so evident; explicit links to this goal are often absent. The results are not analyzed by investigating the improvement in terms of fair development. Notably, there is an almost complete absence of disaggregated data, precluding determination of whether the aid has reached the target groups and what improvements have been achieved. Gender issues are left in

the background, even if gender is the most relevant among inequalities. At first glance, the recommendations attach great importance to marginalized groups, but often, rather than recommendations, they are generic wishes to respect human rights or to promote inclusion; only a few cases specify the need to focus on gender issues. The report on Mozambique stands out for proposing precise directions for interventions well-linked both to results and to social justice orientations (Budget MOZ). The links to experiences are absent in the remaining recommendations that do not reflect many of the problems observed, in the areas of access to services, limitation of rights, and other discrimination phenomena. The reference to inequalities is lacking even for those initiatives, just launched but of great innovative value, such as those aimed at marginalized groups or that have created new tools for the defense of rights (Debito ETT, Emergenza AFG RLB, Art Gold ALB, Ong RAG). The three analysis criteria showed in the documentation examined that any issues of equity, even when found, were not highlighted and failed to recover visibility even in the recommendations. Based on the results of the documentation examined these criteria could be used also in the evaluation design and implementation:

- Identify and include the different categories of stakeholders, from the recipients to the providers, throughout the entire evaluation process. Verify how participation was managed during the activity.
- Recognize the complexity of context to understanding inequalities across different groups. Context analysis should focus more on cultural aspects to identify characteristics that can affect program implementation.
- Include a local expert, belonging to the marginalized group, in the evaluation team. Provide indicators disaggregated by sex, ethnicity, age and geographic location, to understand the situation before and after the program. Verify if accessibility or use of services has improved and if the worst-off groups have benefited from intervention less than the best-off.

Conclusions

These difficulties in recognizing and evaluating the profiles of equity reopen the perennial debate on the theory-practice gap. That is why Chelimsky's suggestions promoted a Forum to increasing dialogue between theory and practice (Chelimsky, 2013). Why, despite the available evaluation approaches presented in the first part of the study, have none been used? One possible interpretation is that given by the distinction proposed by Donaldson and Lipsey between "program theory," the model that hypothesizes the sequences to achieve change, and "evaluation theory," the concretely applied approach (Donaldson et al., 2006). If the program theory dominates (Coryn et al., 2011), and the use of simpler linear models such as the Logframe Matrix is prevalent, the interpretation, in a reductive manner, leads to a search for only the expected results. Instead, those results that are less defined or less measurable are neither sought nor recognized.

This could explain the scarce sensitivity to inequalities, cultural aspects, and, in general, issues of social equity. All the reports examined generally refer to program theory and numerous evaluation tools, but do not frame them in a specific approach. Those few reports that specify the chosen method, be it participatory, realistic, or human development, do not clarify its application. The absence of a precise evaluation approach for the cases examined, however, is not only due to the difficult dialogue between theoretical profiles and concrete application, precisely because methods sensitive to equity are not only practicable but practiced. These methods have elaborated the theory based on experiences positively employed in the field. The aforementioned approaches do not exclude the use of program theory; only that it cannot be of linear causality and it must be built through the stakeholder engagement, which characterizes the whole evaluation process. The results of interest are not so much those expected but those discovered along the way. These methods are therefore sensitive to the peculiar cultural features of

the context, suitable to grasp the situations of exclusion and asymmetries of power.

Therefore, the explanation should perhaps be sought elsewhere, either in the lack of theoretical training (Christie, 2003) or in training that is not broad enough to allow making a choice, as shown by various researches (Rog, 2015), or in practice geared more towards accountability. The rigid application of technocratic approaches, without questioning the structural differences of power, neglects a complexity of contexts (Ofir, 2013). A further limit could be caused by the ToR that generally, besides the widely used DAC criteria, do not require other criteria referable to human rights and gender equality. Yet, even the DAC criteria would allow interpretations compatible with equity-oriented evaluations (Bamberger et al., 2011; Peersman, 2014; UNEG, 2014). In actual fact, in the case under consideration, there were indications in the ToR that could have been used, such as requests to verify the level of participation, gender equality, or human development, but were ignored, or interpreted reductively. These are other missed opportunities.

We, therefore, conclude that the evaluation in these reports, though running into relevant issues, does not interpret them in an equity perspective, and does not make that necessary step forward to express itself proactively to achieve that public good, which lies at the heart of its professional mission. Perhaps an appropriate question is the possible role of evaluation, which cannot be neutral concerning values. The essential question in evaluation practice is "which political positions and whose values" should be enhanced (Greene et al., 2004, p. 99). Evaluation can no longer simply verify what has happened, but must ask to what groups it has been useful, which groups have been excluded, and to what extent has inequity been addressed (Kirkhart, 2016, p.116). In recent years, the responsibility of evaluators has been emphasized (Schwandt, 2018) and observations have been made that an ethical deficit can be highlighted if the aspects of social differences are not considered (Picciotto, 2015).

Evaluation should always be attentive and sensitive to social justice issues, not only when programs expressly refer to it (Agrawal et al., 2015; Rosenstein et al., 2015). Even if evaluation cannot remove the structural causes of inequalities, it can do a lot, highlighting them and proposing useful indications to contrast them in recommendations. According to Stame, "valuing for improvements can make a difference" (Stame, 2018, p. 440). The road to equity is still long and more work needs to be done in evaluation practice. The new Agenda for sustainable development can be a good opportunity (UN, 2015).

Acknowledgements

Special thanks for their support and helpful comments to Nicoletta Stame and Laura Tagle. Thanks also go to other members of the Laboratorio di Valutazione democratica, the officials of the IX unit of Italian Ministry of Foreign Affairs and anonymous peer reviewers.

References

Abma, T. A. (2006). The practice and politics of responsive evaluation. *American Journal of Evaluation*, *27*(1), 31-43.

Agrawal, R. & Rao, B. (2015). Evaluations as catalysts in bridging developmental inequalities. In R. Rist, F. P. Martin, & A. M. Fernandez (Eds), *Poverty, inequality, and evaluation: Changing perspectives* (pp. 24-38). Washington, DC: World Bank Group.

American Evaluation Association. (2011). *Public statement on cultural competence in evaluation*. Available at http://www.eval.org/ccstatement.

Bamberger, M., & Segone, M. (2011). *How to design and manage equity-focused evaluations*. New York, NY: UNICEF.

Bossuyt, J., Shaxson, L. & Datta, A. (2014). *Assessing the uptake of strategic evaluations in EU development cooperation: Study on the uptake of learning from Europe Aid's strategic evaluations into development policy and practice*. Brussels: EuropeAid.

Bowman, N., Dodge-Francis, C., & Tyndall, M. (2015). Culturally responsive indigenous evaluation: A practical approach for

evaluating indigenous projects in tribal reservation contexts. In S. Hood, R. Hopson, & H. Frierson (Eds.), *Continuing the journey to reposition culture and cultural context in evaluation theory and practice* (pp. 335-360). Greenwich, CT: Information Age Publishing.

Bryson, J., Patton, M., & Bowman, R. (2011). Working with evaluation stakeholders: A rationale, step-wise approach and toolkit. *Evaluation and Program Planning*, *34*, 1-12.

Canadian International Development Agency. (2004). *CIDA evaluation guide*. Ottawa.

Canadian Journal of Program Evaluation (CJPE). (2015).. *Special issue*, *30*.

Carden, F., & Alkin, M. C. (2012). Evaluation roots: An international perspective. *Journal of MultiDisciplinary Evaluation*, *8*(17), 108-118.

Carden, F. (2013). Evaluation, not development evaluation. *American Journal of Evaluation*, *34*(4), 576-579.

Cartland, J., Ruch-Ross, H. S., Mason, M., & Donohue, W. (2008). Role sharing between evaluators and stakeholders in practice. *American Journal of Evaluation*, 29(4), 460-477.

Center for Evaluation Innovation (CEI), Institute for Foundation and Donor Learning (IFDL), Dorothy A. Johnson Center for Philanthropy, & Luminare Group. (2017). *Equitable evaluation framing paper.* Equitable Evaluation Initiative. Available at www.equitableeval.org.

Center for Global Development (CGD). (2006). *When will we ever learn? Improving lives through impact evaluation*. Washington, DC.

Chambers, R. (1995). Paradigm shifts and the practice of participatory development. In N. Nelson & S. Wright (Eds.), *Power and participatory development: Theory and practice* (pp. 30-42). London, UK: Intermediate Technology Publications.

Chelimsky, E. (2013). Balancing theory and practice in the real world. *American Journal of Evaluation*, *34*(1), 91-98.

Chiappero-Martinetti, E., (2010). *Verso nuovi concetti e nuove misure di progresso e sviluppo umano*. Available at www.umanam-ente.org.

Chouinard, J. A., & Cousins, J. B. (2009). A review and synthesis of current research on cross-cultural evaluation. *American Journal of Evaluation*, *30*(4), 457-494.

Chouinard, J. A., & Hopson, R. (2015). A critical exploration of culture in international development evaluation. *Canadian Journal of Program Evaluation*, *30*(3), 248-276.

Christie, C. A. (2003). Editor's notes. *New Directions for Evaluation*, *97*, 1-6.

Cornwall, A., & Gaventa, J. (2001). Bridging the gap: Citizenship, participation and accountability. *PLA Notes*, 40, 32-35.

Coryn, C. L., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, *32*(2), 199-226.

Council of the European Union (2015). *Conclusions on culture in the EU's external relations with a focus on culture in development cooperation*. OJ C 417, 15.12.2015.

Cousins, J., B., & Whitmore, E. (1998). Framing participatory evaluation. *New Directions for Evaluation*, *80*, 5-23.

Cracknell, B. (2000). *Evaluating development aid: Issues, problems, solutions*. London, UK Sage.

Cram F., & Mertens, D. M. (2016). Negotiating solidarity between indigenous and transformative paradigms in evaluation. *Evaluation Matters—He Take Tō Te Aromatawai*, *2*, 161-189.

Cullen, A. E., Coryn, C. L., & Rugh, J. (2011a). The politics and consequences of including stakeholders in international development evaluation. *American Journal of Evaluation*, *32*(3), 345-361.

Cullen, A. E., & C. Coryn (2011b). Forms and functions of participatory evaluation in international development: A review of the empirical and theoretical literature. *Journal of MultiDisciplinary Evaluation*, *7*(16), 32-47.

Davies, R., & Dart, J. (2005). *The 'Most Significant Change (MSC) technique': A guide to its use*. Available at http://mande.co.uk/docs/MSCGuide.pdf .

D'Hollander, D., Marx, A., & Wouters, J. (2014). *Integrating human rights into development cooperation: A comparative*

assessment of strategies and practices of donors. Paper no. 15. Leuven: Leuven Centre for Global Governance Studies.

De Toma, C. (2010). *Outreach toolkit*. Availabe at http://www.cso-effectiveness.org.

Dean-Coffey, J. (2018). What's race got to do with it? Equity and philanthropic evaluation practice. *American Journal of Evaluation*, *39*(4), 527-542.

Dijkstra, G., & A., de Kemp (2015). Challenges in evaluating budget support and how to solve them. *Evaluation, 21*(1), 83-98.

Donaldson, S. I. (Ed). (2013). *Emerging practices in international development evaluation. Evaluation and Society*. USA: Charlotte, NC: Information Age Publishing.

Donaldson, S., & Lipsey, M. W. (2006). Roles for theory in contemporary evaluation practice: Developing practical knowledge. In I. Shaw, J. C. Greene, & M. M. Mark (Eds.), *The Handbook of Evaluation: Policies, Programs, and Practices* (pp. 56-75). London, UK: Sage.

Earl, S., Carden, F., & Smutylo, T. (2001). Outcome mapping: Building learning and reflection into development programmes. Ottawa, Canada: International Development Research Centre.

Estrella, M. (2000). An introduction. In M. Estrella et al. (Eds.), *Learning from Change: Issues and experiences in participatory monitoring and evaluation* (pp. 1-14). London, UK: Intermediate Technology Publications.

European Commission (2012). *Budget support guidelines*. European Commission, DEVCO A2.

European Commission (2015a). *2014 report on the implementation of the EU plan of action on gender equality and women's empowerment in development (2010-2015)*.

European Commission (2015b). *Evaluation of EU support to gender equality and women's empowerment in partner countries. Final report volume 1: Main report*.

European Commission (2015c). *Gender equality and women's empowerment: Transforming the lives of girls and women through EU external relations 2016-2020*. SWD (2015) 182 final.

Evalsed (2003). *Guide to the evaluation of socioeconomic development, sourcebook 2*. Avaliable at http://ec.europa.eu/regional_policy/sources/docgener/evaluation/evalsed.

Ferreira, F. H. G., & Ravallion, M. (2008). *Global poverty and inequality: A review of the evidence*. Policy Research Working Paper Series. World Bank.

Fetterman, D. (1994). Empowerment evaluation. *Evaluation Practice*, *15*(1), 1-15.

Fetterman, D. (1996a). Reflections on emergent themes and next steps empowerment evaluation. In D. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation, knowledge and tools for self-assessment and accountability* (pp. 379-384). Thousand Oaks, CA: Sage.

Fetterman, D. (1996b). Empowerment evaluation: An introduction to theory and oractice. In D. Fetterman, S. J. Kaftarian, & A. Wandersman (Eds.), *Empowerment evaluation, knowledge and tools for self-assessment and accountability* (pp. 3-46). Thousand Oaks, CA: Sage.

Fitzpatrick, J. L. (2012). An introduction to context and its role in evaluation practice. *New Directions for Evaluation*, *135*, 7-24.

Frierson, H. T., Hood, S., & Hughes, G. B. (2010). A guide to conducting culturally responsive evaluations. In J. Frechtling (Ed.), *The 2010 user-friendly handbook for project evaluation* (pp.74-96). Arlington, VA: National Science Foundation.

Fuentes-Nieva, R., & Galasso, N. (2014). *Working for the few: Political capture and economic inequality*. Briefing Paper, Oxford, Oxfam International.

Fuhr, H., Gagnon, J., & S. Reinert, E. (2018). A historical overview of development paradigms. In OECD (Eds.), *Perspectives on global development 2019: Rethinking development strategies* (pp.131-183). Paris: OECD Publishing.

Greene, J. (2005). Context. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 82-84). Thousand Oaks, CA: Sage.

Greene, J. (2012). Values-engaged evaluations. In M. Segone (Ed.), *Evaluation for equitable development results* (pp.192-206). UNICEF.

Greene, J., Millett, R., & Hopson, R. K. (2004). Evaluation as democratizing practice. In M. Braverman, N. Constantine, & J. K. Slater (Eds.), *Foundations and evaluation:*

Contexts and practices for effective philanthropy (pp. 96-118). San Francisco, CA: JosseyBass.

Guijt, I. (2010). Accountability and learning. In J. Ubels, N. A. Acquaye-Baddoo, & A. Fowler (Eds.), *Capacity development in practice* (pp. 277-291). London, UK: EarthScan.

Hilhorst, T., & Guijt, I. (2006). *Participatory monitoring and evaluation: A process to support governance and empowerment at the local level*. Amsterdam: Royal Tropical Institute.

Hood, S. (2001). Nobody knows my name: In praise of African American evaluators who were wings. *New Directions for Evaluation*, *92*, 31-44.

Hood, S. (2004). A journey to understand the role of culture in program evaluation: Snapshots and personal reflections of one African American evaluator. *New Directions for Evaluation*, *102*, 21-37.

Hood, S. (2005). Culturally-responsive evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 96-100). Thousand Oaks, CA: Sage.

Hopson, R. K. (2014). Justice signposts in evaluation theory, practice, and policy. *New Directions for Evaluation*, *142*, 83-94.

Hopson, R. K., Kirkhart, K. E., & Bledsoe, K. B. (2012). Decolonizing evaluation in a developing world: Implications and cautions for Equity-focused evaluations. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 59-82). New York, NY: UNICEF.

House, E.R. (1993). *Professional Evaluation: Social impact and political consequences*. Newbury Park, CA: Sage.

House, E.R. (2004). Intellectual history in evaluation. In M. Alkin (Ed.), *Evaluation roots* (pp.218-224). Thousand Oaks, CA: Sage.

House, E. R., & Howe, K. R. (2000a). *Deliberative democratic evaluation checklist*. Available at http://wmich.edu/evalctr/checklists/dd.

House, E. R., & Howe, K. R. (2000b). Deliberative democratic evaluation. In K. E. Ryan & L. DeStefano (Eds.), Evaluation as a democratic process: Promoting inclusion, dialogue, and deliberation. *New Directions for Evaluation*, *85*, 3-12.

International Council on Human Rights Policy (ICHRP). (2012). *No perfect measure: Rethinking evaluation and assessment of human rights work*. Available at http://www.ichrp.org/files/reports/68/181_evaluating_hr_work_report.pdf.

Independent Evaluation Office (IEO). (2015). *How to manage gender-responsive evaluation*. UN Women.

International Fund for Agricultural Development (IFAD). (2015). *Evaluation manual*. Rome: IFAD.

Johnston-Goodstar, K. (2012). Decolonizing evaluation: The necessity of evaluation advisory groups in indigenous evaluation. *New Directions for Evaluation*, *136*, 109-117.

Joint Statement by the Council and the representatives of the Governments of the Member States meeting within the Council, the European Parliament and the Commission on European Union Development Policy (2005). *The European consensus*. 14820/05, 22 November 2005, Brussels.

Kabeer, N. (2010). *Can the MDGs provide a pathway to social justice? The challenge of intersecting inequalities*. Brighton: Institute of Development Studies.

Kirkhart, K. E. (2010). Eyes on the prize: Multicultural validity and evaluation theory. *American Journal of Evaluation*, *31*, 400-413.

Kirkhart, K. E. (2011). Culture and influence in multi-site evaluation. *New Directions for Evaluation*, *129*, 73-85.

Kirkhart, K. E. (2016). Equity, privilege, and validity: Traveling companions or strange bedfellows? In S. Donaldson & R. Picciotto (Eds.), *Evaluation for an equitable society*. Charlotte, NC: Information Age.

LaFrance, J., & Nichols, R. (2010). Reframing evaluation: Defining an Indigenous evaluation framework. *Canadian Journal of Program Evaluation*, *23*, 13-31.

Lee, K. (2004). *The importance of culture in evaluation: A practical guide for evaluators*. Denver, CO: The Colorado Trust.

MacDonald, B. (1974). Evaluation and the control of education. In B. MacDonald, & R. Walker (Eds.), *Innovation, evaluation, research, and the problem of control* (pp. 9-22). Norwich, UK: SAFARI Project, University of East Anglia.

MacDonald, B. (1978). *Democracy and evaluation*. Public address at the University of Alberta Faculty of Education, 17 October, 1978.

Mansuri, G., & Rao, V. (2013). *Localizing development: Does participation work?* Washington, DC: World Bank.

Mathison, S. (2005). Preface. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. xxxiii-xxxv). Thousand Oaks, CA: Sage.

McBride, D. F. (2011). Sociocultural theory: Providing more structure to culturally responsive evaluation. *New Directions for Evaluation*, *131*, 7-13.

Melamed, C. (2012). *Putting inequality in the post-2015 picture*. London, UK: ODI (Overseas Development Institute).

Mertens, D. (2007). Transformative paradigm: Mixed methods and social justice. *Journal of Mixed Methods Research*, *1*, 212-225.

Mertens, D. (2010). Social transformation and evaluation. *Evaluation Journal of Australasia*, *10*(2), 3-10.

Mertens, D. (2012). When human rights is the starting point for evaluation. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 25-38). New York, NY: UNICEF.

Mertens, D. (2016). Assumptions at the philosophical and programmatic levels in evaluation. *Evaluation and Program Planning*, *59*, 102-108.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (1999). *Linee guida per la lotta alla povertà*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2009). *Salute Globale: Principi Guida della Cooperazione Italiana – 2014-2016*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2010a). *La valutazione in itinere ed ex-post dell'APS- Linee guida*.

Ministero degli Affari Esteri (2010b). *Linee Guida per uguaglianza di genere e empowerment delle donne*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2010c). *Linee guida per l'introduzione della tematica della disabilità nell'ambito delle politiche e delle attività della Cooperazione Italiana*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2010d). *Linee guida per la cooperazione italiana sulla democratica ownership*.

Ministero degli Affari Esteri (2010e). *La cooperazione italiana allo sviluppo nel triennio 2010–2012, Linee–Guida e indirizzi di programmazione*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2012a). *Linee Guida sui Minori*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2012b). *Piano efficacia*.

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo, (2012c). *Good humanitarian donorship initiative. Principles and good practice of humanitarian donorship (2012-2015*).

Ministero degli Affari Esteri, Direzione Generale per la Cooperazione allo Sviluppo (2014). *Relazione annuale al Parlamento sull'attuazione della politica di cooperazione allo sviluppo nel 2013*.

Morra, L., Imas, R., & Rist, R. (2009). *The road to results: Designing and conducting effective development evaluations*. World Bank.

Newman, D. L., & Brown, R. D. (1996). *Applied ethics for program evaluation*. Thousand Oaks, CA: Sage.

OECD (2009). *Italy: Development Assistance Committee peer review*. Paris: OECD Publications.

OECD (2012). *Evaluating budget support, methodological approach*. OECD/DAC Network on Development Evaluation.

OECD/DAC (1997). *Evaluation of programs promoting participatory development and good governance*. Paris: OECD/DAC.

OECD/ DAC (2010). *Quality standards for development evaluation*. Paris: OECD.

Ofir, Z., & Kumar, S. (2013). Evaluation in developing countries: What makes it different? In S. I. Donaldson, T. Azzam & Conner (Eds.), *Emerging practices in international evaluation* (pp.11-24). Charlotte, NC: Information Age Publishing.

OHCHR (2004). *Human rights and poverty reduction: conceptual framework*. New York, NY: United Nations.

Ottier, A. (2005). Participatory evaluation in the context of cbpd: theory and practice in international development. *Canadian*

Journal of Program Evaluation, *20*(1), 123-148.

Patton, M. Q. (1997). *Utilization-focused evaluation: The new century text*. Thousand Oaks, CA: Sage.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. London, UK: Sage.

Peersman, G. (2014). Evaluative criteria. *Methodological Briefs Impact Evaluation*, *3*. UNICEF.

Peersman, G., & Rogers, P. (2016). Addressing equity issues in monitoring and evaluation. *Development Bulletin*, *77*, 37-42.

Picciotto, R. (2015). Bringing inequality back in from the cold: Toward a progressive evaluation model. In R. Rist, F. P. Martin, & A. M. Fernandez (Eds.), *Poverty, inequality, and evaluation: Changing perspectives* (pp. 253-277). Washington, DC: World Bank Group.

Podems, D. (2010). Feminist evaluation and gender approaches: There's a difference? *Journal of MultiDisciplinary Evaluation*, *6*(14), 1-17.

Poverty Analysis Discussion Group. (2012). *Understanding poverty and wellbeing: A note with implications for research and policy*.

Rog, D. (2015). Infusing theory into practice, practice into theory: Small wins and big gains for evaluation. *American Journal of Evaluation*, *36*(2), 223-238.

Rogers, P. (2012a). Methodological challenges in using programme theory to evaluate pro-poor and equity-focused programmes. In M. Segone (Ed.), *Evaluation for equitable development results* (pp. 142-171). New York, NY: UNICEF.

Rogers, P. (2012b). Introduction to impact evaluation. *Impact Evaluation Notes*, *1*. Interaction.

Rosenstein, B., & Desivilya, H. (2015). Editor's note. *New Directions for Evaluation*, *146*, 3-7.

Samuels, M., & Ryan, K. (2011). Grounding evaluations in culture. *American Journal of Evaluation*, *32*(2), 183-198.

Schwandt, T. (2015). *Evaluation foundations revisited: Cultivating a life of the mind for practice*. Stanford, CA: Stanford University Press.

Schwandt, T. (2018). Acting together in determining value: A professional ethical responsibility of evaluators. *Evaluation*, *24*(3), 306-317.

Segone, M. (2012a). *Evaluation for equitable development results*. New York, NY: UNICEF.

Segone, M. (2012b). Evaluation to accelerate progress towards equity, social justice and human rights. In Segone (Ed.), *Evaluation for equitable development results* (pp.8-12). New York, NY: UNICEF.

Seigart, D. (2005). Feminist evaluation. In S. Mathison (Ed.), *Encyclopedia of evaluation* (pp. 154-157). Thousand Oaks, CA: Sage.

Sen, A. (1999). *Development as freedom*. Oxford, UK: Oxford University Press.

Sen, A. (2005). Human rights and capabilities. *Journal of Human Development and Capabilities*, *6*(2), 151-166.

SenGupta, S., Hopson, R., & Thompson-Robinson, M. (2004). Cultural competence in evaluation: An overview. *New Directions for Evaluation*, *102*, 5-19.

Solheim, E. (2012). Introduction. In OECD, *Evaluating development activities: 12 lessons from the OECD DAC* (pp.7-8). Paris: OECD.

Stake, R. E. (1975). *Program evaluation: Particularly pesponsive evaluation*. Occasional paper series no. 5, Western Michigan University.

Stake, R. E. (2004a). How far dare an evaluator go toward saving the world. *American Journal of Evaluation*, *25*(1), 103-107.

Stake, R. E. (2004b). *Standards-based and responsive evaluation*. Thousand Oaks, CA: Sage.

Stame, N. (1998). Evaluation in Italy: experience and prospects. *Evaluation*, *4*(1), 91-103.

Stame, N. (2018). Strengthening the ethical expertise of evaluators. *Evaluation*, *24*(4), 438-451.

Start, D., & Hovland, I. (2007). *Tools for policy Impact: A handbook for researchers*. ODI.

Stern, E. (2006). Contextual challenges for evaluation practice. *Sage Handbook of Evaluation* (pp. 292-314). London, UK: Sage.

Stern, E. with contributions from Altinger, L., Feinstein O., Marañón M., Nils-Sjard Schultz, N. & Steen Nielsen N. (2008). *Thematic study on the Paris Declaration, aid effectiveness and development*

effectiveness. Copenhagen: Ministry of Foreign Affairs of Denmark.

Thompson-Robinson, M., Hopson, R., & SenGupta, S. (2004). In search of cultural competence in evaluation: Toward principles and practices. *New Directions for Evaluation*, *102*, 39-50.

Trésor (2014). *Lignes de crédit en faveur des PME-PMI en Tunisie et dans les territoires palestiniens. Unité d'Evaluation des Activités de Développement*. Available at www.tresor.economie.gouv.fr/evaluation-des-activites-de-developpement.

United Nations Department of Economic and Social Affairs (UNDESA). (2013). *Inequality matters: Report of the world social situation 2013*. New York, NY: UN.

United Nations Educational, Scientific and Cultural Organisation (UNESCO). (2001). *UNESCO universal declaration on cultural diversity*. Available at: https://www.refworld.org/docid/435cbcd64.html.

United Nations High Level Panel Report on Post-2015 Development Agenda. (2013). *A new global partnership: Eradicate poverty and transform economies through sustainable development*. New York, NY: UN.

United Nations Task Team on the Post 2015 UN Development Agenda. (2012). *Addressing inequalities: the heart of the 2015 agenda and the future we want for all*. Thematic think piece, ECE, ESCAP, UN DESA, UNRISD, UN Women.

United Nations Women. (2014). *Gender mainstreaming in development programming*. New York, NY: UN Women.

United Nations Development Programme (UNDP). (2009). *Handbook on planning, monitoring and evaluating for development results*. New York, NY: UNDP.

United Nations Development Programme (UNDP). (2013). *Humanity divided: Confronting inequality in developing countries*. New York, NY: UNDP.

United Nations Development Programme (UNDP). (2014). *(HDR), Sustaining progress: Reducing vulnerabilities and building resilience*. New York, NY: UNDP.

UNDP, UNFPA, UNICEF, WFP (2011). *Joint Meeting of the Executive Boards Equity: narrowing the gaps to push for achievement of the Millennium Development Goals*.

UNDP, UNOPS (2009). *Regional level. Governance and Local Development. Art Gold 2 Albania*.

UNICEF, UN Women. (2013). *Global thematic consultation on the post-2015 development agenda: Addressing inequalities–Synthesis report of global public consultation*. UNICEF & UN Women.

United Nations (UN). (2011a). *The evaluation policy of UNDP*. DP/2011/3.

United Nations (UN). (2011b). *Delivering as one for development, one UN programme*. Albania: UN.

United Nations (UN). (2015). *Transforming our world: The 2030 agenda for sustainable development*. Available at A/RES/70/1. http://undocs.org/A/RES/70/1.

United Nations Evaluation Group (UNEG). (2008). *Ethical guidelines for evaluation*. New York, NY: UNEG.

United Nations Evaluation Group (UNEG). (2013). *Impact evaluation in UN agency evaluation systems: Guidance on selection, planning and management*. New York, NY: UNEG.

United Nations Evaluation Group (UNEG). (2014). *Integrating human rights and gender equality in evaluations*. New York, NY: UNEG.

United Nations Evaluation Group (UNEG). (2016). *Norms and standards for evaluation*. New York, NY: UNEG.

United Nations Research Institute for Social Development (UNRISD). (2010). *Combating poverty and inequality*. Geneva: UNRISD.

Valutazioni, retrieved February 26, 2015, from http//www. cooperazione/Leggi e procedure/Valutazioni.

Vogel, I. (2012). Review of the use of Theory of Change in International Development: Review report. DFID.

Weiss, C. H. (1983). The stakeholder approach to evaluation: Origins and promise. *New Directions for Program Evaluation*, *17*, 3-14.

World Health Organisation. (WHO). (2013). *WHO evaluation practice handbook*. Geneva: World Health Organization.

Woolcock, M. (2014). *Culture, politics, and development*. Policy Research, Working Paper 6939.

World Bank. (2006). *World development report, equity and development*. World Bank,Oxford University Press.

World Bank. (2013). *Inclusion matters: The foundation for shared prosperity*. Washington, DC: World Bank.