


Power dynamics in international development evaluations: A case study of the Girls Education Challenge programme

**Author:**Rosie O. Emerson¹ **Affiliation:**¹Department of Education,
University of Massachusetts,
Lowell, United States of
America**Corresponding author:**Rosie Emerson,
rosiemondy@gmail.com**Dates:**

Received: 10 Jan. 2020

Accepted: 06 July 2020

Published: 27 Aug. 2020

How to cite this article:Emerson, R.O., 2020, 'Power
dynamics in international
development evaluations:
A case study of the Girls
Education Challenge
programme', *African
Evaluation Journal* 8(1), a459.
<https://doi.org/10.4102/aej.v8i1.459>**Copyright:**© 2020. The Authors.
Licensee: AOSIS. This work
is licensed under the
Creative Commons
Attribution License.

Background: Low-income countries receive millions of dollars of aid each year in support of international development programmes. These programmes often have a requirement for evaluation. Evaluators are therefore uniquely placed to contribute to the social and economic development of these countries by conducting useful evaluations. This study elaborates on the power dynamics involved in international development evaluations so that evaluators can be better positioned to conduct impactful evaluations.

Objectives: The objective of this study was to explore how power is configured and distributed amongst key stakeholders and how power imbalances impact evaluations.

Method: The study utilized the Girls Education Challenge (GEC) programme as a case study. GEC was a multi-million-dollar program that supported 37 education projects in 18 countries in Africa and South Asia. Interviews were conducted with 13 evaluators and 10 programme representatives - staff from organizations that were part of the GEC programme.

Results: The study concluded that donors wield significant power over evaluations and there are few avenues for less powerful stakeholders to speak truth to power. At best, donors can help to increase the quality and utilisation of evaluations. At worst, they can hinder culturally responsive evaluation practices. Furthermore, the status quo favours international evaluators and utilises local researchers merely as hired hands – overseeing the logistics of data collection.

Conclusion: The main implication of this study is that evaluators need to conduct formal or informal power analyses in order to identify power asymmetries and potential power sharing opportunities and strategies.

Keywords: culturally responsive evaluations; international development evaluations; power; programme evaluation; Africa; South-East Asia; case study.

Introduction

The World Bank (2017) defines low-income countries as those with a gross national income (GNI) per capita of \$1025 or less in 2018. Low-income countries have been the target of development aid for decades. For instance, in 2018 approximately \$55 billion aid was sent to low-income countries compared with about \$215 million to high-income countries (World Bank Indicator n.d.). Development aid is often used to fund international development programmes – donor-funded programmes aimed at the economic, social and political development of low-income countries. These programmes typically have an evaluation requirement. Evaluators are therefore in a unique position to contribute to the development of low-income countries, speak truth to power and transform lives (Naidoo 2013).

However, low-income countries present socially, politically and economically complex environments to carry out evaluations (Chouinard & Cousins 2013). The significant dependence on donor funds (Takyi-Amoako 2012) engenders a situation where evaluations are considered high-stake activities that could result in the termination of much-needed funding (Azzam & Levine 2015). In addition, evaluation is often viewed as being externally imposed and there is sometimes resistance or compliance without real buy-in (Bhola 2003). Moreover, evaluation involves a complex system of accountabilities, which is further complicated by imbalances of power and control. There is accountability to donors and sponsors whose primary interest in evaluation is in making sure their investment was well spent (Chouinard & Cousins, 2013; Carden, 2009; Carden, 2013; Horton, 1999; Horton & Mackay, 2003). There is also accountability to local governments and project beneficiaries

Read online:Scan this QR
code with your
smart phone or
mobile device
to read online.

(Chouinard & Cousins, 2013). Evaluators face tensions in meeting the conflicting needs and interests of diverse stakeholders (Chouinard & Cousins 2013). Unfortunately, when faced with such tensions, evaluators often implement evaluations whose goals and methods favour the most powerful stakeholders, that is, donors (Bamberger 1999; Cullen, Coryn & Rugh 2011).

Power asymmetries in international development evaluations have not received much attention in the evaluation literature. It is assumed that collaborative arrangements inevitably result in greater inclusion or pro-poor policy change (Gaventa, 2006). Therefore, development discourse often talks about participatory research without paying sufficient attention to the power relations within and surrounding collaborative arrangements (Gaventa, 2005). However, as evaluators, we must critically examine power dynamics in and reflect on whether our engagement re-legitimizes the status quo or challenges power relationships that contribute to patterns of exclusion and social injustice (Brugnach & Dewulf, 2017; Gaventa, 2005).

Therefore, in this article, the author argues that power asymmetries are important considerations in international development evaluations. Evaluation is not objective or value neutral. Evaluators need to take responsibility for their positioning, understand whose interests are being served by their work and reflect on how the outcomes they are measuring might be sustaining an unjust status quo (Greene 2001; Trimble et al. 2012). Greene (1997) went as far as to state that advocacy in evaluation is inevitable, as evidenced by whose questions we answer, who sets the criteria for determining merit or worth and whether we leave programme assumptions unchallenged or not.

To further the exploration of power in international development evaluations, this study focuses on the Girls Education Challenge (GEC) programme as a case study. The remainder of this article is structured as follows. Firstly, readers are introduced to the GEC programme. Next, relevant literature on power asymmetries in collaborations is provided, followed by the methodology used in the study. The results of the study are then discussed, starting with the characteristics of the three key GEC stakeholders, their sources of power (or powerlessness) and their ways of dealing with power asymmetries. Positive and negative impacts of power asymmetries in the GEC evaluation are also discussed. Finally, implications for evaluation practice are discussed. In particular, the author encourages researchers and programme stakeholders to conduct formal or informal power analyses in order to explore power-sharing opportunities.

Description of the Girls Education Challenge programme

The first phase of the Girls Education Challenge (GEC) programme was implemented between 2012 and 2017. The GEC was a £355m programme that supported 1 111 320

marginalised girls with improved learning outcomes (Coffey 2017a, 2017b, 2017c). Marginalised girls were defined as girls aged 6–19 years who had not been enrolled, or had dropped out of school or were in danger of doing so (Coffey 2016). The GEC programme was funded by the United Kingdom's (UK) Department for International Development (DfID) and supported 37 projects in 18 countries across sub-Saharan Africa and South Asia. Funding was administered in three funding windows: the first window funded projects that were large and well established in order to scale these initiatives, the second window funded innovative educational initiatives and the third window funded sustainable, commercial models that required matching funds from private sector partners. The GEC programme sought to improve retention, attendance, enrolment and learning of disadvantaged girls, and it was the largest global fund dedicated to girls' education (GEC n.d.). The second phase of GEC (GEC-Transition or GEC-T) provided additional funding to several of the GEC projects to address issues related to transition. The data collection for this study was initiated immediately after programmes had completed their baseline evaluations for GEC-T. Participants were therefore asked to draw on their experiences with both GEC and GEC-T.

The programme evaluation was managed by a fund manager (FM; PricewaterhouseCoopers) and an evaluation manager (Coffey International) (Coffey 2016). The FM oversaw the day-to-day operations of the programme, whereas the evaluation manager was responsible for designing and implementing a rigorous monitoring and evaluation (M&E) framework to assess the effectiveness and impact of individual projects and the GEC as a whole.

Each GEC project was required to contract external evaluators to conduct baseline, midline and endline evaluations. The evaluation was designed to be rigorous and followed a highly prescriptive approach. Programmes were issued with 300 pages of comprehensive guidelines, which included everything from logframe templates and an overview of monitoring, evaluation and learning to guidelines on designing learning outcomes assessment tools. Programmes were also issued a 55-page reporting template describing the specific sections of the report, tables and appendices that had to be included.

All projects worked towards the same high-level GEC outcomes of improved enrolment, retention, attendance and learning for marginalised girls. Evaluation data for all projects also included intervention and control areas. Projects were required to use a representative, longitudinal household survey of target and control communities and/or the longitudinal tracking of school-based cohorts and structured qualitative research. The evaluation manager provided technical support and guidance to GEC projects to ensure that their M&E frameworks and data collection strategies were fit for the purpose. This high level of standardisation was meant to ensure rigour and to allow for comparisons and generalisable observations.

Literature review: A critical perspective of power

According to Brugnach and Dewulf (2017:3), power is defined as 'the capacity of social actors to influence decisions'. Dahl (1957:202–203) defined power as follows: 'A has power over B to the extent to which he can get B to do something that B would not otherwise do' or it is power that prevents somebody from doing what he or she wants to do (Bachrach & Baratz 1962). Ran and Qi (2019:4) defined power as the 'potential ability of controlling or influencing others (individuals, groups, [or] organizations)'.

To fully understand the concept of power, especially in collaborations such as the GEC programme, this study draws on the literature on interorganisational collaborations, which Gray (1989) defined as a process between interdependent organisational actors who negotiate answers to shared concerns; collaborative governance, which is defined as multiorganisational arrangements where diverse stakeholders from various sectors are involved in collective decision-making processes to achieve shared goals (Ansell & Gash 2008; Brugnach & Dewulf 2017; Ran & Qi 2018); and interorganisational domains where different organisations perceive themselves to be connected to common issues (Hardy & Phillips 1998). Much of the literature on collaboration has emphasised the benefits (e.g. Alter 1990; Alter & Hage 1993; Gray 1989; Nathan & Mitroff 1991). There is a strong underlying assumption of equity, fairness and balancing of interests in collaborations (Gray 1985). However, the present article draws on research that addresses power asymmetries in collaborative governance (e.g. Gray 1985; Gray & Hay 1986; Hasenfeld & Chesler 1989; Vangen & Huxham 2005; Ran & Qi 2018; Rose & Black 1985).

The literature talks about different sources of power: formal authority or structural power (the legitimate right to make decisions, control the agenda and frame the problem), resource control or instrumental power (the ability to deploy resources) and discursive legitimacy (actors who are understood to speak legitimately for issues and organisations) (Altheide 1988; Dutton & Duncan 1987; Gray & Hay 1986; Hardy & Phillips 1998; Lukes 2005; Phillips & Hardy 1997; Purdy 2012). Power can also come from differences in access to knowledge (Brugnach & Dewulf 2017). It is also important to remember that power relationships in collaborative networks are dynamic – they evolve and vary over time.

Researchers have suggested several ways to deal with power asymmetries in collaborations. One of the proposed solutions is power sharing (Ansell & Gash 2008; Berkes 2010; Ehler 2003; Gray 1989; Grindle 2004; Huxham & Vangen 2000; Jentoft, Van Son & Bjørkan 2007; Moynihan 2009; Purdy 2012; Winer & Ray 1994). Power sharing 'is a process of sharing responsibility for decision making and actions among

stakeholders in collaboration' (Ran & Qi 2018:837). Ran and Qi (2018) discussed six factors that promote effective collaboration and beneficial power sharing:

1. Trust in the institutional system, which can be built through regulations, contracts and guarantees that all help to reduce uncertainty.
2. Stakeholders are more willing to invest time and effort in power sharing when the mission is long term rather than exigent.
3. The level of mutual consent, reciprocity and trust is lower in a mandated network than in voluntary networks.
4. Power sharing is more effective when stakeholders have successful previous collaboration experience and capacities in negotiation, strategy building, visioning and professional knowledge.
5. The less diffuse power sources are, the less effective power sharing is.
6. Participants are more willing to invest time and energy involved in sharing power when the benefits outweigh the costs. For instance, less powerful stakeholders may voluntarily choose to give up some of their power in exchange for less accountability.

Other researchers are less optimistic about power sharing. For instance, Hardy and Phillips (1998) stated that power sharing can result in the loss of control over direction of change, greater time and effort to manage relationships, and increased risk of escalation of conflict. They instead focused on discursive legitimacy as a more viable option for less powerful stakeholders (Phillips & Hardy 1997). Purdy (2012) agreed and suggested using coalitions to expand participation and augment discursive power. Discursive power seems promising but is unfortunately difficult to identify because it involves looking beyond the visible manifestations of power and deeply analysing the dominant discourse, how influence is being exercised and by whom (Brisbois & Loe 2016). Power does not always manifest in overt ways; sometimes less powerful stakeholders comply with actions that they think more powerful counterparts want to see (Hardy & Phillips 1998). Gaventa (2005) defined invisible power as internalised powerlessness where the status quo seems normal. Invisible power shapes meaning and determines what is acceptable. Critical social theorists also state that modern societies promote one dominant way of thinking and that society needs to constantly reflect and critique these dominant ways of thinking by analysing people's roles and experiences within these modern systems (Freeman & Vasconcelos 2010). Fay (1987) also discussed the theory of false consciousness, which states that the oppressed:

[H]ave internalized the values, beliefs, and even world view of their oppressors ... [and] willingly cooperate with those who oppress them in maintaining those social practices that result in their oppression. (p. 107)

The insidious nature of power, especially discursive power, can make it challenging for less powerful stakeholders to recognise it, combat it and build their own discursive legitimacy.

Other researchers emphasise the need to understand how power works before engaging in questions related to power sharing. For instance, Gaventa (2005) suggested the power cube approach that has been used in power analysis workshops with donor agencies in international development contexts. His framework looks at three dimensions of power: place (global, national and local), spaces (closed, invited and claimed or created) and power (visible, hidden and invisible). Purdy (2012) offered a framework where each source of power (formal authority, resources and discursive legitimacy) is mapped onto the arenas for power use, namely, participants, process design and content. This framework helps to expose how power can be used during a collaborative process and conversely how the process influences and shapes the exercise of power (Purdy 2012). The author suggested that assessments of power should be done collectively and openly. Reed (2008) argued for the highly skilled facilitation of stakeholder engagement processes. Brisbois and Loe (2016:22) also offered some aspects we need to attend to that we can understand how power might be shifted or shared. More importantly, they advised that we consider how and by whom collaborative agendas are set; the financial, technical and institutional capacities of actors and how they are utilised; the knowledge, information and perspectives that are used and valued; and the dominant societal values in the context in question.

Methodology

This qualitative study focused on the GEC programme as a case study. The sample for the study included all the evaluators and programme representatives who agreed to participate.

Data collection

The Institutional Review Board determined that this study was exempt from full review due to the minimal risk posed to study participants. The researcher shared an invitation letter with the GEC point of contact who then contacted all evaluators and evaluands to invite them to participate in the interviews. The recruitment e-mail was sent to 112 people, which included 61 programme representatives (staff from organisations that participated in the GEC programme) and 51 evaluators. In total, 23 participants agreed to participate, which included 13 evaluators and 10 programme representatives (see Table 1 for participant demographics). In most cases, the programme representatives who participated in the study were M&E officers within their organisation. The researcher e-mailed these participants the consent form prior to conducting interviews. No written consent was obtained in order to maintain their privacy and anonymity. Only verbal consent was required.

All interviews were conducted over the phone or Skype and were audio-recorded. Prior to starting the interview, the researcher asked whether participants had any questions

regarding the consent form. Verbal consent was then provided to proceed with the interview. Each interview lasted for about one hour and covered the following topics: ways in which evaluations have been used in GEC programmes; factors influencing evaluation use; and knowledge, skills and attitudes needed to conduct successful evaluations in low-income countries. The recordings were transcribed and later analysed.

Data analysis

Transcripts were imported into NVivo 12 software program and analysed using constant comparison analytical methods (Savin-Baden & Major 2013). During the first phase of coding (open coding), text was highlighted and codes were assigned. Similar codes were applied to similar ideas. During open coding, any excerpts that touched on stakeholders, such as the donor, implementing organisation, community or government, were coded into a code titled 'power'. At this stage, the researcher did not pay much attention to what exactly the excerpts meant or whether they would be significant. During the open coding stage, it was important to simply gather all information on power without bias or selectivity and then subject all the data to further scrutiny. The second phase of coding (axial coding) involved grouping together codes that conveyed similar ideas or themes. The themes were then further analysed and interpreted, which resulted in more nuanced consolidation as patterns became more apparent. All the findings of this study are reported anonymously; the names of specific countries are not mentioned to minimise the possibility of identifying respondents.

Ethical consideration

This article followed all ethical standards for research without direct contact with human or animal subjects.

Results

Girls Education Challenge stakeholders, their sources of power and their response to power imbalances

Donor (United Kingdom's Department for International Development) and fund manager (PricewaterhouseCoopers)

Programmes' main interaction with the donor was through the FM, who maintained close contact and communication with organisations throughout the programme implementation and wielded significant power and influence over the evaluation process. The donors' power was derived from their right to make decisions about the programme implementation and evaluation process. They also had control over financial resources and stipulated the terms of receiving funding. Furthermore, their leadership of a large and globally visible project gave them discursive power. The donors used their power to mandate strategic but generic implementation directions for a multi-country,

Table 1: Projects represented in this study.

Participants	GEC project size in terms of funding (in Euros)	Project location	Participant demographics
Evaluators			
1	> 10m	East Africa	International research firm with local offices. The firm has extensive experience in conducting GEC evaluations. Technical leadership provided by expatriates/international researchers.
2	> 10m	East Africa	
3	1m – 5m	East Africa	
4	> 10m	East Africa	European evaluator with 20 years' experience as an independent consultant. Worked in collaboration with a local data collection firm.
5	1m – 5m	East Africa	Local evaluation firm. Principal investigator has 20 years' experience and strong relationships with government officials. Girls Education Challenge evaluation team comprised nine people, four of whom were from the United States of America or the United Kingdom.
6	1m – 5m	East Africa	
7	1m – 5m	East Africa	
8	1m – 5m	Southern Africa	International research firm with regional offices. The firm not only evaluates projects but also implements development projects. The GEC evaluation team was led by international and/or internationally trained researchers. Data analysts were local.
9	> 10m	East Africa	North American consultant who owns a small research firm that has been in operation for over 20 years. The evaluator has extensive experience working on development evaluations and works with local data collection firms. The GEC team was supported by 12 international researchers with experience working in Africa.
10	> 10m	Southern Africa	
11	1m – 5m	Southern Africa	
12	1m – 5m	East Africa	Local evaluation firm that is an association of over 100 members, most of whom are drawn from the local university. Project principal investigator is a university lecturer with 10 years' experience in evaluation. The GEC evaluation team comprised 20 senior researchers who were all nationals. Research assistants were recruited locally.
13	> 10m	Asia	Principal investigator is a North American professor with no previous experience in conducting evaluations in developing countries. He is an independent consultant working with an international research firm. The team worked with local data collection firm.
Programme representatives			
14	1m – 5m	West Africa	Large international for-profit corporation.
15	< 1m	East Africa	Small international NGO
16	1m – 5m	East Africa	Large international NGO
17	6m – 10m	West Africa	Large international NGO
18	> 10m	East Africa	Large international NGO (partnering with local NGOs)
19	> 10m	Southern Africa	Large international NGO
20	> 10m	Southern Africa	Large international NGO
21	1m – 5m	East Africa	Small local NGO
22	1m – 5m	East Africa	Large international NGO (partnering with local NGOs)
23	1m – 5m	Asia	Large international NGO

GEC, Girls Education Challenge; NGO, non-governmental organisation;

multi-organisation effort; to focus on learning from project experiences; and to push for a prescriptive evaluation design to allow testing of core set of outcomes in different organisational contexts.

In total, 14 evaluators and programme representatives expressed the belief that the FM was primarily interested in the accountability role of the evaluation. Two evaluators explained that the funding for GEC came from UK taxpayers, and thus the donor (DfID) wanted to be able to share favourable feedback about the impact of the United Kingdom in improving educational outcomes for marginalised girls:

'I think that DfID wanted to see big results in a short timeframe because they're a bilateral donor, and they have to report to the Parliament. Wouldn't it be great if they had this transformative effect across the globe on girls' learning outcomes?' (Participant 3, Evaluator, Male)

Consequently, the donor focused heavily on measuring learning outcomes. Learning was defined as increased oral fluency and increased math scores as measured by performance on Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA). Early Grade Reading Assessment was designed by the Research

Triangle Institute (RTI) in 2006 to measure and report on students' acquisition of five early reading skills: letter-sound identification, invented word reading, oral reading fluency, listening and reading comprehension (RTI International 2009). Building on the success and demand for EGRA, RTI designed EGMA in 2008 to develop an assessment of early grade mathematics competencies (Platas et al. 2014). These tests have been used extensively by international development organisations to assess math and reading skills.

Implementing organisations

The implementing organisations varied in size. Three were small non-governmental organisations (NGOs) for whom the GEC grant was the largest grant they had ever received. The other organisations were large international organisations with significant levels of funding. Implementing organisations also had different ways of working in low-income countries. For instance, representatives from three international NGOs stated that they did not have in-country offices; rather, they applied for grants and worked through local implementing partners to deliver the programmes and activities. In contrast, representatives from two international organisations stated that they have local offices and they carry out the work themselves and in partnership with

other local organisations. Regardless of the size, all the organisations implemented GEC projects with multiple complementary interventions. This generalist approach allowed the organisations to take advantage of different funding opportunities and increased their chances of receiving large grants, such as the GEC grants.

With respect to programme evaluation, three programme representatives stated that one of the main challenges they face is retention of staff working for their organisations. In low-income countries, development-related jobs are often the most lucrative ones and demand for highly qualified people far surpasses supply, which makes it difficult for organisations to retain qualified staff. Furthermore, when contracting external evaluators, implementing organisations struggled to distinguish between high- and low-quality evaluators. Two programme representatives expressed the wish that the donor should maintain a list of vetted or preferred evaluators similar to how they maintain a list of preferred suppliers.

The implementing organisations also had different levels of M&E capacity. Four representatives from large international NGOs stated that they had the required financial and human resources to conduct several internal evaluations and rigorous monitoring. For these organisations, external evaluations were simply additive, in that they helped to confirm, contradict or elaborate on what the organisations already knew. These individuals also stated that they valued evaluations highly and had systems and processes in place to learn from evaluations. On the contrary, two representatives from the small NGOs stated that they had limited internal capacity and relied heavily on external evaluations to provide them with critical information on what was and was not working effectively within their programmes. From this study, it appeared that smaller organisations were much more reliant on external evaluations as a source of information and yet had far fewer resources to build robust knowledge management systems.

All the organisations that participated in this study were primarily interested in implementing quality projects for the benefit of vulnerable populations. These organisations were keen on making the most of their grant money, and three programme representatives questioned the wisdom and ethics of spending such a significant amount of money on a rigorous and demanding evaluation when valuable lessons could be gleaned from smaller, cheaper evaluations that would free up more resources to help even more people.

Organisations were also interested in developing and cementing relationships with communities and national governments to facilitate their development work. For instance, one organisation put in place community feedback mechanisms to ensure communities benefitted from the evaluation. Another organisation tried to involve the government in various stages of the evaluation to secure their continued buy-in and support. A third organisation

modified its programming to also benefit boys in the community so as to garner increased local support for their interventions.

Organisations' response to power: Attitudes towards the FM varied. Organisations' experience with their FM depended on the cultural competence of their FM point of contact. Five programme representatives noted that the FM staff sometimes lacked cultural knowledge and experience and did not show sensitivity to the context. For instance, one representative said:

'In the initial negotiations, there was a smart young statistician who was leading the negotiations on the PwC [PricewaterhouseCoopers] side. I don't think she'd ever stepped foot in Africa. She had an incredibly unrealistic view of what was possible. She had this kind of "take it or leave it" attitude.' (Participant 19, Programme representative, Female)

However, in one South Asian country, a programme representative mentioned that the FM was extremely collaborative, flexible and helpful. In that country, the FM supported all the GEC projects and encouraged collaboration and learning amongst them. Although most of the programme representatives expressed that the GEC evaluation was very top-down and the FM exercised excessive control over the evaluation process, this participant did not believe the evaluation was prescriptive; rather, she felt that the evaluation provided enough room for organisations to define their goals and objectives, thereby ensuring that organisations' information needs were prioritised.

Many of the organisations tried to push back against the FM's power and influence, especially regarding measurement of learning outcomes. The purposes, uses and limitations of EGRA are outlined in the EGRA Toolkit (RTI International 2009). These guidelines state that EGRA may be used for diagnostic purposes to improve reading instruction, but should not be used as a high-stake accountability measure to arrive at funding decisions (RTI International 2009:16). However, in GEC, EGRA and EGMA were used to determine funding decisions (Miske & Joglekar 2018). A system of Payment by Results (PbR), based on performance on EGRA and EGMA, was put in place. In this system, project performance was measured at midline and endline to assess whether the treatment groups had achieved pre-set PbR targets. If they failed to meet the targets, 10% – 20% of the organisation's budget was withheld. If a programme exceeded the target, they received a bonus. Several projects failed to meet their PbR targets and lost funding.

Four programme representatives stated that their organisations contested the appropriateness of using EGRA and EGMA as the measure for learning. One organisation was able to bring in experts to argue their case with evidence. Ultimately, the organisation was allowed to use national exams in addition to the required EGRA and EGMA. However, this was not deemed to be an ideal resolution because it placed enormous burdens on field teams and beneficiaries.

Some organisations also issued a joint statement expressing concern about the management of GEC projects. They raised a number of issues, including excessive evaluation demands, which were over-stretching resources and detracting from project implementation, collection of excessive amounts of data when only a small portion of it was going to be used, timing of midline evaluations too close to the baseline evaluations (6 months in some cases), which provided no time to learn from baselines or to effect noticeable change, and donor insistence on collection of data on disability despite the potential for further stigmatisation and organisations' lack of expertise in the area of data collection on disability. Two programme representatives stated that their concerns went largely unaddressed, except for PbR, which was removed in the second phase of GEC.

In general, implementing organisations submitted to the power of the donor and did what they could to meet all the requirements, no matter how onerous they deemed them. One evaluator said:

'For some programs, they're terrified of the evaluation and they're just like, "God, just write to the template please and make it work", because they are complicated and at some level maybe beyond the skill set of that office or that team.' (Participant 2, Evaluator, Male)

Two organisations went as far as changing their programming to meet donor expectations. For instance, one of the programmes has a special focus on securing children's rights and previously did not have any interventions that directly influenced teaching and learning. However, with the emphasis on learning outcomes as a measure of programme success, they included interventions that directly impact teaching and learning.

Another consequence was a cultural shift within one organisation, which started prioritising performance over learning. This particular organisation hired a new M&E manager who was strong quantitatively but not a good fit in terms of leading a learning organisation, because the organisation felt they needed someone who could handle rigorous evaluations more than they needed someone who could help institutionalise learning.

Evaluators

In total, 13 evaluators participated in this study. Some were independent consultants who were contracted by evaluation firms to lead the evaluation. Others were full-time staff members in established research firms. The research firms also varied. Some were small, with less than 10 full-time staff members, and others were large evaluation firms that consistently undertook large, complex evaluations. In most cases, technical leadership and oversight of the evaluation was conducted by a team of international researchers (from the United States of America, the United Kingdom, Canada or Australia). These international evaluators often worked in partnership with local research firms that were responsible for the logistics of data collection. The relationship between

international and local firms was sometimes problematic because of trust and contracting issues. For example, in one project, issues between a UK research firm and a local data collection firm resulted in a 9-month delay because the programme had to rebuild the bridges between the external evaluator and the local data collector.

The primary interest of evaluators was to apply their research skills and knowledge on interesting, challenging and meaningful projects. They also experienced financial gain, as they were paid for their services. However, four evaluators mentioned that the funding for their evaluations was not commensurate with the level of effort required, and thus they found themselves providing many hours of free labour.

The evaluators unanimously reported that this was the most demanding evaluation they had ever participated in, as they were not accustomed to receiving so much guidance on evaluations. All the evaluators were taken aback by the highly prescriptive and demanding nature of the evaluation. They stated that the Terms of Reference did not go into detail about the work involved, and the guidelines and expectations kept changing. Furthermore, in their previous work, clients had often given evaluators discretion over the evaluation design and reporting. However, they were now in a situation where they had to adhere to strict guidelines. The contracting arrangements were also designed in such a way that even though evaluators were hired by programmes, their reports were signed off by the FM. Evaluators shared instances where the organisation or client was pleased with the report, but the evaluators still found themselves going through numerous rounds of feedback and iterations because the FM was not satisfied. Two evaluators stated that there was a sense the report was for the donor rather than for the organisation:

'They handed down a set of guidance documents about the evaluation and how it should be conducted. They handed down a template report and we addressed questions to that. Once we got down into the actual tool design and setting up the project and also doing the evaluations, of course clients had their two cents. Their input came in, but it came in in a secondary fashion. It was definitely around the edges where their input was relevant.' (Participant 5, Evaluator, Male)

Evaluators' response to power: In some instances, the evaluators pushed back against the FM's power. For instance, in one case, the evaluators could not go to certain places because of conflict and instability. The FM initially did not understand why they could not go to these areas and suggested working with local individuals in those regions, but the evaluator was adamant about avoiding those areas entirely. Even the programme experienced implementation challenges in those areas, and one staff member died whilst trying to implement the programme, which highlighted to the FM just how grave the situation was. With time, the FM became more receptive to the concerns raised by the evaluator because the evaluator had years of experience (and thus had greater contextual

knowledge) in that country. In another instance, an evaluator pushed back on the number of revisions he was asked to make to the report. He also pushed back on requests to increase the sample size. He stated that his identity as an older, white man with over 20 years' experience in international development evaluation helped him to push back:

'Some of the local PwC [PricewaterhouseCoopers] people are African and are younger than me and I think I exploit the gray-haired old white man visiting Africa to hold a conversation which I'm not proud of but I know that they can't contradict me directly. In that culture it's too difficult. I can say things that other people might not be able to say. I think if I were a lot younger, if I was a woman, it would be different.' (Participant 3, Evaluator, Male)

It was apparent that evaluators' background experiences and demographics (race, gender and age) helped them to push back against the FM's power and influence. However, in general, evaluators ultimately acquiesced to the FM's requirements. Some of the evaluators were unable to cope with the demands and rigour of the evaluation and were either replaced or decided not to bid on any other GEC projects. Others became extremely adept in following the requirements and undertook several GEC evaluations. The overall impression amongst evaluators was that the GEC evaluation was a 'survival of the fittest'.

It is apparent that in international development programmes, such as the GEC, where diverse stakeholder groups collaborate, power asymmetries exist and without explicit attention to power sharing, evaluations and programmes are impacted.

Positive effects of power imbalances on Girls Education Challenge evaluation

Donor-influenced evaluation use

The FM forced programmes to learn from evaluation by putting in place review and adaptation meetings, which routinised reflection on the lessons from evaluations. These meetings were attended by the FM, the implementing organisation and any other implementing partners. Furthermore, as part of the reporting template, the FM included sections where the programme had to respond to the evaluator's findings. These measures encouraged serious consideration of the findings and helped to inculcate a learning attitude amongst programmes.

Donors contributed to evaluation capacity building

In GEC evaluation, the donor motivated both programmes and evaluators to care about evaluation quality and therefore played a key role in capacity building. Extensive guidelines were provided on proper sampling and rigorous study designs, and programmes were highly involved in all aspects of the evaluation. In some cases, going through the evaluation process increased the knowledge of both the evaluators and evaluands. Furthermore, two organisations hired new or additional

staff with greater technical expertise. The donor also influenced expectations of evaluator competencies. Programmes were keen to hire highly qualified external evaluators because the evaluations had high stakes.

Negative effects of power imbalances on Girls Education Challenge evaluation

Inappropriate study designs

The FM and implementing organisations had different interests. The FM prioritised a standardised way of aggregating the findings across the programmes and assessing the overall impact of the GEC programme. The programmes, although sympathetic to the FM's desire for generalisation, did not believe that aggregation was particularly useful, as context matters and lessons do not always transfer from one place to another. Implementing organisations were most interested in learning about their specific projects and wished the FM had focused on quality, whilst giving them latitude to decide how to evaluate and what kind of report they would like. The tension between generalisability and specificity ultimately favoured the FM, who wielded more power.

The FM pushed for the use of quasi-experimental approaches with clear treatment and control groups that were tracked over time. Some evaluators argued that the methodologies were not always relevant to the context:

'I would say that the expectations were very, very high. Unrealistically high about the kind of data that was viable to collect in the field, but particularly – and this was the real sticking point – the time frame in which change was anticipated.' (Participant 4, Evaluator, Female)

Some programmes were implemented in conflict-afflicted zones or included pastoral communities where beneficiaries were constantly on the move, which made it extremely difficult to conduct a longitudinal study with cohort tracking. Furthermore, because of the high levels of poverty in these countries, there were ongoing development projects being conducted by the government, other NGOs, bilateral funders and others. Therefore, finding a true control group was extremely challenging. Some of these areas were also prone to disease or natural disaster. For instance, one of the GEC countries was affected by an Ebola outbreak, which greatly limited the effectiveness of the programme. Another country experienced a drought midway through the programme. The study participants did not feel the FM balanced the desire for rigour with the contextual realities of working in low-income countries. Furthermore, the evaluation budget was rarely commensurate with the work involved. Five participants shared that funding was a huge challenge.

Four interview participants were critical of the use of EGRA and EGMA. Firstly, the program representatives stated that using local tests would have enhanced local credibility of the evaluation findings and engendered greater interest, understanding and use of those findings by local

stakeholders. Secondly, programmes and evaluators were given guidelines on the tests but were personally tasked with contextualising the tests, which was outside the scope of their education, responsibilities and training. As such, some evaluators struggled to come up with appropriate tests for the context. One evaluator said:

'It was so off, contextually. Read the story. Somebody walks the dog with a leash down the road. Really? My African colleagues, they were like, "First of all, people don't see dogs as pets. Secondly, we don't even know what a leash is" [laughter].' (Participant 2, Evaluator, Male)

Thirdly, administering the tests was extremely challenging because evaluators needed local data collectors who were well versed with local languages and cultures. In many cases, these enumerators were not highly educated and were administering these learning assessments for the first time. Consequently, they made many mistakes, which were costly because the evaluations sometimes erroneously understated programme effectiveness, resulting in penalties.

Fourthly, nine participants mentioned that the narrow focus on learning outcomes was misplaced and, in some cases, did not fully reflect the programme's success or mission. One programme witnessed radical improvements in the community's mindset towards girls' education and in girls' willingness to report human rights violations. However, because their learning outcomes declined over the course of the programme, the programme was deemed unsuccessful.

Fifthly, two participants expressed that the focus on learning assessments pushed programmes to focus on short-term results rather than on long-term systemic results. Programmes spent more time justifying their ratings and quantifying their work rather than engaging in qualitative reflection and inquiry regarding their programmes. One evaluator said:

'I guess if I were to say one critical thing about the GEC, it was that myopic focus on short-term results. Honestly, they have the constituents and they want to be able to hoist the flag after three years and say, "Look at what we did". It doesn't serve the longer term because we're talking about education systems.' (Participant 3, Evaluator, Male)

A significant implication of the donor's focus on numbers was the subordination of qualitative inquiry. Programmes and evaluators spent much time and energy in getting accurate data on learning measures that they felt they had few resources left to focus on the qualitative aspects of the evaluation. Although the evaluations were all designed to utilise mixed methods, programmes followed the donor's cues and prioritised the numbers. Consequently, two participants stated that there was a huge missed opportunity to uncover the reasons why some interventions worked, for whom, in what contexts and under what circumstances. One evaluator said:

'Donors and programs were not prioritizing exactly the same thing. The baseline report we put together, we felt like we did a really thorough job and some aspects were really dug into and

these were things that the program was very excited about. Then you find that really when it comes to reporting to the fund manager, no, no, no. They just want it simpler. [laughs] Like this is too much. [laughs] It leads to questions that they don't necessarily want to – that are incompatible with what they want to do when they're aggregating the data.' (Participant 10, Evaluator, Male)

Extractive evaluation

Seven participants stated that the GEC evaluation was highly extractive. Data collection involved large sample sizes (thousands of students in treatment and control groups). Several evaluation tools were also used, which proved burdensome to field teams and beneficiaries. Two participants expressed that the FM did not have a thoughtful design that took the ethics of evaluation into consideration. Requiring so much data from marginalised people and, in some cases, children with disabilities was deemed excessive. There were also no deliberate efforts to share data back to communities in a manner that may offset the time burden of data collection:

'At the moment it [*the evaluation*] is very extractive. I understand we have UK funders and they basically want to know that their money is doing what they thought it would do, but again, if we think about development as a model which is at its best like a transfer of resources, equity, and building relationships, just taking knowledge out of communities and putting it up to donors is not a very empowering or equitable way of focusing some of those findings. It becomes even more potentially problematic when the evaluation is done by a bunch of Westerners and foreigners coming in doing this research, and then leaving again, and not presenting it back or not working with the communities and following up.' (Participant 20, Programme representative, Female)

Evaluators and organisations alike expressed the wish that they were better able to disseminate the findings to local communities and governments. However, evaluators were required to use reporting templates which resulted in evaluation reports that were over 200 pages long. Two evaluators doubted that anyone, besides the FM, read the reports because of the length. Furthermore, the demanding nature of the evaluation left little time to develop abridged versions of the report that could be more easily understood by local stakeholders, which likely limited the utility of the reports. At the writing of this article, one of the evaluators e-mailed the researcher with an update that they had just completed the midline evaluation of the second phase of GEC (GEC-T) and that their final report was over 300 pages long. Evidently, some of the issues uncovered in this study persist.

Discussion and implications

This study has helped to further elaborate the context of international development evaluations, particularly the power dynamics involved in evaluation. The status quo represents a situation where donors wield significant power and influence over evaluations and other stakeholder groups have little or no avenues to challenge this power. In the GEC programme, evaluators were asked to assess effectiveness based on measures

and designs that were largely determined by donors. For the most part, evaluators did not attempt to challenge the donor's power and instead engaged in a 'survival for the fittest' approach where some simply dropped out and others became adept in following the rules. Greene (2001) called on evaluators to understand whose interests are being served by their work and Naidoo (2013) emphasised evaluators' potential to speak truth to power and transform lives. Then the question arises, 'what should evaluators do in situations where significant power asymmetries exist?' The answer to this question warrants a separate article. However, the implication of this study is the need for evaluators to conduct a formal or informal power analysis when they engage in evaluation, which involves the following questions: 'who wields power and how do they use their power?', 'what impacts do power asymmetries have?' and 'can I as an evaluator create or advocate for the creation of intentional avenues for less powerful stakeholders to speak truth to power and to determine the evaluation agenda?' This is a daunting prospect and, some might argue, goes beyond our duty as evaluators. However, it is not impossible. Noblit and Jay (2010), for instance, used critical race theory to guide the evaluation and to speak truth to power by developing a counter-narrative to the story rooted in White values.

Conclusion

In conclusion, this study sought to better understand power dynamics in international development evaluations by focusing on the Girls Education Challenge (GEC) programme as a case study. The major finding was that donors wield significant power over evaluations and less powerful stakeholders have few avenues to speak truth to power. The main limitation of this study is that local stakeholders and beneficiary communities were not interviewed. As such, the study cannot shed light on sources of power (or powerlessness) at the local level. Future studies should include these local actors. Further, case studies highlighting power-sharing strategies in international development evaluations would help us as a field to more effectively confront and address power asymmetries.

Acknowledgements

Competing interests

The author declares that no competing interests exist.

Authors' contributions

I declare that I am the sole author of this research article.

Funding information

This research received no specific grant from any funding agency in the public, commercial or not-for-profit sectors.

Data availability statement

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Disclaimer

The views and opinions expressed in this article are those of the author and do not necessarily reflect the official policy or position of any affiliated agency of the author.

References

- Alter, C., 1990, 'An exploratory study of conflict and coordination in interorganizational service delivery systems', *Academy of Management Journal* 33(3), 478–502. <https://doi.org/10.5465/256577>
- Alter, C. & Hage, J., 1993, *Organizations working together*, vol. 191, Sage. Newbury Park, CA
- Altheide, D.L., 1988, 'Mediating cutbacks in human services: A case study in the negotiated order', *Sociological Quarterly* 29(3), 339–355. <https://doi.org/10.1111/j.1533-8525.1988.tb01257.x>
- Ansell, C. & Gash, A., 2008, 'Collaborative governance in theory and practice', *Journal of Public Administration Research and Theory* 18(4), 543–571. <https://doi.org/10.1093/jopart/mum032>
- Azzam, T. & Levine, B., 2015, 'Politics in evaluation: Politically responsive evaluation in high stakes environments', *Evaluation and Program Planning* 53, 44–56. <https://doi.org/10.1016/j.evalprogplan.2015.07.002>
- Bachrach, P. & Baratz, M.S., 1962, 'Two faces of power', *The American Political Science Review* 56(4), 947–952. <https://doi.org/10.2307/1952796>
- Bamberger, M., 1999, 'Ethical issues in conducting evaluation in international settings', *New Directions for Evaluation* 82, 89–97. <https://doi.org/10.1002/ev.1140>
- Berkes, F., 2010, 'Devolution of environment and resources governance: Trends and future', *Environmental Conservation* 37, 489–500. <https://doi.org/10.1017/S037689291000072X>
- Bhola, H.S., 2003, 'Social and cultural contexts of educational evaluation: A global perspective', in T. Kellaghan & D.L. Stufflebeam (eds.), *International handbook of educational evaluation*, pp. 397–415, Kluwer, Boston, MA.
- Brisbois, M.C. & de Loë, R.C., 2016, 'Power in collaborative approaches to governance for water: A systematic review', *Society & Natural Resources* 29(7), 775–790. <https://doi.org/10.1080/08941920.2015.1080339>
- Brugnach, M., Craps, M. & Dewulf, A.R.P.J., 2017, 'Including indigenous peoples in climate change mitigation: Addressing issues of scale, knowledge and power', *Climatic Change* 140(1), 19–32. <https://doi.org/10.1007/s10584-014-1280-3>
- Carden, F., 2009, *Knowledge to policy: Making the most of development research*, International Development Research Centre, Sage, Ottawa.
- Carden, F., 2013, 'Evaluation, not development evaluation', *American Journal of Evaluation* 34(4), 576–579. <https://doi.org/10.1177/1098214013495706>
- Chouinard, J.A. & Cousins, J.B., 2013, 'Participatory evaluation for development: Examining research-based knowledge from within the African context', *African Evaluation Journal* 1(1), 9. <https://doi.org/10.4102/aej.v1i1.43>
- Coffey, 2016, Baseline report – Strategic partnerships window, viewed 8 September 2019, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/494679/Baseline-Strategic-Partnerships-Window-13January2016.pdf
- Coffey, 2017a, Endline evaluation report step change window, viewed 8 September 2019, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/700963/Endline-report-Girls-Education-Challenge-Step-Change-Window.pdf
- Coffey, 2017b, Endline evaluation report strategic partnerships window, viewed 8 September 2019, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/701007/Endline-report-Girls-Education-Challenge-Partnership-Window.pdf
- Coffey, 2017c, Endline evaluation report innovation window, viewed 8 September 2019, from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/700823/Girls-Education-Challenge-Fund-Innovation-Window-Endline-Report.pdf
- Cullen, A.E., Coryn, C.L. & Rugh, J., 2011, 'The politics and consequences of including stakeholders in international development evaluation', *American Journal of Evaluation* 32(3), 345–361. <https://doi.org/10.1177/1098214010396076>
- Dahl, R. A. 1957, The concept of power. *Behavioral Science* 2, 201–215.
- Dutton, J.E. & Duncan, R.B., 1987, 'The creation of momentum for change through the process of strategic issue diagnosis', *Strategic Management Journal* 8(3), 279–295. <https://doi.org/10.1002/smj.4250080306>
- Ehler, C.N., 2003, 'Indicators to measure governance performance in integrated coastal management', *Ocean & Coastal Management* 46, 335–345. [https://doi.org/10.1016/S0964-5691\(03\)00020-6](https://doi.org/10.1016/S0964-5691(03)00020-6)
- Fay, B., 1987, *Critical social science: Liberation and its limits*, Cornell University Press, Ithaca, NY.
- Freeman, M. & Vasconcelos, E.F.S., 2010, 'Critical social theory: Core tenets, inherent issues', *New Directions for Evaluation* 127, 7–19. <https://doi.org/10.1002/ev.335>
- Gaventa, J., 2005, 'Reflections on the uses of the "power cube" approach for analyzing the spaces, places and dynamics of civil society participation and engagement', Prepared for Dutch CFA Evaluation 'Assessing Civil Society Participation as Supported In-Country by Cordaid, Hivos, Novib and Plan Netherlands.

- Gaventa, J., 2006, 'Finding the spaces for change: A power analysis', *IDS Bulletin* 37(6), 23–33. <https://doi.org/10.1111/j.1759-5436.2006.tb00320.x>
- Gray, B., 1985, 'Conditions facilitating interorganizational collaboration', *Human Relations* 38(10), 911. <https://doi.org/10.1177/001872678503801001>
- Gray, B., 1989, 'Collaborating: Finding common ground for multiparty problems', Jossey-Bass, San Francisco, CA.
- Gray, B. & Hay, T.M., 1986, 'Political limits to interorganizational consensus and change', *The Journal of Applied Behavioral Science* 22(2), 95–112. <https://doi.org/10.1177/002188638602200204>
- Greene, J.C., 1997, 'Evaluation as advocacy', *Evaluation Practice* 18(1), 25–35. [https://doi.org/10.1016/S0886-1633\(97\)90005-2](https://doi.org/10.1016/S0886-1633(97)90005-2)
- Greene, J.C., 2001, 'Evaluation extrapolations', *American Journal of Evaluation* 22(3), 397–402. <https://doi.org/10.1177/109821400102200317>
- Grindle, M.S., 2004, 'Good enough governance: Poverty reduction and reform in developing countries', *Governance* 17, 525–548. <https://doi.org/10.1111/j.0952-1895.2004.00256.x>
- Hardy, C. & Phillips, N., 1998, 'Strategies of engagement: Lessons from the critical examination of collaboration and conflict in an interorganizational domain', *Organization Science* 9(2), 217–230. <https://doi.org/10.1287/orsc.9.2.217>
- Hasenfeld, Y. & Chesler, M.A., 1989, 'Client empowerment in the human services: Personal and professional agenda', *Journal of Applied Behavioral Science* 25, 499–521. <https://doi.org/10.1177/002188638902500413>
- Hay, K., 2010, 'Evaluation field building in South Asia: Reflections, anecdotes, and questions', *American Journal of Evaluation* 31(2), 222–231. <https://doi.org/10.1177/1098214010366175>
- Horton, D., 1999, 'Evaluation in low income countries: An introduction', *Knowledge, Technology & Policy* 11(4), 5–12. <https://doi.org/10.1007/s12130-999-1001-9>
- Horton, D. & Mackay, R., 2003, 'Using evaluation to enhance institutional learning and change: Recent experiences with agricultural research and development', *Agricultural Systems* 78(2), 127–142. [https://doi.org/10.1016/S0308-521X\(03\)00123-9](https://doi.org/10.1016/S0308-521X(03)00123-9)
- Huxham, C. & Vangen, S., 2000, 'Leadership in the shaping and implementation of collaboration agendas: How things happen in a (not quite) joined-up world', *Academy of Management Journal* 43(6), 1159–1175. <https://doi.org/10.5465/1556343>
- Jentoft, S., Van Son, T.C. & Björkan, M., 2007, 'Marine protected areas: A governance system analysis', *Human Ecology* 35, 611–622. <https://doi.org/10.1007/s10745-007-9125-6>
- Kellner, H., 1989, *Language and historical representation: Getting the story crooked*, University of Wisconsin Press, Madison.
- Lukes, S., 2005, *Power: A radical view*, 2nd edn., Palgrave Macmillan, Basingstoke.
- Miske, S.J. & Joglekar, A.B. (2018), 'Using or Misusing the Early Grade Reading Assessment? Examining A Measure of Payment by Results in the Girls' Education Challenge', Wiseman, A.W. (Ed.), *Annual Review of Comparative and International Education 2017 (International Perspectives on Education and Society, Vol. 34)*, Emerald Publishing Limited, pp. 187–202. <https://doi.org/10.1108/S1479-367920180000034026>
- Moynihan, D.P., 2009, 'The network governance of crisis response: Case studies of incident command systems', *Journal of Public Administration Research and Theory* 19, 895–915. <https://doi.org/10.1093/jopart/mun033>
- Naidoo, I.A., 2013, 'Growth and Integration in the evaluation profession: Some perspectives for consideration', *American Journal of Evaluation* 34(4), 572–576. <https://doi.org/10.1177/1098214013493495>
- Nathan, M.L. & Mitroff, I.I., 1991, 'The use of negotiated order theory as a tool for the analysis and development of an interorganizational field', *The Journal of Applied Behavioral Science* 27(2), 163–180. <https://doi.org/10.1177/0021886391272002>
- Noblitt, G. W., & Jay, M. (2010). Against the majoritarian story of school reform: The Comer schools evaluation as a critical race counternarrative. *New Directions for Evaluation*, 2010(127), 71–82.
- Phillips, N. & Hardy, C., 1997, 'Managing multiple identities: Discourse, legitimacy and resources in the UK refugee system', *Organization* 4(2), 159–185. <https://doi.org/10.1177/135050849742002>
- Platas, L.M., Ketterlin-Gellar, L., Brombacher, A. & Sitabkhan, Y., 2014, 'Early grade mathematics assessment (EGMA) toolkit', RTI International, Research Triangle Park, NC, viewed 10 september 2019, from https://ierc-publicfiles.s3.amazonaws.com/public/resources/EGMA%20Toolkit_March2014.pdf
- Purdy, J.M., 2012, 'A framework for assessing power in collaborative governance processes', *Public Administration Review* 72(3), 409–417. <https://doi.org/10.1111/j.1540-6210.2011.02525.x>
- Ran, B. & Qi, H., 2018, 'Contingencies of power sharing in collaborative governance', *The American Review of Public Administration* 48(8), 836–851. <https://doi.org/10.1177/0275074017745355>
- Reed, M. S., 2008, 'Stakeholder participation for environmental management: a literature review', *Biological conservation*, 141(10), 2417–2431. <https://doi.org/10.1016/j.biocon.2008.07.014>
- Rose, S.M. & Black, B.L., 1985, *Advocacy and empowerment: Mental health care in the community*, Routledge and Kegan Paul, Boston, MA.
- RTI International, 2009, *Early grade reading assessment toolkit*, Office of Human Development, World Bank, Washington, DC.
- Savin-Baden, M., & Howell Major, C., 2013, *Qualitative research: The essential guide to theory and practice*, Routledge, Oxon, United Kingdom.
- Takyi-Amoako, E., 2012, 'Comparing power spaces: The shaping of Ghana's education strategic plan', *Compare: A Journal of Comparative & International Education* 42(1), 113. <https://doi.org/10.1080/03057925.2011.629080>
- Trimble, J., Trickett, E., Fisher, C. & Goodyear, L., 2012, 'A conversation on multicultural competence in evaluation', *American Journal of Evaluation* 33(1), 112–123. <https://doi.org/10.1177/1098214011430377>
- Vangen, S. & Huxham, C., 2005, *Aiming for collaborative advantage: Challenging the concept of shared vision*, Advanced Institute of Management Research Paper 015. <http://dx.doi.org/10.2139/ssrn.1306963>
- Winer, M. & Ray, K., 1994, *Collaboration handbook: Creating, sustaining, and enjoying the journey*, Amherst H. Wilder Foundation, St. Paul, MN.
- World Bank Group, 2017, *World Bank country and lending groups: Country classification*, viewed 08 October 2018, from <https://datahelpdesk.worldbank.org/knowledgebase/articles/906519>
- World Bank Indicator, n.d., viewed 06 June 2020, from <https://data.worldbank.org/indicator/DT.ODA.ALLD.CD>