

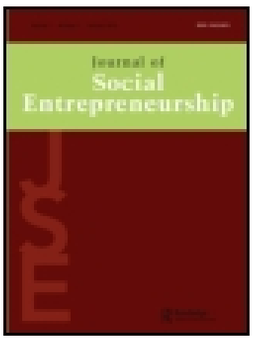
https://dspace.org/

Is Programme Evaluation the same as Social Impact Measurement?

Tsotsotso, Khotso

2020-01-08
Journal of Social Entrepreneurship
https://hdl.handle.net/10539/28801

Downloaded from a DSpace repository. https://dspace.org/



Journal of Social Entrepreneurship

ISSN: 1942-0676 (Print) 1942-0684 (Online) Journal homepage: https://www.tandfonline.com/loi/rjse20

Is Programme Evaluation the Same as Social Impact Measurement?

Khotso Tsotsotso

To cite this article: Khotso Tsotsotso (2020): Is Programme Evaluation the Same as Social Impact Measurement?, Journal of Social Entrepreneurship, DOI: 10.1080/19420676.2019.1683877

To link to this article: https://doi.org/10.1080/19420676.2019.1683877

Published online: 08 Jan 2020.

Submit your article to this journal

Article views: 36

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rjse20

JOURNAL OF SOCIAL ENTREPRENEURSHIP
https://doi.org/10.1080/19420676.2019.1683877

Routledge
Taylor & Francis Group

Check for updates

Is Programme Evaluation the Same as Social Impact Measurement?

Khotso Tsotsotso*

Unversity of Witwatersrand Center for Learning on Evaluation and Results, Johannesburg, South Africa

ABSTRACT

This study provides an analysis of the practical and theoretical differences in Social Impact Measurements (SIM) – as defined and is core to Impact Investing – and Programme Evaluation (PE) used in Monitoring and Evaluation (M&E). The study is the result of an inspiring effort to converge experiences of both Impact Investment practitioners and investment managers, with those of Programme Evaluators and M&E specialists. A meeting during the AEA's (American Evaluation Association) ImpaCon conference in Atlanta in 2017. The effort was to facilitate co-learning in order to improve and grow the Impact Investing industry. The study analyses qualitative survey responses from a purposively selected panel of experts including: experienced SIM practitioners and scholars, impact investment managers, programme evaluators and evaluation scholars. Responses are deductively analysed to provide thematic reactions to the research questions. Even though there is a common theoretical intent to determine intervention worth in both SIM and PE, and a common adherence to principles of evaluative thinking; the study concludes that there are clear theoretical and practical differences in participatory and utilisation approach, efficiency and rigour.

KEYWORDS

Evaluation; impact measurement; impact investing; social impact

Introduction

The notion that private resources and expertise can be 'mobilized', and redirected to provide solutions for social and/or environmental challenges – parallel to preservation and growth of capital; is certainly gaining popularity in the Global South. Akin to this phenomenon, the need for measuring the social impact of these types of investments is equally at the centre stage.

The idea of recognising and measuring the social value of an investment was observed as early as 1928 when the first Screened Investment Fund, (the Pioneer Fund in Boston) was witnessed (Rossi, Freeman, and Lipsey 2004). This phenomenon paved way to a wave of social movements in the global north in this regard. Examples include the introduction of the Programme Related Investments (PRIs) by the Ford

CONTACT Khotso Tsotsotso khotso@newleaders.co.za
*New Leaders Foundation

© 2020 Informa UK Limited, trading as Taylor & Francis Group

Foundation – also called Social Investing, shifting from grants to low interest loans for social programmes in the late 1960; introduction of Socially Responsible Investing (SRI) between 1960 and 1970; emergence of sustainable investing (General Investment Management founded by David Blood of Goldman Sachs) in 1990s and early 2000s; and more recently, coining of the term 'impact investing' by the The Rockefeller Foundation's Bellagio Center in 2007 (Rossi, Freeman, and Lipsey 2004).

The latest event – which is also an inspiration of this study – is Impact Convergence, which took place at the American Evaluation Association Conference in 2016, this was a gathering of global leaders in impact investing and programme evaluation, to discuss ideas on how to collaboratively improve impact measurement. The key outcome of this gathering was the glaring recognition that there is the lack of clarity in practice, and language inconsistencies in how impact investors and programme evaluators deal with impact measurement. There are deepened disparities in precision and an opportunity to identify synergies which would take social and environmental impact consideration to progressive heights.

Although there is varying use of 'semantics' between social impact measurements and programme evaluations, the principles applied may just as well be the same. However, the author hypothesises that there are important methodological differences between social impact measurement and programme evaluation methods in practice. This is notwithstanding the fact that social impact measurement (as practiced in impact investing) and programme evaluation (as practised in social research) share the same conceptual features even though they follow different processes and normally result in different outputs (forms of results).

The main aim of the study is to explore the relationship between *Programme Evaluation* (as normally used to assess the effectiveness of social and environmental programmes) and *Social Impact Measurement* (as practiced in Impact Investing). The study identifies fundamental differences and similarities between the two concepts, which can potentially enhance learning for practitioners in the Impact Investing 'ecosystem'. Finally, the study explores some methodological considerations in the implementation of each of the concepts in real-life settings.

In the following chapter, a detailed conceptual framework of *programme evaluation, social impact measurement* and *impact investing* is provided. This is followed by a description of methods and techniques used to investigate the fundamental differences between PE and SIM. Finally, the Findings chapter 4 provides analysis and discussion of findings of the study, which is succeeded by the Conlusion chapter, providing the conclusion and recommendations made by the author.

Conceptual framework

In order to elevate some of the latest discourse regarding the purpose and approaches of programme evaluation, and to outline a framework for comparing programme evaluation against Social Impact Measurement, the study systemically reviews relevant literature in a manner which guides the reader to easily 'digest' the argument being constructed.

The review resumes by introducing 'Evaluative Thinking' as a prelude to programme evaluation. Then a detailed explanation of 'programme evaluation' is provided,

followed by identifying key principles of programme evaluation. In the literature review, key principles of programme evaluation are subsequently used to explain some broad theoretical considerations and methodologies in conducting programme evaluation.

Finally, the review introduces the concept of Impact investing with specific focus on the practice of Social Impact Measurement as applied in Impact Investing.

Introduction to evaluative thinking

There is a recognition of a paradigm shift taking place in the socio-economic sector away from a predominantly linear-based model of change, to one that is more dynamic, reflective and responsive. This shift is characterised by a growing culture of what has lately been coined 'Evaluative Thinking'. Buckley et al. (2015) define Evaluative thinking as critical thinking applied in the context of evaluation, driven by the attitudes of inquisitiveness and a belief in the value of evidence, that involves: identifying assumptions, posing thoughtful questions, pursuing deeper understanding through reflection and perspective taking, and making informed decisions in preparation for action.

In an effort to explain evaluative thinking in the context of programme evaluation, Patton (2014) draws the critical distinction by pointing out the notion that programme evaluation is an activity inspired by the values and 'behaviours' of evaluative thinking (as suggested by Buckley et al (2015) above. Patton further elaborates that *'Evaluative thinking is way of doing business'*. He makes another critical assertion in that programme evaluation is only useful if the underlying social programme and the respective organisation's culture manifest evaluative thinking. Hereby implying that, the fundamental principles of programme evaluation are rooted in the values and 'behaviours' demonstrated by the concept of evaluative thinking.

In a paper titled *Embracing Evaluative Thinking for Better Outcomes* sponsored by both CLEAR-AA and the Rockefeller Foundation; Griñó et al. (2014) apply these principles in a framework for measuring evaluative thinking, tested in a number of African countries. In this paper, the authors postulate that evaluative thinking is a systematic approach to ensuring the four aspects of programme performance management: (1) what results are expected, (2) how results can be achieved, (3) what evidence is needed to inform future actions, and finally, (4) how results can be improved in the future. The authors make a further argument that evaluative thinking must be embedded in an enabling organisational or programme culture which puts performance improvement at the centre of its belief set. This way, effective 'evaluative thinking' promotes consistent collection of evidence for learning and the examination of the quality of evidence.

Griñó et al. (2014) offer us some practical characteristics of functional evaluative thinking. These are summarised as follows:

- ET requires all stakeholders involved in a project to consciously and constantly reflect on project experience with a view to implementing improvements based on what is learned. Therefore implying continuous learning.

- It is demonstrated in the implementation of well-focussed programmes and in the use of high-quality evaluations that feed into programme and decision making
- It requires that time and resources are allocated for reflection on evaluation findings and in using those findings.

Explaining programme evaluation

A programme is a set of activities supported by a set of resources to achieve a specific and intended goal (Scriven 1998). On the other hand, Evaluation is defined as the 'examination of the worth, merit, or significance of the programme.'

In explaining what programme evaluation is, Patton (1997) makes a crucial distinction between programme evaluation and any other ordinary informal assessment. Patton identifies the fundamental difference to be the fact that programme evaluation requires a systematic approach (following from the values of evaluative thinking defined above) and, is conducted according to a set of specific guidelines. Following Patton's guide, a paper by the U.S. Department of Health translates this into a suggested definition for programme evaluation. The U.S. Department of Health (2011) provides the following definition:

> Programme evaluation is the systematic collection of information about the activities, characteristics, and outcomes of programs to make judgments about the program, improve program effectiveness, and/or inform decisions about future program development.

To this practical definition, the early work of Patton (1997) warns that programme evaluation does not occur in a 'vacuum'; that it should be influenced by the context of the underlying programme. It should therefore follow that programme evaluations should be practical, feasible and contextually relevant and should be within the confines of resources, time, and socio-political context. Finally, Evaluation findings should be used both to make decisions about programme implementation and to improve programme effectiveness.

Even earlier than the work of Patton in 1997, programme evaluation literature had offered useful frameworks to identify programme evaluations from other forms of enquiries. Amongst other contributions to the discourse, Steele (1970) considers different forms of programme evaluation studies of the 1900s and identifies three essential elements of a programme evaluation; criteria, evidence and judgement. Before concluding that programme evaluations are meant to be purposeful (they should precede the desire to make a decision or change the programme) and contributing to present or future programme performance; Steele (1970) explains each of the three elements as follows:

Criteria refer to the programme aspects against which the programme is to be judged. This is essentially the performance or success criteria. At the centre of the criteria, is the set of performance indicators and targets which the programme should achieve.

Evidence is the extent to which the programme meets the criteria.

It follows then that, *Judgement* refers to the extent to which the set criteria is met. This practically speaks to the magnitude of programme effect.

Principles of programme evaluation

In an effort to further differentiate evaluation from other forms of enquiry, the Higher Education Innovation Fund (Ulrichsen 2015) suggests some principles which also resonate with the descriptions offered in this review so far. Whiles these guiding principles are there to ensure high quality, relevance and usefulness of an evaluation, they can be instrumental in identifying a 'true' evaluation from other forms of assessments.

According to Higher Education Innovation Fund (Ulrichsen 2015), the evaluation guiding principles are: (1) *Focus and feasibility* which pertains to how evaluation resources are best used to address key questions and facilitate good decision-making. This principle mainly prevents the lack of focus in evaluation scope which leads to failure to address intended questions. (2) *Usefulness and timeliness*, referring firstly to relevance of the underlying challenge(s). Secondly, this principle refers to the information's contribution to programme improvement, learning and/or decision-making. Finally, the principle requires that the evaluation is appropriately timed to add value to project or organisation. (3) *Credibility, validity and reliability*: this principle generally pertains to the integrity of the evaluation information/evidence provided. Social programmes mostly attract high levels of financial investments and have the potential to change lives, so decisions that are often supported by evaluation evidence must be based on credible, valid and reliable information possible. (4) *Sensitivity*: evaluations recognise the connectedness between the programme and the key stakeholders. In its approach, a programme evaluation applies sensitivity in data collection, analysis and most importantly dissemination of findings. This principle places particular emphasis on the process of evaluation and recognises its direct effect on stakeholder buy-in, participation and use of evaluation results. It puts the evaluator(s) and the evaluator(s)'s interaction with key programme champions at the centre of the evaluation, and less reliance on evaluation findings as the sole agent for change.

Evaluation approaches and theory

While programme evaluation is constantly growing and changing in practice; many experienced evaluators and evaluation scholars are slowly beginning to reach some consensus about the design, approach and methods. One such agreement is the revelation that evaluations ought to be context-specific and therefore, appropriation of approach and methods is depended on the context of both the programme and the evaluation itself. This notion is attested to by authors such as Worthen, Sanders, and Fitzpatrick (1997) and Owen (2004).

Although there is still debate about the correct terminology or semantics surrounding the different designs or approaches, there is definitely some convergence reached (Dart 2004). In his paper on the 'Normative Approaches to Evaluation'; Dart (2004) provides six different approaches to evaluation, namely; (1) Experimental approaches, (2) Testing-objectives approaches, (3) Decision-management approaches, (4) Judgemental approaches, (5) Pluralist-intuitionist approaches, and (6) Theory-driven approaches. The approaches are explained as follows:

Experimental approach. Experimental approaches are inspired by, and better explained in the early work of Popper (1959), Campbell, (1991), and Cook (1966).

Borrowing from the fundamentals of pure research methods; this approach is based on the idea that programme evaluation should generate new knowledge using empirical data from the underlying programme (Dart, 2004). Campbell (1991) emphasises the element of 'internal validity' as central to the experimental approach. Internal validity refers to the extent to which the evaluation methods appropriately measures the intended evaluation subject i.e. the accuracy of the method in measuring the programme effect as stated in the evaluation.

Experimental approaches are built from the Humean Theory of Causality whose basic task is to demonstrate the conjunction in which programme action (X) produces programme outcome (Y) (Popper 1959). A key determinant of the precision of experimental approaches in evaluation is the strength of the design of the comparator (technically referred to as the counterfactual). In this approach, the design and therefore also the implied internal validity depends on the stage and context of the programme being evaluated (Besharati 2014). Pure experiments apply designs like a Randomised Control Trials which is normally suitable before the implementation of the programme (*ex-ante*). This design is regarded as the most valid design as far as internal validity is concerned. In programmes that are evaluated at the end of the programme (*ex-post*); suitable designs are classified as quasi-experimental. While measure of internal validity is still regarded to be exceptional; quai-experimental designs are inherently weaker than pure experiments as both the assumptions and confounding factors are not always easy to overcome.

Among many criticisms of these approaches, it is the fact that experimental design tends to be technically intensive and expensive to apply in most programmes. In recent discourse, a key criticism is the limitation of experimental approaches in the ability of the findings to be 'transferred' to different contexts (external validity). These designs tend to inherit assumptions which are rigidly specific to programme context.

Testing-objectives approach. Tyler (1967) refers to evaluations in this approach as 'educational evaluations'. These evaluations are mainly concerned with the ability to accurately determine the extent to which programme goals are achieved. The evaluations are generally in the form of iterative formative evaluations with the sole purpose of demonstrating programme achievement over time (Guba and Lincoln 1989). The main feature of this evaluation approach is the fact that evaluation results are not shared during evaluation iterations until the programme is completed, thereby preventing programme adjustment during implementation. Evaluation rigour in this approach is attained using iteration (testing and re-testing against objectives) and not through comparing.

This objective-testing approach is not without criticism; at the onset, the lack of programme adjustment during implementation is a challenge in many programmes as both designers are not necessarily sure about envisaged outcomes and would appreciate the possibility to learn and adjust outcomes over the course of the programme. A key weakness of the approach was best explained by Scriven (1967) as he highlighted that this approache is weak in applying 'Judgement' in real-time as a key principle and element of an evaluation.

Decision-management approach. The fundamentals in this approach are best elaborated by Patton's paper on Utilisation Focussed Evaluation (1997). Unlike the 'Testing-objective approach', this approach puts programme managers and their ability to make informed decisions (utilisation) at the centre of the evaluation design. Evaluations are mainly judged on their ability to serve their intended users. Caron (1993) best describes this sentiment as a requirement that evaluation is 'carried out for the benefit of the organisation in that evaluation is a function of management and should therefore facilitate decision-making by management.'

In meeting Patton (2002)'s ethos of Utilisation Focussed Evaluation, evaluations in this approach are meant to match the evaluation design to 'the information and decision needs of primary intended users, taking into account 'other stakeholders, political factors, organisational constraints, project/program history, available resources, and cultural factors of specific evaluation context'.

As attractive and central to use as this approach seems, it inherits a strong criticism in its implementation. Pawson and Tilley (1997) point out that such evaluation carry a strong sense of a Rothschildian vision in them. Translated to the evaluation context; the Rothschild principle defines a relationship between the evaluator and the client (programme management) in which the client states and pays for what 'he' wants, and the evaluator's sole responsibility is to deliver what has been paid for. This notion makes the assumption that evaluation 'Use' is limited to management and does not include programme clients or public at large. It is easy to see how this narrowed view to 'Use' does not necessarily uphold the values of inclusive and democratic development as is advocated for the twenty-first century.

Judgement approaches

This approach is both inspired by and advocated for, by the work of Scriven (1974) – 'Goal-free' evaluation – and Eisner (1985) – Connoisseurial model of evaluation. Central to this approach is the judgement of programme's worth by professional experts. Initially stated evaluation outcomes in Scriven's idea of 'goal-free' evaluations are a 'non-factor' to the evaluation scope. Evaluators have to discover what effect the programme has and match their effects against the needs of those who they affect (Scriven 1974). The key role of the evaluator is to get to the emerging programme effect in the most unbiased manner possible, suggesting limited contact with programme staff. This effect is subsequently matched to the needs of programme beneficiaries. Therefore, the approach puts strong emphasis on evaluator's judgement of programme merit based on empirical evidence. The advantage of this approach is in its ability to identify unintended effects (both negative or positive) which can strengthen stakeholder's programme knowledge (Scriven 1974). This approach avoids a fruitless effort to prove often misguided and vaguely worded programme objectives at the expense of learning. This aspect of the approach is particularly appropriate in situations where evaluation management capacity and therefore, evaluation commissioning capabilities are poor (Bless, Tsotsotso and Gebremichael 2017).

Many evaluators – such as Newman et al. (1995); Shadish and Epstein (1987); Alkin (1972) and Patton (1997) – question the practicality of the 'Judgement approach.'

They unanimously scrutinise the idea of leading a goal-free evaluation and discredit it to a 'fishing expedition'. This thinking and approach to evaluation has not gained popularity among evaluators in practice, even in the twenty-first century. Perhaps this is mainly driven by its lack of promotion of good programme planning and programme performance management principles which are explicitly supported by theory-based approaches (see below).

Pluralist-intuitionist approach

This approach is 'born' from the theories surrounding 'value pluralism' which generally identifies and preserve multiple value perspectives. Within the past decade, driven by maturity in democracy and the promotion of inclusive development, particularly in the global south; this approach has enjoyed growing popularity in the form of participatory monitoring and evaluation (Estrella and Gaventa 1998).

Largely inspired by a liberal ideology, this evaluation approach has been described by authors such as Stake (1967), Guba and Lincoln (1989), Wadsworth (1991) and Smith (1994) as one that is client-centred, ethically-subjectivist and epistemological in design. The design is concerned with constructing a shared 'depiction' of programme reality. Programme success is an interpretation of a consolidated opinion of multiple participants affected by the programme. Less emphasis is given to evaluation validity in favour of social justice. It is a shift from a positivist approach to evidence generation promoted by experimental designs described above, instead, it involves a concerted move towards interpretivist, relativism, naturalistic inquiry, constructivism, and feminist inquiry. These theories are all in support of a rising African value for knowledge creation in the twenty-first century. It is best explained by a prolific Nigerian author Ms. Chimamanda Ngozi Adichie as avoiding the dangers of a 'Single Story'. Evaluations in this approach are mainly qualitative in nature and use individual stories to tell an account of programme effect.

This approach can be particularly weak in situations where a true, definite and factual account of the programme success instead of multiple programme reality is required. Multiple opinions about programme success and/or failure can make it categorically difficult for the evaluator to make concrete recommendations (Dart 2004).

Theory-based approach

Pawson and Tilley (1997) and House (1991) have described how theory-based evaluation approach has gained popularity in the 1900s, even more so in the 2000s as accountability and good programme planning receive attention in social development. This approach requires a clear programme theory model with explicitly stated causal linkages, demonstrating how programme outcomes are to be achieved. Chen (1990) describe theory-based approach as 'specification of what must be done to achieve the desired goals, what other important impacts may also be anticipated, and how these goals and impacts would be generated'.

The approach sets out plausible theories about the 'works' of the programme within a specified context. In an evaluation, empirical data is used to prove or refute the developed programme theory. Proponents of theory-based approach particularly enjoy the idea that this approach not only isolate effect, but also explain the

mechanisms of change in a given context. This is in full response to the criticism of the 'atheoretical' nature of evaluations resulting in an over-simplified input/output type of evaluations, which offer a very limited explanation of the mechanisms of programme-induced change, a phenomenon described by Chen (1990) as a 'Black Box' approach.

Scriven has historically been a strong critic of theory-based approach. In his paper, Scriven (1994) argues that the point of a programme evaluation is not to determine whether the programme works or not, but rather, to establish the merit, worth, quality or value of the programme. And, to do this, one does not need to know how the programme ought to work or why it fails to work, or even what its components are. Scriven further argues that investigating extensively, the mechanism of programme effect can be a costly and unnecessary exercise. Sometimes, responding to specific stakeholder concerns is much more efficient and strategic than 'carefully crafted analytical suppositions.'

One of the other common criticisms of the theory-based approach is led by Hamilton-Smith and Hopkins (1998). They assert that theory-based approach does not adequately focus on the investigation of unintended programme effects. They tend to promote over-indulgence into micro-mechanisms postulated by the underlying theory, at the expense of emerging patterns and mechanisms of change, thereby defeating its primary intent.

Explaining impact investing and social impact measurement

In constant search for development and socio-environmental panacea – most especially for the global south – the past century has experienced growing innovations in the space. More importantly, the role of private sector and private investors in social and environmental development has become more prominent and perhaps more direct during the same period. Certainly, the call for innovative approaches to development funding has matched this call. Private sector is seen to play a central role in generating employment and stimulating entrepreneurship, by making knowledge and technology accessible to the poor, in building social and economic infrastructure, and in addressing climate change challenges of developing countries (Besharati and Tsotsotso 2015).

The term 'impact investments', is believed to have been coined in the 2007 Bellagio conference sponsored by the Rockefeller Foundation. Impact investments are investments (debt, equity, guarantees, mezzanine, and various other asset classes) made to organisations and funds with the intention of generating measurable social and environmental impact in addition to financial return (GIIN 2009).

In describing key elements of Impact Investing, Trelstad (2016) introduces the idea of 'dual-pursuit' in socio-economic development. Dual-pursuit for change involves two models (approaches to socio-economic development): (1) A separate pursuit of financial impact, succeeded by pursuit for environmental impact, or (2) A joint pursuit of both financial and social impact at the same time.

In the first model (*separate pursuit*), the investor first maximises profit through traditional business means. Then, once profit is maximised, the investor uses some of the

profit proceeds to pursue some form of social or environment impact, usually in the form altruistic or philanthropic behaviour.

This model has been the centre and arguably, the 'building block' of the corporate social responsibility (CSR) market. Over the years, this model has undergone many criticisms.

In the second model, however (*Joint pursuit*), which is a reflection of the key elements of Impact Investing, the investor 'optimizes' both a financial return and social and/or environmental impact at the same time. Although not yet conclusive in its description, various sources have reached some convergence around key elements of Impact Investing (GIIN 2009; Finkelman and Huntington 2017).

Intentionality, investment with return expectations, positive impact, impact measurement

For the benefit of this paper, a purposeful attention is paid to the last element (Impact measurement) as the main subject of the study. So far, numerous efforts have been made to standardise approaches to estimating investment effects on social and environmental progression, and in particular to assist investors with their due diligence through metrics, ratings and certification (Flynn, Young, and Barnett 2015).

According to the analysis conducted by Loveridge (2016) in her study of the Donor Working Group's work commissioned by the Donor Committee for Enterprise Development, social impact measurement in Impact Investing has been limited to three broad roles; firstly, an assessment of potential social and environmental impacts, risks and financial returns during the pre-investment stage. Secondly, Measuring financial returns and outputs after investment decisions have been made. And finally, a considerable investment into efforts to standardise investment performance results and measurements. This includes 'building' open and widely shared measurement infrastructure used as a scoring and rating system, and measurement indicator catalogues to estimate social impact of Impact Investments such as the IRIS (Impact Reporting and Investment Standards).

Loveridge (2016) notes some key performance areas which the Impact Investing community is yet to incorporate into their social impact measurements. These are:

- Incorporating rigorous measurement of social and environmental impact into the costs of investments, as distinct from third party sponsors undertaking and/or funding measurement activities.
- Assessing investment criteria other than impact (i.e. efficiency, relevance, sustainability, learning).
- Measuring the effectiveness of intermediaries, who work with investors, to identify investment opportunities and structure financial instruments, and investees to identify opportunities for capital and to become ready for impact.
- Measuring systemic change, beyond the impact to targeted beneficiaries and to the wider market.
- Analysing relationships between different types of returns – social, financial and environmental.

O'Flynn and Barnett (2017) also make a critical note that less has been achieved in addressing a more evaluative understanding of social impact by impact investors. In their study involving the review of 'evaluative' practice among prominent impact investment globally, the pair conclude that while some impact investments show promising uses of social impact measurement methods which meet criteria for evaluative understanding of social impact, each has different strengths and limitations in providing a more robust assessment of impact.

Approaches and methods used in social impact measurements

There are various methods used in the Impact Investing community to assess social impact. While there are many variations of the methods applied by different organisation and investments, they can be grouped into some broad categories. Notably, some methods have been fully or partly adopted from key concepts of programme evaluation. After a considerable search through the brief reports posted on the GIIN website, four broad categories of methods were identified;

1. **Expected Return Method** – a traditional financial method involving a calculation of the weighted average of the likely profits of asset, brought back to present value at a specific rate of return. Thereby estimating the social change in monetary terms.
2. **Mission Alignment Methods (Lean data Approach)** – Monitoring of impact over a period of time, against a set of pre-conceived and predetermined indicators. Indicators are usually determined by the investor or are based on the investors' value system and desired goals. This usual constitutes populating a predefined Metrix over the life of the investment.
3. **Theory of Change and Logic Model Framework** – this is a modification of a Theory of change development into a measurement method. It involves the development of a clear theory of change with clearly stated risks and assumptions, and a comprehensive logframe with indicators plus, development of a monitoring plan and tools to be used over time. The data collection is not limited to a metrix.
4. **Experimental & Quasi-Experimental Methods** – Very seldom applied, but these constitute a full evaluation project with experimental and quasi-experimental methods as explained above.

Whiles the range of methods are known and applied differently, there are potentially unique contextual requirements for Social Impact Measurements in Impact Investing which evaluators may need to be cognisant (Jackson 2013), justifying some deepened studies into the difference between Programme Evaluation and Social Impact Measurement.

Methodology

Data collection method

The initial intention of the study was to apply a three-round Delphi technique in which panellist would be asked a set of questions on three occasions, with the hope

to approach consensus regarding the central question of the study on each round until a reasonable agreement is reached.

However, the researcher found overwhelming levels of saturation (Mason 2010) when reached by the first round of questions. This first round of survey questions constituted a list of five open-ended questions, each paired with some binary questions to test for consensus.

Sample

A panel of 12 experts and experienced practitioners was invited to take part in the study. The sample was primarily selected based on prominence and experience in either programme evaluation or Impact Investing, either as practitioners and scholars or both. 80% of the panel includes experts based within the African continent, whiles the remaining 20% are panellist based offshore but perform most of their work within the continent. While most of the panellist are also associates of the Centre for Learning on Evaluation and Results – Anglophone Africa, the rest of them were identified from conferences and popular impact investing reports such as those by the Rockefeller Foundation or the Industrial Development Corporation based in South Africa. This sample is therefore purposive sample chosen (Trotter 2012).

There appears to be no general agreement about sample size in qualitative studies. While some reports are based on a single-person studies (O'Halloran 2010), other commentators suggest sample sizes ranging from 6 to 30 (O'Halloran 2010). It was felt that an expert panel of 12 respondents should be able to supply varied and detailed accounts for the purposes of this stud.

Data analysis methods

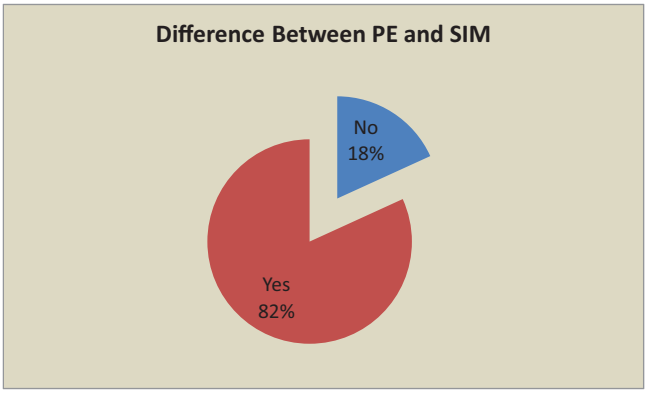
The data were analysed on the basis of a systematic coding, following the approach suggested by Saldaña (2015). This type of analysis consists of a systematic coding (breaking down) of data according to a code list (or coding scheme), in such a way as to identify (practically and theoretically) relevant patterns. The codes are then grouped and synthesised 'up' into (more general) categories, which in turn get linked to more general themes and (theoretical) concepts.

All of the questionnaire responses were read by the researcher and coded using a deductive approach, which allowed the researcher to identify both new and repetitive types of responses emerging naturally from the data (Saldaña 2015). A number of categories were generated from the data coding and under these, all of the data were accounted for, and organised into an excel coding scheme. Each category was assigned a meaningful description linked to literature. This later made it easier for the researcher to interpret the coding scheme into a 'flowing narrative. Two independent other researchers were asked to verify the seeming accuracy of the category system and after discussion with them, minor modifications were made.

Findings

In this section, each of the evaluation questions is addressed and discussed, making references and linkages to current academic and industry discourse.

Is there a difference between Programme Evaluation (PE) and Social Impact Measurement as used in Impact Investing?



An unexpectedly overwhelming majority of the panel of experts indicated a difference between PE and SIM. Interestingly, the sample (the panel of experts) had a fairly balanced representation of respondents who are predominantly experts in Monitoring and Evaluation and those with considerable experience in either Impact Investing or Enterprise development (55%:45%). Although the researcher expected inconclusive results regarding the existence of a disparity between the two concepts; the results were conclusively in favour of a positive difference.

The rest of the survey questions were designed to identify and provide a deeper understanding about causes and drivers of perceived differences between PE and SIM. Therefore, in addressing the rest of the questions, coded responses were instrumental in highlighting emerging themes pertaining to the underlying research questions.

In particular, the study analyses the following; (1) Key differences between PE and SIM (both practical and Theoretical Difference), (2) Methodological consideration Driving the difference, and finally, (3) Some Similarities Between PE and SIM which may foster further collaboration between Evaluator and Social Impact Assessors.

Practical and theoretical differences between PE and SIM

This section analyses the practical and theoretical differences in PE and SIM as deducted from the responses of the panel of experts.

> *'Program evaluation focuses more on the outcomes while social impact measurement focuses on the outputs of a program. Also, the former is more rigorous and requires more specialized skill set compared to the latter.'*

Theoretical differences

According to the panel, there are two main theoretical matters to be considered as key differentiators between PE and SIM, namely, differences in 'Rigour and Validity', and 'application of an established evaluation criteria as opposed to a utilisation focus.'

Rigour and validity. As opposed to social impact measurement (SIM), in programme evaluations, there seems to be a purposeful investment into rigorous evaluation methods with time and effort spent on validating results, with the hope to emphasise accuracy and reliability of the evaluation results. The popularity of experimental methods in PE can certainly be a reasonable indication of this finding (Basheka and Byamugisha 2015). Borrowing from techniques developed over many years by researchers, programme evaluators have adapted and developed appropriate tools to measure evaluation validity.

On the other hand, there is a consensus (among the panel members) that SIM puts less emphasis on validating the results of the impact measurement. In fact, literature related to SIM also explains that more emphasis and time is spent on agreeable indicators and standards well ahead of the underlying impact investment to ascertain potential effect (Steele 1970)

Perhaps the cause/driver of this difference may have more to do with average sizes of 'investments or programme sizes'. In PE, most evaluated social programmes tend to be programmes with large budgets and even larger potential effect (Bamberger, Rugh, and Mabry 2011), thereby receiving interest and attention from a number of critical development parties such as policy-makers whose decision can bring about a systemic change. This way, emphasis on accuracy, validity and reliability should be expected to be a critical issue. This is at least the case in most African countries (CLEAR-AA 2017).

As new investment alternatives in Africa, Impact Investments narratively inherit a perception of high risk with generally low investor risk appetites in a continent with comparatively smaller investment markets. As a result, whiles Impact investment funds may be significantly large, the individual impact investments (underlying businesses/programmes) tend to be small and cater to small communities with specific needs. The attention tends to be on identifying specific business niches. With small individual investments, SIM budgets are limited, thereby leading to limited scope of the impact assessment.

Application of established evaluation criteria vs. Utilisation focus

It is a general consensus that most investees are required to demonstrate potential impact to the investor and/or investment managers as part of the 'investment due-diligence' process in Impact Investing. This usually translates to pre-defined statements of impact (or set thereof) with clearly defined indicators which are normally important

to the investor or investment portfolio as a whole. This notion of thinking through the envisaged impact of the intervention and subsequently becoming strict and focussed in the scope of the impact assessment, follows Patton's (1970) Utilisation focus approach. The logic here is the fact that the focus of the SIM is strictly defined and mostly serves the interests of the investor and/or investment management. And, due to limited SIM budget, the assessments are mostly designed to be narrow (lean) and specific in collecting data.

> *'The reason is that programme evaluation, in most instances, seeks to unpack the 'DAC criteria', and looks into how, why and what with respect to a programme. In some ways, programme evaluation validates a programmes existence. SIM currently refers more to measurement of actual impact achieved through an investment – it is the valuation of the social change achieved, which is then 'paired' with the financial return'*

In programme Evaluation (at least within the last 10 years) – and with the rise of inclusive development and strengthening accountability – there is a concerted effort to shift away from a pure utilisation focus and narrow evaluation approach, in favour of a participatory and more 'democratic' evaluation approach. One of the findings of the CLEAR-AA (2017) is that over 75% of the evaluations are evaluations of multi-stakeholder programmes with varying influences and interests, but whose participation in the evaluation is equally important. Therefore, to cater to all key stakeholders, different aspects of the programme are covered by the evaluation design, i.e. programme evaluations tend to follow an established evaluation criterion such as the DAC criteria which covers all the main facets of an evaluation from the relevance of the intervention to the cost-effectiveness and sustainability of the impact.

> *'Social impact measurement operates largely in the domain of impact investments and social enterprise and to some extent blended finance. It seeks efficiencies through low costs, speed, technology and innovation–and using some of the financial returns to the investor to pay for some measurement costs'*

In light of this, key differences between SIM and PE tend to be characterised by phrases like '*single effect, as opposed to a thorough enquiry*'; 'validation of programme as opposed to fulfilling a Learning agenda'; 'tracking specific indicators as opposed to searching for all possible impact'; 'emphasis on quantifying social measures as opposed to articulation of observed social change'

Practical differences

Additional to the fundamental/theoretical difference discussed above, there were practical or implementation differences described by the panel of experts. The key practical differences are the differences in 'Ease of Measure and efficiency' and 'Specialization and Domain of SIM'.

Ease of measure and efficiency

The development of infrastructure such as the IRIS and the GIIN for Social Impact Measurement 'practitioners' is a fair indication of the sentiments that SIM endures

high levels of emphasis on standardisation and simplification of impact assessment, a sentiment shared generally by the panel of experts. Perhaps it is also driven by limitations in SIM budgets, but also, the recognition of the fact that Impact Investing is a growing field and a core function such as impact assessment should be shared and practiced by all, in a simple and least complicated manner. But, also in a way that would allow the industry to conduct and grade investments. Comparability and grading are important within the investor community.

On the other hand, the emphasis on PE is suggestively different. With the recognition – over years of evaluation experience – that programme context and social nuances are different, there is greater importance attributed to the appropriateness of the resulting programme evaluation design. There is generally a push to apply unique and appropriate design for different programme context. It is therefore no surprise to observe emphasis on Evaluation Capacity Development (ECD) as a driver in M&E. Specific skills such as the ability to construct a programme evaluation framework therefore becomes critical as it caters to, and is indicative of, the importance of contextualising the evaluation process.

Specialising and domain

While programme evaluation is multidisciplinary and not necessarily sectorial in its core focus, the relationship between Impact Investing as a discipline and SIM is inseparable. The existence and practice of SIM is core to the definition of Impact Investing. As such, the choice of methods, indicators and reporting standards tend to be specific to, and appropriate for the investing community. In its assessment, SIM incorporates some imperatives which are not necessarily part of the design of programme evaluations. Imperatives such as investor's risk appetite, investment risk and investor exit strategy, and investment value post-exit are critical imperatives into SIM.

> *'Programme evaluation is an older and more established practice, and also includes more established methodologies that have been accepted in the public and development sectors… Social impact measurement generally came about more recently in development but also in areas where private sector is working towards development objectives such as impact investing'*

Methodological consideration driving the difference

Given the key nuances outlined and discussed above, the study extracted some implied methodological disparities between PE and SIM. The key methodological differences are in: the *Process and Participation levels*; the *Data collection methods and tools* used and finally, the importance of *the causal inference issue* in the design of the impact assessment.

The process and Participation levels refer to the use of the evaluation process in programme evaluation to build partnership and 'championship' with key programme stakeholder, for the main hope to enhance evidence use. This is due to the understanding that the resulting evidence can be usable and appropriate for other audience besides the 'investment management'. Secondly, in PE, it (over the years) has been emerging that evidence use can be optimal when the intended or targeted user is

made a part of process of the evaluation. The evaluation process can trigger changes in programme discourse by managers even before the end results.

In contrast, SIM tends to be non-participatory and Metrix-driven with a clear set of standards against which to compare. As a result, instrument designs process is hardly collaborative as the tools tend to measure standardised indicators. Whiles some SIM practitioners are starting to take a participatory approach to measurement, a fully participatory process is not necessarily prioritised within the community.

Programme evaluation methods for impact evaluation have a significant inclination in favour of establishing a causal relationship between the intervention and the impact. As discussed in the literature above, there is a reach and continuously growing body of knowledge in methods to design evaluations for identifying and explaining the change mechanism of and intervention.

With the exception of a few impact investment portfolios, SIM methods and reporting requirements do not necessarily reflect a direct description of the relationship between the investment and the impact. The emphasis tends to be more on quantifying different social outcomes without any considerable effort to explain the mechanisms of change. With the adoption of the practice of Theory of Change in SIM, perhaps a description of the causal relationship will gain popularity in this domain.

Enabling similarities between PE and SIM

Despite the number and depth of disparities between SIM and PE identified in this study thus far, there were some similarities which emerged from the responses of the panel of experts. This similarities or this set of points for a common ground between evaluators and social impact assessors in impact investing is potentially a critical enabler of the convergence of the two communities. Perhaps it is this common ground which can inspire further peer-leaning and improvement in the two communities.

In analysing the responses from the panel, the following similarities were identified: (1) The *theoretical intent* to determine intervention merits, investment worth and success is a common theme in both PE and SIM. This serves as a core reason to consider collaboration and co-learning between the two communities of practice. (2) The notion of developing theories to describe the mechanism of social change induced by the employed intervention is practiced in both PE and SIM. Whiles the phases and modalities of practice are somewhat different, the fundamental intent regarding the use of a *theory-based approach* is reportedly the same. (3) Although there is comparatively more emphasis on the desire to validate the underlying investment in SIM, there is some overlap regarding the *intention to use* the resulting evidence to improve the intervention in both PE and SIM. Lastly, (4) What was found to be an exact similarity between SIM and PE was a clear behaviour or demonstration of the values that define '*Evaluative Thinking*'. In both practices, there was clear 'exhibition' of virtues such as identification of intervention assumptions, posing thoughtful questions, intention to pursue deeper understanding through reflection and perspective taking, and varying intentions to make informed decisions in preparation for action.

Conclusion

In exploring the differences and similarities in programme evaluation and social impact measurement; the study has identified key theoretical and practical principles of both PE and SIM using existing literature. The study has further used literature to build and provide a framework around which the nuances within which, both SIM and PE can be understood. Among other elements within the framework, the study identifies key elements such as the values surrounding 'evaluative thinking'; appropriateness of evaluation approach; the role of non-funder stakeholders; the levels of rigour in the evaluation or impact assessment process; and the importance of the assessment process itself.

This framework developed is used to ascertain and discuss reported differences between PE and SIM provided by a purposively selected panel of experts. After considerable analysis, the study concludes that there are definite theoretical and practical differences between PE and SIM. The Key difference is in the varying levels of evaluation of assessment rigour applied, and the use of established criteria as opposed to focus on utilisation. Practically, the SIM process is relatively much more simplified and standardised to support comparability with in the industry, and the approaches are found t be more specific to the needs of the investment industry. This is against a more technically intensive and complexity-friendly process of programme evaluation which caters to context specificity and deepened learning. These differences have implied effects on the variation in applied methods in PE and SIM. The design approaches and data methods tools cater to respective nuances between the two practices.

Finally, some commonalities are identified and are purportedly potential drivers for collaboration and co-learning between the two communities. PE and SIM share a theoretical intent to communicate intervention-worth as well as the common application of a theory-based approach. It is also a conclusion of this study that in communities (of evaluators and Investment managers), the concept and values surrounding 'Evaluative Thinking' are fully adhered to.

Disclosure statement

No potential conflict of interest was reported by the author.

ORCID

Khotso Tsotsotso http://orcid.org/0000-0001-7152-6622

References

Bamberger, M., J. Rugh, and L. Mabry. 2011. *RealWorld Evaluation: Working under Budget, Time, Data, and Political Constraints.* Thousand Oaks, CA: Sage.

Basheka, B. C., and A. Byamugisha. 2015. "The State of Monitoring and Evaluation (MandE) as a Discipline in Africa." *African Journal of Public Affairs* 8 (3): 75–95.

Besharati, N. A. 2014. "Platinum and Passes The Impact of Mining Investments on Education Outcomes in South Africa."

Besharati, N., and K. Tsotsotso. 2015. *In Search of the Education Panacea: A Systematic Review and Comparative Meta-Analysis of Interventions to Improve Learner Achievement in South Africa. Unpublished Research Report.* Johannesburg: University of the Witwatersrand.

Bless, B. D., K. Tsotsotso, and E. K. Gebremichael. 2017. "Evaluation Capacity Assessment of the Transport Sector in South Africa: An innovative Approach." *African Evaluation Journal* 5 (1): 1–8. doi:10.4102/aej.v5i1.188.

Buckley, J., T. Archibald, M. Hargraves, and W.M. Trochim. 2015. "Defining and Teaching Evaluative Thinking: Insights from Research on Critical Thinking." *American Journal of Evaluation* 36 (3): 375–388. doi:10.1177/1098214015581706.

Campbell, D. T. 1991. "Methods for the Experimenting Society." *Evaluation Practice* 12 (3): 223–260. doi:10.1016/0886-1633(91)90039-Z.

Caron, D. J. 1993. "Knowledge Required to Perform the Duties of an Evaluator." *The Canadian Journal of Program Evaluation* 8 (1): 59–78.

Chen, H. T. 1990. *Theory-Driven Evaluations*. Newbury, CA: Sage.

CLEAR-AA. 2017. *African Evaluation Database*. Accessed October 22 2019. https://crest2.sun.ac.za/african_evaluation_db/default/african_eval_db_01

Cook, D. L. 1966. *Program Evaluation and Review Technique: Applications in Education*.

Dart, J. 2004. *Six Normative Approaches to Evaluation*. Cremorne.

Eisner, E. W. 1985. *The Educational Imagination*. New York: Macmillan.

Estrella, M., and J. Gaventa. 1998. *Who Counts Reality? Participatory Monitoring and Evaluation: A Literature Review*. Brighton: Institute of Development Studies, University of Sussex.

Finkelman, J., and K. Huntington. 2017. *Impact Investing: History and Opportunity*. Accessed October 22 2019. https://athenacapital.com/wp-content/uploads/Impact-Investing-History-and-Opportunity.pdf

Flynn, J., J. Young, and C. Barnett. 2015. *Impact Investments: A Literature Review*. Brighton: Institute of Development Studies.

Global Impact Investing Network (GIIN). 2009. *New Industry Group Launched to Facilitate For – Profit Investing that Addresses Social and Environmental Challenges*. Accessed October 22 2019. https://thegiin.org/assets/GIIN%20Launch%202009.pdf

Griñó, L., C. Levine, S. Porter, and G. Roberts. 2014. *Embracing Evaluative Thinking for Better Outcomes: Four NGO Case Studies. InterAction and the Centre for Learning on Evaluation and Results for Anglophone Africa (CLEAR-AA)*. Accessed October 22 2019. http://www.interaction.org/document/embracing-evaluative-thinking-better-outcomes-four-ngo-case-studies.

Guba, E.G. and Y.S. Lincoln. 1989. *Fourth Generation Evaluation*. Newbury Park, CA: Sage.

Hamilton-Smith, N., and M. Hopkins. 1998. "Theory-Driven Interviewing: From Theory into Practice." *Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique* 60 (1): 80–105. doi:10.1177/075910639806000103.

House, E. R. 1991. "Realism in Research." *Educational Researcher* 20 (6): 2–9. doi:10.3102/0013189X020006002.

Jackson, E. T. 2013. "Interrogating the Theory of Change: Evaluating Impact Investing Where it Matters Most." *Journal of Sustainable Finance & Investment* 3 (2): 95–110. doi:10.1080/20430795.2013.776257.

Loveridge, D. 2016. *Results Measurement in Impact Investing: A Preliminary Review*. http://www.google.co.za/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&ved=2ahUKEwjn2JWP0b_lAhUNilwKHUHkD6oQFjABegQIChAE&url=https%3A%2F%2Fwww.enterprise-development.org%2Fwp-content%2Fuploads%2FDCED-RM-in-Impact-Investing-Review.pdf&usg=AOvVaw0dql0slFdcpxJl_TksJ81i

Mason, M. 2010. "Sample Size and Saturation in PhD Studies Using Qualitative Interviews." *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* 11 (2): 32–44.

Newman, D. L., M. A. Scheirer, W. R. Shadish, and C. Wye. 1995. "Guiding Principles for Evaluators." *New Directions for Program Evaluation* 1995 (66): 19–26. doi:10.1002/ev.1706.

O'Halloran, K. 2010. "Investigating Argumentation in Reading Groups: Combining Manual Qualitative Coding and Automated Corpus Analysis Tools." *Applied Linguistics* 32 (2): 172–196.

O'Flynn, P., and C. Barnett. 2017. *Evaluation and Impact Investing: A Review of Methodologies to Assess Social Impact* (No. IDS Evidence Report; 222). *IDS*.

Owen, J. M. 2004. "Evaluation Forms: Toward an Inclusive Framework for Evaluation Practice." In *Evaluation roots: Tracing Theorists' Views and Influences*, 356–370.

Patton, M. Q. 1997. *Utilization-Focused Evaluation: The New Century Text*. 3rd ed. Thousand Oaks, CA: Sage.
Patton, M. Q. 2002. "A Vision of Evaluation that Strengthens Democracy." *Evaluation* 8 (1): 125–139.
Patton, M. Q. 2014. *Evaluation Flash Cards: Embedding Evaluative Thinking in Organizational Culture*. St. Paul, Minnesota: Otto Bremer Foundation.
Pawson, R., and N. Tilley. 1997. "An Introduction to Scientific Realist Evaluation." In *Evaluation for the 21st Century: A Handbook*, edited by E. Chelimsky and W. R. Shadish, 405–418. Thousand Oaks, CA, USA: Sage Publications, Inc.
Popper, K. 1959. *The Logic of Scientific Discovery*. New York: Harper and Row.
Rossi, P., H. Freeman, and M. Lipsey. 2004. *Evaluation: A Systematic Approach*. 6th ed. Thousand Oaks, California: Sage Publications.
Saldaña, J. 2015. *The Coding Manual for Qualitative Researchers*. Thousand Oaks, California: Sage.
Scriven, M. 1974. "Pros and Cons About Goal-Free Evaluation." In *Evaluation in Education* edited by W. J. Popham, 34–67. Berkeley, CA: McCutchan.
Scriven, M. 1998. "Minimalist Theory of Evaluation: The Least Theory that Practice Requires." *American Journal of Evaluation* 19 (1): 57–70. doi:10.1177/109821409801900105.
Scriven, M. 1994. "Evaluation as a Discipline." *Studies in Educational Evaluation* 20 (1): 147–166. doi:10.1016/S0191-491X(00)80010-3.
Shadish, W. R., Jr, and R. Epstein. 1987. "Patterns of Program Evaluation Practice among Members of the Evaluation Research Society and Evaluation Network." *Evaluation Review* 11 (5): 555–590.
Smith, R. M. 1994. Unfinished liberalism. *Social Research* 61 (3), 631–670.
Stake, R. E. 1967. "The countenance of educational evaluation. Department for Exceptional Children, Gifted Children Section." *Teacher's College Record*, 68: 523–540.
Steele, S. M. 1970. "Program Evaluation: A Broader Definition." *Journal of Extension* 8 (2): 5–17.
Trelstad, B. 2016. "Impact Investing: A Brief History." *Capitalism and Society* 11 (2): Article 4.
Trotter, R. T. II, 2012. "Qualitative Research Sample Design and Sample Size: Resolving and Unresolved Issues and Inferential Imperatives." *Preventive Medicine* 55 (5): 398–400. doi:10.1016/j.ypmed.2012.07.003.
Tyler, R. W. 1967. *Changing Concepts of Educational Evaluation*. Chicago, IL: Rand McNally.
Ulrichsen, T. C. 2015. *Assessing the Economic Impacts of the Higher Education Innovation Fund: A mixed-Method Quantitative Assessment*. Bristol: Higher Education Funding Council for England.
U.S. Department of Health and Human Services Centers for Disease Control and Prevention. 2011. *Office of the Director, Office of Strategy and Innovation. Introduction to Program Evaluation for Public Health Programs: A Self-Study Guide*. Atlanta, GA: Centers for Disease Control and Prevention,
Wadsworth, Y. 1991. *Evaluation on the Run*. Melbourne: ARIA.
Worthen, B. R., J. R. Sanders, and J. L. Fitzpatrick. 1997. *Educational Evaluation: Alternative Approaches and Practical Guidelines*. 2nd ed. New York: White Plains.