



# Randomised Control Trials as a Dead End for African Development

Seán Mfundza Muller\*

## Abstract

Randomised Control Trials (RCTs) have been presented as a method for identifying interventions that will improve the lives of residents of economically poorer countries. I argue that RCT-driven development policy is unlikely to achieve even its narrower objectives and, instead, is likely to hinder the attainment of long-term economic and social development objectives in African countries. The article examines two fundamental methodological problems: first, that proselytisers of this approach, despite claims to the contrary, have not addressed the challenge of extrapolating from experimental results to policy interventions; second, that while the use of RCTs is framed as an objective scientific approach to policymaking, it systematically smuggles in the ideological and other biases of the researchers involved. These claims are illustrated with three sets of examples: controversial studies that involved cutting off water to poor households in Nairobi and randomising exposure to a Christian missionary programme in the Philippines; studies on civil servant absenteeism in Kenya and India cited in the 2019 Nobel award; and two case studies from South Africa pertaining to labour market and educational interventions where RCTs distorted the policy process with negative consequences. The conclusion is that the RCT approach is a dead end for African development.

**Keywords:** African development; development economics; epistemic hierarchy; evidence-based policy; randomised control trials

---

\* Johannesburg Institute for Advanced Study (JIAS), University of Johannesburg, South Africa. Email: [academic@Seanmuller.Co.Za](mailto:academic@Seanmuller.Co.Za)

## Résumé

Les essais contrôlés randomisés (ECR) ont été présentés comme une méthode permettant d'identifier les interventions qui améliorent la vie des populations des pays économiquement pauvres. Je soutiens qu'il est peu probable qu'une politique de développement axée sur les ECR atteignent ces objectifs, même les plus restreints d'entre eux, et qu'elle risque plutôt d'entraver la réalisation des objectifs de développement économique et social à long terme dans les pays africains. L'article examine deux problèmes méthodologiques fondamentaux : premièrement, malgré les affirmations contraires, les prosélytes de cette approche n'ont pas relevé le défi de l'extrapolation des résultats expérimentaux aux interventions politiques ; deuxièmement, même si le recours aux ECR est présenté comme une approche scientifique objective de l'élaboration de politiques, il intègre systématiquement les préjugés idéologiques et autres des chercheurs impliqués. Ces affirmations sont illustrées par trois séries d'exemples : des études controversées qui impliquaient de couper l'eau aux ménages pauvres de Nairobi, et de randomiser l'exposition à un programme missionnaire chrétien aux Philippines ; des études sur l'absentéisme des fonctionnaires au Kenya et en Inde citées dans le prix Nobel 2019 ; et deux études de cas de l'Afrique du Sud concernant des interventions sur le marché du travail et l'éducation dans lesquelles les ECR ont faussé le processus politique, avec des conséquences négatives. La conclusion est que l'approche ECR est une voie sans issue pour le développement de l'Afrique.

**Mots-clés :** développement africain ; économie de développement ; hiérarchie épistémique ; politique fondée sur des données probantes ; essais contrôlés randomisés

## Introduction

Randomised Control Trials (RCTs) are purported to be an important tool in reducing world poverty and contributing to economic and social development. On close scrutiny, the basis for that claim turns out to be remarkably weak. Worse, and in marked contrast to the hype about this methodological approach, there are substantive reasons to believe that the use of RCTs is likely to be *harmful* to the long-term prospects and well-being of African people and African countries. In reaching this conclusion, the article brings together a number of different strands of prior work.<sup>1</sup> The first section discusses some fundamental methodological limitations of RCTs and the resultant intellectual inconsistency of the proponents of this approach. This shows how the 'randomista' project is flawed even on its own terms. The second section then discusses how pre-existing views

and biases, whether ideological or about how social and economic processes work, contradict the framing of RCTs as a neutral scientific endeavour. The third section illustrates these arguments by critically analysing three sets of examples: controversial recently published studies on missionary programmes and economic outcomes in the Philippines and cutting off water to poor households in Nairobi in order to encourage payment for services; studies on civil servant absenteeism in Kenya and India cited in the 2019 Nobel Memorial Prize award; and two brief case studies from South Africa relating to labour market and schooling experiments. The final section brings these together to make the case that RCT-driven policy is a dead end for African development.

## **An Overview of Methodological Limitations**

RCTs are a method for obtaining quantitative estimates of causal effects. Their use for this purpose in medicine, most particularly in pharmaceutical trials, is well known. The deployment of RCTs to address social and economic questions is not a new phenomenon but the scale and influence of the approach in development economics in recent decades is without precedent. Although medicine is often cited as providing justification for this approach elsewhere, there is significant contestation about the merits of the evidence hierarchy in medicine that is created by denoting RCTs as a ‘gold standard’ of methodological approaches (Concato, Shah and Horwitz 2000; Rothwell 2005; Cartwright 2007). The application of such methods outside the narrow domain of pharmaceutical research and regulatory processes is also far from straightforward. And there are various respects in which such experiments in social and economic domains may be unethical (Baele 2013; Ziliak and Teather-Posadas 2016; Abramowicz and Szafarz 2020; Hoffmann 2020).

If we were to put ethical arguments aside, which will be the case for the remainder of this article, one could argue that RCTs are ‘mostly harmless’ (Angrist and Pischke 2009) as a different methodological emphasis in intellectual inquiry. That line of argument, however, is undermined by the fact that the use of RCTs in economics has been accompanied by, or premised on, assertions of methodological superiority in the policy realm and a deliberate effort to obtain influence over policy decisions (Banerjee 2007; Banerjee and Duflo 2009; 2011). Such a project seeking policy influence was recently awarded the 2019 Nobel Prize in economics (Nobel Media 2019); its proponents are referred to as ‘randomistas’ (Ravallion 2009, 2018; Deaton 2010).

The basic argument of the randomista project goes something like this:

- Reliable, quantitative estimates of causal effects are needed to make the right policy decisions.
- The assumptions required by other methods to obtain such estimates are implausible.
- Simple analysis using an RCT can identify the effects of policy interest without requiring prior knowledge.
- Therefore RCTs herald a 'credibility revolution' (Angrist and Pischke 2010) in economics and should be prioritised by policy-makers seeking simple answers to important questions.

Although there is nothing in this argument that pertains specifically to development or developing countries, its greatest influence has been on research related to the challenges faced by such countries and their people. On the face of it, that is because the primary proponents of the approach had the problems of developing countries as their intended focus; the institutions they subsequently created, such as the Abdul Latif Jameel Poverty Action Lab (J-PAL), the International Initiative for Impact Evaluation (3ie), Innovations for Poverty Action (IPA) and the Development Impact Evaluation unit (DIME) at the World Bank, inherited that emphasis. A more critical perspective suggests that the approach gained greatest influence in development also because of the role of the aid industry and the relative inability of developing countries to resist such overtures.

Every component of the randomista argument outlined above is contested and has been the subject of substantive methodological criticism – see the surveys by Deaton and Cartwright (2018) and Muller (2014, 2021). Here I will mention one fatal contradiction at the core of advocating RCTs for development policy. Even if RCTs do actually identify the causal effects they seek to, direct policy relevance requires going from that empirical finding in an experimental population to recommending an intervention in a broader population that is the one policy-makers are concerned with.<sup>2</sup> The main obstacle to doing so is that the causal effect of an intervention rarely, if ever, exists in isolation: it depends on the characteristics of the people and the context in which it is implemented. Or in other words: the causal effect of an intervention depends on its *interaction* with other factors. Those factors may vary across populations and geographical space, so that within and across countries the causal effect is different. Furthermore, they may vary across time, so that even in the experimental population there would be a different outcome at some point in the future.

While this is quite intuitive and was pointed out by the first authors to systematically consider the problem in relation to social experiments (Cook and Campbell 1979), it presents a serious dilemma for randomistas. If the

causal effect depends on other factors, then to extrapolate it to a different population requires knowing what the relevant factors are and having data to see how they differ across the two populations. But a key component of the methodological argument in econometrics in favour of RCTs is that assumptions about causal structure are not credible and that RCTs offer ‘simple’ answers to policy questions.

The above counterargument can be formalised (Muller 2015) to show that by simply asserting (rather than establishing) that the results of RCTs apply to broader populations, the randomistas endorse an assumption with the exact same structure as the one they reject when it comes to non-experimental methods. At best this is intellectually inconsistent, at worst it is fundamentally dishonest – either way it constitutes a fatal flaw at the heart of the randomista project (Muller 2020).

The extent of the problem has not been seriously or adequately acknowledged – as is reflected in the continued policy recommendations and promises about policy relevance made by randomistas. However, defenders of the approach have cited a number of indirect solutions to the problem.

Muller (2021b) provides a comprehensive itemisation of these defences and explains why none of them succeed. One purported solution of particular interest here is that practitioners of randomised trials use their ‘expertise’ to assess when and whether experimental results can be applied in other populations. The invocation of qualitative expertise, not least within the broader stance of randomistas in which they imitate ‘rigorous scientific practice’, is particularly notable given the historical disdain shown by economists towards qualitative methods and claims based on individual experience rather than ‘data’, ‘models’ and ‘formal analysis’. But it also does nothing to resolve the intellectual contradiction, since if qualitative expertise can be used to decide prospects of extrapolation why could it not also be used to decide the prospects of identifying causal effects? Doing so would render not just RCTs redundant but also econometric methods more broadly.

Proposing the qualitative expertise of randomistas as a solution to the extrapolation problem also draws attention to another dangerous characteristic of their project: not only are RCTs placed at the top of a hierarchy of evidence, but randomistas seek to place *themselves* at the top of a hierarchy of knowledge (an ‘epistemic hierarchy’). Both these dimensions of the randomista project are especially salient given the history of African economic and social development, in which foreign experts of a certain type imposed policies with little regard for local circumstances or local expertise. Many such experts were economists or were implementing neoclassical economist-derived theories of development.

## Ideologically Infused Experiments

The privileging of economists' views over others with arguably greater expertise is therefore hardly a new phenomenon – in the past it has led to accusations of 'economics imperialism' within the social sciences. However, the promise of simple answers based on 'scientific experiments' combined with a well-funded push for influence (Bédécarrats, Guérin and Roubaud 2019) marks the randomista project as possibly even worse in this respect than its predecessors. In recent times, and likely in response to such criticism, the randomista project has made a concerted effort to present itself as being locally driven or locally informed. The popular and policy literature it has produced seeks to emphasise consultation with local organisations and governments in selecting interventions and the role of such institutions in implementing experiments. Yet the corresponding academic literature, which is the ultimate source of the randomista authority, reflects little local influence and there has been no independent corroboration of other claims.

Being at the top of a hierarchy of expertise is a sure way of securing policy influence, but there is no *prima facie* reason to believe that an academic in the United States running an RCT has more knowledge of a local health system than, for instance, a competent doctor who has worked in that system for decades. Linked to this is a further epistemological problem with RCTs, which has received little attention but is profoundly important for developing countries seeking to determine appropriate strategies and trajectories for social and economic development.<sup>3</sup> The problem is this: *the very choice of an intervention on which to base an RCT is itself the result of a pre-existing conception of how the world works and how it should work.*

Consider the following example. A policy-maker in the Busia district in Kenya is concerned about learning outcomes in state schools because a relatively small proportion of children attain basic competency in literacy and numeracy. The policy-maker asks a randomista at J-PAL for assistance in improving outcomes, and the randomista, naturally, proposes that an RCT or multiple RCTs be run to establish 'what works'.<sup>4</sup> But where does the intervention that will form the basis for the RCT come from? In other words, how does the randomista decide on what is worth trialling by experiment?

One approach might be to try something that appears to have worked elsewhere. But this approach is questionable, since under the full absurdity of the randomista approach nothing can be said to have worked unless it is verified by the results of an RCT. Thus, in the base case the randomista must draw an intervention from the set of interventions they believe might

work.<sup>5</sup> From a purely methodological perspective this is interesting because one of the main methodological arguments made within neoclassical economics for the randomista project was premised on scepticism of *ex ante* causal knowledge (Angrist and Pischke 2010). Yet the mere choice of an intervention imposes researcher beliefs in at least three respects:<sup>6</sup>

1. In determining the set of interventions that may work in theory
2. Determining the subset of 1. that are considered practically feasible
3. Prioritising the possibilities in 2. in order to reach a single intervention or, at best, a handful of interventions.

To continue with the schooling outcomes example, suppose that either the policy-maker or researcher has evidence of high teacher absenteeism and this is deemed to be a likely cause of poor outcomes. What experimental intervention might one institute? A researcher who considers public employees in developing countries to be inherently lazy may favour a punitive incentive system based on increased monitoring. If that is practically infeasible because of resistance from teachers or other stakeholders, a reward-based system may be the next best option. On the other hand, a researcher who believes that underresourced and low-quality work environments negatively affect motivation may propose an intervention that substantially increases aggregate school resources.

Notice that each researcher's preferred experimental intervention may not be even in their counterpart's set of possible or theoretically effective interventions. That will also carry over to interpretation of the results of any RCT. The first researcher may interpret RCT evidence of ineffectiveness of a resource-based intervention as merely confirming what they already expected, whereas the second may interpret it as reflecting the fact that increases in school resources may take time to have an impact. Similarly, the second researcher may interpret the ineffectiveness of an incentive-based intervention as reflecting the fact that absenteeism is caused by other factors, whereas the first may interpret it as reflecting a need to alter the design of the incentive mechanism.

The concern, then, in the present context of RCT-driven development efforts is that materially important subjectivity is being smuggled into policy recommendations in developing countries through notionally objective experimental analysis. A possible response to this concern is to cite randomistas' claims, in some of their policy literature, that they select interventions from the set *proposed* to them by local partners. As has already been noted above, there is no independent corroboration of these kinds of claims about the interactions between these (mostly North American)

researchers and their local counterparts. And the types of interventions randomistas typically implement have a distinctly economic flavour – as is discussed further in the next section.

Even if it were true that such interventions were selected from a set of possibilities identified by local organisations, this may not improve matters much. There are two likely reasons. First, a self-selection effect: organisations with whom randomistas establish medium-term relationships of the kind needed for such experiments are likely to share various aspects of the latter's outlook. Second, if the range of possible interventions considered by local organisations is sufficiently broad, the mere act of selecting one will play the greater role in determining the character of the intervention.

The broader point is that, contrary to the manner in which randomistas portray these studies, there is nothing neutral about RCTs: the interventions chosen for testing are the outcome of decisions by researchers conducting experiments and the institutions that fund them. The experimental interventions will, therefore, reflect preconceived notions of how the world works and what solutions should be considered plausible. Where policy is determined on the basis of findings from such studies, the reliance on such experiments imposes unwarranted blinkers on development policy. Or, to phrase the matter in the foundational terms (Robbins 1932) of modern neoclassical economics, randomistas purport to be doing positive economics and thereby conceal deeply normative aspects of their empirical work and policy recommendations.

### **Illustrative Examples from the Nobel Award and South African Case Studies**

The preceding sections emphasised two particular problems with the randomista approach to development – that context-dependence undermines extrapolation and that RCTs smuggle in pre-existing conceptions of appropriate policy under the guise of science. This article now considers some specific examples to illustrate more concretely these abstract points and the dangers of the randomista project for African (and other developing) nations. The first is two very recent controversies pertaining to an experiment that cut off water to non-paying residents of poor areas in Nairobi, and another that randomised exposure to Protestant missionaries in the Philippines. The second is a strand of work by the 2019 economics Nobel winners on absenteeism among civil servants in developing countries. Finally, two case studies from South Africa are discussed, where the randomista approach has had and continues to have a harmful influence on the policy process, decisions and the allocation of resources.



***Recent controversies: less water, more missionaries for the poor in developing countries***

Despite the increasing sophistication in which the problems of the randomista approach have been cloaked, it is not hard to find new examples to illustrate the troubling attitudes that lie behind such experimentation. Only two that have been recently prominent are discussed below, but there are many similarly concerning examples.

In 2020, researchers published a working paper in the National Bureau of Economic Research (NBER) series which presented the results of an experiment that involved entirely cutting off the water supply to residents of poor areas ('slums') in Nairobi to increase payment rates (Coville *et al.* 2020). The NBER is a North American-based, invitation-only research institution focusing on empirical work in economics. Although its working papers are not peer-reviewed and the ability to publish therein is based first and foremost on professional networks rather than substantive merit, according to Google Scholar the working paper series is the most-cited single outlet in economics.

After the working paper's publication, controversy erupted on social media that called into question the ethics of cutting off water to poor, vulnerable people and communities. A response by three of the authors (Coville, Galiani and Gertler 2020) sought to emphasise the role of the local water authority, essentially seeking to exculpate themselves of responsibility through a claim of local agency. The authors insinuated that the idea to disconnect non-paying households or compounds was that of the local authority and, furthermore, that their role in proposing an experiment with alternative approaches effectively reduced the extent of disconnections over the period studied.

The study falls squarely within the randomista network: one of the authors worked for DIME, the study received funding from J-PAL and DIME (among others) and its international ethical clearance was provided by the 'Innovations for Poverty Action International Review Board'. Exemplifying a point made above, the authors' subsequent elaboration of context seeks to emphasise the role of local partners yet such information was wholly absent from the original NBER working paper. Aside from failing to establish that claim credibly, much as the broader literature fails in this domain too, the authors also neglected to elaborate on their role or that of the World Bank as the lender. They summarise the situation as follows (Coville *et al.* 2020:1):

These interventions grew out of a policy crisis faced by the Nairobi City Water and Sewerage Company, and a 6-year engagement between the utility and research team. To help expand access to water and sanitation services, the Kenyan Government had used support from the World Bank to subsidise and greatly expand physical connections from housing compounds to the water and sanitation system and achieved near universal levels in urban slums. On the measure of increasing access, the programme was highly successful. However, a culture of nonpayment became rampant.

Given the role of the World Bank as financier – on terms left unspecified – and the fact that one of the authors worked at the Bank, it seems plausible that the ‘6-year engagement’ referred to may not have been voluntary on the part of the utility nor one characterised by equal power relations.

The discussion of the institutional challenge faced by the water utility, both in the working paper and in the subsequent response to criticism, is remarkably unsophisticated. The crux of the argument was that since the water utility’s revenue was inadequate to meet its supply and maintenance costs, the extreme act of cutting off water entirely to non-paying poor households was warranted. The rationale being that if the utility’s broader service provision was undermined then more poor households would be harmed in the long run by allowing the situation to continue. However, nowhere in the study do the authors consider ability to pay, the moral and legal rights of citizens to access a certain basic quantity of clean water, or the role of public subsidies. Given that these are among the predominant concerns in the literature on utilities that provide public services (whether privately or publicly operated), their omission reflects, at best, shoddy scholarship. More than that, it illustrates the points made above: how the prior beliefs of researchers and funders distort the perceived set of options available and thereby impose normative beliefs that have marked consequences for welfare and distributional consequences.

In the face of scathing public criticism, the working paper in question was withdrawn, though without providing clearly stated reasons for doing so. A revised version was published a year later (Coville *et al.* 2021) with a new title – ‘Financing municipal water and sanitation services in Nairobi’s informal settlements’ rather than ‘Enforcing payment for water and sanitation services in Nairobi’s slums’ – and with a seven-page appendix that addresses ‘ethical considerations’. The later version now includes numerous Kenyans in the acknowledgements section. The appendix also claims that disconnections were considered only after it was ascertained that alternative measures had been tried or the utility had ruled out alternatives (such as prepaid meters). Nowhere does the revised version acknowledge that any of the changes were made in response to criticism. It is debatable whether

such ex post rationalisations and elaborations should be treated as credible, not least since there was no obstacle to including such information in earlier working paper versions. It is perhaps also notable that the appendix reveals that an earlier study design intended to estimate the elasticity of demand for sanitation services was rendered infeasible because of high take-up rates – much research time and effort would therefore have been lost had the authors not found another experiment to run.

In the same year, a paper was published in the discipline's top-ranked *Quarterly Journal of Economics* in which the researchers sought to examine whether religiosity has an effect on social and economic outcomes (Bryan, Choi and Karlan 2020). Specifically, the study examined whether randomised exposure to the programme of a particular (Protestant) Christian missionary organisation (International Care Ministries) led to an improvement in social and economic outcomes. The programme reportedly 'consists of three components – Protestant Christian theology, values and character virtues ... health behaviors ... and livelihood (i.e., self-employment) skills ... taught over 15 weekly meetings' (Bryan *et al.* 2020). The paper examines a range of primary and secondary religious and economic outcomes, assessed six months and thirty months after the end of the programmes. The headline finding is that 'Six months after the programme ended, treated households have higher religiosity and income ...'. On that basis, the paper concludes with the claim that, 'this church-based programme may represent a method of increasing non-cognitive skills and reducing poverty among adults in developing countries'.

Again, the institutional associations of the authors place them squarely within the randomista project, in that they had affiliations to IPA and J-PAL. This paper had also first appeared in the NBER working paper series two years earlier (Bryan, Choi and Karlan 2018) but had not generated much controversy at the time. Furthermore, the content of this paper also reflects a major scholarly failing. The authors locate their study in relation to the Western debate that relates to Max Weber's proposition about the advantageous nature of the 'Protestant work ethic'. At best the relevance of that debate for the economic development of presently less wealthy countries needs to be established. At worst it is of little or no relevance even to questions of religiosity and economic outcomes in such countries because the primary context Weber had in mind was so different: a subset of European countries in the sixteenth and seventeenth centuries. In either case, what is unarguable is that the authors omit entire literatures on historical and current missionary activity in developing countries. Furthermore, the catalogue of such historical experiences contains numerous examples in

which missionaries played a pivotal role in facilitating the exploitation of the ancestors of current residents of developing countries. Indeed, although there are certainly nuances that merit attention, compelling cases have been made that in some contexts Christian missionary activity was one of the sources of subsequent economic ‘underdevelopment’ (Rodney 1972).<sup>7</sup>

In addition to this blatant failure to properly or even-handedly locate the intervention that the paper studies in its appropriate historical context, the conclusion has little support from its actual findings. The finding that the income of participants increases is not corroborated by any of the other measures of economic outcomes or perceptions of relative economic status. In the body of the paper the authors acknowledge that ‘it is possible that the income result is a purely random Type I error’ – a statistically significant finding that is spurious and arises from testing a large number of alternate hypotheses. Although the authors go to some lengths to dispel this possibility, they are unable to do so convincingly. Furthermore, the income effect is no longer statistically significant after thirty months. An impartial assessment of the empirical findings may quite reasonably conclude that there are no consistent, significant effects of the missionary programme – in stark contrast to the authors’ conclusion.

These papers illustrate not only ethically questionable conduct but also the kind of researcher perspectives that inform the selection and design of experimental interventions. Despite the prestige of the authors and the outlets in which these papers were published, some may nevertheless wish to argue that these examples are exceptions to the norm. In the first subsection below, this article seeks to pre-empt such arguments by demonstrating how one subset of studies cited in the 2019 Nobel Memorial Prize award also exemplifies the criticisms outlined above. In the second subsection, this article discusses two detailed examples from South Africa, which pertain to schooling and labour market outcomes and which illustrate how RCTs can lead to deeply flawed policy processes, a waste of effort and the misdirection of public resources.

### ***Nobel researchers and ignoble civil servants: absenteeism, monitoring and contract teachers***

The 2019 Nobel award identifies a number of noteworthy components of work by the recipients of the award (Abhijit Banerjee, Esther Duflo and Michael Kremer), one of which is ‘a series of papers in the early 2000s, [wherein] Duflo and Banerjee, along with their various co-authors, began a systematic exploration of how to address teacher absenteeism’ (Nobel Media 2019:17). The three papers concern:

1. An experiment that started in 2003 and ran for thirty months using camera-based monitoring and financial incentives (salary bonuses or deductions) in fifty-seven, single-teacher 'nonformal education centres' run by the local NGO Seva Mandir, in Udaipur, Rajasthan, India (Duflo, Hanna and Ryan 2012).
2. An experiment in which Seva Mandir monitored assistant nurse midwife attendance, linked to financial incentives (wage 'fines and punishments'), at public rural health subcentres in Rajasthan (Banerjee, Duflo and Glennerster 2008).
3. An experiment that assigned additional contract teachers, who were paid one-quarter of the usual compensation, to 140 primary schools in three districts of the Western Province of Kenya in 2005 and 2006, funded by the World Bank and implemented by the NGO International Child Support (Duflo, Dupas and Kremer 2015).

Each of these studies could be the subject of a detailed critique, but for present purposes this article focuses on their troubling commonalities. It draws also on the framing of the broader research agenda by Banerjee and Duflo (2006) and the Nobel Committee (Nobel Media 2019).

A first concern is that none of these studies seeks to establish the *reasons* for, or causes of, civil servant absenteeism. This is rather bizarre given that the intention is to test interventions that will resolve the problem. Of course, it is not impossible that one might find solutions to a problem without understanding its causes. But for intellectual and practical purposes it would seem prudent to make some effort to consider what lies behind the policy problem that one intends to 'solve'. One possible explanation for this remarkable omission is that the authors of these studies already had a belief about the causes of absenteeism, which they elected to leave implicit. A hint as to what these beliefs might be is provided by some of the adjectives used to describe absenteeism, such as 'delinquency' and 'shirking' – suggesting a view that absenteeism is caused by laziness or disregard for public service obligations.

In addition to the failure to consider reasons for absenteeism, there is a corresponding failure in the ostensible reasons provided for caring about absenteeism. Although it would be quite acceptable to simply begin with the premise that absenteeism is a problem in and of itself, whatever the reasons for it, the authors of these studies seek to go further. For instance, the ostensible motivation for the nurse incentive scheme was that free proximate public rural health facilities were not being utilised by the very poor (Banerjee *et al.* 2008:487–88). Yet the authors did nothing to actually establish their hypothesis that the reason for non-use of these

services, in favour of private providers, was absenteeism of the staff at the centres: they simply segue from speculation to strong assertions with little actual substantiation.<sup>8</sup>

A second concern relates to the nature of the interventions that were considered and implemented by the authors of these studies. The obvious commonality between all three is: monitor and incentivise. To many mainstream neoclassical economists this may appear unobjectionable, but only because they share an approach in which any problem that pertains to individual behaviour ought to be addressed by the behavioural toolkit of their discipline. This approach can be found in models of individual behaviour, such as principal–agent models and models of individual utility maximisation. In these models of human beings as ‘homogenous globules of [selfish] desire’ (Veblen 1898), it simply may be assumed that civil servants will seek to shirk their responsibilities.<sup>9</sup> And if one assumes that shirking is the consequence of such individual considerations, it follows in the economic worldview that one ought to institute systems that monitor and incentivise (punish and reward).<sup>10</sup> Yet on inspection one discovers that the authors of these studies have done nothing to justify their focus on individual decisions relative to the structural causes of absenteeism.

In an approach characteristic of the entire randomista literature, such structural factors are given a paragraph of consideration in one overview study. Banerjee and Duflo (2006:129–30) belatedly note that it is:

... worth recognizing that the working conditions faced by the providers in many of the studies reported here are hardly ideal and that this might partly explain why providers do not always respond to the incentives ... [in some of the circumstances faced by an auxiliary nurse-midwife in rural Rajasthan] most people would probably be tempted to stay at home at least occasionally. ?

But the authors then go on to say that the fact that teachers in the camera experiment responded to incentives shows that ‘at least for them, improving attendance is within reach’. Yet this argument still fails to give structural concerns their due. It may be that the threat of withholding up to 50 per cent of a teacher’s meagre salary is sufficient to compel them to attend classes more regularly, but that proves nothing about the reasonableness of such a demand. It may only prove that the desperation of the teacher for that amount causes them to overcome the obstacles to attendance. Whether such a teacher is capable of then delivering a good education and whether such a situation is sustainable, is an entirely different matter. And both those considerations are crucial for the ultimate objective, which is improving the educational experience and outcomes of the residents of developing countries.

A third concern is that the thrust of all these studies is to downplay the significance of resources for service delivery and the societal outcomes of this in the countries in question. This is concerning because it is not, in fact, warranted by the analysis that has been conducted. Consider, for instance, the following line of argument in relation to the nurse monitoring scheme:

[the results] show that ensuring that nurses come to work is a low priority for the local health administration and that incentive systems are quickly undermined if there is insufficient political will to enforce them. Given this, resources don't seem to be the main limitation and so pumping more money into the system without attendant reform to reduce absenteeism, as is currently planned under the recently launched National Rural Health Mission, will not solve the underlying problem. (Banerjee *et al.* 2008:488)

The logic in this argument does not hold up to scrutiny. As we have seen, the authors did not establish the reasons for absenteeism. Furthermore, they did not establish why there was insufficient will to enforce the mechanism they had created. An entirely plausible hypothesis is that absenteeism is high because of low salaries and under-resourced facilities that create poor working conditions. Of course, like the authors' own version, this is speculation. But it is least as plausible as the authors' own interpretation and yet has diametrically opposite policy implications. The difference between the two positions seems to arise entirely from the *ex ante* beliefs of different researchers and the wholesale omission of structural factors by the randomistas. It is in this sense that randomistas seek, wittingly or unwittingly, to impose their unsubstantiated prejudices on development policy through supposedly scientific methods.

### *Two illustrative examples from south africa*

As a final set of illustrations, two specific examples from South Africa are described below: the misleading use of an RCT to make the case for an employment tax incentive; and the contribution of RCTs and their proponents to the continued neglect of systemic contributors to poor education outcomes.

#### *The employment tax incentive: using an rct to distort the policy debate<sup>11</sup>*

In the mid-2000s the South African government invited a group of economists, subsequently known as 'the Harvard Panel', to advise on the country's economic policy (Center for International Development 2008). One proposal that emanated from this initiative was for an employment tax incentive for firms, which was aimed at reducing the extraordinarily high



national unemployment rate (Levinsohn 2008). Underlying the proposal was a conceptualisation of unemployment as resulting, at least to a significant degree, from the price of labour being too high. That view had long been contested by trade unions, leading to a polarised situation involving academics siding with different vested interests: one side framed unions as seeking to privilege their members at the expense of the unemployed, while the other side framed business as seeking to destroy collective action in order to better exploit workers. Unsurprisingly, the proposed tax incentive was opposed by trade unions.

The original analysis that had proposed the incentive acknowledged that the question of how responsive employment is to wages is an empirical one and that therefore more evidence was needed to substantiate any incentive and determine its characteristics. Although there was some evidence that the National Treasury and the academics involved already believed that an incentive was desirable, two studies were conducted in order to inform the decision: one was a computable general equilibrium (CGE) analysis (Burns, Edwards and Pauw 2010) and the other a randomised trial of a wage subsidy voucher given to job-seekers.<sup>12</sup> The randomised trial was conducted by academics with links to the National Treasury and funding support from 3ie. The 'policy influence plan' submitted to the funder (Unknown 2011) shows that the researchers anticipated unions as an obstacle to the impact of their findings – clearly expecting a positive result.

The working paper with the details of the study and its findings was published only after Parliament had approved the Employment Tax Incentive Bill (Levinsohn *et al.* 2014a). However, prior to this decision the local researcher who was running the experiment published a number of articles in the popular press arguing for the adoption of the incentive based on the positive findings of the study (Rankin 2012, 2013). The National Treasury also cited the study in its presentation to Parliament. Yet parts of the full working paper that was published later are more cautious about what can be claimed and scrutiny of the study details shows that the RCT provides little, if any, insight into the core policy question.

Among the reasons why the claim that the RCT findings supported the implementation of the national incentive was false are: that the voucher intervention bore little resemblance to the intended incentive; the experimental population was not nationally representative; additional evidence from within the experiment itself (Levinsohn *et al.* 2014a) did not support the claim that a lower wage was the mechanism behind the higher employment rate of voucher-holders; and any positive effect could have been the consequence of a competitive effect among workers that would disappear when the intervention was scaled up.



This example illustrates the points made in the preceding sections. The policy claims based on the RCT were not appropriate given the study's limitations. Furthermore, the researchers showed a clear bias in favour of the policy. One even worked for a consultancy company that provided services to labour brokers who would benefit directly from the incentive.<sup>13</sup> Yet by leveraging the dubious scientism and epistemic authority associated with the randomista project, an RCT was used to endorse a policy that committed the government to billions of Rands of tax incentives for the private sector with little discernible impact on unemployment, and reduced public resources further during a period of fiscal consolidation.

*RCTs and selective denial of systemic contributors to poor education outcomes*

In addition to and overlapping with the studies of absenteeism discussed above, many of the studies cited in the 2019 Nobel award concern educational experiments. The vast majority of these are concerned with interventions that either do not materially increase resources available to schools, or – as in the hiring of low-paid contract teachers – do so in a manner that undermines the wages or power of incumbent teachers. This follows a longer tradition in the economics of education of denying or downplaying the relevance of fiscal resources (on the basis, incidentally, of econometric findings that are not credible by randomista standards).

Given that South African academic economics is arguably a largely imitative enterprise (Muller 2017), it is unsurprising that both these stances have been reproduced in that community. The researchers who currently dominate the education policy space in South Africa produced a report on 'binding constraints in education' that, remarkably, did not list resources as a binding constraint (Van der Berg *et al.* 2016). This is even more remarkable since South Africa is regularly ranked the most unequal country in the world, with high rates of unemployment, poverty and violence and an education system for black South Africans that until 1994 was infamously designed to produce 'hewers of wood and drawers of water'.

The justification for this remarkably extreme position is two-fold. First, one of the authors previously claimed that South African education expenditure was high relative to other countries (Van der Berg 2007) and therefore resources could not be a cause of poor outcomes. Second, in studies done using non-experimental econometric methods the authors and their collaborators apparently failed to find evidence that resources had a significant impact on outcomes. The view that resources are unimportant dovetails with a negative view of teachers, school management and trade unions – that it is not the inequities bequeathed by apartheid

that cause poor educational outcomes but the inefficient management of adequate resources.

RCTs fit neatly into this stance since, as has been the case internationally, they focus attention on non-structural issues, resource optimisation and deficit models of developing country civil servants. Unsurprisingly, then, these researchers and their similarly minded collaborators in the Department of Basic Education have taken enthusiastically to these methods – even though acceptance of the randomista claim about credibility would render much of their prior work non-credible. Examples include poorly thought-out interventions such as randomly sending study guides to schools in one province and then fishing for statistical significance (Taylor and Watson 2015), along with somewhat more carefully considered larger-scale projects to test teacher training or early grade reading interventions.

While the small group of researchers conducting these studies claim that the extrapolation problem is '[not] serious enough to call the method into question' (Fleisch 2008:10), the accompanying analysis suggests an inadequate grasp of the fundamental problem. It perhaps bears mentioning that, given the current enthusiasm for RCTs, adopting this research method serves both the researchers' academic publishing aspirations and desire for policy influence, regardless of whether it serves the public interest.

What has been particularly striking about the recent turn to RCTs in South African basic education policy is that educationalists and civil society activists who championed reading interventions for decades were ignored by government. But economists with little, if any, direct knowledge of the education system who advocated early grade reading interventions on the back of 'scientific' experiments rapidly got to the point of having their recommendations reproduced word for word in the president's State of the Nation Address.<sup>14</sup>

Linked to this is a broader phenomenon in which researchers who enhance their epistemic status using RCTs are also given more authority to inform policy using different methods on separate policy questions – in areas where other researchers may have greater, longer-standing knowledge and expertise. This further compounds an arguably distorted epistemic hierarchy in which academics and others who draw expertise from research are given almost all the weight assigned to non-political inputs, whereas 'experts of practice' – such as teachers – are ignored except to the extent that their expertise is filtered through the former's research, analysis and anecdote. And all of this, as with the employment tax incentive RCT that misled the public and policy-makers, occurs under the broader narrative of 'evidence-based policy'.

## Randomised Control Trials as a Dead End

The focus of this article has been on presenting a methodological critique of the randomista project broadly on its own terms. It is possible of course to approach the project from a number of other, different, critical angles. One is to argue that the project's failures and limitations may be traced to a fundamental flaw in its motivating question, as summarised in the Nobel award: 'understanding why some countries are poor [requires] empirically identify[ing] important sources of inefficiency and policies to address them' (Nobel Media 2019:6). The glaring error of presumption therein is that national poverty and cross-country inequality are driven by 'inefficiency' as opposed to a host of other structural and historical factors.

A notable thread running through these and other papers is the contrast that is drawn between government as an implementing agent and non-governmental organisations. The thrust is that a range of interventions in global South countries are effective when implemented by non-government institutions and involve non-public servants or temporary employees in the public service. Analogous conclusions have been drawn by other researchers in the same methodological genre as the randomistas (Bold *et al.* 2018). But while these studies present such findings as novel and insightful, the extent to which they are is questionable.

Many, perhaps most, public institutions in developing countries can be characterised as under-resourced, under-capacitated bureaucracies overwhelmed by their socioeconomic obligations, with corresponding, stultifying institutional cultures that have become embedded over time. That smaller organisations, which can select their own commitments and be funded accordingly, sometimes perform better than the state is not especially surprising. Nor, more importantly, is this an especially useful observation. The broad implication, if one takes these studies at face value, is that where possible one should seek to bypass the state in order to stimulate development, reduce poverty and so forth. Yet such a conclusion verges on incoherence.

- What would it mean to scale up private initiatives to fulfil public duties? Where would sufficient and sustainable resources come from if not the public purse?
- If resources come from the public purse, how would private organisations be held duly accountable?
- And once proper accountability is imposed for public duties and resources, in what sense would these organisations differ from the state bureaucracy they were purported to replace?

The proponents of RCT-led development have largely side-stepped such difficult questions by simply avoiding the implied endpoint of their own studies. Instead, they propose either policies that seek to outsource public functions at the margin or to introduce monitoring and incentive systems that remake public institutions in the image of crude neoclassical economic models populated by selfish utility maximisers.<sup>15</sup>

The conclusion that arises from the preceding methodological analysis is that the main problem with the randomista project for those concerned with development is not its methodological intolerance (Harrison 2013) *per se*. It is that an undue emphasis on RCTs for policy purposes is methodologically unsubstantiated, smuggles in ideological and epistemic bias, distracts from important questions and in doing all this diverts scarce intellectual and societal resources and political will towards projects that will rarely deliver on even their narrow promises. The randomistas appear to be driven by a 'missionary zeal' (Bardhan 2013) and by the belief that they are the chosen ones to save the denizens of developing countries from poverty with an 'incredible certitude' (Manski 2011) about their findings that is not warranted. The combination of these factors poses a real danger to developing countries that have limited capacity to resist a well-resourced, externally-driven project to determine their policies across a wide range of areas.

Compounding this, as noted by many critics, is that the randomista project focuses both research and policy on narrow questions and interventions that lend themselves to RCTs, rather than on those that are most important for developing countries. The deliberate pursuit of medium- and long-term structural change through a process of learning, which has characterised the development paths of most now-wealthy nations, is outside the scope of the randomista project. So while there has been an attempt by randomistas to frame their stance as one of hope rather than pessimism regarding the prospects of major improvements in developing countries, this is disingenuous. The randomista project is premised, mostly implicitly but occasionally explicitly, on a fundamental pessimism about developing countries achieving the economic improvements of developed countries.

The randomista project can be seen as an extreme manifestation of imperialistic tendencies among economists, premised on dubious claims about economics as a scientific activity. In that light, it is notable that even those who have endorsed the proclamation of a 'credibility revolution' have been silent about what this implies for all past policy advice by economists. If indeed it is true that RCTs are required for credible causal estimates then all past policy claims by economists who have used the implausible

assumptions of other methods must have been inappropriate at best, or harmful at worst (Muller 2023). The zealousness of the randomista project manifests in the argument that the hubris of economists who preceded it was fundamentally flawed but ‘this time is different’. Close scrutiny of the project suggests otherwise. This time may be worse.

For all the above reasons, this article adopts a stronger stance than that of many critics of RCT. Not only will the widespread adoption of and reliance on this method fail to yield the benefits promised by the randomista project, it is likely to hinder the attainment of long-term improvements in the prospects and wellbeing of the residents of developing countries. This remains true even as the external drivers of the project seek to nurture local adherents to, and notionally independent proselytisers of, the methodological ideology. The neglect of structural issues, even at the institutional level, in favour of myopic efforts to tweak individual behaviour is not only unlikely to make a dent in the large-scale challenges faced by developing countries, but in many cases the resultant proposals can be shown to be implausible and inadequately substantiated for even the narrow problems they identify. Properly located, methodologically and epistemically, RCTs should play at most a small role in informing the policy decisions of developing countries. If given the authority and power sought by the randomistas, RCTs will be a dead end for African development. Whatever factors have hindered the attainment of greater progress in African countries since independence, there is no reason to believe that RCTs will address or circumvent them. The challenge for African countries remains to set out, as other countries have done historically, an alternative path to the new missionary complex that has been spawned by the randomista project.

## Notes

1. For the sake of brevity and exposition references are kept to a minimum; more extensive references can be found in other works (Muller 2014, 2015, 2020), which the present article builds on with an eye to concerns particularly related to African development.
2. This problem is widely referred to as the problem of ‘external validity’, following Cook and Campbell (1979) who contrasted it with the problem of identifying a causal effect (‘internal validity’). It is also referred to, among other names, as the ‘generalisability’ or ‘transportability’ problem.
3. There are some cogent critiques of popular conceptualisations of the notion of ‘development’ but the term is used here in a broad, relatively unobjectionable manner to refer to improvement in the wellbeing and prospects of those within a country – without requiring any particular presumption of what improvement might mean. The term ‘developing countries’ is used in a similar vein.

4. If the randomistas were to propose some other approach presumably they would also need to suggest that someone else, who is actually an expert in that approach, be consulted.
5. In some places, randomistas have made much of their consultation with local partners in deciding which intervention to test. Aside from the fact that there is little independent evidence of this, it is quite clear even in such accounts that the researchers do not agree to run interventions that they believe are unlikely to be effective. Furthermore, the nature of, and rationale for, the vast majority of interventions is evidently economistic in nature.
6. This point, and the subsequent education example, is elaborated on in greater detail in Muller (2022).
7. Many such accounts are outside the methodological paradigm of neoclassical economics, which is often the basis (implicitly or explicitly) for excluding them from consideration in such analyses. Within the mainstream economics literature, the econometric challenges in inferring causal relationships between missionary activity and economic development remain the subject of debate (Jedwab, Meier zu Selhausen and Moradi 2022).
8. The authors also do not establish whether the absence of nurses from their offices on the days they were supposed to be there meant they were not in fact working (Banerjee, Duflo and Glennerster 2008:492, fn3).
9. In fact, there is nothing about even the neoclassical model that requires the assumption of selfishness and that assumption has been relaxed in theoretical work over the years. Nevertheless, the assumption of selfishness and maximisation of pecuniary interests and non-work time has been the workhorse of empirical analysis in the mainstream of the discipline since at least the middle of the twentieth century.
10. The resonance with Foucault's 'Discipline and Punish' is somewhat interesting given that all the work considered concerns public institutions.
11. The discussion in this subsection is an abbreviated version of Muller (2021).
12. The nature of CGE studies is such that they effectively assume the answer to the primary question ('Would a publicly funded reduction in the wage causally increase employment?') and model the sensitivity of outcomes to other assumptions; in that sense they are rather uninteresting, and unhelpful, for making the main policy decision.
13. <https://web.archive.org/web/20130621010241/http://prophet.co.za:80/People.aspx>
14. And it is perhaps no coincidence that the dominant demographic in the former group were black women whereas the latter were predominantly white men trained at the university that was the intellectual heart of apartheid.
15. In other instances, some randomista initiatives seek to engage with governments to influence them on a range of fronts, whether to implement experiments, to rely on the results of experiments as a primary evidence base for policy, or to implement or change policies in response to experimental results. Of course, this simply returns us to the more fundamental question as to why randomistas should be given special epistemic status in the policy domain.

## References

- Abramowicz, M., and Szafarz, A., 2020, 'Ethics of Randomized Controlled Trials: Should Economists Care about Equipoise?', in Bédécarrats, F., Guérin, I. and Roubaud, F., eds, *Randomized Control Trials in the Field of Development: A Critical Perspective*, Oxford: Oxford University Press.
- Angrist, J.D., and Pischke, J-S., 2009, *Mostly Harmless Econometrics*, Princeton: Princeton University Press.
- Angrist, J.D., and Pischke, J-S., 2010, 'The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics', *Journal of Economic Perspectives*, Vol. 24, No. 2, pp. 3–30.
- Baele, S.J, 2013, 'The Ethics of New Development Economics: Is the Experimental Approach to Development Economics Morally Wrong?', *The Journal of Philosophical Economics*, Vol. VII, No. 1, pp. 1–42.
- Banerjee, A., 2007, *Making Aid Work*, Cambridge (MA): MIT Press.
- Banerjee, A. and Duflo, E., 2006, 'Addressing Absence', *Journal of Economic Perspectives*, Vol. 20, No. 1, pp. 117–32.
- Banerjee, A., and Duflo, E., 2009, 'The Experimental Approach to Development Economics', *Annual Review of Economics*, Vol. 1, pp. 151–78.
- Banerjee, A., and Duflo, E., 2011, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, New York: Public Affairs.
- Banerjee, A., Duflo E., and Glennerster, R., 2008, 'Putting a Band Aid on a Corpse: Incentives for Nurses in the Indian Public Health Care System', *Journal of the European Economic Association*, Vol. 6, No. 2–3, pp. 487–500.
- Bardhan, P., 2013, 'Little, Big: Two Ideas about Fighting Global Poverty', *Boston Review*, May.
- Bédécarrats, F., Guérin I., and Roubaud, F., 2019, 'All that glitters is not gold the political economy of randomised evaluations in development', *Development and Change*, Vol. 50, No. 3, pp. 735–762.
- Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., and Sandefur, J., 2018, 'Experimental Evidence on Scaling up Education Reforms in Kenya', *Journal of Public Economics*, Vol. 168, December: pp. 1–20.
- Bryan, G., Choi, J., and Karlan, D., 2018, 'Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes', w24278, Cambridge, MA: National Bureau of Economic Research.
- Bryan, G., Choi, J., and Karlan, D., 2021, 'Randomizing Religion: The Impact of Protestant Evangelism on Economic Outcomes', *The Quarterly Journal of Economics*, Vol. 136, No. 1, pp. 293–380.
- Burns, J., Edwards, L. and Pauw, K., 2010, 'Revisiting wage subsidies: How pro-poor is a South African wage subsidy likely to be?', *Development Southern Africa*, Vol. 30, No. 2, pp. 186–210.
- Cartwright, N., 2007, 'Are RCTs the Gold Standard?', *BioSocieties*, Vol. 2, No. 1, pp. 11–20.



- Center for International Development, 2008, *CID South Africa Growth Initiative*. <http://www.cid.harvard.edu/southafrica/>. Accessed 7 July 2017.
- Concato, J., Shah, N., and Horwitz, R., 2000, 'Randomized Controlled Trials, Observational Studies and the Hierarchy of Research Designs', *New England Journal of Medicine*, Vol. 342, No. 25, pp. 1887–92.
- Cook, T.D., and Campbell, D.T., 1979, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, New York: Houghton Mifflin.
- Coville, A., Galiani, S., and Gertler, P., 2020, 'A Comment on the Ethical Issues Twitter Discussion' in "Enforcing Payment for Water and Sanitation Services in Nairobi's Slums" (Version 2), 8 August 2020 (<https://drive.google.com/file/d/1nVGR4qlhWt2EcNiy6d02By3e4kptVSer/view>). Accessed 22 November 2021.
- Coville, A., Galiani, S., Gertler, P., and Yoshida, S., 2020, 'Enforcing Payment for Water and Sanitation Services in Nairobi's Slums', w27569, Cambridge, MA: National Bureau of Economic Research.
- Coville, A., Galiani, S., Gertler, P., and Yoshida, S., 2021, 'Financing Municipal Water and Sanitation Services in Nairobi's Informal Settlements', 9725, Washington, DC: World Bank.
- Deaton, A., 2010, 'Instruments Randomization and Learning about Development', *Journal of Economic Literature*, Vol. 48, No. 2, pp. 424–55.
- Deaton, A., and Cartwright, N., 2018, 'Understanding and Misunderstanding Randomized Controlled Trials', *Social Science & Medicine*, Vol. 210, pp. 2–21.
- Duflo, E., Dupas, P., and Kremer, M., 2015, 'School Governance, Teacher Incentives and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools', *Journal of Political Economy*, Vol. 123, pp. 92–110.
- Duflo, E., Hanna, R., and Ryan, S.P., 2012, 'Incentives Work Getting Teachers to Come to School', *American Economic Review*, Vol. 102, No. 4, pp. 1241–78.
- Fleisch, B., 2008, *Primary Education in Crisis: Why South African Schoolchildren Underachieve in Reading and Mathematics*, Cape Town: UCT Press.
- Harrison, G.W., 2013, 'Field Experiments and Methodological Intolerance', *Journal of Economic Methodology*, Vol. 20, No. 2, pp. 103–17.
- Hoffmann, N., 2020, 'Involuntary Experiments in Former Colonies: The Case for a Moratorium', *World Development*, Vol. 127: 104805.
- Jedwab, R., Meier zu Selhausen, F., and Moradi, A., 2022, 'The Economics of Missionary Expansion: Evidence from Africa and Implications for Development', *Journal of Economic Growth*, Vol. 27, No. 2, pp. 149–92.
- Levinsohn, J., 2008, 'Two Policies to Alleviate Unemployment in South Africa', Center for International Development Working Paper 166.
- Levinsohn, J., Rankin, N., Roberts, G., and Schöer, V., 2014a, 'Wage Subsidies and Youth Employment in South Africa: Evidence from a Randomised Control Trial', Stellenbosch Economic Working Papers 02/14.
- Levinsohn, J., Rankin, N., Roberts, G., and Schöer, V., 2014b, A Youth Wage Subsidy Experiment for South Africa, 3ie Impact Evaluation Report 15 ([http://www.3ieimpact.org/media/filer\\_public/2014/08/20/ie15-rankin-youth\\_wage\\_subsidy.pdf](http://www.3ieimpact.org/media/filer_public/2014/08/20/ie15-rankin-youth_wage_subsidy.pdf)).



- Manski, C.F., 2011, 'Policy Analysis with Incredible Certitude', *Economic Journal*, Vol. 121, No. 554, pp. F261–89.
- Muller, S.M., 2014, 'Randomised Trials for Policy: A Review of the External Validity of Treatment Effects', PhD dissertation, University of Cape Town.
- Muller, S.M., 2015, 'Causal Interaction and External Validity: Obstacles to the Policy Relevance of Randomized Evaluations', *The World Bank Economic Review*, Vol. 29, Suppl 1, pp. S217–25.
- Muller, S.M., 2017, 'What Does an (South) African Economics Look Like?', *Proceedings of the 2017 Scholarship of Teaching and Learning (SoTL) in the South Conference*, Johannesburg, 24–27 July, 2017, pp. 87–97.
- Muller, S.M., 2020, 'The Implications of a Fundamental Contradiction in Advocating Randomized Trials for Policy', *World Development*, Vol. 127, 104831.
- Muller, S.M., 2021a, 'Evidence for a YETI? A Cautionary Tale from South Africa's Youth Employment Tax Incentive', *Development and Change*, Vol. 52, No. 6, pp. 1301–1342.
- Muller, S.M., 2021b, 'Randomised Trials in Economics', in Kincaid, H and Ross, D., eds, *The Modern Guide to Philosophy of Economics*, New York: Edward Elgar Publishing.
- Muller, S.M., 2022, 'The Unacknowledged Normative Content of Randomised Control Trials in Economics and Its Dangers', in Badei, S. and Grivaux, A., eds, 2022, *The Positive and the Normative in Economic Thought*, New York: Routledge.
- Muller, S.M., 2023, 'Is economics credible? A critical appraisal of three examples from microeconomics', *Journal of Economic Methodology*, Vol. 30, No.2, 157–175, DOI: 10.1080/1350178X.2023.2202682
- Nobel Media, 2019, *Advanced Information*.
- Rankin, N., 2012, 'A Common Approach', *Financial Mail*. <http://www.financialmail.co.za/fm/2012/05/24/a-common-approach>.
- Rankin, N., 2013, 'Give Youth Wage Subsidy a Chance', *Mail&Guardian*. <http://mg.co.za/article/2013-11-01-give-youth-wage-subsidy-a-chance>.
- Ravallion, M., 2009, 'Should the Randomistas Rule?' *Economists' Voice*, Vol. 6, No. 2, pp. 1–5.
- Ravallion, M., 2018, 'Should the Randomistas (Continue to) Rule?', Center for Global Development Working Paper 492. <https://www.cgdev.org/publication/should-randomistas-continue-rule>
- Robbins, L., 1932, *An Essay on the Nature and Significance of Economic Science*, London: Macmillan and Co.
- Rodney, W., 1972, *How Europe Underdeveloped Africa*, London: Bogle-L'Ouverture Publications.
- Rothwell, P.M., 2005, 'Subgroup Analysis in Randomised Controlled Trials: Importance, Indications and Interpretation', *The Lancet*, Vol. 365, pp. 176–86.
- Taylor, S., and Watson, P., 2015, *The Impact of Study Guides on "Matric" Performance: Evidence from a Randomised Experiment*, Stellenbosch Economic Working Papers.
- Unknown, 2011, POLICY INFLUENCE PLAN – Grant Number OW2.044 – University of the Witwatersrand. 3ie International Initiative for Impact Evaluation.

- Van der Berg, S., 2007, 'Apartheid's Enduring Legacy: Inequalities in Education', *Journal of African Economies*, Vol. 16, No. 5, pp. 849–80.
- Van der Berg, S., Spaull, N., Wills, G., Gustafsson, M., and Kotzé, J., 2016, *Identifying Binding Constraints in Education*, Stellenbosch: RESEP.
- Veblen, T., 1898, 'Why Is Economics Not an Evolutionary Science?', *The Quarterly Journal of Economics*, Vol. 12, No. 4, pp. 373–397.
- Ziliak, S.T., and Teather-Posadas, E.R., 2016, 'The Unprincipled Randomization Principle in Economics and Medicine', in DeMartino, G. F. and McCloskey, D., eds, *The Oxford Handbook of Professional Economic Ethics*, Oxford: Oxford University Press.