

## Randomised Trials as a Dead-end for African Development

---

**R**andomised controlled trials (RCTs) are supposedly 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 could in fact be *harmful* to the prospects and well-being of Africans and African countries.

This short article elaborates on these concerns as follows.<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 on its own terms*. The second section then discusses how pre-

**Seán M. Muller**

School of Economics  
University of Johannesburg  
South Africa

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 final section briefly illustrates these arguments with two important examples from South Africa where, I suggest, the randomista approach has done, and continues to do, significant harm.

### **An overview of methodological limitations**

RCTs are a method for obtaining quantitative estimates of causal

effects, and their use for drug trials in medical contexts is well-known. The deployment of RCTs to address social and economic questions is not straightforward and may even be unethical in many cases (see Hoffmann, this issue). Beyond that, however, one could argue that as a different methodological emphasis in intellectual inquiry, RCTs are ‘mostly harmless’ (Angrist and Pischke 2009). But what is characteristic of the dramatic increase in the use of this method in economics are assertions of methodological superiority in the policy realm and a deliberate effort to obtain influence (Banerjee 2007; Banerjee and Duflo 2009; 2011). It is this latter project that was recently awarded the 2019 Nobel Prize in Economics (Nobel Media 2019) and the proponents of which I refer to, following others (Ravallion 2009; 2018; Deaton 2010), as ‘randomistas’.

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

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

Every component of the argument is contested and has been the subject of substantive criticism, but for present purposes I focus on one fatal contradiction at the core of advocating RCTs for development policy.

Even if RCTs do actually identify causal effects, direct policy relevance requires going from an empirical finding in an experimental population to recommending an intervention in a broader population that is the one policymakers are concerned with.<sup>2</sup> And 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 people and 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 time as well as geographical space, so that even in the experimental population there would be a different outcome five years later.

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 the methodological argument for RCTs is that assumptions about causal structure, used by non-experimental methods to obtain supposedly causal estimates, are not credible.

This argument can be formalised to show (Muller 2015) 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).

Although the extent of the problem has not been adequately acknowledged – in the sense of refraining from making policy recommendations or promises about policy relevance – some indirect solutions have been proposed.<sup>3</sup> One of these of particular interest is that practitioners of randomised trials use their 'expertise' to assess when/whether experimental results can be applied in other populations. The invocation of qualitative expertise, not least within a broader stance of imitative scientism, will surely interest many readers who are familiar with the historical disdain shown by economists towards qualitative methods and claims

based on individual expertise rather than 'data', 'models' and 'econometrics'. 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 prospects of identifying causal effects? Doing so would render not just RCTs redundant but 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').

Of course, the privileging of economists' views over others with arguably greater expertise is not a new phenomenon – in the past it has led to accusations of 'economics imperialism'. However, the promise of simple answers based on 'scientific experiments' combined with a well-funded push for influence marks the randomista project out as even worse in this respect than its predecessors.

### **Ideologically-infused experiments**

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 American academic 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 problem with RCTs that has received little attention, but is profoundly important for developing countries seeking to

determine appropriate strategies and trajectories for social and economic development.<sup>4</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 policymaker 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 policymaker asks a randomista for assistance in improving outcomes and the randomista, naturally, proposes that an RCT, or multiple RCTs, be run to establish ‘what works’.<sup>5</sup> But where does the intervention that will form the basis for the RCT come from?

One answer might be to say: ‘let us try something that appears to have worked elsewhere’. But this begs the question, since under the full absurdity of the randomista approach nothing can be said to have worked unless that 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>6</sup>

From a purely methodological perspective this is interesting because, as noted above, the methodological motive for the randomista project arose from scepticism of *ex ante* causal knowledge. Yet the mere choice of an intervention imposes researcher beliefs in at least three respects:

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 order to select a single intervention or, at best, a handful of interventions.

To focus on our chosen example, suppose that either the policymaker or the 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? Researcher A 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, Researcher B – who believes that under-resourcing and low-quality work environments negatively affect motivation – may propose an intervention that substantially increases school resources.

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

The broader point is that there is nothing neutral about RCTs: the

interventions chosen for testing are the outcome of decisions by researchers conducting experiments and institutions funding them, and will therefore reflect their preconceived notions of how the world works and what solutions should be considered plausible.

## Two examples from South Africa

Two specific examples from South Africa illustrate the salience of these arguments and the dangers of the randomista project for developing countries: 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 educational outcomes. Both examples are discussed extensively in separate articles.

### The employment tax incentive: using an RCT to distort the policy debate

In the mid-2000s the South African government invited a group of economists, subsequently known as ‘the Harvard Group’, 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 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 already some evidence that the National Treasury and many academics involved 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 et al. 2010) and the other a randomised trial of a wage subsidy voucher given to job seekers.

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, and I do not discuss that work further here.

The randomised trial was conducted by academics with links to the National Treasury and funding support from 3ie, which along with the Jamaal Abdul Latif Poverty Action Lab (J-PAL) and Poverty Action, is one of the main international organisations funding the use of RCTs for development research. 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 only published after Parliament had approved the Employment Tax Incentive Bill (Levinsohn et al. 2014). However, prior to the decision the local researcher running the experiment published a number of articles in the popular press arguing for 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 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 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. 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 at a time when it was implementing fiscal consolidation.

### **RCTs and selective denial of systemic contributors to poor educational outcomes**

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 largely an imitative enterprise (Muller 2017), it is unsurprising that both these stances have been reproduced locally. The researchers who currently dominate this policy space in South Africa produced a report on 'binding constraints in education' that did not list resources as a binding constraint (van der Berg et al. 2016). This in a country regularly ranked the most unequal 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 merely 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 et al. 2017:

10) it is evident that they do not have a grasp of the fundamental problem outlined above. 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>7</sup>

Linked to this is how researchers who enhance their epistemic status using RCTs are also given more authority to inform policy using different methods to address separate policy questions on which other researchers may have greater, long-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 while '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 policymakers, occurs under a broader narrative of 'evidence-based policy'.

## RCTs as a dead-end

The main problem with the randomista project for those concerned with development, then, 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 resources and political will toward projects that will rarely deliver on even their narrow promises. The randomistas appear to be driven by a 'missionary zeal' (Bardhan 2013) 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. And the combination poses a real danger to developing countries that have limited resources to resist a well-resourced project to determine their policies.

Compounding this is that, as noted by many critics, 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 that has characterised the development paths of most now-wealthy nations is outside the scope of the randomista project (see Chelwa, this issue). 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, that is disingenuous. The randomista project is premised, mostly implicitly but occasionally

explicitly, on a fundamental pessimism about developing countries achieving the economic improvements of their predecessors.

Although the connections need to be elaborated in more detail, 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 using the implausible assumptions of other methods must have been inappropriate at best, or harmful at worst. The zealotry of the randomista project manifests itself in the argument that the hubris of economists that preceded it was fundamentally flawed but ‘this time is different’. Close scrutiny of the project suggests otherwise: ‘this time is worse’.

For all the above reasons, I adopt a stronger stance than many critics of RCTs: not only will the widespread adoption 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 well-being of the residents of developing countries. Properly located methodologically and epistemically, RCTs would play at most a small role in informing 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, those. 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 congealed around the randomista project.

## Notes

1. For the sake of brevity and exposition I keep references to a minimum; more extensive references can be found in other work on which the present paper is based (Muller 2014b; 2014a; 2015; 2020) and shorter versions of some of these arguments can be found in Chelwa and Muller (2019) and Chelwa, Hoffmann and Muller (2019a; 2019b).
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 as the ‘generalisability’ or ‘transportability’ problem.
3. In a forthcoming book chapter (Muller forthcoming) I discuss a range of efforts to address the problem, including replication and machine learning, and explain why they are inadequate.
4. There are some cogent critiques of popular conceptualisations of the notion of ‘development’ but I use the term in a broad, relatively unobjectionable manner here to refer to improvement in the well-being and prospects of those within a country – without requiring any particular presumption of what improvement might mean.
5. 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.
6. In some places, randomistas have made much of their consultation with local partners in deciding what 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 economic in nature.

7. And it is perhaps no coincidence that the dominant demographic in the former group were black women while the latter are predominantly white men trained at the university which was the intellectual heart of apartheid.

## References

- Angrist, J.D. and J.-S., Pischke, 2009, *Mostly Harmless Econometrics*, Princeton NJ: Princeton University Press.
- Angrist, J.D. and J.-S., Pischke, 2010, ‘The credibility revolution in empirical economics: how better research design is taking the con out of econometrics’, *Journal of Economic Perspectives* 24 (2): 3–30.
- Banerjee, A. and Duflo, E., 2011, *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, Public Affairs.
- Banerjee, A.V., 2007, *Making Aid Work*, Cambridge MA: MIT Press.
- Banerjee, A.V. and Duflo, E., 2009, ‘The experimental approach to development economics’, *Annual Review of Economics* 1: 151–78.
- Bardhan, P., 2013, ‘Little, big: two ideas about fighting global poverty’, Boston Review, 20 May.
- Burns, J., Edwards, L. and Pauw, K., 2010, ‘Revisiting wage subsidies: how pro-poor is a South African wage subsidy likely to be?’, available at: <http://www.opensaldru.uct.ac.za/handle/11090/73>.
- Center for International Development, 2008, ‘CID South Africa Growth Initiative’, available at: <http://www.cid.harvard.edu/south-africa/>, accessed 13 September 2014.
- Chelwa, G., Hoffmann, N. and Muller, S., 2019a, ‘How randomised trials became big in development economics’, *The Conversation*, 9 December, available at: <https://theconversation.com/how-randomised-trials-became-big-in-development-economics-128398>.
- Chelwa, G., Hoffmann, N. and Muller, S., 2019b, ‘Randomised trials in economics: what the critics have to say’, *The Conversation*, 10 December,



- available at: <https://theconversation.com/randomised-trials-in-economics-what-the-critics-have-to-say-128538>.
- Chelwa, G. and Muller, S., 2019, 'The poverty of poor economics', Africa is a Country, 17 October, available at: <https://africasacountry.com/2019/10/the-poverty-of-poor-economics>.
- Cook, T.D. and Campbell, D.T., 1979, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, Wadsworth.
- Deaton, A., 2010, 'Instruments, randomization, and learning about development', *Journal of Economic Literature* 48 (2): 424–55.
- Fleisch, B., Taylor, S., Schoer, V. and Mabogoane, T., 2017, 'The value of large-scale randomised control trials in system-wide improvement: the case of the reading catch-up programme', *South African Journal of Education* 37 (1): 1–13.
- Harrison, G.W., 2013, 'Field experiments and methodological intolerance', *Journal of Economic Methodology* 20 (2): 103–17.
- 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 Schoer, V., 2014, 'Wage subsidies and youth employment in South Africa: evidence from a randomised control trial', Stellenbosch Economic Working Papers 02/14, available at <http://www.econ.sun.ac.za/wpapers/2014/wp022014>.
- Manski, C., 2011, 'Policy analysis with incredible certitude', *Economic Journal* 121 (554): F261–F289.
- Muller, S.M., 2014a, 'The external validity of treatment effects: an investigation of educational production', PhD thesis, University of Cape Town.
- Muller, S.M., 2014b, 'Randomised trials for policy: a review of the external validity of treatment effects', working paper 127, SALDRU.
- Muller, S.M., 2015, 'Causal interaction and external validity: obstacles to the policy relevance of randomized evaluations', *World Bank Economic Review* 29 (suppl. 1): S217–S225.
- Muller, S.M., 2017, 'What does an (south) African economics look like?', in *Proceedings of the Scholarship of Teaching and Learning (SoTL) in the South Conference*.
- Muller, S.M., 2020, 'The implications of a fundamental contradiction in advocating randomized trials for policy', *World Development* 127.
- Muller, S.M., forthcoming, 'Randomised trials in economics', in H. Kincaid and D. Ross eds. *The Modern Guide to Philosophy of Economics*, Edward Elgar.
- Nobel Media, 2019, 'Advanced information', available at <https://www.nobelprize.org/prizes/economic-sciences/2019/advanced-information/>.
- Rankin, N., 2012, 'A common approach', *Financial Mail*, 24 May.
- Rankin, N., 2013, 'Give youth wage subsidy a chance', *Mail and Guardian*, 1 November.
- Ravallion, M., 2009, 'Should the Randomistas rule?', *Economists' Voice*.
- Ravallion, M., 2018, 'Should the randomistas (continue to) rule?', Center for Global Development Working Paper 492.
- Taylor, S. and Watson, P., 2015, 'The impact of study guides on "matric" performance: evidence from a randomised experiment', *Stellenbosch Economic Working Papers* 13 (15).
- Unknown, 2011, 'Policy influence plan – grant number OW2.044 – University of the Witwatersrand', available at [http://policyimpact-toolkit.squarespace.com/storage/pip-examples/PIP\\_OW2-044.pdf](http://policyimpact-toolkit.squarespace.com/storage/pip-examples/PIP_OW2-044.pdf), accessed 17 June 2018.
- van der Berg, S., 2007, 'Apartheid's enduring legacy: inequalities in education', *Journal of African Economies* 16 (5): 849–80.
- van der Berg, S., Spaull, N., Wills, G., Gustafsson, M. and Kotzé J., 2016, 'Identifying binding constraints in education', RESEP report for the Presidency, 24 May.

