



**QUEEN'S  
UNIVERSITY  
BELFAST**

## **The future promise of RCTs in Education: Some reflections on the Closing the Gap project**

Connolly, P. (2017). The future promise of RCTs in Education: Some reflections on the Closing the Gap project. In A. Childs, & I. Menter (Eds.), *Mobilising Teacher Researchers* (pp. 197-206). Routledge.

**Published in:**  
Mobilising Teacher Researchers

**Document Version:**  
Peer reviewed version

**Queen's University Belfast - Research Portal:**  
[Link to publication record in Queen's University Belfast Research Portal](#)

**Publisher rights**

© Taylor & Francis Group.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

**General rights**

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact [openaccess@qub.ac.uk](mailto:openaccess@qub.ac.uk).

# **The Future Promise of RCTs in Education: Some Reflections on the Closing the Gap Project**

Paul Connolly\*

(\*Centre for Evidence and Social Innovation, Queen's University Belfast)

## **Introduction**

In reading the preceding chapters it is encouraging to note that the debates surrounding the use of randomised controlled trials (RCTs) in education are beginning to move on. As noted by Menter and Thompson (Chapter 3), it is regrettable that one of our main methodology textbooks in education within the UK is still encouraging our next generation of educational researchers to believe that RCTs are 'fundamentally flawed' and that they 'belong to a discredited view of science as positivism' (Cohen *et al.*, 2011: pp. 314 & 318). However, the debates covered in this volume reflect a much more constructive and nuanced engagement with the issues and challenges faced by the use of RCTs in education. Moreover, using the *Closing the Gap: Test and Learn* project (CtG) as a case study has allowed for a meaningful exploration of the potential and limitations of RCTs as they have been employed in practice.

CtG has been a significant project in the context of the UK but it is not unique. CtG is not one trial but comprises seven core parallel trials. There is now a significant tradition of successfully completing trials of this size, and larger, in the UK that have involved various forms of collaboration with schools. Indeed, when these current CtG trials were underway

(mid-2015), our own research found that there had already been 746 RCTs published internationally since 1980, with a quarter of these (26%) involving over 1,000 participants (Connolly, 2015). Within this, we found 83 unique RCTs in the UK and Ireland that had been successfully completed and published. However, it is precisely because CtG is not unique in this regard that it makes it much more valuable as a practical case study for exploring the use of RCTs in practice in education.

The purpose of this chapter is to draw out and reflect upon some of the key issues arising from the preceding chapters; using CtG as a vehicle to highlight several more general points regarding the role of RCTs in educational research. In doing this it is important to declare from the outset that I am not an impartial commentator on this debate. Rather, as Director of the Centre for Evidence and Social Innovation (CESI) at Queen's University Belfast and founding Director of the new National Campbell Centre for the UK and Ireland, I am a strong advocate for the use of RCTs in education. Indeed, CESI has either completed and/or is currently in the process of running over 50 RCTs in the fields of education and social care. However, my advocacy of the greater use of trials is not uncritical and we have learnt much over the last decade regarding the issues and challenges facing the use of RCTs in practice. It is this experience, elaborated more fully elsewhere (Connolly *et al.*, 2017), that I will seek to draw upon in the discussion that follows.

### **The Role of RCTs in Educational Research**

The rather trenchant nature of debates that have taken place over the last decade surrounding the use of RCTs in education, as illustrated by Cohen *et al.* (2011), are more a reflection of broader political rather than methodological concerns. Politically, it is the case that RCTs have tended to assume a privileged position in various policy circles over the last decade or

so, especially in the US (Lather, 2010). The prevalent discourse of RCTs as representing the ‘gold standard’ of educational research has been particularly damaging and has raised quite legitimate concerns regarding the potential marginalisation of other forms of research, as highlighted by Menter and Thompson (Chapter 3). In this regard, Higgins (Chapter 5) is quite right to stress that the RCT is just one of many tools available to educational researchers, designed to do a very specific job. As he also makes clear, the RCT is an absolutely necessary but not sufficient research design for drawing conclusions about the effectiveness of education practice.

The basic logic of the RCT is compelling and illustrates why they are necessary when attempting to understand the effectiveness of educational interventions. Ultimately, if we are interested in determining whether an intervention has had a tangible impact on improving the learning of students, we need some measure of where they started (the baseline or pre-test) and then where they have progressed to (the post-test). It is only by tracking the actual change in educational outcomes of the students (whether in terms of their knowledge, understanding and/or skills) that we will have any objective measure of whether the intervention has led to change amongst the participating students and the magnitude of that change. However, educational outcomes amongst students are likely to change over a period of time and for a wide variety of reasons. If we just tracked changes in a cohort of students participating in an educational intervention then we would have no basis for determining how much of that change has been due to the effects of the intervention and how much may have been due to the plethora of other intervening factors that will have impacted upon them but have nothing to do with the intervention. It is for this reason that we need a control group of similar students who we can also track over the same period of time. Having a control group allows us to determine whether the changes experienced by the students participating in the

intervention are above and beyond those that we would have expected to have happened in any case (as represented by the control group).

Of course, the validity of this comparison of students in the ‘intervention group’ with those in the ‘control group’ rests on the assumption that we are comparing like with like. More specifically, the basic premise of the RCT is that we can be as sure as possible that the only systematic difference between the two groups of students is that one has received the educational intervention whilst the other has not. It is because of this premise that we can then conclude with some confidence that any differences in the changes experienced by both groups must be due to the effects of the intervention as all other potential factors will be broadly balanced across the two groups. This is why the random allocation of students to intervention and control groups (whether individually or as whole classes or schools or other type of grouping) is so critical to an RCT as the process of random allocation, assuming the sample size is large enough, will tend to ensure that all other potentially mediating factors are balanced across the two groups and that they are thus well matched. It is for this reason that random allocation is thus not just regarded as an ‘optional extra’ but is fundamental to the design of an RCT. As soon as this element of the design is compromised then there is a strong likelihood that biases are introduced to the trial. It is also for this reason that we tend to be obsessive about the process of randomisation and with reporting it clearly, in detail and entirely transparently.

However, whilst the logic of the RCT design allows us to draw robust conclusions regarding whether a particular educational intervention has had an effect, on its own it can tell us very little about *why* that effect has occurred (or has not occurred as the case may be). Such questions regarding why an intervention has had an impact or not are equally important in the study of the effectiveness of educational practice. However, they require very different

methods that shift the focus towards: the perspectives and experiences of participants; the complex sets of social processes associated with the intervention; and the contexts within which they occur. In this sense, if RCTs are the ‘gold standard’ for determining whether an intervention has had an effect, then qualitative methods – and particularly in-depth ethnographic case studies – are the ‘gold standard’ for helping us understand causal mechanisms and processes and thus for understanding why. In making this point, it is worth noting that Higgins’ argument about RCTs being a necessary but not sufficient method for drawing conclusions about effective educational practice is equally applicable to these other methods. Thus, for example, it is not difficult to find evaluative studies in education that have sought to understand the effectiveness of a programme using solely qualitative methods. The fundamental problem here is that in the absence of evidence from an RCT or equivalent design, we have no way of knowing whether the intervention has actually been effective in the first place. Whilst our qualitative research may have produced very plausible and compelling theories for making sense of the impact of the educational intervention, these may be completely misguided in explaining outcomes that do not actually exist (Bonnell *et al.*, 2012).

### **The Rhetoric and Reality of RCTs**

Beyond the issue of what role RCTs perform within the wider context of educational research, there is the more fundamental question regarding whether it is actually possible to undertake RCTs in education and, if so, how far they can be meaningfully led by or involve the active participation of teachers. As several chapters in this present volume have documented, the present case study of CtG clearly demonstrates that not only are RCTs quite possible in the field of education but there was little resistance found towards them and, on the contrary, there appears to be a ‘considerable appetite’ amongst schools (Chapter 6). This

certainly resonates strongly with our own experience at the Centre for Evidence and Social Innovation at Queen's University Belfast where we have run a large number of trials in schools over the last decade. Whilst some work is required in talking through the basic logic underpinning RCTs, once this is done we have also found very little resistance on the part of teachers and schools. Indeed, our experience is that the emphasis on effectiveness and outcomes tends to align closely with the everyday concerns of teachers. Teachers thus tend to be continuously reflecting upon their particular pedagogical approaches and seeking out new and potentially more effective methods. They also tend to be natural experimenters and engaged in an ongoing process of comparing the results obtained from adopting one method with those attained previously with another. Perhaps most significantly, we have come across very few teachers who have been afraid of asking the question 'what works?' or who have been concerned with research efforts to determine the effectiveness of particular approaches.

None of this is to suggest that teachers or schools are unaware of the complexities associated with such issues. When a teacher asks the question 'what works?' they will be clearly aware that any answer cannot simply be applied to each and every student in their class but is about the effectiveness of specific approaches *on average*. They will also tend to be very clear that the effectiveness of educational programmes and interventions is likely to vary across contexts and from one group of students to the next. This is why they will often be concerned with asking and seeking to answer the more specific question of: 'what is most likely to work best for my particular class or year group?' All these points apply, equally, to the many parents and students we have engaged with over the last decade. Very few people we have worked with have failed to appreciate the complex and context-specific nature of how students learn and develop. However, they also recognise that there is a need to start somewhere and that, however partial and fallible the evidence might be, it is helpful to learn

from what appears to have been effective (or not) elsewhere. These concerns regarding the notion of ‘what works’, and the inability to take a grounded and pragmatic approach to the evidence claims generated through trials, does therefore appear to be a peculiarly academic one.

Beyond this, the experience of CtG also clearly suggests that the role of teachers and schools within RCTs need not necessarily be relegated to objects of research but that they can be active participants and co-producers of such research. This was particularly the case for the subgroup of schools involved in designing and conducting mini-RCTs. As outlined in several of the previous chapters, with the support of small grants from NCTL, these ‘Early Adopters’ planned and successfully completed 50 mini trials. As would be expected, such efforts on the part of the schools required capacity building and support but clearly demonstrated the potential of teachers to play a strong and collaborative role in the use of RCTs in education. Moreover, and as Menter and Thompson note in referring to this subgroup of schools: those ‘with the most advanced understanding of experimental methods also developed the most sophisticated contextual qualitative methods for a closer understanding of the reasons for positive or negative effect sizes’ (Chapter 3, p X).

However, and in contrast to the relative success of these mini-RCTs, Menter and Thompson have raised some concerns regarding the limited room for flexibility or creativity regarding the larger trials run through the CtG programme. The key point here is that there is no reason why teachers and schools (and parents and students) cannot play a full role as active participants and co-producers of RCTs, even large trials. Having said this, there is a need to understand the nature of trials and thus where in the research process such contributions can be made. In this sense, and based upon our own experience within the Centre for Evidence and Social Innovation, it is quite possible to work collaboratively with teachers and schools,

and with parents, students and local communities, to jointly identify priorities and the focus of an educational intervention and to develop and pilot that intervention. We have also shown how it is possible to work together to jointly identify the outcomes to be measured and the specific design to be employed for the RCT. Moreover, we regularly work collaboratively with key stakeholders to make sense of and interpret the findings that arise from our trials and to agree the best mechanisms for reporting these.

Thus, flexibility and innovation can be built into the entire process of a trial – from initial conception and design through to its execution and then reporting – and it is also equally possible for teachers and schools to be centrally involved as co-leaders of this. However, the one element of a trial where it is necessary to maintain a standard approach is in the actual delivery of the intervention concerned. In this regard, whilst there is complete flexibility in relation to what intervention is developed, once it has been agreed it does need to be delivered as originally intended. To understand the reason for this we need to return to the basic logic of a trial as described earlier. If we were to find a difference between outcomes amongst students in the intervention group compared to the control group at post-test, then we can only attribute this difference to the effects of the intervention if we are clear about what that intervention has been. This, in turn, requires us to identify what the key elements of the intervention should be and then to ensure that these are adhered to for the duration of the programme. Only when this has been done can we be confident in generating claims regarding the effects of a particular educational approach. Moreover, the whole point of RCTs is to identify which interventions tend to be more effective for which groups of students and in which contexts so that those that have been found to be effective can be replicated by others. This is simply not possible if there is complete flexibility and where teachers for a given intervention can adapt it as they wish. In such circumstances, whilst an

intervention may have been found to be effective overall, there will be no way of knowing which practices amongst the wide variety that emerged from the teachers adapting the intervention in differing ways, were those that were critical in achieving the effects.

This, then, is why there tends to be an emphasis within RCTs on specifying exactly what the nature of the intervention is that is to be evaluated, often through setting out its core elements in a programme or manual, and then in measuring how well the intervention was actually delivered against these original intentions. There are, however, three caveats to this focus on what has tended to be termed ‘programme fidelity’. The first is that where interim findings suggest that an intervention may be actually causing harm then there is an ethical imperative to stop the intervention at that point; a practice that resonates with some of the issues raised by Fancourt (Chapter 8). The second point is that this focus on specifying the nature of the intervention does not, in itself, require that the intervention is strictly prescribed. It is quite possible, and valid, to develop an intervention that is only based on a broad set of principles or values to be adhered to or a few key parameters within which teachers or schools must work. So long as these are then followed, a RCT design will still be capable of determining whether the application of such principles or values is sufficient to have an effect. The third and final point to note is that there is an important role to be played by RCTs in formative evaluation. Higgins sets this out clearly, in Chapter 5, in relation to the different emphases of pilot, efficacy and effectiveness trials. In this sense, considerable room for flexibility and innovation remains if RCTs are seen as just specific stages of a longer-term process of developing and refining educational interventions. Whilst there is a need to adhere to the core elements of an intervention at the point that it is being delivered and its effectiveness evaluated, once the findings are known there is every opportunity to adapt and refine the intervention subsequently.

## **The Design and Administration of RCTs**

One of the key lessons we have learnt through our work in undertaking RCTs through the Centre for Evidence and Social Innovation over the last decade has been the importance of starting small and building up from pilot studies and efficacy trials towards eventual larger-scale RCTs. Some of the early trials we were commissioned to run were funded by The Atlantic Philanthropies through its Disadvantaged Children and Youth Programme on the island of Ireland. The emphasis of this programme was on encouraging a shift towards early intervention and prevention by funding a relatively large number of NGOs and community-based organisations to develop and deliver evidence-based interventions. One of the requirements of the funding was that the interventions needed to be evaluated through the use of an RCT so that, over time, the Disadvantaged Children and Youth Programme would contribute to the development of a robust evidence base to inform future efforts and investments.

Further details of the Programme and evaluations of its impact over the last decade are available on The Atlantic Philanthropies' website at:

<http://www.atlanticphilanthropies.org/themes/children-youth>. There is little doubt that the Programme, overall, has achieved significant impact and has led to the prioritisation of early intervention and prevention and also a focus on outcomes-based accountability by the respective governments in Northern Ireland and the Republic of Ireland. However, there were also some clear lessons to be learnt that arose particularly from the early stages of the Programme where there was a focus on developing new interventions against tight timeframes and then moving directly to relatively large-scale RCTs to evaluate their effectiveness. This, in turn, led to concerns regarding such issues as: the premature evaluation of interventions before they had been sufficiently developed; the selection of inappropriate

outcomes and measures for RCTs that did not adequately reflect the actual aims of the intervention(s) concerned; and difficulties experienced at times in managing large-scale trials and ensuring ongoing engagement of schools and their commitment to either delivering the programmes as intended and/or their willingness to continue to participate in the trials to the end.

Through the experience we have gained over the last decade, we have learnt the importance of beginning small; of taking our time in working with teachers and schools and other educational providers to better understand the nature of the problems and issues that we wish to focus on and in identifying and developing interventions that are most appropriate to address these. We have also learnt how important it is to develop interventions that are evidence-based and theoretically-informed and of how useful logic models are in helping to develop a shared understanding not only of the nature of the issues at hand but also of the theory of change that will underpin the intervention seeking to address these. Perhaps most importantly, and as indicated above, we have learnt the importance of starting with small pilots and efficacy trials to test ‘proof of concept’ before beginning to scale up to larger trials; not least of all because there is a world of difference between delivering an intervention in a tightly-controlled and well-supported handful of schools to attempting to roll the intervention out across a whole region or nationally.

My reading of the previous chapters suggests that some of these same issues are also evident in the development and delivery of the CtG programme. For example, there appears to have been relatively little time provided for the development and piloting of the seven core interventions before subjecting them to large-scale RCTs. There are also concerns raised regarding the appropriateness of the two core outcome measures applied uniformly to all the interventions and an acknowledgement by Churches *et al.* (Chapter 2) that some of these

interventions were never intended to achieve change in relation to these two particular outcomes. There is also a lack of clarity regarding the randomisation processes adopted for the main trials. Whilst it is explained that schools were organised into seven ‘pools’ corresponding to the seven interventions and then randomly allocated to control and intervention groups within these pools, Menter and Thompson note that: ‘some schools, post allocation, were allowed to trade with other schools within their alliance for their preferred intervention’ (Chapter 3, p. X). Moreover, and as Cordingley *et al.* report (Chapter 4), once schools were allocated to control or intervention groups, they were free to select which of their classes were to participate in the trial. If correct, both these processes undermine the fundamental principle of randomisation as outlined earlier and are likely to have introduced biases to the respective RCTs. Moreover, and again an issue that is fairly typical of large-scale trials that are not tightly controlled, attrition appears to have been relatively high, ranging from 22 to 48 per cent of pupils across the trials.

Overall, and considering the above points, it is not surprising to find that six of the seven large-scale trials found no effects. Given the space restrictions of the present volume, it is clearly not possible for the chapter authors in this present volume to report the full findings of the various trials here. However, the main report of the CtG programme available online does not provide much additional information (Churches, 2016). In this regard, there are clear and extensive guidelines that exist – through the internationally-developed and agreed CONSORT Statements – for how to properly report trials (see: <http://www.consort-statement.org>). In light of the earlier discussion regarding the basic logic of trials, it should come as no surprise to note that such guidelines require accurate, comprehensive and transparent reporting of all key aspects of a trial to allow for a full assessment of the reliability and validity of the findings reported from an RCT. Such guidelines include the

need to report: the exact processes used for randomisation; the precise levels of attrition and where they occurred in the process; data on the characteristics of the intervention and control groups at pre-test to allow assessment of their comparability; the statistical methods used for the main analysis and any additional analyses planned; full details on the outcomes used that should all be pre-specified; and specific information on each of the interventions and how they were actually delivered with sufficient details to allow replication. In relation to CtG, whilst the main report includes some of these details, key information is currently missing. It is noted however that, at the time of writing, the main report had only recently been published and it is therefore hoped that there are plans to report the findings of each of the trials more fully in due course.

## **Conclusions**

It is clear from the chapters in this book that CtG has been an important programme nationally. It has provided further evidence that RCTs are not only entirely possible in education but that there is little resistance to them and, on the contrary, there is a significant appetite amongst teachers and schools to participate in them. The evidence from CtG, especially from the impressive number of mini-trials conducted, is that schools can also play an active role in co-designing and undertaking pilot trials themselves. However, and as also noted, there are also lessons to be learnt regarding the particular approach adopted to the design and delivery of the CtG programme, especially in relation to the tight time pressures involved and the decision to move directly to relatively large-scale trials for interventions that, for the most part, had not previously been subject to trials. This latter point is important in that whilst six of the seven core interventions were well established in terms of their delivery in schools, they had not benefited from the process of pilot testing or efficacy trials. This is a significant limitation as many of the problems and issues faced by the current CtG

trials (e.g. ill-defined outcomes; high levels of attrition; inconsistencies in the randomisation processes; and problems associated with fidelity of delivery) could have been identified and addressed early and, most importantly, before the considerable investment and expense incurred through large-scale trials. In this regard, and just as we have learnt through our own engagement in RCTs over the last decade, starting small and working carefully through the stages of pilot tests and efficacy trials before subjecting interventions to larger-scale effectiveness trials not only helps mitigate many of these problems but also creates a more appropriate context for the development of meaningful collaborative partnerships between researchers, schools, parents and students where RCTs can be genuinely co-produced.

This latter point is critical and reflects the way in which RCTs are increasingly becoming high stakes tests of interventions. In this regard, positive findings from an RCT can be extremely beneficial in providing robust evidence to secure further funding and support for a specific intervention. However, and conversely, there is a growing trend for particular interventions to be evaluated by one RCT and then for their future to rest upon a positive result from that trial. In cases where the findings from the RCT are inconclusive and/or provide no evidence of effectiveness, there is the risk that further funding and support for that intervention will simply be withdrawn. This makes it all the more imperative that we avoid the pressure to move too quickly and prematurely to large-scale RCTs. However, it also emphasises the need to contextualise the findings of individual trials within the wider evidence-base provided by other trials of that intervention or similar interventions.

With the increasing number of RCTs being completed in education, this is where systematic reviews and meta-analysis can play an increasingly important role. Even where we are dealing with a clearly-defined intervention that has been subject to several RCTs, meta-analyses still tend to reveal considerable variations in the size of the effects found from one

trial to the next. Such variations partly reflect differences in the interventions themselves as well as variations in the differing contexts within which the intervention has been delivered and also the differing characteristics of the students it has been delivered to. However, they also partly reflect simple randomness between samples. There will be cases therefore where an intervention is likely to be effective but where, simply because of the random way in which a particular sample has been chosen for an RCT, the findings of that trial are inconclusive or even negative. At the very least, this is where it is important not to dismiss a specific intervention based on the findings of just one trial but to contextualise the trial findings through a broader meta-analysis that can pool the data from several trials of the same (or similar) interventions. However, the promise of systematic reviews and meta-analyses is much more than this. With the ever-increasing number of RCTs now being published there are more opportunities to pool data from several trials of similar interventions to begin to analyse the variations that exist. This, in turn, will provide greater opportunities to begin to assess in which contexts particular interventions tend to be more effective and also for what subgroups. Moreover, we will increasingly be able to note the key elements of each intervention and use meta-analyses to assess what particular elements, and combinations of elements, of interventions are most significant in achieving positive outcomes and, again, in what contexts and for whom?

## **References**

Bonnell C., Fletcher, A., Morton, M., Lorenc, T. and Moore, L. (2012) 'Realist randomised controlled trials: A new approach to evaluating complex public health interventions', *Social Science & Medicine*, 75 (12): 2299–306.

Churches, R. (2016) *Closing the Gap: Test and Learn, Research Report Winter 2016*.

London: National College for Teaching and Leadership. Available:

[https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/495580/closing\\_the\\_gap\\_test\\_and\\_learn\\_full\\_report.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/495580/closing_the_gap_test_and_learn_full_report.pdf)

Cohen, L., Manion, L. & Morrison, K. (2011) *Research Methods in Education*, 7<sup>th</sup> Edition,

London: Routledge.

Connolly, P. (2015) The Trials of Evidence-Based Practice in Education. Keynote address,

British Educational Research Association Annual Conference, 15-17 September,

Queen's University Belfast. Available at:

<https://www.youtube.com/watch?v=svuMXIAsaCE>

Connolly, P., Biggart, A., Miller, S., O'Hare, L. & Thurston, T. (2017) *Using Randomised*

*Controlled Trials in Education*. BERA/Sage Research Methods in Education Series.

London: Sage.

Lather, P. (2010) *Engaging Science Policy: From the Side of the Messy*. New York: Peter

Lang.