Sheffield Hallam University

Randomised Controlled Trials and the Interventionisation of Education

BURNETT, Cathy <http://orcid.org/0000-0001-6087-244X> and COLDWELL, Michael <http://orcid.org/0000-0002-7385-3077>

Citation:

Copyright and re-use policy

Randomised Controlled Trials and the Interventionisation of Education

Cathy Burnett & Mike Coldwell

Abstract

Since the 1990s there have been repeated calls for the systematic use of randomised controlled trials (RCTs) to inform educational decision-making. The advent of the Education Endowment Foundation (EEF) – described as England's What Works Centre for Education – in 2011 has made this a reality in England: by 2020, over a third of English schools were involved in such trials. Despite much debate about the value and role of RCTs, less attention has been paid to one specific effect of such trials, a phenomenon we refer to as interventionisation. This article uses two examples, focused on language and literacy education and teacher professional development, to demonstrate how increased use of trials may work to 'interventionise' education through channelling the focus of innovati on and development to tightly structured interventions and generating a series of narrowing effects. It argues that a broad view of research and a diversity of methodologies is needed not only to generate rich understandings of educational practice, but to develop and sustain educational provision that is fit for a dynamic world and which responds to the challenges and opportunities presented in complex educational contexts.

Key words: Randomised controlled trial, educational interventions, education, EEF, evidence-based practice

Introduction

Scrutiny and critique of research evidence in education has intensified over the past two decades resulting in various calls for more information on 'what works' for schools, teachers and learners (Marsh, 2005, Tooley & Darby, 1998). This has led to repeated calls for the widespread use of randomised controlled trials (RCTs) and associated systematic reviews as a key source of evidence to inform practice. RCTs are experimental designs within which a group is randomly split into (at least) two subgroups, one (or more) of which - the 'intervention group' - receives some form of new input (teaching, medicine, resource) with the rest - the 'control group' - carrying on as usual[1]. An outcome such as an aspect of mathematics learning is defined and a measure created (such as a test) and the difference in results using the measure between the groups indicates whether and to what extent the new input appears to be more effective in relation to this outcome (the 'effect size'). The logic of the RCT is that the randomisation processes reduce systematic differences between the intervention and control groups, providing evidence that the intervention causes the improved outcome. Systematic Reviews synthesise the findings of a range of evidence but usually give greatest weight to evidence from RCTs.

During the 1990s and 2000s, while methodological skirmishes occurred between proponents of RCTs (e.g. Oakley, 2006; Taylor-Fitzgibbon, 2003) and critics (e.g. Hammersley, 2008; Pawson & Tilley, 1997), RCTS in education across the world were generally few and far between, with one exception being health education. In recent years, however, calls for evidence-based practice have been highly influential in shaping educational policy and practice in many jurisdictions and this has led to widespread dissemination of the results of RCTS to schools through, for example, the What Works

---

[1] Sometimes two or more interventions are compared without a non-intervention group, but this is the usual model

Clearinghouse in the USA (IES, n.d.) and – more recently – school and teacher networks (Nelson & Campbell, 2019). In England, RCTs remained largely marginal until the influential intervention of Goldacre and his argument for greater use of RCTs in education (Goldacre, 2013) and the creation of the Education Endowment Foundation (EEF) in 2011, which has funded large numbers of RCTs in educational contexts (with over a hundred published at the time of writing). As a result, the number of education-focused RCTs conducted in England has grown rapidly, with claims that a third of English schools are now involved in such trials, either as partners in research design or as participants (The Economist, 2018). While educational research more broadly continues to draw on a diversity of methodologies, this major investment in funding combined with effective dissemination means that RCTs have gained considerable influence in tandem with the drive for evidence-based practice. EEF, for example, provides easy access for schools via its website to information about the effectiveness of various programmes as measured by randomised controlled trials.

In this article we argue that one of the effects of the expansion of RCTs, and their promotion by government bodies and other influential advocates, is an increased focus on interventions as a route to educational improvement at the expense of alternative approaches. To be studied using RCT approaches, policies and practices must be clearly defined, and those most tightly structured – for example, those that are routinised and time limited – typically show the strongest statistical effects in RCTs and can consequently appear to be disproportionately successful in relation to less easily defined programmes of development. As such, not only may they be attractive to schools, but researchers seeking funding may frame educational initiatives in ways more likely to demonstrate effects in RCTs – that is, to frame initiatives as interventions. In this article therefore we explore how a disproportionate focus on RCTs may be supporting increased use of interventions which may lead to what we refer to as *interventionisation* in and of education, which, as we expand below, we define as a set of narrowing effects produced by using interventions as a key strategy for educational improvement. We argue that an over-emphasis on RCTs not only limits what can be learned from the wide range of available educational research, but may shape the nature of educational development and innovation in ways that are problematic. Our critique has two key thrusts. Firstly, interventionisation may narrow the focus of teaching and learning, and secondly, by so doing, perpetuate a research agenda that marginalises research that uses more open-ended approaches to understanding education. Our argument is cautionary, as there is to our knowledge no extant research that has measured the relative use of interventions in recent years, or that has looked broadly at the impact of interventions on school practices and culture. However, we do suggest that the concerns raised generate important questions for educators, educational researchers and policy makers.

In developing this argument, we begin by locating our concerns in relation to wider debates about the use of RCTs and interventions in education before focusing on two areas of educational practice that have received considerable investment in RCTs in recent years: language and literacy education, and professional development in education.

Randomised Controlled Trials: Ongoing debates

The value of RCTs in education has been the focus of much critique, particularly given their increasing use and claims about their status as the 'gold standard' (Oakley, 2006) of research evidence. Arguments fall broadly into three groups, the first two of which – linked to validity and applicability – are framed by a concern with understanding causation. The third – focussed on ethics – at least in part moves beyond this.

Firstly, in relation to validity, there is a set of arguments that questions the ability of RCTs to provide evidence of causation, even on their own terms. Cartwright (2019) argues that, since 'blinding' (whereby the experimenters, deliverers and recipients of interventions do not know whether they are involved in the experimental or control group/condition) is usually not possible in educational RCTs, a host of issues can emerge including 'implicit and explicit bias in evaluating results; effects of expectation of pupils, teachers, parents, and peers; unintended differential treatment of pupils' (p10). In addition, Harrison and McCaig (2017) signal a range of other potential issues including unmeasured negative consequences that may outweigh those measured as part of the intervention and

complexity (of the problem or the social context) which can make RCT designs impractical and causal inference suspect.

On a different note, Wrigley and colleagues (Wrigley, 2018; Wrigley & McCusker, 2019) have argued that the primacy of RCTs and Systematic Review (SR) is based on an inaccurate and simplistic view of science. In particular, they argue that it is flawed to see a scientific orientation as requiring experimental methods, when in fact a variety of approaches are used including: 'reflective observation, intelligent noticing, trial and error, and even intuition' (p.2). Following the critical realist approach of Bhaskar, and echoing arguments made by realist evaluators Pawson and Tilley (1997), Wrigley and McCusker argue that since the world operates as an open system, closed system approaches (where researchers try to isolate change processes, for example using experimental designs) only ever have limited application, particularly in social research. From this perspective, a range of approaches to investigate the social world are required in relation to causal explanation. Realists like Pawson and Tilley argue that strong causal theory is crucial here, which RCT evidence (on its own, at least) is ill-suited to contribute: a point taken up by others aiming to make stronger judgements in relation to RCT evidence. For example Joyce (2019, p.17) notes that 'other methods, including qualitative methods that aim to explain complex, contextually embedded phenomena' are better suited to the task.

The second set of arguments relate to the wider applicability of evidence from RCTs. Phillips (2019, p.6), makes a key point in relation to the applicability of RCT evidence that was uncovered in the 1980s: that increasing the internal validity of a trial reduces its wider applicability, so 'in general, internal and external validity were at odds, in that steps taken to strengthen or maximise one of them were likely to decrease or weaken the other.' Cartwright (2019, p11) proposes therefore that RCTs need to be complemented by a range of evidence including 'psychological, educational, and sociological theory, evidence of what has worked elsewhere, other studies, and knowledge of your own pupils and setting.' (p.11)

A third set of arguments relate to ethical concerns linked to the different experiences of treatment and control groups. Notwithstanding the difficulties of identifying control groups within education (given the diversity of educational practices and attainment levels across regions, schools and classes), providing different experiences for different learners will disadvantage some of them: those in the control group if the intervention is successful and those in the treatment group(s) if not (Oakley, 2006). One response from medical research is to only conduct RCTs if experts are genuinely uncertain as to which course of action is best (referred to as equipoise – see Beauchamp and Childress, 2009), and another is to be confident that the intervention to be trialled will be beneficial (Bonell et al., 2011). However, as Fives et al. (2015) argue, these conditions can be mutually exclusive within educational and other social research, and professionals' duty of care may be at odds with requirements for scientific rigour.

These three areas of debate relate to the logic of RCTs in understanding the social world and their consequences, to the extent to which RCTs in themselves make a contribution to knowledge about education. This article however speaks to wider debates about the RCTs and the development of educational practice. It is worth emphasising that we do not write from a position of opposition to RCTs in principle; indeed, one of us leads RCT evaluations as part of the EEF evaluators' panel and has previously argued that, when underpinned by clear theory and accompanied by process evaluation focussing on implementation in specific contexts, RCTS provide some of the strongest forms of evidence in relation to understanding causal mechanisms at the level of an individual educational programme (Author2 & colleague, 2018). Our argument is that the fundamental shift in research efforts in education towards RCTs, and their effective promotion by policymakers and other key influencers, has had the effect of deflecting attention from the results of other forms of research and that this in itself is generating not only a partial, and arguably impoverished, evidence base for education. This in turn may limit the range of educational approaches and educational ambitions that gain influence amongst schools, school leaders and teachers. Our concern therefore is with what

happens as RCTs enter relations with wider educational practices, with what – in effect – they *do* within complex educational systems.

These arguments build on the idea that methods actively work to sustain certain ways of knowing and doing. As Law writes, 'The argument is no longer that methods *discover* and depict realities. Instead it is that they participate in the *enactment* of those realities' (Law, 2004, p.45). Building on similar ideas, Biesta (2010; 2016) argues that evidence-based education, with its emphasis on causal relations, works to narrow educational systems and reduce the agency of teachers and learners: as the results of RCTs are disseminated in an education system that values 'hard evidence' over insights generated through other evidence forms, certain practices may gain greater legitimacy than those that are impossible to measure through RCTs. Our contribution to this debate is that the availability of funding for conducting RCTs, and their effective promotion as a key route to evidence-informed practice, funnels research and development into certain kinds of educational programmes and instantiations of educational practice, which we, following EEF, refer to as *interventions*, and that this process may have significant implications for education, a set of effects that we refer to as *interventionisation*.

Interventions
By interventions we refer to structured programmes of activities, which typically involve a clearly defined set of materials, practices and/or routinised teaching sequences that are often time-bounded. Interventions are adopted by schools to address an identified need, relating to a specific group of pupils or an area for development across a school. Their purposes range, for example, from addressing the acquisition of skills or knowledge (e.g. metacognition; aspects of reading), to encouraging specified behaviours (e.g. school attendance; pupils' self-regulation), to addressing attitudes or aspiration (e.g. aspiration towards a career in science). RCTs can, in theory at least, be used to measure the impact of any kind of educational programme, and indeed the EEF has been open to trialling a wide range of approaches through their programme, underpinned by diverse theoretical and educational standpoints. Importantly for our argument, however there is some evidence that highly structured interventions are more likely to have greater effect sizes in randomised trials than less easily codified approaches or initiatives. For example, one of the world's foremost proponents of experimental evidence review argued in a 2018 blog that 'effective programs are usually well-specified, with clear procedures and materials' (Slavin, 2018), and indeed the process of preparing a programme for trial often involves developing a series of materials, rubrics or training that standardise or proceduralise approaches. Of course, however well-codified, interventions will always be enacted in messy, complex educational contexts, inflected by various practices, values and purposes, and the extent to which it is possible to bound initiatives *in practice* has been the subject of much critique (Moss, 2012; Schwandt, 2005). However, our focus is not so much on how interventions are enacted or the value of interventions per se. Instead we are interested in the implications for practice of the widespread implementation of interventions – and the translation of loosely structured or open-ended approaches *into* interventions – fuelled by the availability of funding for RCTs. Our concern is that treating RCTs as a key source of evidence requires a focus on interventions, and that this in turn may lead to an over-emphasis on interventions as a key route to educational improvement which, in turn, may have negative or limiting effects. Our interest is in what happens as RCTs- and the interventions they require- come into relation with other aspects of educational practice.

In studies of institutional, commercial and industrial practices, Law explores how ways of knowing and doing are upheld through heterogeneous networks of human and non-human actors, networks that include research processes as well as ideas, processes, equipment, media, tools, people and so on (e.g. Law, 2004; Law & Mol, 2008). Certain understandings and practices are performed and held in place by these networks, or hinterlands (2008, p.242), networks that work to perform stability as actors hold each other in place. Law and Ruppert's (2013) theorisation of research methods as 'devices' is particularly helpful in thinking about the relational effects associated with RCTs. Devices are, for Law and Ruppert (2013), structured 'teleological arrangements' that 'assemble and arrange the world in specific social and material patterns' (p.230). These devices can be social in nature, and can

produce intended and explicit outcomes. RCTS – as devices – produce knowledge about effective interventions within an experimental paradigm, and this intended knowledge is made public via published analysis plans and evaluation protocols. However, as Law and Ruppert note, devices may also produce 'unrecognised and unintended social consequences' (Law & Ruppert 2013, p.231).

In what follows we explore possible unrecognised and unintended consequences with reference to two examples: language and literacy education; and teacher professional development. In each case, we review critically the kinds of insights that can be – and are being – generated through RCTs, and consider how and why an over-emphasis on interventions may be problematic. Our evidence base is inevitably selective and we do not attempt a comprehensive overview of the contribution of RCTs to knowledge in either of these areas. We do, however, identify a number of questions and possible trends that are worthy of further investigation. The first example highlights what can be lost as interventions emerge as a key route to school improvement. The second charts how a focus on intervention can funnel attention in a broad area of educational support to a highly specific model. Together they illustrate how an emphasis on RCTs may produce a series of narrowing effects, effects that may intensify as they interface with other aspects of educational practice.

Interventionising education: two examples

a) *Language and Literacy Provision*

In this example we consider the scope of language and literacy RCTs within the Education Endowment Foundation's work since its inception in 2011, to illustrate how one organisation's programme of research can influence significantly the development of aspects of school curriculum and pedagogy. Language and literacy have always been high stakes for schools, not least in England where attainment in literacy is central to school accountability. It is therefore unsurprising that language and literacy have been the subject of numerous EEF-funded RCTs. At the time of writing (November 2019), EEF had funded 33 completed randomised controlled trials tagged as addressing language and literacy, representing approximately one third of all 98 projects funded. The projects subjected to trials vary considerably in approach and reflect diverse underpinning assumptions about literacy and language learning. Here we do not comment on the range or quality of those interventions nor the validity of the trials, nor expand further on the methodological challenges associated with RCTs in the area of language and literacy (see Moss, 2012). Our focus is on what may be *omitted* if, as RCTs become the prime source of evidence for educational improvement, interventions become a key strategy for educational improvement.

Our first point is that while language and literacy education is shaped by research and practice reflecting a variety of paradigms (Ellis & Smith, 2017; Hall, 2013), an emphasis on interventions sidelines certain understandings about language and literacy, and by implication certain pedagogical approaches. This is because approaches deriving from certain perspectives fit the logic of RCTs more easily than others. A phonics programme for example, reflecting a cognitive-psychological perspective on literacy (see Hall, 2002), may be devised and described for use in different locations, and tests exist that map onto expected outcomes. Indeed, at the time of writing, at least 19 of the EEF's completed projects (linked to a variety of topics) appeared to reflect a cognitive-psychological perspective (although underpinning theorisations are not always explicit). Approaches emanating from other perspectives, however, are less amenable to RCTs. From a sociocultural or sociopolitical perspective, for example, literacy is embedded in personal, social, political and civic life, and literacy education involves developing learners' communicative repertoires through valuing and building on the litera*cies* they bring to school (Cummins, 2000). There are numerous examples of approaches that have built on such understandings (e.g. Comber and Kamler, 2005; Souto Manning & Yoon, 2019). However, these are difficult – perhaps impossible – to codify for the purposes of a trial. In contrast to a routinised intervention, these approaches are loosely framed: they involve tailored working with learners' interests, experiences and preferences, the agility to respond to opportunities for reading and writing as they emerge through ongoing activity, and a sensitivity to relationships between literacies and community, power place and identity. Such approaches are therefore unlikely to be submitted to

trial (as the philosophical underpinnings of RCTs are at odds with those that underlie such approaches) and – if they were – packaging them as time-focused, transferable interventions would dilute their value (as standardisation would detract from localised and situated pedagogies). Addressing a broad area for development such as language or literacy through interventions, then, may ignore insights or recommendations generated through a range of research and close down opportunities for professional reflection and action.

It is worth noting here that some EEF-funded projects *have* explored open-ended approaches and indeed some of these have generated encouraging results from trials. One RCT, for example, investigated the impact on children's attainment of a teacher professional development programme based on Alexander's (2011) approach to dialogic teaching (Jay et al., 2017). However, the process of framing an open-ended approach based on principles for practice rather than a teaching sequence brings certain challenges, and some of the subtleties and complexities of what is being advocated can be erased during the process of dissemination. Snell and Lefstein (2018), for example, explore how dialogic approaches may disadvantage some learners when operationalised in particular ways. This is not to suggest that sensitive teachers cannot adapt dialogic approaches to address such concerns, but rather that appropriate responses are hard to codify for the purpose of transferable interventions, and that the process of codifying approaches as interventions undermines – or at least downplays- the teacher's role in responding flexibly to learners' needs over the short and long term.

Our second point is that increasing reliance on interventions (with all the associated costs of development and evaluation) may help sustain what is effectively an outdated literacy curriculum. Literacy in everyday life has changed and diversified considerably in recent years and looks set to continue to do so as our relations with each other and the world are increasingly digitised (Gillen, 2014). Communicative practices in contemporary life involve using a range of modes and media. They require not only continually evolving sets of skills (in creating, navigating and making sense of texts on and off-screen), but a critical engagement with what texts do within complex networked environments, an awareness of associated possibilities (and risks) for participation in civic, economic, political and social activity, and engaging with literacies as distributed (rather than individual) activity (Lankshear & Knobel, 2011). It has long been argued therefore that literacy education should recognise that literacy is evolving and diversifying and that literacy education needs to be agile in responding to a shifting communicative landscape (e.g. Burnett & Merchant, 2018). However, at the time of writing, none of the language and literacy programmes subjected to trial by EEF had explored aspects of reading and writing on screen. While this may be partly due to curriculum pressures – the National Curriculum for English in England focuses only on paper-based reading and writing – we suggest that one reason for this is that RCTs rely on stability in what should be learned. This is partly because they use standardised tests that must be devised and validated over extended time periods and are therefore unlikely to be agile enough to gauge progress in a shifting communicative landscape. While tests have been devised to capture aspects of digital literacies (Coiro & Kennedy, 2011), these are not – and arguably could not be – sufficiently sophisticated to capture the skills, attitudes and dispositions associated with diverse and evolving communicative practices. Moreover, for the reasons explored above, if open-ended approaches that support flexible responses to an evolving communicative context are translated into interventions – written into lesson plans, resources and so on – they assume continuities that are at odds with a changing communicative context. For these reasons, we propose that RCTs, within which interventions are a crucial component, may help sustain a curriculum that fails to recognise the changing nature of literacy.

In summary, this example illustrates how the growth of RCTs and the interventions whose impact they measure may limit what can be known, and by implication what is subsequently recommended to schools and teachers. Specifically it may freeze literacy at a time when language and literacy in everyday life continue to diversify. It is beyond the scope of this article to demonstrate the actual effects of the extensive funding of RCTs and how far and in which ways interventionisation manifests in language and literacy provision. However, this discussion does signal some potential areas of concern. It illustrates how the interventionisation associated with RCTs may lead to learning and

teaching experiences that are less rich and less relevant than they would have been if informed by a wider range of research and by more open-ended professional reflection.

b) Professional Development

In this section, we trace how a perspective that sees Professional Development (PD) through an RCT lens can lead to prioritisation of some particular forms of PD over others. In the PD field, there is an ongoing concern about the lack of evidence for impact of PD, which we define here as the activities (including training, coaching, self-study, etc.) that teachers engage in to develop as professionals, and the learning that accrues from them. A number of arguments have been made to account for this. These include the difficulty of attributing causal inferences to professional learning activities due to their complexity (e.g. Clarke & Hollingsworth, 2002; Desimone, 2009). For example, Opfer and Pedder (2011, p.376) state that many previous authors of reviews of PD literature have argued 'that the problem stems, in part, from researchers employing simplistic conceptualisations of teacher professional learning that fail to consider how learning is embedded in professional lives and working conditions.' Other arguments centre on inadequate models of change processes involved: for example, Colleagues and Author2 (2018) examine how key models used in the field can provide only partial explanations of change processes, and Korthagen (2017) focuses on the inadequacy of the representations of teacher learning used by PD researchers.

There has been a range of responses to these issues. Of note, Opfer and Pedder (2011) argue for a complex systems approach; Korthagen (2017) presents a model of 'professional development 3.0' taking into account the highly personalised and values-led approaches of individual teachers; and Kennedy, A. (2005; 2014) and Sachs (2011) foreground the importance of understanding the purposes of PD. An approach taken by Desimone (2009) and developed in a review of PD reviews by Cordingley et al. (2015) is to build a set of recurrent features (often called 'core features' - Desimone, 2009; Kennedy, M., 2016; Wayne et al., 2008) of PD that appear to be related to positive teacher and student outcomes.

The increasing focus on RCTs, however, has led to a further set of arguments and responses that come to a different set of conclusions. For researchers working with RCTs, the diagnosis of the problem is that there are too few high quality RCTs to help understand the likely effectiveness of PD on student outcomes (Wayne et al., 2008). Part of this argument relates to the set of core features identified by Cordingley et al. (2015). Sims and Fletcher-Wood (2018, 2020) argue that the empirical base for these core features is weak, citing evidence from three RCTs that included all of the features but did not find any impact on student outcomes. Kennedy, M. (2016) also presents a nuanced argument that the underlying active components of the features need to be considered carefully.

Kennedy, M.'s (2016) review is particularly pertinent to our argument, since she found that effect sizes for RCTs of PD interventions were lower when compared with RCTs focussed on other educational changes. She argues that the reason for this is that even in its simplest form, PD is 'a three-step process: PD alters teachers' knowledge, which in turn alters their practices, which in turn alters student learning' (p.960). This offers several points of potential 'slippage'; for example, those conducting the PD may be poorly trained; and teachers or pupils may not respond as expected.

There are a number of differing responses to this point from those working within the experimental tradition. One is to accept the lower threshold for effect sizes than for other studies (aiming to avoid Type 2[2] errors), as advocated by Kennedy, M. (2016). A second is to argue that the logic of some forms of PD – those with a curriculum and pedagogical purpose – may not be amenable to RCT methods, since trials require teacher compliance, consistent implementation and uniformity of approaches whereas the logic of some forms of PD actually requires teacher creativity, variation, and diversity of approaches (see Boylan & Demack, 2018).

---

[2] The type 2 error, in statistics, is to wrongly reject a positive conclusion - a 'false negative'.

A third response is to look for forms of PD that are likely to have higher effect sizes, and it is this response that is most significant for our argument. Taking this approach has led in recent times to strong support for one form of PD of this type: teacher coaching, which tends to have higher effect sizes in meta-analyses of RCTs. Kraft et al. (2018)'s review of RCTs of teacher coaching interventions defines coaching as 'individualized, time-intensive, sustained over the course of a semester or year, context specific, and focused on discrete skills' (p.548), and finds a range of differing outcomes in different contexts, with an overall picture that coaching programmes as interventions - in situ, linked to other support - can be beneficial for teacher and pupil outcomes, although the evidence for the latter is less clear. Sims and Fletcher-Wood (2018, 2020) go further still, drawing on this review and Wayne et al.'s (2007) review to advocate for a specific form of coaching - instructional coaching - based on an examination of the coaching approaches with the strongest effect sizes (Allen et al., 2011; Allen et al., 2015).

In summary, this example lays out a gradual funneling of focus of PD via a lens that prioritises RCT evidence to hone in on ever more specific forms of PD: those that are most amenable to trial (as articulated by Boylan and Demack, 2018) and that are best described as interventions. Approaches with multiple and complex purposes - such as Sachs' (2011) transformative PDL for 'reimagining' - are shed along the way.

Discussion: interventionisation
In this section, we draw on our examples to build our conceptualisation of interventionisation, expanding on the two thrusts of our argument: firstly, that interventionisation may lead to a narrow view of education that underplays the complexity of educational practice with implications for curriculum, pedagogy, professionalism and the purposes of education; and secondly, that it may perpetuate a research agenda that not only marginalises other kinds of research, but has potential to distort or distract from educational approaches that are loosely framed or organic, such as those underpinned by sociocultural or socio-political perspectives (see language and literacy discussion) . These two effects, we suggest, are mutually self-sustaining. In relation to the first thrust of our argument, these examples illustrate a series of effects prompted by an increased emphasis on interventions associated, in part, with the widespread use and promotion of RCTs. For approaches to be tested through RCTs they must be codified into time-bounded, clearly described and transferable programmes - in effect they must be translated into interventions. Our examples illustrate a series of assumptions presupposed by such interventions, which can be summarised as follows:

- Knowledge, understanding and skills are measurable and fixed in nature (e.g. literacy involves a fixed set of skills rather than an evolving set of practices);
- Educational objectives are easily defined and bounded (e.g. language and literacy skills are generic and easily described);
- The teaching and learning process usually involves the routinised delivery and receipt of a package of specified work (e.g. that sequences of literacy teaching and learning can be successful irrespective of social and cultural context);
- Therefore, the teacher operates as a deliverer of focussed, time bound packages that have clearly measurable outcomes (e.g. PD is best achieved through supported acquisition of specific techniques) who learns best via highly structured PD approaches (such as instructional coaching)
- And the learner is a receiver of a package of learning that can be neatly transferred into outcomes (e.g. reading can be taught to young children through training in phonics).

Built into these assumptions are certain understandings of causality that are at odds with an understanding of education a complex practice. Interventions as a component of RCTs require an assumption of educational practice as proceduralised and, as Hammersley (2001, p105) notes, it is likely that "where a process cannot be proceduralised, seeking to reduce it to procedures will lead to distortion". A focus on proceduralisation does not take account, for example, of the extensive body of

work that has explored teaching as a social and material process (e.g. Fenwick and Landri 2014), the complex and shifting identities and agency of teachers and learners (e.g. Alsup, 2006; González et al., 2005; Priestley et al., 2015), and schools as organisations inflected by multiple political, spatial, historical, social and economic imperatives (e.g. Nespor, 1997; Youdell, 2011). Collins[3], (1992) working in Science and Technology Studies, argues that even in the hard sciences a proceduralised - or what he calls "algorithmical" - model misrepresents practice; rather he suggests an "enculturational" model, which requires "professional judgement that can only be acquired through close interaction with competent members of a community" (Bohlin, 2016, p.38). Collins and Kusch's (1998) concepts of polimorphic and mimeomorphic actions are useful here: whilst more complex practices, involving judgment and criticality, are polimorphic, proceduralised processes are mimeomorphic. Bohlin (2016) uses these concepts to critique certain educational research practices (systematic review) and we may go further: the practice of teaching – as explored through extensive research into the lives of teachers, pupils and schools- is polimorphic in character, whereas interventions within RCTs treat it as mimeomorphic.

With these ideas in mind – and returning to the idea of the device- RCTs play a part in upholding certain ways of both understanding and performing educational practice, doing so alongside multiple other actors, such as standardised tests, systematic reviews and available funding. Importantly, for our argument, these actors also include interventions. Each of these works to reinforce the legitimacy of the others as ways of understanding and operating in the world. RCTs, interventions and the kinds of assumptions about educational practice listed above perform, in effect, one another, creating a sense of stability, neutrality and objectivity. From this perspective, the narrowing effects of interventions described above are not just indicative of the impact of interventions – they themselves work to sustain particular understandings and practices in education. Hence a school may be far more willing to invest time and effort in an intervention that has been seen to have had demonstrable impact through RCTs, than to work with more complex or nuanced insights developed through qualitative research.

Of course these sets of relations, as both examples illustrate, do not represent the only ways of knowing and doing education in the contemporary context and, as Law (2008, p.242) argues, 'those realities are only real in *particular networks or systems of circulation*.' However, the effects of interventionisation are political. The move to frame 'best evidence' only in terms of randomised controlled trials limits what can be known, and by implication what can subsequently be recommended to schools and teachers and used to support professional dialogue and development. RCTs may operate as devices within loose networks of actors to hold particular ways of knowing in place, and which are simultaneously held in place themselves. Our point here is that interventionisation may not only be an effect of an increased focus on RCTs, but an emerging phenomenon produced through a series of relational effects, which helps to sustain an educational reality that is hostile to understandings generated through other research paradigms.

Conclusion

In this article, we have argued that a decisive turn to RCTs in generating evidence to inform practice risks a disproportionate emphasis on interventions as a key route to educational improvement. This in turn may help to sustain certain educational approaches, understandings and beliefs as opposed to others, generating narrowing effects linked to teaching and learning, the role and identity of teachers and learners, the curriculum, the values and purposes of education more broadly – and to the kinds of research that gain influence. We propose that RCTs – and the interventions that are central to them- are not in themselves problematic. However they may manifest – and act - in certain ways in an environment where funding for educational research and development is sparse, when a certain kind of evidence is tied to school accountability, and when knowledge about education is moving in new ways as brokered by new alliances between academics, policy makers and practitioners. An increase in interventions therefore may lead to what we term interventionisation, whereby interventions become seen as the key route for developing practice, and generate a series of unintended narrowing effects in education more broadly. More research needs to be done to explore whether- and how far -

interventionisation plays out in practice, and its possible effects on the scope and vision of education in different sites. We need to know, for example, how far schools are focusing on interventions as routes to educational improvement, and the significance of this for the experience of teachers and learners and for educational practice in local contexts – how far do narrowing effects manifest, for example, and what is enabled or marginalised as a result. And, if – as hypothesised here – such effects are produced through shifting assemblages of RCTs, policy, policy actors, educators and so on, then we need to know more about the effects of interventions in different jurisdictions, regions and schools, where they come into relation with diverse policies and practices.

Recently the growth of interventions has started to raise concerns even amongst strong advocates of RCTs. Proponents have improved their design to include more nuance and to enable them to deal with complexity in more sophisticated ways. For example, Bonnel et al.'s (2012) work on realist RCTs foregrounds complexity, context and a focus on causation, and the literature on Implementation Science is helping to develop stronger process evaluation. Such approaches recognise that programmes and interventions are situated in context and are enabling RCTs to generate more nuanced conclusions. Moreover, partly due to noticing that RCT evidence – like other research – is rarely used by teachers (Author2 et al., 2017; Nelson & O'Beirne, 2014), the EEF has begun, at the time of writing, to run trials that focus not on new programmes but the everyday practices of teachers, in particular the Teacher Choices trials (Coe, 2019). Our argument, however, does raise questions about such developments. Whilst imaginative developments in using RCTs are welcome, their contribution needs to be evaluated in relation to their performative effects in complex educational systems. There is a risk that, through programmes such as The Teacher Choices trials, interventionisation encroaches even more closely on teachers' work, in ways that oversimplify professional practice. The first Teacher Choices trial, for example, investigates lesson starters, comparing use of retrieval quizzes with discussions, inevitably narrowing the focus of possible ways of beginning a lesson. This may well generate useful learning, but care needs to be taken for this learning not to drown out learning from other sources.

In response to the growth of RCTs, there have been many calls to engage with a greater range of research methodologies that provide different forms of warrant and different ways of understanding educational practice – approaches that, in effect, acknowledge and explore the polymorphic nature of educational practice. These include ethnographic research which explores the social and cultural life of educational communities, critical and scientific realist research that investigates how generative causal mechanisms are enacted in context, and interpretivist research that considers how individuals and groups make meanings and construct their social worlds. Our argument is that this methodological diversity is not only important because it enriches professional dialogue and extends educational ambition, but also because it intervenes differently in practice and as such, may work against the interventionisation of practice and its narrowing effects.

References

Author2 et al. (2017).
Author2 & colleague (2018).
Alexander, R. (2011). Towards dialogic teaching: Rethinking classroom talk. Dialogos.
Allen, J., Pianta, R., Gregory, A., Mikami, A., Lun, & J. (2011). An interaction-based approach to enhancing secondary school instruction and student achievement. Science, 333(6045), 1034-1037.
Allen, J., Hafen, C., Gregory, A., Mikami, A., & Pianta, R. (2015). Enhancing secondary school instruction and student achievement: replication and extension of the my teaching partner-secondary intervention. Journal of Research on Educational Effectiveness, 8(4), 475-489.
Alsup, J. (2006). Teacher identity discourses: Negotiating personal and professional spaces. Mahwah, New Jersey: Lawrence Erlbaum.
Archer, M. (1995). Realist social theory: The morphogenetic approach. Cambridge University Press.

Beauchamp, T. L., & J. F. Childress. (2009). Principles of biomedical ethics. 6th ed. Oxford University Press.

Biesta, G. (2010). Why 'What Works' still won't work: From evidence-based education to value-based education. Studies in the Philosophy of Education, 29,491-503.

Biesta, G. (2016). Improving education through research? From effectiveness, causality and technology to purpose, complexity and culture. Policy Futures in Education, 14(2), 194–210.

Bohlin, I. (2016) Conceptualising Epistemological Barriers to the Synthesis of Educational Research: A Challenge to Science and Technology Studies. In: Elmgren, M., Folke-Fichtelius, M., Hallsén, S., Román, H. and Wermke, W. (Eds) *Att ta utbildningens komplexitet på allvar*, Upsalla: Uppsala University Library.

Bonell, C. P., J. Hargreaves, S. Cousens, D. Ross, R. Hayes, M. Petticrew, & Kirkwood, B.R. (2011). Alternatives to randomisation in the evaluation of public health interventions: Design challenges and solutions. Journal of Epidemiology and Community Health, 65(7):582–587.

Bonell, C., Fletcher, A., Morton, M., Lorenc, T., & Moore, L. (2012). Realist randomised controlled trials: A new approach to evaluating complex public health interventions. Social Science & Medicine, 75(12), 2299-2306.

Boylan, M., Coldwell, M., Maxwell, B. & Jordan, J. (2018). Rethinking models of professional learning as tools: a conceptual analysis to inform research and practice. Professional development in education. 44(1), 120-139.

Boylan, M. & Demack, S. (2018). Innovation, evaluation design and typologies of professional learning. Educational Research, 60(3), 336-356.

Burnett, C. & Merchant, G. (2018). *New media in the classroom: rethinking primary literacy*. London: Sage.

Cartwright, N. (2019). What is meant by 'rigour' in evidence-based educational policy and what's so good about it? Educational Research and Evaluation, DOI: 10.1080/13803611.2019.1617990.

Clarke, D. & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. Teaching and Teacher Education, 18(8), 947-976.

Coe, R. (2019). Teacher Choices Trials - our new approach to researching questions teachers want answers to. [Blog] 7th September 2019. Available at https://educationendowmentfoundation.org.uk/news/eef-blog-teacher-choices-trials-our-new-approach-to-researching-questions/. Last accessed: 18th May 2020.

Coldwell, M., & Maxwell, B. (2018). Using evidence-informed logic models to bridge methods in educational evaluation. Review of Education, 6(3), 267-300.

Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., ... & Burns, H. (2017). Evidence-informed teaching: An evaluation of progress in England. Research report. London: Department for Education.

Collins, H. (1992). Changing order: replication and induction in scientific practice, 2nd ed. Chicago: The University of Chicago Press.

Collins, H., & Kusch, M. (1998). The shape of actions: what humans and machines can do. Cambridge, Mass.: The MIT Press.

Coiro, J., & Kennedy, C. (2011). The Online Reading Comprehension Assessment (ORCA) Project: Preparing students for Common Core State Standards and 21st century literacies. University of Connecticut.

Comber B., & Kamler B (2005). Turn-around pedagogies: Literacy interventions for at-risk students. Primary English Teaching Association.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., & Coe, R. (2015). Developing great teaching: Lessons from the international reviews into effective professional development. Teacher Development Trust.

Cummins, J. (2000). Language, power and pedagogy: Bilingual children in the crossfire. Multilingual Matters.

Desimone, L. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. Educational Researcher, 38(3), 181–199.

Dickens, J. (2016). Philanthro-philes – meet the donors propping up our education system. Schools Week June 17th, 2016. Available at: https://schoolsweek.co.uk/the-rise-and-rise-of-the-philanthro-philes/ last accessed 20th January 2020.

The Economist (2018). England has become one of the world's biggest education laboratories. The Economist March 31st 2019, available at: https://www.economist.com/britain/2018/03/31/england-has-become-one-of-the-worlds-biggest-education-laboratories last accessed 22nd July 2019.

Ellis, K. & Smith, C. (2017). Assessment, teacher education and the emergence of professional expertise. Literacy, 51(2), 84-93.

Fenwick, T. & Landri, P. (Eds.) (2014). Materialities, textures and pedagogies. Abingdon, Oxon: Routledge.

Fives, A., Russell, D., Canavan, J., Lyons, R., Eaton, P., Devaney, C., Kearn, N. & O'Brien, A. (2015). The ethics of randomized controlled trials in social settings: Can social trials be scientifically promising and must there be equipoise?, International Journal of Research & Method in Education, 38(1),56-71.

Gillen, J.(2014). Digital literacies. Routledge.

Goldacre B. (2013). Building Evidence into Education. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/193913/Building_evidence_into_education.pdf (accessed 5th March 2014).

González, N., Moll, L. C., & Amanti, C. (eds.) (2005) *Funds of Knowledge: Theorizing Practices in Households, Communities, and Classrooms*, Mahwah, NJ: Lawrence Erlbaum.

Hall, K. (2002). *Listening to Stephen Read: Multiple Perspectives on Literacy*. Open University Press.

Hall, K.(2013). Effective literacy teaching in the early years of school: a review of evidence. In J. Larson, & J. Marsh (Eds.), The Sage handbook of early childhood literacy, 2nd edn. (pp.523-540). Sage.

Hammersley, M. (2001). On'systematic'reviews of research literatures: a'narrative'response to Evans & Benefield. *British educational research journal*, 27(5), 543-554.

Hammersley, M. (2008). Paradigm war revived? On the diagnosis of resistance to randomized controlled trails and systematic review of education. International Journal of Research and Method in Education, 31(1), 3-10.

Harrison, N. & McCaig, C. (2017). Examining the epistemology of impact and success of educational interventions using a reflective case study of university bursaries. British Educational Research Journal, 43(2), 290-309.

IES (n.d.). What Works Clearinghouse. https://ies.ed.gov/ncee/wwc/). Accessed 11th November 2019.

Jay, T., Willis, B., Thomas, P., Taylor, R., Moore, N., Burnett, C., Merchant, G. & Stevens, A. (2017). Dialogic Teaching - Evaluation report and executive summary. Education Endowment Foundation, London.

Joyce, K. (2019). The key role of representativeness in evidence-based education. Educational Research and Evaluation, DOI: 10.1080/13803611.2019.1617989

Kennedy, A. (2005). Models of continuing professional development: A framework for analysis. Journal of In-service Education. 31(2), 235-250.

Kennedy, A. (2014). Understanding continuing professional development: the need for theory to impact on policy and practice. Professional Development in Education, 40(5), 688-697.

Kennedy, M. (2016). How Does Professional Development Improve Teaching? Review of Educational Research, 86(4), 945-980.

Korthagen, F. (2017). Inconvenient truths about teacher learning: towards professional development 3.0. Teachers and Teaching, 23(4), 387-405.

Kraft, M.A., Blazar, D. & Hogan, D. (2018). The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. Review of Educational Research, 88(4), 547-588.

Lankshear, C. & Knobel, M. (2011). New literacies: Everyday practices and social learning. Open University Press.

Law, J. (2004). After Method: Mess in social science research. Routledge.

Law, J. & Mol. A. (2008). The actor-enacted: Cumbrian sheep. In C. Knappett, & L. Malafouris (Eds.), Material Agency (pp.57-88). Springer

Law, J., & Ruppert, E. (2013). The social life of methods: Devices. Journal of Cultural Economy, 6(3), 229-240.

Lortie-Forgues, H. & Inglis, M. (2019). Rigorous large-scale educational RCTs are often uninformative: Should we be concerned? Educational Researcher, 48(3), 158-166.

Marsh, R. (2005). Evidence-based practice for education? Educational Psychology, 25(2), 701-704.

Moss, G. (2012). Literacy policy and English/literacy practice: Researching the interaction between different knowledge fields. English Teaching: Practice and Critique, 11(1), 104-120.

Nelson, J. & Campbell, C. (2019). Using evidence in education. In A. Boaz, H. Davies, A. Fraser & S. Nutley (Eds.), What works now? Evidence-informed policy and practice (pp.131-150). Policy Press.

Nelson, J, & O'Beirne, C. (2014). Using evidence in the classroom: What works and why? NFER.

Nespor, J. (1997). Tangled up in school: politics, space, bodies and signs in the educational process. Mahwah, New Jersey:Lawrence Erlbaum.

Oakley, A. (2006). Resistances to 'new' technologies of evaluation: Education research in the UK as a case study. Evidence & Policy 2(1), 63-87.

Opfer, D. & Pedder, D. (2011). Conceptualizing teacher professional learning. Review of Educational Research, 81, 376.

Pawson, R. & Tilley, N. (1997). Realistic evaluation. Sage Publishing.

Phillips, D.C. (2019). Evidence of confusion about evidence of causes: Comments on the debate about EBP in education. Educational Research and Evaluation, DOI: 10.1080/13803611.2019.1617980

Priestley, M., Biesta, G., Robinson, S. (2015). Teacher agency: An ecological approach. London: Bloomsbury.

Sachs, J. (2011). Skilling or emancipating? Metaphors for continuing teacher professional Development. In S.Mochler, & J. Sachs (Eds.), Rethinking educational practice through reflexive inquiry, (pp.153–167). Springer.

Schwandt, T. (2005). The centrality of practice to evaluation. American Journal of Evaluation 26(1), 96-105.

Sims, S & Fletcher-Wood, H. (2018). Characteristics of effective teacher professional development: What we know, what we don't, how we can find out. Working paper, available at: https://improvingteaching.co.uk/wp-content/uploads/2018/09/Characteristics-of-Effective-Teacher-Professional-Development.pdf last accessed 2nd July 2019.

Slavin, R. (2018). Succeeding Faster in Education. [Blog] October 4th 2018, available at https://robertslavinsblog.wordpress.com/2018/10/04/succeeding-faster-in-education/ last accessed 27th July 2019.

Souto Manning, M. & Yoon, H.S. (2019). Rethinking Early Literacies: Reading and Re-writing Worlds. Routledge.

Snell, J. & Lefstein, A. (2018). ''Low Ability,'' Participation, and Identity in Dialogic Pedagogy. American Educational Research Journal, 55 (1), 40-78.

Taylor-Fitzgibbon, C. (2003). Strengthen the science in social science. Social Science Teacher 32(2), 6-9.

Tooley, J. & Darby, D. (1998). Educational research: a critique: A survey of published educational research: report presented to OFSTED. Office for Standards in Education.

Wayne, A.J., Yoon, K.S., Zhu, P., Cronen, S. & Garet, M.S. (2008). Experimenting with teacher professional development: Motives and methods. Educational Researcher, 37(8), 469-479.

Wrigley, T. (2018). The power of 'evidence': Reliable science or a set of blunt tools? British Educational Research Journal, 44(3), 359–376.

Wrigley, T., & McCusker, S. (2019). Evidence-based teaching: a simple view of 'science'. Educational Research and Evaluation, DOI: 10.1080/13803611.2019.1617992

Youdell, D. (2011). School trouble: Identity, power and politics in education. Abingdon, Oxon: Routledge.