



Randomized control trials in education (RCTs): What is their contribution to education theory about teaching?

Yael Shalem

School of Education, University of the Witwatersrand, Johannesburg, South Africa
Yael.shalem@wits.ac.za
<https://orcid.org/0000-0001-6710-7150>

Francine De Clercq

School of Education, University of the Witwatersrand, Johannesburg, South Africa
Francine.declercq@wits.ac.za
<https://orcid.org/0000-0001-9209-059X>

(Received: 6 December 2021; accepted: 28 June 2022)

Abstract

Random Controlled Trials (RCTs) have become one of the most sought-after approaches to impact evaluations of large-scale educational interventions in developed and developing countries. In this paper, we examine the contribution of RCT-based evaluations of large-scale early grade interventions to education theory about teaching. After a brief introduction of the development context of RCT-based evaluations, we examine the research model of RCTs in education and some of the knowledge claims made by RCT scholars, with specific attention to their claims about changing modes of teaching. We then introduce, briefly, five multi-pronged interventions to improve early grade reading in three developing countries (India, Kenya, and South Africa). Finally, we discuss two key educational ideas about teaching supported by these early grade interventions and locate them in education theory about teaching. Our argument is that these ideas about teaching are not new; they are debated by education researchers and because RCTs' evaluation research does not provide empirical analysis of these ideas, it cannot be integrated by teacher educators and education researchers into knowledge about teaching and teacher education and development. Teaching is not seen as an empirical object to be theorised by this massive growing research field. If collaboration and dialogue were to emerge between development economists, education researchers, and teacher educators, RCTs' findings of educational interventions could contribute to what is already known in educational theory about teaching.

Keywords: RCTs, large-scale interventions, teacher education and development, education theory, teacher knowledge about teaching

Introduction

Randomised Control Trials (RCTs) have become the gold standard research method that influences decision-makers and funders on the most effective and cost-effective large-scale interventions to be funded. In the field of education, large-scale interventions (henceforth interventions) aim to add reading resources and textbooks, improve alignment between teaching and the curriculum, monitor classroom instruction, time on task, and curriculum coverage, and strengthen accountability. In addition, and important for our paper, interventions are increasingly experimenting with different forms of teacher training, including instructional coaching, and developing unique teaching material in the form of scripted and/or semi scripted lesson plans (SLPs). Underlying these latter “treatments”, as commonly termed by RCTs scholars, is a model of teacher development but also a theory (or theories) of teaching.

In this paper we do not focus on the model of teacher development used by interventions (De Clercq & Shalem, 2014; Shalem & De Clercq, 2019). Our aim is to examine the contribution of RCT-based evaluations of early-grade interventions in education theory about teaching. We begin with a brief introduction to the development context of RCT-based evaluations. Second, we examine the research model of RCTs in education. Third, we examine knowledge claims made by RCT scholars to show that, although the general opinion is that RCT knowledge claims are confined to the correlation between variables in a specific intervention, some RCT scholars make far more extensive and wide-ranging knowledge claims, including changing human behaviour. In the field of education, this includes changing modes of teaching. Fourth, we introduce briefly five multi-pronged interventions to improve early grade reading in three developing countries (India, Kenya, and South Africa). Fifth, we discuss two key educational ideas about teaching supported by these interventions and locate them in education theory about teaching. We want to show that these ideas about teaching are not new; they are debated by education researchers and, with due collaboration, between development economists, education researchers, and teacher educators, RCT findings of educational interventions could contribute to what is already known in research on knowledge about teaching. We argue that if educational ideas promoted by RCT research about teaching are placed within existing education theories about teaching, a meaningful conversation can begin between development economists, teacher educators, and education researchers. This conversation will also improve the external validity of RCT-based evaluations.

We contend that research that declares that “poor quality of learners’ learning correlates strongly with poor quality of teachers’ teaching” (Bunyi et al., 2012, p. 5), and that is so well funded for the purpose of collecting massive amount of data, should support education theory about teaching, so that future generations of teachers can learn teaching practices that have been shown to be successful and appropriate for complex educational environments. In its current form, RCT-based evaluations contribute robust evidence about the impact of certain treatments on learners’ outcomes. But, as we will show, despite the recurring claim that a specific mode of teaching (supported by learning and teaching materials and coaching) makes

a significant difference to learners' results, there is no way in which teacher educators and education researchers can integrate the RCT findings into knowledge about teaching and teacher education and development. This is because *teaching* is not seen as an empirical object to be theorised by this massive growing research field.

Context and development of RCTs in the mid-1990s

By the mid-1990s, international agencies became seriously concerned that most interventions in various public sectors did not work or did not have a substantial or sustained impact on the ground. The rise of international large-scale assessments that followed a global change in the culture of assessment supplied new comparative performance data and put pressure on countries, in particular poor and middle-income ones, to develop better policies and programmes in line with their assessment results (Addey et al., 2017). Many governments in the global south came under pressure because of their ineffective and poorly performing education system, a pressure increased by the recommendations to provide quality education for all coming from Education for All and more recently by the Millennium goals and the Sustainable Development Goals. The subsequent growth in education enrolment and participation came, not unexpectedly, with a simultaneous decline of educational quality in many countries of the global south. Determined to alleviate the education inequalities of disadvantaged communities, the international agencies (United States Agency for International Development, United Nations Educational, Scientific and Cultural Organization and the UK Department for International Development) became resolute that they should allocate funds only to programmes that could be backed up by large-scale research evidence of their effectiveness and their specific modality (de Souza Leão & Eyal, 2019).¹ The growing trend became investments in programmes that used what we might call a systems lens to deal with education problems at scale and improve learning levels for all (Gibbs et al., 2021). The partnership between Governments or Non-Governmental Organizations working with disadvantaged communities and researchers using RCTs was a perfect match.

This context gave rise to the involvement of development economists in advising and evaluating the impact of social policy interventions while at the same time strengthening their evaluation tools such as RCTs (Banerjee, Duflo et al., 2016; Banerjee & Duflo, 2009; Duflo et al., 2016). The data base index of RCT-based evaluations (Impact Evaluation Repository (IER) created by the International Initiative for Impact Evaluation included close to 700 evaluations by 2012, growing to 2500 by 2014, and to 4205 by 2015 (Cameron et al., 2016; Sabet & Brown, 2018). The sectors of health, nutrition, and population, education and social protection constitute 65% of all RCT-based evaluations in the IER data base (Sabet & Brown, 2018).

The largest share (34.4%) of studies was conducted in sub-Saharan Africa. There was a rapid massive growth of international evaluation agencies such as the Abdul Latif Jameel Poverty Action Lab (JPAL), the World Bank's Development Impact Evaluation Initiative (DIME) and

1 <https://www.usaid.gov/evaluation/usaid-program-effectiveness>

the Strategic Impact Evaluation Fund (SIEF) at the World Bank (Cameron et al., 2016). JPAL, based at the Massachusetts Institute of Technology, conducted more than 1000 randomised evaluations in different development sectors in more than 80 countries.² In addition, these statistical/quantitative evaluations came to dominate at international development conferences; they were widely published in many academic journals for their large-scale quantitative research findings and their influence on social development policy and research funding continues to grow. RCTs came to take “a larger place in the policy conversation at the turn of the century and received substantially more funding from donor organizations and local governments” (Banerjee, Banerji et al., 2016, p. 2).

How do RCTs in education evaluations work?

RCT-based evaluations in education set out to find answers to questions such as “Which version of a given educational treatment programme seems to produce the largest increase in learners’ outcomes” and “Can it work in other contexts or be scaled up?” In the case of improving educational outcomes, interventions are about testing different combinations of mechanisms that seem suitable to changing and improving teachers’ practice by relying on a specific model of change, and on previous ideas about teaching. The confidence in the knowledge claims produced after the testing is determined by quantitative variables measured at the beginning, mid-way, and at the end of an intervention, with quantifiable evidence, most often the percentage of learning gains achieved for a certain multi-pronged treatment (Bhide et al., 2018) or part of it. By replicating these interventions in different geographical contexts and testing difference in modality, RCT scholars hope to provide knowledge about how to change teachers’ practice. Their model of teacher development is based on the idea that, by providing teachers with support and accountability measures, their new repertoire of practice will be expanded, and teachers will be encouraged “to incorporate practices of new and effective lesson strategies” (Fleisch et al., 2016, p. 158).

RCTs: Claims to knowledge

Development economists argue that, through randomised experiments, they can test treatment interventions, and find out which components of the interventions are truly necessary and should be adapted by policy makers (and by other researchers or practitioners in their respective fields). Inspired by what they have read or researched, and by previous randomised experiment projects, development economists test different variations of interventions that did not exist before (Banerjee, Duflo et al., 2016), with the view to testing and changing human behaviour (2016).³ RCT-based evaluations have big ambition for grand change in the poor communities in the world which they seek to influence by identifying “scalable”

2 de Souza Leão and Eyal (2019) report that when JPAL was founded in 2003, “it consisted of 4 affiliated Professors and conducted 33 projects. By 2017, there were 161 affiliated Professors and they were involved in 902 evaluations in 72 countries (p. 384).

3 For example, “If citizens vote for candidates based on their ethnicity or caste is that because of very strong preferences, clientelistic networks, or a combination of weak preferences and no alternative information on candidate quality? Do people only value what they pay for? How important are liquidity constraints, as opposed to lack of information or low human capital, in explaining poor child health and low business profitability in low-income families?” (2016, p. 18)

innovations (the next cell phone), or change “systems” (health care) or reform institutions (democracy) (Banerjee, Duflo et al., 2016).

Their first claim to knowledge suggests the intention to have large transformative influence by advising policy makers on educational interventions that could change attitudes and behaviours towards better teaching practices and better learners’ outcomes that are critical to the survival of poor communities in developing countries.

The second aspirational claim to knowledge associated with RCT-based evaluations is for “external validity.” In the words of Athey and Imbens (2016),

external validity is concerned with generalizing causal inferences, drawn for a particular population and setting, to others, where these alternative settings could involve different populations, different outcomes, or different contexts. (In Banerjee, Duflo et al., 2016, p. 8)

Using production functions, statistical regressions and/or hierarchical data modelling, and RCT-based evaluation, researchers look to identify the factor/s within a treatment that has/have the highest correlation with improved learners’ outcomes (de Souza Leão & Eyal, 2019). For an effective treatment to be justified and adopted, a particular multi-pronged intervention must be replicated (in the same basic intervention program or similar enough) across different contextual conditions and time periods. Replications across different contexts are intended “to increase the operational value of the multi-pronged treatment for policy makers” (who can choose the best combinations of mechanisms across slightly different multi-pronged treatments) (Banerjee, Duflo et al., 2016, p. 17).

The literature on external validity of RCTs research is vast. The problems of generalising to other time periods, other countries, etc., are issues that have been discussed extensively (Banerjee, Duflo et al., 2016; Deaton, 2010; Duflo et al., 2006; Muller, 2015). Claims for external validity are made with care since they attract strong criticism within the RCT research community itself. They are contrasted by those who state that interventions that work in one context cannot be recommended to poor communities from different contexts because evaluation studies cannot take account of the specific characteristics of communities, nor explain the reasons behind a behavioural change (Gibbs et al., 2021; Jones et al., 2009; Tomlinson et al., 2015). As long as RCT-based evaluations and other related research studies do not explain (and most RCT scholars say that they are not meant to explain) how the identified variables of an intervention (and their best modality) work to improve results, their external validity is constrained. Theoretically, this should pose limitations on advising policy makers about the benefits of any evaluated interventions.

The third aspirational claim of RCT-based evaluations deals with the confidence in the accumulative power of RCTs in advocating a different model of teacher development. An example of this is found in the Kenyan and South African studies (see later). RCT-based evaluations were brought to these countries in times of desperation for a new model of teacher development. For a long time, different forms of interventions were trialled but they

brought about little change in the system as a whole and in teacher practice more specifically (Bertram, 2011; De Clercq & Shalem, 2014; Hoadley, 2016; Shalem & De Clercq, 2019). South African advocates of RCT testing of interventions went as far as claiming that RCTs provide “a sound basis for building powerful knowledge on large-scale reform of instruction in the Global South” (Fleisch & Schöer, 2014, p. 11). This claim of powerful knowledge, that can predict, explain, and enable us to envisage alternatives by providing the best understanding of the social world (Young and Muller, 2013), was made at the beginning of sets of interventions trialled on a large group of mainly poor schools in South Africa with the view to improving learning and teaching of reading in these schools.

We acknowledge these aspirations. We also acknowledge their important contribution to a model of teacher development. Moreover, we acknowledge the value of RCTs in evaluating educational interventions in developing countries because of the immense importance of improving the teaching of learners from poor communities. We acknowledge debates on external validity. Our serious concern, however, is about the weak integration of RCT-based evaluations with education theory about teaching. We concede that RCT-based evaluations show that small-scale training interventions have not made much difference to teacher practice in developing countries. But, because proponents of RCT-based evaluations aspire to change human behaviour, and, in education more specifically, to improve teaching practice and learners’ outcomes, the knowledge about teaching accumulated from this research needs to be brought not only to policy makers but into the educational arena as well. Before engaging with this argument, we need to describe, briefly, the interventions. In the section that follows, we foreground and locate in educational debates two educational ideas about teaching that these interventions are known for as well as their educational assumptions about how teachers teach and how learners learn.

RCTs in education in three developing countries

The five interventions aimed at improving reading⁴ of poorly performing learners in three developing countries are Pratham’s Read India, the Primary Mathematics and Reading programme (PRIMR) as well as Tusome in Kenya, and the Gauteng Primary Literacy and Mathematics Strategy (GPLMS) as well as the Early Grades Reading Study (EGRS) in South Africa. RCT-based evaluations of these interventions report some success, measured by an increase in testing scores that are shown by RCTs to be correlated with specific components of the respective interventions. Our aim is to introduce each intervention and its impact (increase in learning outcomes) but not to examine in detail the strategies and implementation over time nor to interrogate the detailed results.

India

Cognisant of the fact that school systems are not well designed to address the needs of learners who have not gained early-learning skills, the Pratham Read India programme developed a reading deficit programme to give such learners a chance to catch up. The

4 And, in some, mathematics was targeted as well.

Pratham-trained teachers and/or community volunteers worked to teach the Pratham curriculum either in or outside school hours or in summer camps. The programme targeted what it saw to be the root of the learning crisis by transforming the Grade 3 and 5 structures and introducing two components that were subsequently assessed by JPAL evaluators: 1) “teaching-at-the-right-level” (TaRL); and 2) the use of Pratham-trained volunteers rather than teachers to implement the programme in/outside of the schools (Banerjee, Banerji et al., 2016). A sample or template of semi-SLPs was also provided for each reading task at the five distinct language ability levels. The approach to teaching reading followed a balanced approach of back-to-basic phonics and whole language.

After reaching about 33 million children by 2007, this remedial early grade learning programme aimed to contribute to knowledge on strategies under which effective pedagogy can be brought to scale. The intention was to change government policy and practice in schools (Banerji & Chavan, 2016) and institutionalise the Pratham’s methodology that was said to be cost effective and easy to scale up (Banerjee et al., 2007). It was tested in a slightly different manner in three states: in the first state, the intervention was led by teachers supported by government supervisors, while, in the second state, the reading methodology was taught by Pratham volunteers during school time and this is what led to the greatest improvement of learners’ reading results (Banerjee et al., 2017). A third version of the programme, that involved teachers implementing themselves the programme in their classroom, failed. This is because teachers reported a conflict between Pratham’s back-to-basics phonics pedagogy and the grouping of learners of same reading ability on the one hand, and the national Indian primary school curriculum emphasis on coverage, on the other. Teachers preferred to account through official channels for their coverage of the national curriculum rather than change their pedagogy and organisation of learners in compliance with the non-compulsory Pratham’s approach. The best impact is recorded in Uttar Pradesh in 2013–4 with trained volunteers (not teachers) teaching in two learning camps (10-day and 20-day each) during school time, leading to increased test score gains of 0.7 to 1.0 on average (Banerjee et al., 2017). By now this programme⁵ has expanded its reach beyond India and is found in many African governments such as those of Cote d’Ivoire, Nigeria, Ghana, Kenya, and Zambia.⁶

Kenya

PRIMR in Kenya was an intervention begun in 2011 that targeted early grades and involved basic instructional materials such as English/Kiswahili learners’ books with ongoing instructional support and coaching to teachers. According to the Research Triangle Institute (RIT) evaluators, PRIMR foregrounds “starting at the basic level” and “utilizing explicit pedagogy” (Piper et al., 2015, p. 80), mediated by teachers’ guides with semi-scripted daily

⁵ <https://www.teachingattherightlevel.org/tarl-in-action>

⁶ The programme was not implemented by teachers in the schools of Bihar and Uttarakhand. Attempts to institutionalise the programme in classrooms failed because teachers felt more accountable to curriculum advisers for curriculum coverage than to this alternative pedagogical intervention. Sharma and Deshpande (2010) paraphrase what teachers told them in interviews: “[T]he materials are good in terms of language and content. The language is simple and the content is relevant . . . However, teaching with these materials require patience and time. So, they do not use them regularly as they also have to complete the syllabus” (Banerjee et al., 2017, p. 89).

lesson plans and teachers being coached by curriculum support officers (Piper, Simmons-Zuilkowski, et al., 2018). The rationale behind this kind of intervention was that

[m]any teachers do not understand how children learn how to read and that they lack the pedagogical skills needed to teach this skill. Interventions that train teachers in the components of literacy acquisition, appropriate pedagogical skills, and appropriate use of teaching-learning materials have reported positive gains in children learning how to read. (Bunyi et al., 2012, p. 5)

This intervention also follows the balanced approach to literacy learning “with attention to both decoding skills and interpretation” (Piper et al., 2014, pp. 13–14). PRIMR impact saw a moderate increase in learners’ reading scores (0.73 to 1.29 SD respectively for English and Kiswahili). The largest effect size came from teachers’ guides with lesson plans (Piper, Simmons-Zuilkowski, Dubeck et al., 2018, p. 333). These results sustained, and in most cases, strengthened in the four-year intervention.

Following the success of PRIMR, the Kenyan Ministry of Education decided to take it to scale in 2015 with a similar literacy programme, known as Tusome, for the first three grades of schooling (Piper et al., 2015; Piper et al., 2018; Piper, Simmons-Zuilkowski, Dubeck et al., 2018; Piper, Sitabkhan et al., 2018). The idea behind this national programme was to build capacity in government structures to support and monitor effectively this new instructional programme. Teachers guides (aligned with the national curriculum), and coaching were also used to explicate the sequence of lessons and reading activities. Piper, DeStefano, Kinyanjui et al. (2018) reported that Tusome’s impact doubled or tripled the Kenyan literacy benchmarks (0.6 to 1.0 standard deviations respectively on English and Kiswahili learning outcomes). But although reading comprehension, the most difficult subtask, improved in Kiswahili and English, it remained the lowest level of improvement in relation to other reading sub-tasks. Less-poor learners benefited more since they had more school resources and this reaffirms the notion that equitable outcomes demand inequitable allocation of support resources as these scholars pointed out.

They found that the instructional change in teachers’ practice was lower than expected, partly because of weak utilisation of the internal project monitoring information to target instructional support. In the evaluation of teachers’ use of teacher guides, it was noted that “most of these modifications (59 percent) negatively impacted the quality of the lesson” and “more should be done to support teachers in understanding the activities in each lesson and to encourage the teachers to use the instructional methodology” (Piper, Sitabkhan et al., 2018, p. 23).

South Africa

The South African GPLMS in Gauteng Province, the first system-wide intervention targeting teachers of poorly performing primary schools, focused on improving “the instructional core.” Teachers were provided with detailed SLPs that specified the *what*, *when*, and *how* of teaching specific curriculum content, as well as quality learning materials. Teachers were

trained by coaches who modelled what good practice looked like, and who reflected with teachers on their enactments of the lesson plans. Similar to what happened in Kenya, this intervention followed the balanced approach to teaching reading that combines phonics and whole language. As Fleisch and Schöer (2014) put it, “There is recognition in the Simple Literacy Approach that primary school children move from ‘learning to read’ to ‘reading to learn’, ‘reading for a purpose’ and ‘reading for pleasure’” (p. 2).

This intervention relies on a particular modality of changing teachers’ practices. It was expected that, by following SLPs and by working with coaches (once a month), teachers’ repertoire of good practices would expand. As in many of these interventions, the idea is that more repeated teacher practice of the new practices (with the help of coaches and lesson guides) will increase their understanding of the curriculum topics and this will encourage them to go back to their classroom to “incorporate practices of new and effective lesson strategies” (Fleisch et al., 2016, p. 158).

Programmes such as the EGRS are used to develop teachers, to facilitate the expansion of their knowledge and skills, to contribute to their growth and competency and to implement instructional strategies that enhance the teachers’ content knowledge, pedagogy and teaching styles. (Motilal & Fleisch, 2020, p. 5)

A 2012 evaluation of the GPLMS intervention was done through Regression Discontinuity Design and Difference in Differences studies and it revealed some, but not substantial, changes. Low increase in learners’ reading scores was found and this meant that schools below the 40% cut off benefited most. It is generally concluded that the GPLMS led to improvement of treated schools by 19.3 percentage points from 2008 to 2012 (Fleisch & Schöer, 2014, p. 6). Subsequently, three EGRS quasi-experiments were introduced in different provinces over a period of two years. In these different permutations, the structured pedagogic intervention continued to rely on the so-called triple cocktail⁷ of fully SLPs (prescribing/guiding the order of the content, the pacing of their teaching), learner resources, and training/coaches. On-site coaching was found to be more successful than “just-in-time” centralised teacher training (Kotze et al., 2019) or virtual coaching (Cilliers et al., 2020). The RCT-based evaluation pointed out the important value of on-site teacher coaching (0.24 SD); “[b]etween 10 to 20% more learners surpassed reading fluency at the end of Grade 2 as a result of the coaching intervention” and the reading gap between non-fee and quintile 5 schools closed by about one-fifth (Taylor, 2019, p. 11). A second EGRS classroom observation study on early grade reading in Grade 1 (Department of Basic Education, 2017) noted improvement in the form of more assessment, more speaking and writing, more print visibility, and better learner pacing and time-on-task. The argument put forward was that a virtual coach is less able to monitor, model, and correct the more difficult teaching practices, and that on-site coaches are better able to relate to teachers, win their trust, encourage accountability, and, most importantly, allow for more reflection.

7 This concept was first used and developed by Brahm Fleisch who was instrumental in initiating the GPLMS (see Fleisch & Schöer, 2014)

If the participating teachers were given an opportunity to explain the reading strategies, talk to coaches about them and reflect with other teachers, there would be significantly greater awareness and understanding of the strategies. (Sailors and Price, 2015, p.124 in Motilal & Fleisch, 2020, p. 9)

Key educational ideas supported by these large-scale interventions and debates in education theory about teaching

The five interventions we have reviewed here rely on similar educational ideas about teaching and its underlying notion of learners' learning. They emphasise individualised teaching (by grouping learners, even if in different organisational forms), explicit instruction, as well as a balanced approach of back-to-basic phonics and whole language. In what follows, we focus on the first two ideas and locate them in debates in education theory about teaching. Our main aim is to show that the ideas are not new; they are contested and are without empirical evidence concerning these ideas of teaching. Also, RCT-based evaluations cannot contribute to how education theory about teaching has been theorised for decades. We believe that this is a missed opportunity.

The first key educational idea for teaching, used by the Pratham Read India programme, is “[t]eaching-at-the-right-level” (known as TaRL). In South Africa, a similar idea is applied to one of the teaching activities referred to as Group Guided Reading. This term is used to describe the organisation and grouping of learners for some period of the day or part of the school year “not according to their age, but according to what they know—for example, by splitting the class, organizing supplemental sessions, or reorganizing children by reading level—and match[ing] the teaching to the level of the students” (Banerjee et al., 2017, p. 84). An assessment tool is used to group learners: five different language levels are used to distinguish different ability levels.⁸ This intervention advocates more dedicated time to basic skills and group work with “plenty of reading material at the children’s level and simple tools to track progress and give attention to children who need help” (Banerji & Chavan, 2016, p. 463). Regular one-on-one assessment of learners is a key element to track progress even though this idea is reported to be foreign to the formal Indian system. “The pedagogy became more structured and more formal, with an emphasis on frequent testing” (Banerjee et al., 2017, p. 84).

Organisation of learners by some or other form of ability assessment is not a new idea and has been debated for decades. These debates are ignored by the interventions and their RCT-based evaluations and, even more so, the idea is presented as common sense (Banerjee et al.,

8 The literacy components are beginning level, letter and phoneme recognition, reading words, reading paragraphs, and reading a short story (<https://www.teachingattherightlevel.org/the-tarl-approach/classroom-methodology/reading>). These levels have increasingly sophisticated dimensions of language comprehension and expression and reading comprehension as well as fluent/creative writing. They each consist of activities such as reading aloud, discussions, phonetic games, vocabulary exercises, mind-mapping and writing (Banerji & Chavan, 2016).

2017). This is despite the fundamental tension that underpins this idea and that makes it difficult to apply.

One interesting take on the tension involved in grouping learners goes back to Darling-Hammond's work on accountability where she points to the tension between the "best" and the "equal" principles. The former principle refers to the idea that each learner is "entitled to receive the education that is best for him or her." The latter principle refers to the idea that each learner "is entitled to receive an education at least as good as (equal to) that provided for others" (Darling-Hammond, 1989, p. 65). During the 1970s, sociologists of education argued that this kind of classification brought about alienation and labelling. Individualised teaching is, ideally, dealing with these tensions but it demands high levels of skilled teaching, sensitivity, and judgment, let alone cost. Bernstein (1990) whose sociological theory of education developed in view of the inequality of educational achievement between middle- and working-class children, discussed what he called educational forms of 'repair', and he shows that each form of 'repair' embodies its own curriculum and/or pedagogical trade-off. Some depend on the availability of certain economic conditions of possibility while others give preference to some educational priorities at the expense of others (Allais et al., 2019). The above-mentioned educational theorists point to social and pedagogical tensions associated with teaching-at-the-right-level in the sense of the grouping of learners according to their ability.

Conceptually, teaching-at-the-right-level in a mixed classroom (as is the case in Kenya and South Africa but not in India) is the most challenging form of teaching because it requires a complex set of curriculum and pedagogical decisions, both at system and classroom levels. The Pratham Read India programme foregrounds the idea of recognising and separating learners with different abilities in the class and (in some parts of the lesson also by ways of diagnosis) ensuring that teachers can tailor their teaching according to their learners' different reading abilities. But what variations of individualised teaching the grouping of learners give rise to, how they are enacted, what kinds of decisions teachers make, and how teachers manage coverage as well as individuation, are not explained, much less explicated. The Kenyan and the South African RCT-based results suggest that the most vulnerable schools remain so and this suggests that teachers find the idea of individuation very difficult. Hence, what will be crucial evidence to collect and analyse for new generations of student teachers and for educational research more broadly are case studies (from such a large sample of schools), that demonstrate what strong and confident teachers do to individuate teaching and whether this is correlated (or not) with their knowledge of content, use of time (for example), or selection of certain activities over others.

The second educational idea or concept embedded in the Kenyan and South African interventions is that of explicit instruction. "Explicit instruction" in the Kenyan intervention includes a focus on the five reading skills, the use of graded readers, "moving teachers away from using whole-class oral repetition" (Piper et al., 2014, pp. 13–14); the "sequential step-by-step manner that reinforces the five components of reading" (Piper et al., 2015, p. 72) and

“teacher and student interactions” (Piper, Simmons-Zuilkowski, Dubeck et al., 2018, p. 325). As can be seen below, “explicit instruction” is described as a curriculum organisational issue.

Teachers’ guides were provided in Kiswahili and English, including structured lesson plans for English, Kiswahili and mathematics. The guides, developed for the first and second grades, included one lesson per day for the full school year. The Kiswahili and English lessons focused on the explicit instruction of early reading skills, such as letter sounds, blending, reading comprehension methods and writing activities. (Piper, Simmons-Zuilkowski, Kwayumba et al., 2018, p. 112)

In the above extract the term “explicit instruction” could mean that an enacted curriculum tool in the form of a teachers’ guide makes the content to be taught (here, it is early reading skills) explicit to teachers. This is, of course, not the same as making the teaching of those reading skills explicit to learners (Shalem, 2018; Shalem et al., 2016). “Instructional core” is the term for explicit instruction promoted by the South African interventions (GPLMS and EGRS). Its educational roots are in the work of Richard Elmore on school reform.

Instructional practice is broadly defined as the set of interactions that occur at the level of the instructional core, that is, the relationship between a teacher and a learner in the presence of knowledge. (City et al. 2009 and Cohen et al. 2003 in Rincón-Gallardo & Fleisch, 2016, pp. 381–382)

Ideas we found associated with instructional core include a focus on the reading skills (Motilal & Fleisch, 2020); the use of graded readers; repertoire of practices and daily and weekly routines (Fleisch & Schöer, 2014); interaction between teachers and learners; appropriate expectations from learners of lower socio-economic environment; habituation; systematic completion of tasks; reorganization of classroom space (near teacher desk, on carpet, etc); and small groups working in a different space (individualization) (Fleisch & Dixon, 2017).

Some of the terms used to describe explicit instruction refer to what curricula specify as well as to what lesson plans (if used) can specify in more detail (for example, repertoire of practices, daily and weekly routines, and focus on the reading skills). Other terms are simply too broad and common-sense (interaction between teachers and learners, and the systematic completion of tasks). As mentioned earlier, we found one paper linked to an RCT-based evaluation of a South African intervention that aims to understand the mechanism that explains the shift in teachers’ teaching and that introduces ideas from Foucault to explain the use of routines and activities with specific emphasis on habituation and use of space (Fleisch & Dixon, 2017). This is a productive attempt made by two education researchers to explain and explicate explicit instruction.

However, here is the issue with the term *explicit instruction*: different curriculum forms prescribe and specify (make explicit), to a different degree, the selection and sequencing of the knowledge to be taught. South Africa is noted for its history of curriculum reform that has, over time, succeeded in making the selection and sequencing of knowledge more explicit

(Hoadley, 2018). But curriculum specification is not the same as explicit instruction. Moreover, no theory of teaching advocates implicit forms of instruction. Rather, different approaches to teaching make explicit different aspects of teaching *because* they rely on different learning theories. To explain this point, we locate the idea of explicit instruction in two very different education theories of teaching.

The first theory follows Lave and Wenger's (1991) situated learning theory (and that of other social constructivists) that argues that, by being directly involved with real world objects in authentic contexts and experimenting with them, learners come to understand the use of things in the world and the meanings of ideas embedded in them. Concepts associated with this view include terms such as participation in social practice, learning-by-doing, and communities of practice. In this theory, ideas about teaching foreground designing learning activities, creating an interactive learning environment, and the gradual transition of learners from peripheral to complex forms of participation. The role of teachers is to bring learning out into the open and this includes making explicit learners' opinions and ideas—i.e., their voice. Teaching is expected to be individualised, useful, and relevant to the life of learners. *Activity* forms a crucial representational resource. The whole language approach to reading has affinity with this view (Pearson, 2004). The teacher is expected to diagnose and interpret learners' actions, allow different reactions to texts, and, over time, help them to participate in a wider system of literacy practices. In this view, *explicit instruction* would refer to the activities that teachers design, the variety of forms of participation they encourage, and their acceptance of the plurality of meanings and of the variety of ways of knowing.

The second theory of teaching draws on a social realist view of knowledge that emphasises relations between concepts, strong demarcation between school and everyday knowledge, and the role of evaluative criteria in education transmission (Bernstein, 1990). Activity and mediation are means of transmission of specialised knowledge, procedures and rules; they demarcate what is expected to be known and how, and are used to assess levels of proficient performance with an emphasis on what is not achieved. This approach is associated with systemic functionalists' emphasis on direct teaching of the formal properties of language (Clark, 2005) and of phonics. According to this theory, the teacher is expected to provide instructions that explain what aspect/s of reading is/are practised by the activity, identify and correct misunderstandings of the meaning of the text exhibited by learners, make decisions on what ideas, language structures, rules, and meanings to elaborate (as well as when and how to elaborate those), as well as transmit knowledge criteria (Hoadley, 2018) about what is particularly important and why, what ideas belong together and which do not, and which ideas are correct and which are false. In this view, *explicit instruction* would refer to accuracy and correctness of knowledge, teachers' explanations of concepts, and the transmission of evaluative criteria of what counts.

The huge potential of evidence for research to be conducted by both education researchers and development economists on teaching located in such educational debates is not mined by RCT-based evaluations; the different types of *explicit instruction* and when, during the lesson, teachers use them, are not explicated. This means that new generations of student

teachers (and teachers in professional development courses) can continue to rely only on the conceptual debate and on the small case studies that already exist. Yet, RCT-based evaluations may have contributed more to knowledge of teaching if this debate on explicit instruction were acknowledged seriously by development economists and donors, if the enactment of SLPs were examined systematically and forms of modelling by coaches explicated, and if the knowledge needed to enact the teaching-at-the-right-level were investigated and modalities of individualised teaching explicated. These evaluations may also have contributed better knowledge about teaching that could explain and/or enable alternatives about making explicit teaching techniques and evaluative criteria or about gradually becoming a participant of the social practice of reading, and what affordances for learning in practice teachers need. It is possible that analysis of data on teaching would help transcend the polarity that structures current debates in education theory about teaching. Knowledge of teaching would mean further empirical verification or falsification of the knowledge of these two different schools based on specific evidence to be collected from the thousands of teachers teaching in these interventions or from a sample of these. Lack of discussion and engagement with these important issues by RCTs make their pragmatic empirical evaluations a danger if they are used exclusively to advise policymakers and donors since it could be undermined by what is already shown by education theory.

Conclusion

The literature on RCT-based evaluations we report on provides a lot of educational data used to evaluate what has or has not worked (generally by measuring short and sometimes medium-term learning gains). It does not, however, refer nor attempt to engage explicitly with a theory of teaching embedded in these interventions. Nor does it show what the evaluations can add or change about what education researchers and teacher educators already know from its rich conceptual tradition.

A response to our request could be that RCT-based evaluations are not meant to explain or answer these kinds of questions. It is argued that, to understand how and why an intervention works and what the generative mechanisms for change are, small qualitative studies are required (Fleisch & Dixon, 2017). The truth is that these are few and far between and that RCT scholars of educational interventions refer mainly to other RCT studies by replicating what has been shown to work or not. Research work is shared among quantitative researchers who work for donor agencies, economists who deal with large educational data, and, increasingly, with policy makers. Education researchers as well as teacher educators are not seen as potential interlocutors with whom to discuss these issues and theorise the complexities involved in teachers' knowledge about teaching (and learning to teach). Ambivalences and disagreements that characterise educational knowledge for teaching appear to be ignored as if they are of little interest. Some go so far as to argue that educational knowledge about teaching is merely common sense.

How do you get a bureaucracy to make a *common-sense change* that has a very strong chance of being beneficial—like not totally ignoring students who have fallen behind

and instead offering them a path to catching up? (Banerjee et al., 2017, p. 95, emphasis added)

One could also argue that we cannot expect economists to engage with education theory about teaching and that education researchers and teacher educators have not come to the party because of some prejudice against these kinds of studies. This may be true. The question, however, is not about this. If the reservoir of knowledge developed from RCT-based evaluations is mainly system change-focused, and if the reservoir is developed and stored within this narrow stratum of similar researchers only, how will that knowledge be relayed to education researchers and teacher educators? How will the RCTs' claim to knowledge of changing teacher practice, that, as we show above, is a complex conceptual issue and is subject to all sorts of conditions of possibility, support education theory so that it can be used to develop better teaching and learning?

What is needed is not only combining rigorous quantitative and qualitative research but, more importantly, combining conceptual and empirical research of concepts such as teaching-at-the-right-level, systematic teaching, teaching/learning-by-doing, etc., since this will assist in building education theory about teaching for teacher education research from RCT-based evaluations. After all, it is the teachers who are expected to carry on with the educational ideas that drive the treatments of intervention and future generations of teachers can only benefit from appropriate explication of knowledge about teaching methods in general and of teaching reading in particular. Without these kinds of foci or analyses and as long as the findings and instructional regimes used in the interventions and their underlying theory of change are not subjected to educational theorisation, teacher educators and researchers of teacher knowledge will convey the findings as something to emulate (or ignore). Clearly this approach cannot contribute to professional knowledge scholarship of learning and teaching, and to teacher development.

We sum up this discussion with the question, “Are RCT-based evaluations in education conducted to influence policy makers and donor agencies or do they intend to be of any value to teacher education and development research?” If the answer is the former, how can they then be of real systemic value?

References

- Addey, C., Sellar, S., Steiner-Khamsi, G., Lingard, B., & Verger, A. (2017). The rise of international large-scale assessments and rationales for participation. *Compare: A Journal of Comparative and International Education*, 47(3), 434–452. <https://doi.org/10.1080/03057925.2017.1301399>
- Allais, S., Cooper, A., & Shalem, Y. (2019). Rupturing or reinforcing inequality? The role of education in South Africa today. *Transformation: Critical Perspectives on Southern Africa*, 101, 105–128.

- Banerji, R., & Chavan, M. (2016). Improving literacy and math instruction at scale in India's primary schools: The case of Pratham's Read India program. *Journal of Educational Change, 17*(4), 453–475. <https://doi.org/10.1007/s10833-016-9285-5>
- Banerjee, A., & Duflo, E. (2009). *The experimental approach to development economics*, (pp. 151–178) [Working Paper 14467]. National Bureau of Economic Research. <https://www.nber.org/papers/w14467>
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukherji, S., Shotland, M., & Walton, M. (2016). *Mainstreaming an effective intervention: Evidence from randomized evaluations of "Teaching at the Right Level" in India* (No. w22746.). National Bureau of Economic Research.
- Banerjee, A., Banerji, R., Berry, J., Duflo, E., Kannan, H., Mukerji, S., Shotland, M., & Walton, M. (2017). From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives, 31*(4), 73–102. <https://doi.org/10.1257/jep.31.4.73>
- Banerjee, A., Cole, S., Duflo, E., & Linden, L. (2007). Remedying education: Evidence from two randomized experiments in India. *The Quarterly Journal of Economics, 122*(3), 1235–1264.
- Banerjee, A., Duflo, E., & Kremer, M. (2016). *The influence of randomized controlled trials on development economics research and on development policy*, 1–76, paper for "The State of Economics, The State of the World" Conference proceedings volume. <https://economics.mit.edu/files/16473>
- Bernstein, B. (1990). *Class, code and control: The structuring of pedagogic practice* (vol. iv). Routledge.
- Bertram, C. (2011). What does research say about teacher learning and teacher knowledge? Implications for professional development in South Africa. *Journal of Education, 52*, 5–26.
- Bhide, A., Shah, P. S., & Acharya, G. (2018). A simplified guide to randomized controlled trials. *Acta Obstetrica et Gynecologica Scandinavica, 97*(4), 380–387. <https://doi.org/10.1111/aogs.13309>
- Bunyi, G., Cherotich, I., & Piper, B. (2012). *Primary Math and Reading (PRIMR) Program: Kenya*. Research Triangle Institute International.
- Cameron, D. B., Mishra, A., & Brown, A. N. (2016). The growth of impact evaluation for international development: How much have we learned? *Journal of Development Effectiveness, 8*(1), 1–21. <https://doi.org/10.1080/19439342.2015.1034156>

- Cilliers, J., Fleisch, B., Taylor, S., & Thulare, T. (2020). *Can virtual replace in-person coaching? Experimental evidence on teacher professional development and student learning in South Africa*. Research on Improving Systems of Education [Working Paper 20/050].
https://assets.publishing.service.gov.uk/media/5f9052afe90e072ca19b9c27/RISE_WP-050_Cilliers_etal.pdf
- Clark, U. (2005). Bernstein's theory of pedagogic discourse: Linguistics, educational policy and practice in the UK. *English Teaching: Practice and Critique*, 4(3), 32–47.
- Darling-Hammond, L. (1989). Accountability for professional practice. *Teachers College Record*, 91(1), 59–79.
- De Clercq, F., & Shalem, Y. (2014). Teacher knowledge and employer-driven professional development: A critical analysis of the Gauteng Department of Education programmes. *Southern African Review of Education*, 20(1), 129–147.
- de Souza Leão, L., & Eyal, G. (2019). The rise of randomized controlled trials (RCTs) in international development in historical perspective. *Theory and Society*, 48(3), 383–418. <https://doi.org/10.1007/s11186-019-09352-6>
- Deaton, A. (2010). Instruments, randomization, and learning about development. *Journal of Economic Literature*, 48(2), 424–455. <https://doi.org/10.1257/jel.48.2.424>
- Department of Basic Education. (2017). *The Second Early Grade Reading Study. Classroom observation study: Grade 1*.
<https://www.education.gov.za/Programmes/EarlyGradeReadingStudy.aspx>
- Duflo, E., Banerjee, A., & Kremer, M. (2016). *Randomized controlled trials, development economics and policy making in developing countries* [PPT].
<http://pubdocs.worldbank.org/en/394531465569503682/Esther-Duflo-PRESENTATION.pdf>
- Duflo, E., Glennerster, R., & Kremer, M. (2006). *Using randomization in development economics research: A toolkit* [Technical Working Paper 333]. National Bureau of Economic Research. <http://www.nber.org/papers/t0333>
- Fleisch, B., & Dixon, K. (2019). Identifying mechanisms of change in the Early Grade Reading Study in South Africa. *South African Journal of Education*, 39 (3),
<https://doi.org/10.15700/saje.v39n3a1696>
- Fleisch, B., & Schöer, V. (2014). Large-scale instructional reform in the Global South: Insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education*, 34(3), 1–12.

- Fleisch, B., Schöer, V., Roberts, G., & Thornton, A. (2016). System-wide improvement of early-grade mathematics: New evidence from the Gauteng Primary Language and Mathematics Strategy. *International Journal of Educational Development, 46*, 157–174.
- Gibbs, E., Jones, C., Atkinson, J., Attfield, I., Bronwin, R., Hinton, R., Potter, A., & Savage, L. (2021). Scaling and ‘systems thinking’ in education: Reflections from UK aid professionals. *Compare: A Journal of Comparative and International Education, 51*(1), 137–156. <https://doi.org/10.1080/03057925.2020.1784552>
- Hoadley, U. (2016). *A review of the research literature on teaching and learning in the foundation phase in South Africa*. Research on Socioeconomic Policy (ReSEP) [Working Papers: 05 /16.] Department of Economics, University of Stellenbosch, RSA.
- Hoadley, U. (2018). *Pedagogy in poverty: Lessons from twenty years of curriculum reform in South Africa*. Routledge.
- Jones, R., Jones, R. O., McCowan, C., Montgomery, A. A., & Fahey, T. (2009). The external validity of published randomized controlled trials in primary care. *BMC Family Practice, 10*(1), 5. <https://doi.org/10.1186/1471-2296-10-5>
- Kotze, J., Fleisch, B., & Taylor, S. (2019). Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa. *International Journal of Educational Development, 66*, 203–213. <https://doi.org/10.1016/j.ijedudev.2018.09.004>
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.
- Motilal, G. B., & Fleisch, B. (2020). The triple cocktail programme to improve the teaching of reading: Types of engagement. *South African Journal of Childhood Education, 10*(1). <https://doi.org/10.4102/sajce.v10i1.709>
- Muller, S. (2015). Causal interaction and external validity: Obstacles to the policy relevance of randomized evaluations. *The World Bank Economic Review, 29* (suppl 1), S217–S225. <https://doi.org/10.1093/wber/lhv027>
- Pearson, D. (2004). The reading wars. *Educational Policy, 18*(1), 216–252. <https://doi.org/10.1177/0895904803260041>
- Piper, B., DeStefano, J., Kinyanjui, E. M., & Ong’ele, S. (2018). Scaling up successfully: Lessons from Kenya’s Tusome national literacy program. *Journal of Educational Change, 19*(3), 293–321. <https://doi.org/10.1007/s10833-018-9325-4>

- Piper, B., Jepkemei, E., & Kibukho, K. (2015). Pro-poor PRIMR: Improving early literacy skills for children from low-income families in Kenya. *Africa Education Review*, 12(1), 67–87. <https://doi.org/10.1080/18146627.2015.1036566>
- Piper, B., Simmons-Zuilkowski, S., Dubeck, M., Jepkemei, E., & King, S. J. (2018). Identifying the essential ingredients to literacy and numeracy improvement: Teacher professional development and coaching, student textbooks, and structured teachers' guides. *World Development*, 106, 324–336. <https://doi.org/10.1016/j.worlddev.2018.01.018>
- Piper, B., Simmons-Zuilkowski, S., Kwayumba, D., & Oyanga, A. (2018). Examining the secondary effects of mother-tongue literacy instruction in Kenya: Impacts on student learning in English, Kiswahili, and mathematics. *International Journal of Educational Development*, 59, 110–127. <https://doi.org/10.1016/j.ijedudev.2017.10.002>
- Piper, B., Simmons-Zuilkowski, S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the PRIMR Initiative. *International Journal of Educational Development*, 37, 11–21. <https://doi.org/10.1016/j.ijedudev.2014.02.006>
- Piper, B., Sitabkhan, Y., Mejia, J., & Betts, K. (2018). *Effectiveness of teachers' guides in the global south: Scripting, learning outcomes, and classroom utilization*. Research Triangle Institute Press. <https://doi.org/10.3768/rtipress.2018.op.0053.1805>
- Rincón-Gallardo, S., & Fleisch, B. (2016). Bringing effective instructional practice to scale: An introduction. *Journal of Educational Change*, 17(4), 379–383. <https://doi.org/10.1007/s10833-016-9288-2>
- Sabet, S. M., & Brown, A. N. (2018). Is impact evaluation still on the rise? The new trends in 2010–2015. *Journal of Development Effectiveness*, 10(3), 291–304. <https://doi.org/10.1080/19439342.2018.1483414>
- Shalem, Y. (2018). Scripted lesson plans—What is visible and invisible in visible pedagogy? In B. Barrett, U. Hoadley & J. Morgan (Eds.), *Knowledge, curriculum and equity: Social realist perspectives* (pp. 183–199). Routledge.
- Shalem, Y., & De Clercq, F. (2019). Teacher development and inequality in South Africa: Do we have now a theory of change. In N. Spaull & J. Jansen (Eds.), *Why inequalities in SA education persist? A study of the present situation and future possibilities* (pp. 243–261). Springer.
- Shalem, Y., Steinberg, C., Koornhof, H., & De Clercq, F. (2016). The what and how in scripted lesson plans: The case of the Gauteng Primary Language and Mathematics Strategy. *Journal of Education*, 66, 1–24.

- Taylor, S. (2019). *How can learning inequalities be reduced? Lessons learnt from experimental research in South Africa African Schooling: The Enigma of Inequality*. <https://doi.org/10.13140/RG.2.2.26339.91689>
- Tomlinson, M., Ward, C. L. L., & Marlow, M. (2015). Improving the efficiency of evidence-based interventions: The strengths and limitations of randomised controlled trials. *South African Crime Quarterly*, 51(0), 43. <https://doi.org/10.4314/sacq.v51i1.5>
- Young, M., & Muller, J. (2013). On the powers of powerful knowledge. *Review of Education*, 1(3), 229–250.