FELIX MARINGE (ED)

SYSTEMATIC REVIEWS OF
**RESEARCH
IN BASIC
EDUCATION**
IN SOUTH AFRICA

CHAPTER 2

Interventions to improve learner achievement in South Africa: A systematic meta-analysis

Neissan Alessandro Besharati¹, Brahm Fleisch and Khotso Tsotsotso

Abstract

The paper summarises a systematic review of 165 effects of 37 education interventions implemented in South Africa post-1994, by public and private education agencies. The review appraises 35 impact evaluations and conducts a comparative meta-analysis of rigorous experimental and quasi-experimental studies implemented in South Africa's schooling sector. The meta-analysis reveals many locally-produced studies, which had not been captured in previous international reviews, and highlights some of the programmes and policies that had the most impact on language, mathematics and science learning in South African public schools. It also explores the various contextual factors and internal design features that influence the magnitude of the effects reported in various impact studies. The paper provides researchers, funders, governments and education managers with a synthesis of 20 years of empirical evidence from South Africa that can inform programming and policies for improved learning outcomes in developing countries.

Keywords: South Africa, Education, Meta-Analysis, Systematic Review, Impact Evaluations, Effect Size

Introduction

Though a middle-income emerging economy, South Africa is a country that suffers from an extremely weak public school system. This exacerbates other national development challenges, such as unemployment, poverty and inequality and racial divides. Both international and regional learner assessments consistently show South Africa's underperformance, not only in comparison to the rest of the world, but also to lower-income countries on the African continent (Fleisch, 2008; Spaull, 2013).

The historical legacy from the apartheid era was racially segregated schooling systems, designed to privilege the white minority and limit the aspirations and potentials of the black majority. With the ushering in of the democratic dispensation in 1994, the government had the arduous task of undoing decades of structural inequality and amalgamating the previously "white", "Indian",

¹ This review is based on a chapter of Besharati's (2016) PhD thesis at the University of the Witwatersrand

“Coloured” and “black” schools within one schooling system. Numerous reforms were introduced, such as the South African Schools Act (1996), Integrated Quality Monitoring System (IQMS), the national school feeding scheme and a National Qualifications Framework. The national curriculum was revamped several times (1997, 2002, 2008) with little impact on the bimodal pattern of achievement (Fleisch, 2008; Spaull, 2013).

The results from the national public external examination provide an important signal about learner knowledge and ability to enter further education and the work force (Kanjee, 2007). In an effort to assess learning at earlier points in the school system, the Department of Basic Education initially introduced the Systematic Evaluations (SEs) and later the Annual National Assessments (ANAs). Some provinces, such as the Western Cape, have also been implementing their own systematic learner assessments in the schools. It was assumed that these assessment processes would contribute to education improvement through a combination of external and internal accountability (Elmore, 2008).

Notwithstanding twenty years of education reforms, the international and national assessments consistently showed that the majority of learners in previously black schools continued to perform significantly lower than learners in previously white schools (Bloch, 2009; Spaull, 2013). At secondary level, the learning gaps are particularly wide in the subject areas of mathematics and sciences.

Numerous studies (Besharati, 2014; Botes & Mji, 2010; Taylor & Prinsloo, 2005; Simkins, 2010) have highlighted how language has been a serious obstacle to the learning of more complex mathematical and scientific concepts. The issue becomes more complicated when 83% of South African learners’ home language is one of the nine African languages, but would switch to English or Afrikaans as their medium of instruction from Grade 4 onward. Linked to this is the findings (Van der Berg, 2008; Gustafsson, 2007; NPC, 2011; Taylor, 2011) that social-economic status (SES), such as household poverty, race, location, parents’ educational background and other community factors, are also very highly correlated to learner performance.

Aside from the broader social-economic factors, constraints within the education system contribute to unequal learning outcomes. Many schools, especially in the rural and poor parts of the country, lack adequate infrastructure, facilities, materials and resources to facilitate effective teaching and learning. Schools attended by children from poor and working-class families are more likely to be staffed with unmotivated and ineffective teachers (Christie, Butler & Potterton, 2007; Deacon & Simkins, 2011; Taylor, 2011). Although the majority of South African educators have been trained under the post-1994 dispensation (CDE, 2011), there is evidence that suggests that many have poor content knowledge and weak pedagogical skills (Spaull, 2011; Taylor, 2009). Absenteeism, accountability, productivity and professionalism are serious concerns in the South African schooling system

Chapter 2

(Chisholm, Hoadley, & Kivilu, 2005; NPC, 2011; Taylor, 2011) that extend from teachers and school managers all the way to the educational authorities at district and provincial level.

Notwithstanding the grim picture illustrated above, education has always been a high priority in South Africa's national development planning processes (NPC, 2012), as this is closely linked to other social and economic problems affecting the country. South Africa contributes 5-6% of its GDP every year to the education sector (OECD, 2008). The majority of financing has been undertaken by the government, which utilises around 20% of national budget and 30-50% of provincial budgets every year on education spending (National Treasury, 2015). Education remains a high priority in the Government's 2030 National Development Plan, as well as the Department of Basic Education Strategic Action Plans.

The private sector has also played a very prominent role in South Africa's education sector. Acknowledging the importance of a well-educated and capable young workforce for the success of businesses, the domestic corporate sector invests the largest portion of its social investments (around R3-5 billion a year) in education (Trialogue, 2015). Corporate social investments in South Africa's education sector are ten time larger than the aid provided by the traditional donors, such as DFID, GIZ, UN, and World Bank (Besharati, 2015). Numerous public-private partnerships, such as the Joint Education Trust (JET), the Business Trust, the National Business Initiative (NBI) and the latest National Education Collaboration Trust (NECT), have also been established to coordinate stakeholders in addressing the education challenges of the country.

Over the past two decades, national and provincial governments, foreign donors, private companies, foundations, NGOs and universities, have piloted and implemented a range of diverse initiatives to improve the situation of South Africa's education sector. Where resources have been limited, interventions have been typically narrower and more focused. Government and bigger donors, on the other hand, have often recognised the complexity and the multi-layered deficiencies in the education system; thus they have often taken on more comprehensive whole school development (WSD) programmes. These, however, have been fairly expensive initiatives and viewed by some as "scattered gun shot" (Bloch, 2009).

There is a growing body of research internationally that provides insights into what works to improve education (Conn, 2017; Glewwe et al., Hattie, 2009, 2014; Krishnaratne, White, & Carpenter, 2013; McEwan, 2015; Kim et al. 2019). What this literature shows is that there are numerous types of education interventions aimed at achieving a variety of outcomes. Some, like school feeding, might achieve a number of simultaneous objectives, such as incentivising enrolment/retention and improving nutrition – thus cognitive abilities of learners (i.e., Conn, 2017). While this literature is useful as a starting point, given the unique history and

challenges of South African education, we need to build on evidence of what works to improve learning outcomes in South African schools.

The focus of the following meta-analysis, is the interventions specifically aimed at improving learning outcomes and achievement (Table 1). Eight categories of interventions have been defined for the purpose of conducting a systematic review of South Africa's education sector. These categories include interventions targeted directly at learners; teachers' development and incentives for teachers; the provision of educational materials; management and governance; school infrastructure; change in policies; interventions that attempt to provide integrated school reform. The typology is inevitably limiting, subjective and reductionist as these divisions are not as clear-cut as they appear to be, as most education programmes and policies operate in a complex manner and are often the combination of different interventions and approaches, and the interplay between them. The classification of interventions used in studies like this, can greatly affect the final results and conclusions of systematic reviews, as was found also by Evans and Popova (2015).

Table 2.1 *Categories of education interventions*

<i>Learner-targeted support</i>	Typically implemented by the private sector and civil society groups for a small group of selected learners, they consist of supplementary enrichment classes or tutoring by specialised teachers, often done after school hours, on weekends or during the school holidays. Another category of learner-focused interventions can be scholarships, bursaries and support to disadvantaged-but-promising learners to be placed in high-end schools or special courses.
<i>Teacher development initiatives</i>	Interventions implemented to address teacher quantity, quality and performance. This can include the provision of additional teachers and assistants for a short or long period of time. Teacher development is usually divided into pre-service (PRESET) training and in-service training (INSET). The latter consists of upgrading of qualifications, knowledge or skill-sets through different models of training implemented over weekends/evenings, school holidays or several months of block-release college/university programmes. Many teacher development programmes often also have a component of classroom visits, coaching, monitoring and experience sharing between participants.
<i>Learning & teaching study material (LTSM)</i>	Materials, technology and other resources provided to schools to enhance teaching and learning experience. Examples of such can be textbooks, workbooks, study guides, lesson plans, science-kits, computers, multi-media accessories, educational toys, reading books, etc.
<i>Management & governance</i>	Interventions aimed at strengthening school management and leadership, often done in the form of training and coaching of principals and school management teams (SMTs). These are often geared to putting new systems in place to improve school functionality and to promote accountability, efficiency and good governance. These can include the introduction of specific incentives and prizes for achievement and performance.
<i>Infrastructure & facilities</i>	The classical <i>hardware</i> investments, such as building, upgrading, expanding and refurbishing schools, classrooms, administration blocks, toilets, science labs, sporting facilities and libraries. Although not always as obvious as some of the above interventions, the atmosphere, cleanliness, size, lighting, order and other environmental aspects of a school can have an impact on the learning processes.

Chapter 2

<i>Structural reforms, policies & incentives</i>	These are interventions aimed at improving the broader education systems that support the schools. These can include programmes for district development, introduction of specific language policy, systematisation of learner assessments, performance rewards, decentralisation of functions and other structural changes to improve management and accountability. These are normally interventions done by, or with, the government's Department of Education at national and sub-national level.
<i>Community & family involvement</i>	These acknowledge the important role parents and community play in the learning process inside and outside the formal schooling process. Such initiatives extend programme components to other non-school stakeholders and broader community actors. These include interventions aimed at strengthening school governing boards (SGBs), which often include parents and other local authorities.
<i>Integrated School Development</i>	These are complex multi-layered programmes that combine three or more of the above interventions into a holistic strategy to improve school functionality and, ultimately, learning results. These are usually organised in the form of whole school improvement models that are built on prior research and specific theoretical frameworks, piloted in a few areas and then scaled to many schools, with large injections of funding and cooperation of various partners.

Challenges in evaluating education interventions

In the past two decades, there has been an abundance of projects and initiatives implemented by different institutions to improve learning results in South African schools. Each of these interventions may have been evaluated with different methods and approaches, and with different degrees of rigour – thus successes or the failures of each programme have been reported in a very heterogeneous manner.

A number of evaluations undertaken in South Africa's education sector have utilised the case study approach (Mouton et al., 2013), often due to the limitations with data, methodological expertise, budget, time and context in which they were implemented. Although, in many cases, such studies were the only option, these still carried value and contributed knowledge to the field. Nonetheless, the bulk of South Africa's education evaluations relies heavily on qualitative approaches which, on one hand, provide depth, context and texture, but on the other hand, by nature, are more subjective and susceptible to bias and personal experiences of the evaluator, the client and the respondents (Bamberger, 2009). This poses questions with regard to the independence and scientific validity of many of the agency-commissioned evaluations (Roche & Kelly, 2005). Many of the education evaluations commissioned by funding agencies and corporate social investors (Trialogue, 2015), focus on the reporting of inputs (how much was spent), activities (what was done) and, occasionally, outputs of a programme (i.e. number of beneficiaries). Few evaluations are done looking at the impact of initiatives at the level of outcomes, such as real changes (behavioural, institutional, social-economic) in the recipient population. Evaluations capturing no or negative results of interventions often go unpublished or are concealed from public debate, while positive evaluations are used by donor and implementing agencies as marketing

material. Because of their methodological heterogeneity, qualitative studies are also more difficult to compare between each other and assess against a common scale.

The quantitative education evaluations have utilised statistical indicators to empirically measure and compare change in outcome of interest, most commonly *learner achievement* (Schollar, 2015). This is traditionally measured through the indicators of *test scores* or *pass rates*. In South Africa, most learning assessments are designed and implemented by experienced evaluation and research organisations, such as the Human Sciences Research Council, JET Education Services and Eric Schollar and Associates. Some researchers, however, prefer to use state-run standardised universal learner assessments, such as National Senior Certificate examination results and Annual National Assessment (ANA) results (for 2011 to 2014).

Like much of the developing world, most education evaluations in South Africa have been non-experimental. Many studies, popularly referred to as impact evaluations, are actually case studies, where beneficiary outcomes are observed only after the intervention. More advanced evaluations conduct baselines and are able to measure change before and after intervention, but these are relatively rare (Bamberger, Rao, & Woolcock, 2010). Some studies have gone to the extent of creating a comparison group and, in best-case scenarios, have measured pre- and post-results of both programme and control group. A range of literature also exists in South Africa (i.e., Fleisch, 2006; Gustafsson, 2007; Hunt et al., 2010; Simkins & Paterson, 2005; Van der Berg, 2008) that utilises more complex econometric models to control for different variables and estimate effects of interventions, school and community factors on learning outcomes.

In the past decade, there has been a rise in the use of experimental approaches for the evaluation of education interventions in developing countries (Duflo, Glennerster & Kremer, 2008; McEwan, 2014) and also in South Africa (i.e. Fleisch et al., 2011; Louw, Muller & Tredoux, 2008; Schollar, 2015; Taylor & Watson, 2013). However, randomised control trials (RCTs) often face ethical, political and practical constraints, as they need to be built in at the design stage in the beginning of an intervention. Impact evaluations of South African government policies have been difficult to conduct as interventions have often been rolled out in all schools of the country, leaving no space for a counterfactual. This challenge was encountered, for instance, by Coetzee and Van der Berg (2012), as they attempted to evaluate the impact of South Africa's school nutrition programme.

Because of the above constraints, what has dominated much of the impact evaluations in South Africa's education sector has been the use of descriptive and quasi-experimental approaches. By and large, the most common counterfactual evaluations used in South Africa's education sector have been the combination of Difference-In-Difference (DID) designs, with some form of matching techniques – including Propensity Score Matching (PSM; Besharati, 2014; Hobden & Hobden,

Chapter 2

2009).² Other South African researchers have tried to emulate RCTs, by using other evaluation approaches, such as natural experiments (Gustafsson & Taylor, 2013) or cut-off point design (Fleisch & Schoer, 2014; Fleisch et al., 2016).

In South Africa, government agencies, private philanthropies and academia have tried to address the deficiencies in the education sector by reviewing the evidence emerging from the numerous policies and programmes implemented over the years. The JET Education Services, a not-for-profit education agency established in the early 1990s, has been one of the forerunners in the use of research and evaluation to inform education programming in South Africa. Zenex Foundation, a prominent education grant-maker committed to evidence-based programming, commissioned education evaluators (Roberts & Schollar, 2006) to conduct a meta-evaluation on the interventions aimed at improving language, mathematics and science learning outcomes in South Africa. A similar follow-up review on the Zenex-funded school programmes was conducted again by the same authors in 2011. Building on some of their previous scholarly work, Taylor, Fleisch and Schindler (2008) conducted a review for the Presidency on the changes in South Africa's education system since 1994. In the same year, the Organisation for Economic Cooperation and Development (OECD) conducted a large review of national education policies implemented in South Africa (OECD, 2008).

In more recent years, Sayed, Kanjee and Nkomo (2013) published a review of some of the most prominent programmes to improve the quality of education in South Africa, compiling lessons learnt from the Education Quality Improvement Partnership (EQUIP), the District Development Support Programme (DDSP), IMBEWU I and II, Quality Learning Programme (QLP), Learning for Living (LfL), the Integrated Education Programme (IEP) and the Khanyisa School Development Programme.

One of the most comprehensive review efforts in South Africa's education literature was conducted by Mouton et al. (2013) in order to inform the Zenex Foundation of the next phase of school development programming. The review synthesised evidence from hundreds of evaluation reports, journal articles, books and previous meta-evaluations that looked at the effectiveness of education development programmes. The report provided the Foundation with a framework to analyse the literature and make evidence-based decisions on their future school development programme. As part of the review, Mouton et al. (2013) provided an evaluation framework which illustrated the range of studies that spanned from experimental, theory-based evaluations (TBE) to case studies. Although containing

² DID is a design technique that approximates an experimental design by measuring the differential effects of a treatment between two groups but in a natural experiment. PSM is a technique that allows a researcher to calculate an effect of an intervention by comparing two groups, one having received the intervention compared with one that did not by matching the groups on key variables.

many elements of systematic reviews, Mouton et al.'s review (2013) did not strictly follow the systematic review protocols outlined by the Campbell Collaboration, or by the PRISMA (www.prisma-statement.org) for systematic reviews. Mouton's review included quality appraisal of the primary studies, identifying these as "credible" and "moderately credible" to draw out evidence and lessons to inform future programming. The approach Mouton used could best be characterised as a narrative review, as it provided qualitative analysis, depth, texture, context and elaboration on the *mechanisms of change* of school development in South Africa. Mouton et al.'s review did not, however, provide an empirical comparator between impact results (i.e., effect sizes) from the different evaluation studies, therefore making it more difficult for policy-makers and investors to assess the most cost-effective among the competing options. In order to take Mouton et al.'s (2013) review a step further, the following chapter presents a systematic meta-analysis of the major education interventions implemented in South Africa from 1994 to 2017, integrating statistical analysis of effect sizes with qualitative analysis of the interventions and the evaluations conducted on them.

Method

Multiple Treatment Meta-Analysis

In the following review, aimed at comparing effectiveness of different types of education programmes implemented in South Africa, a specific type of methodological approach multiple treatment meta-analysis (MTMA) was used. MTMA has been elaborated upon by statisticians such as Caldwell (2005), Lu and Ades (2006) and Salanti (2012), and applied also in other meta-analyses of education interventions in developing countries (Conn, 2017; McEwan, 2015).

In contrast to standard meta-analysis, multi-treatment meta-analysis assesses and compares different types of interventions aimed at achieving the same outcome (i.e. improved learner results). This method helps to answer policy questions and assists decision-makers when there are multiple options on what investments to make and which type of interventions to favour. In MTMA, heterogeneity is expected, therefore a random effects model that excludes the calculation of the average pooled effect size, is used (Figure 2.2).

Individual impact evaluations of different interventions are tested against different populations, which can be more or less responsive to the treatment (Cranney et al., 2002), depending on context, time and other factors. One key feature of MTMA is that it tries to emulate a large multi-treatment experiment by utilising a common control group. Comparisons are made through observational means, by assessing one intervention in relationship to another intervention, or a null situation. This model is based on the assumption that the comparator used to assess the effectiveness of the various interventions is somewhat similar and

consistent. In our particular case the common control group is South African public schools running business as usual and affected by the same social, economic, political and environmental factors present in the country. Although the various experimental evaluations have occurred in the same country's educational system, it needs to be acknowledged that the contexts do vary, based on the provincial/local characteristics and the historical period in which the intervention and the evaluation occurred. Figure 2.1 presents a graphic illustration of the MTMA that was conducted in this study.

In line with standard practice (Campbell, 2015; <http://www.prisma-statement.org>), the following MTMA of South Africa's education sector was conducted by multiple reviewers, following all the key steps of standard systematic reviews and meta-analyses. It is elaborated upon in more detail in the succeeding sections.

Systematic review protocol

Inclusion criteria. The studies that were included in the following systematic review met these specifications:

- **Study designs.** Rigorously performed experimental and quasi-experimental impact evaluations, ideally with pre- and post-measures for both treatment and control groups. Sample sizes were relatively large to allow for enough statistical power of the results.
- **Participants.** Only studies conducted in South Africa's schooling sector, both primary and secondary phases (grades 1 to 12), post-1994. Both programme and control groups consisted of learners or schools from South Africa. This was to ensure a certain level of consistency with regard to context, social, economic, political, and systematic confounders.
- **Interventions.** Any type of interventions aimed at improving learning outcomes. These ranged from teachers-based, learner-focused materials and resources, management- or system-oriented, whole school development. See Table 2.1.
- **Outcomes.** Learner achievement, measured through test-scores or pass-rates. Particular focus was given to the domains of language, mathematics and science, but other learning results were also included.

Identification strategy

As with any systematic review, the study endeavoured to uncover all the published and unpublished materials in order to reveal studies that would meet the above requirements. These included academic papers, journal articles and research and evaluation reports from government, corporations, NGOs and consultants – and any other grey literature or papers hidden in “file drawers”.

Search terms and key words utilised included: *impact, *evaluation, *South Africa, *education, *effects, *learning, *achievement, *mathematics, *science,

*language, *programme, *outcome, *projects, *school, *teacher, *numeracy, *literacy – or a combination of these.

In the literature search the following sources³ were consulted:

- databases and libraries of specialised agencies;
- academic databases and journals;
- collegial networks and informal communication with prominent evaluators and researchers in South Africa's education sector;
- other South African reviews, meta-evaluations, primary studies and edited books; and
- previous international reviews and meta-analyses on education in developing countries.

Appraising quality and bias in the studies

Once all the studies that met the inclusion criteria had been identified and short-listed from the literature, one of the most complex tasks was to decide on their quality. The subsequent exercise was the screening of individual evaluations on methodological rigour and design quality, to decide whether they should be included, or not, in the meta-analysis. The tension always lies between not leaving out any useful and important piece of evidence, while also making sure that a few bad studies do not negatively influence the overall findings of the meta-analysis (Hombrados & Waddington, 2012). In deciding which studies are methodologically sound, careful assessments needed to be made about internal and external validity, construct measures and statistical errors.

In the medical field, meta-analyses restrict themselves almost entirely to rigorously implemented RCTs. In the social sciences and in education policy, however, as discussed earlier, the field is dominated by quasi-experiments, which can present numerous different threats to validity and risk of biases. These can be mitigated if treatment/control allocation rules are clear (Hansen et al., 2011), selection bias has been appropriately addressed and programme implementation and causality are carefully controlled. Whether big or small, any bias, assumptions, errors or problems in a particular study need to be carefully examined and reported, so that these can be taken into account in the meta-analysis and used as controls in the subsequent sensitivity and moderator analysis.

More than 50 different threats to validity exist in both experimental and quasi-experimental studies, but, in the current meta-analysis on the interventions in South Africa's education sector, the following aspects have been given special attention in the evaluation quality appraisal process:

³ For a detailed list of sources, refer to Bisharati (2017).

Chapter 2

- **Selection bias.** If the counterfactual is not strong, or if programme and control group are practically and significantly different. The experiment therefore becomes compromised from the start.
- **Implementation bias.** If the programme was not implemented correctly according to plan, or if the two groups received essentially different treatments from different service providers. Implementation bias can also include spillover between treatment and control group and contamination of the samples by other interventions.
- **Data and testing bias.** If a large portion of the data is unreliable, inaccurate or missing. Large amount of attrition occurring, or no-show of participants between pre- and post-tests. This can also include the presence of ceiling effects or floor effect in the data-gathering process.
- **Evaluator bias.** If the authors of the evaluation had a conflict of interest (political or financial), were directly involved in the implementation of the programme and might have had a vested interest in the results of the study. This compromises the independence of the study.
- **Motivational bias.** If there is a strong presence of Hawthorne or John Henry effect, where the participants of either the treatment or control group behave differently, as they are conscious of being observed and have a vested interest in the results of the evaluation.

In education evaluation there is often a tension between trying to avoid selection biases and contamination biases. To allow for similar characteristics in programme and control groups, schools and learners are often matched based on proximity of location; however, this can also increase the likelihood of control participants taking advantage of the intervention when they are not supposed to. These challenges have also been illustrated by Schollar (2015) in his RCT of the Primary Math Research Project (PMRP) in Limpopo Province. These highlight again the many practical and ethical complications of conducting experimental evaluations in real-life scenarios, such as in the education policy field.

Many tools and instruments, developed by many organisations conducting systematic reviews, exist to assess risk of bias; however, rating studies are generally very controversial, subjective and discouraged as a source of reliable quality assessment (Conn, 2017; Wilson et al., 2011). Assessing the quality of studies ultimately remains a judgement call based on the expertise of the reviewer, both in statistics and the content matter. Having acknowledged all the limitations, this systematic review used an adapted version of the Cochrane risk of bias tool (Higgins et al, 2011). After a careful assessment by the review team, each study was rated against the five major bias areas discussed above. A score, from 1 (minimum) to 5 (critical), was assigned to each study, signalling the degree of bias and problems detected.

Most of the studies that were included in this meta-analysis had pre- and post-scores for both treatment and control groups, allowing them all to integrate the DID method in the evaluation design. The consistent use of the double difference method reduced some of the biases by combining the time variable (before and after) with the counterfactual variable (participants and non-participants). Any inherent differences between programme and control groups, because of observed or unobserved differences, was significantly reduced, assuming that these differences remained constant over time (Blum, Krishnan, & Legovini, 2010). This is the case with most studies and interventions included in this meta-analysis, as the learner population was affected by the same systemic and external influences in the same country during the same period of time.

Data extraction and follow-up. The effect size, which was used in this meta-analysis, is Hedge's *g*, which is an enhanced version of the *standardised mean difference (SMD)* and *Cohen's d*. To compute this effect size, what is essentially required is mean scores, standard deviations and the sample sizes for both treatment and control groups in both pre- and post-tests (as most of our primary studies were DIDs). These statistics also allow for the calculation of the confidence intervals and standard errors for each effect size.

In many studies, the sample sizes of the participants would not be the same in the pre- and post-tests, because of learner attrition, thus posing a threat to validity in the implementation of the experiment. Other times, pre- and post-tests were not done with the same cohort of participants, but rather longitudinally with the same grade level. In such cases, the reviewers decided either to use the average sample sizes, or to use only the values from the individuals that appeared in both pre- and post-tests. These situations are explained when discussing the merits of each individual evaluation design (see Besharati, 2016). When experimental studies were conducted using cluster sampling techniques (i.e. randomising at school or class level), standard errors and confidence intervals of effect sizes were adjusted, taking into account the clusters, to ensure more precision and fair estimates of results in the comparative meta-analysis.

The required statistics were reported in their entirety in only some of the studies that were reviewed. In most of the cases, the standard deviations were not reported, therefore the reviewers had to contact the original authors to request the missing data. At other times, standard deviations would be re-calculated manually, by going through the raw data provided by the author of the primary study. Following up with the original authors also provided the reviewers with an opportunity to seek further clarification on the evaluations, the interventions and other important details of the studies. Except for the Dinaledi evaluation, which was conducted by World Bank economists based in Washington DC, following up with the original authors on clarifications and data requests was relatively easy, as they were all

based in South Africa and easily accessible through informal e-mail, telephone communication and face-to-face discussion.

The data collection and data cleaning; the risk of bias rating; and the coding of the various studies was conducted by a team of reviewers (the primary researcher and two research assistants) to allow for double-checking of the accuracy of all the data throughout the process. In some complex situations, the reviewers had to debate how to handle a particular case and take decisions on which data to use. Once the required statistics were checked and extracted from all the studies, effect sizes and confidence intervals were calculated utilising Comprehensive Meta-Analysis (CMA) statistical software package.

For a couple of studies, numerous attempts were made to contact the original authors of the papers to seek the missing statistics. However, no response was received and these studies unfortunately had to be excluded from the meta-analysis.

Results

Qualitative overview of studies and interventions

At the end of the systematic literature search, the eligibility screening and data extraction phases, the review yielded a grand total of 35 studies, 37 interventions and 165 effect sizes (Table S1 online only), based on impact measured against different learning outcomes (language, mathematics, science, etc.), or within different sub-groups (i.e. classes, cohorts, years, locations, testing instruments, intervention combination and dosage). Similar to McEwan's (2014) and Conn's (2017) reviews of education interventions, many experimental evaluations included in this meta-analysis had multiple treatment arms. Each primary study, in fact, reported anything from 1 to 21 effect sizes. If an intervention had more than one effect size due to different learning outcomes or sub-groups being assessed, these were each calculated separately and the cases would be identified: for instance as Plus-Time A, Plus-Time B, Plus-Time C, and so forth. Some studies assessed more than one intervention (i.e., Baumgartner et al., 2012; Besharati, 2014; Taylor et al., 2017) and a number of studies were also found for the same intervention.

It is interesting to note that, of the 301 impact studies on education interventions in low- and middle-income countries captured in the six international reviews discussed by Evans and Popova (2015), only one of the South African studies listed above (Louw et al., 2008) was found in only two of the international reviews (Conn, 2017; Glewwe & al., 2014). This points to a serious publication bias present in most international systematic reviews, that gives more prominence to academic studies conducted by Northern scholars (Conn, 2017). The international education reviews, conducted between 2013 and 2016, appear to have overlooked a large portion of the knowledge produced locally in South Africa, which is often

published as evaluation reports by governmental, research and private education agencies active in the sector.

For analytical purposes, all the interventions included in this review have been grouped into (a) government programmes and policies; (b) private sector projects; and (c) academic experiments. The full report of the systematic review (Besharati, 2016) contains a detailed narrative section discussing the various South African education programmes and policies. It also includes a commentary on the quality, strength, weaknesses and limitations of the studies that were conducted for each of the interventions. For each of the primary studies, the potential biases and threats to validity are discussed and the methodological quality of the evaluation is rated, based on the risk of bias scale discussed earlier.

Some of the studies had substantial risk of bias and other methodological problems that precluded them from being included in the subsequent quantitative meta-analysis. A further screening of the above studies was conducted to exclude any evaluation that had high (4) or critical risk of bias (5). This left only the well-implemented RCTs and strong quasi-experimental studies, such as the ones using RDD, PSM or other techniques that took care of selection bias. Evaluations with credible results were normally conducted with large samples and therefore had low standard errors and small confidence intervals.

The screening process now left the meta-analysis with *26 credible studies of 28 different interventions*, which covered overall *110 different effect sizes*, as the main cases and units of analysis. This allowed for a legitimate comparison of results that had a certain level of reliability; and of studies and interventions that were placed on the same level plane, with similar contexts and evaluation parameters.

The funnel plot (Figure 2.1) illustrates the effect sizes against the sampling errors in all the studies in the meta-analysis, once the ones with high risk of bias had been removed. From the funnel plot, we can see that most studies are at the top, therefore the majority of the results reported are statistically significant, although some studies, with a standard error below 0,1, provide slightly less reliable results than the evaluations done with larger sample sizes.

Chapter 2

Figure 2.1 *Funnel plot SA education meta-analysis*

Funnel plots normally help in assessing the degree of publication bias. However, we can also see that there is a similar number of South African studies that show positive, as well as negative, effects on both sides of the zero effect line. The symmetrical nature of the funnel plot highlights that both studies reporting negative results, as well as positive results, were captured in this meta-analysis and thus there is a fairly limited presence of publication bias. There are, however, a number of major outliers at the right hand side of the plot, which indicate extremely and unusually high effect sizes. Most of these cases emerge from Schollar's evaluations (2002, 2005, 2014, 2015), therefore results from these studies need to be assessed with more caution.

Interpreting the effect sizes

Once all the studies included in the review were systematically screened for quality and appropriately short-listed, the statistical meta-analysis could begin through the aid of CMA software. The results of meta-analyses are normally represented through forest plots that illustrate effect sizes of the various treatments, together with their respective confidence intervals, which is reflective of their standard error. The statistical results of the South African education meta-analysis produced a grand total of 110 effects sizes, ranging from interventions with a lowest (negative) Hedge's g of $-0,28$ to a highest of $+1,21$.

While researchers are concerned with statistical significance, policy-makers are interested in educational significance and with the magnitude of the impact. The effect sizes give an indication to non-academic readers, policy-makers, investors and the general public as to whether the interventions have a substantial and practical impact on real-life population outcomes. The question is whether the effect of an education intervention is large enough to warrant replication, transferability to

public policy, extension and scale to the rest of the district, province or country. How do we know if an effect size is meaningful or not for education policy?

There are some general scales introduced by meta-analysis experts, such as Cohen (1988) and Lipsey and Wilson (2001), which provide a rough indication of whether effect sizes should be considered big or small. These guidelines are widely used in the social sciences; however, they are merely a 'rule of thumb' and should not be used religiously. Bloom et al. (2008) argue, in fact, that there are no universal guidelines for interpreting effect sizes. They suggest, rather, to develop empirical benchmarks, which are contextualised in relationship to past research on the same type of interventions, population and outcome measures.

Effect sizes vary tremendously from sector to sector and education programmes are known to have lower effect sizes than in other fields (McEwan, 2014). Different education analysts argue that interventions need to be taken seriously in the policy space, once they demonstrate effect sizes over 0,25 (Coe, 2002), 0,4 (Hattie, 2009) and 0,59 (Marzano, 2003). What has also become clear is that different magnitudes of effect can be expected from different types of education interventions, with learner-targeted one-to-one tutoring, for example, producing much bigger effect sizes than whole school development programmes (Slavin & Smith, 2009).

The way evaluations are conducted and the methodologies that are used also matter greatly in meta-analysis. Lipsey et al. (2012) have, for example, discovered that impact studies using researcher self-developed testing instruments tend to yield higher effect sizes than evaluations which use broad standardised state-run assessments (such as the South African NSC exams). Whether RCTs or other quasi-experimental techniques are used could also have an influence on the results achieved during an impact evaluation.

When conducting experiments in the education sector, Bloom et al. (2008) have also highlighted that effect magnitude may vary depending on the school levels in which the programmes have been implemented. The education and child developmental literature (Bloom et al., 2008) shows that, even in the absence of an intervention, a child is expected to improve his or her learning abilities year after year as a natural course of life.

Although these effect sizes may change in different countries and social contexts – and with different subject outcomes being assessed, the trajectory of learning is clearly the same, where larger gains occur at lower grades and at younger ages while less learning occurs at older ages and at the end of the formal schooling process. This also highlights, again, the importance of using appropriate control groups and good counterfactual evaluation when assessing the impact of interventions in the education sector, in order not to be misled by large learning gains occurring anyway in early grades of the schooling system (i.e., Fleisch et al., 2015).

Context, population, and learning outcomes all play a major role in affecting the results of impact studies and meta-analyses. The vast majority of field experiments and systematic reviews in the education sector are still conducted in the United States and in the industrialised world. Thus, one of the best ways to benchmark the results from the South African impact evaluations is to assess the results against one another, as these were all implemented under similar circumstances and population contexts.

Clemens, Radelet and Bhavnani (2004) also explain that the time at which the impact of the intervention is measured can make a significant difference in the magnitude of the effect. Most interventions measure large effect sizes, during and shortly after they are implemented, due to the excitement and attention given to the beneficiaries during the project life-span (Schollar, 2013). But, several years after the project has finished, a natural process of *decay* or *compound* is expected to occur (Harris, 2009), therefore impact can look very different if measured in the short-term (right after the intervention), medium-term or long-term (several years later). Almost all the effect sizes included in this meta-analysis have been measured at the end of the final year the intervention was administered. The only exception was in Schollar (2015), where the effect size of the PMRP workbooks was measured also three years later (in 2010), with effects of the programme still high, though lower than when measured right after the treatment was provided in 2007.

From the preliminary results of this South African education meta-analysis, we can see that the same intervention (PMRP workbooks) evaluated with the same technique (RCT) in two different studies yield strikingly opposite results, where the *Back to the Basics Workbooks* feature in the same meta-analysis (see forest plot above) as the top most impactful intervention, as well as the third least effective among the 110 effect sizes reviewed. This specific case will be discussed in more detail later in the paper, but overall this goes to further prove that evaluation design and context does matter a lot and effect sizes within a meta-analysis can be confounded by many other factors beyond the simple treatment intervention.

Exploring heterogeneity through moderator analysis

Having acknowledged that effect size is highly contingent on population and study design, it may not be appropriate to accept the results of the meta-analysis at face value. What would possibly be more useful is to explore, in more depth, the heterogeneity within these effects, so as to better understand which interventions seem to be most impactful within specific contexts and circumstances.

In the early stages of the review, all studies that met the inclusion criteria were subsequently coded one-by-one: on the characteristics of the interventions of the population, on the outcome measures and on the evaluation design. Each of these elements would potentially become *predictor variables*, or what is known in meta-

analysis as *moderators*, which would be used to analyse heterogeneity in results. The various moderators would be statistically analysed to see if they had any influence on the effect sizes (as well as the standard errors) emerging from the studies.

To the extent possible, the reviewers tried to code as many moderators as possible in numerical format to allow for regression analysis. Some binary categorical variables could be converted into dummy variables, but others had to remain in categorical format and better analysed through ANOVA and sub-group analysis. Data was entered, checked and organised – first in Microsoft Excel and later into SPSS and CMA – for more complex analyses. The variables coded as part of the review are illustrated in Table S2 (online only), with the right column of the table providing an example from the Besharati (2014) study.

Once all the outcome, intervention, context and evaluation variables were appropriately coded for each study in the meta-analysis, meta-regression, meta-ANOVA and sub-group analysis were used in order to conduct a more in-depth moderator analysis. The next sections will present and discuss the results of this exploration.

Meta-regression and meta-ANOVA

The first step in the moderator analysis was to conduct a meta-regression to observe which of the many studies, interventions and contextual variables (listed in Table 2.4) had a strong influence on the meta-analysis results. Previous education systematic reviews by Conn (2017) and McEwan (2014), which conducted meta-regression as part of moderator analysis, have also highlighted that effect sizes may vary significantly once you control for other factors (simultaneous treatments, intervention features, context, methodology).

When conducting a meta-regression, all the same assumptions, checks and rules which apply to a normal multivariate regression also need to be taken into consideration: large samples, linearity, normality, multi-collinearity, independence of cases, homoscedasticity. As explained previously, the meta-analysis contained 110 valid cases of effect sizes extracted from 26 high-quality studies of 28 interventions. If the meta-regression was conducted using the interventions or studies as the unit of analysis, these would possibly be too small a sample to yield enough statistical power to infer conclusions. Also, the various effect sizes reported in each study were referring to different outcome measures and, therefore, would not be appropriate to collapse under one average effect size per study.

On the other hand, if the meta-regression is conducted at the level of effect sizes ($n=110$), there would be more statistical power; however, a lot of key assumptions of regression analysis could be violated. In our South African meta-analysis, one study cluster may yield anything from 1 to 21 effect sizes. This means that the effect sizes are actually *nested* within the studies and therefore may suffer from the same standard errors, sampling, biases, quality concerns and variations caused

Chapter 2

by context and methodological design. The individual cases used in the regression model are therefore not independent, but, in fact, strongly correlated with other cases from the same study cluster.

Considering this challenge, a separate meta-regression was conducted using both effect sizes and studies as units of analysis. If results were consistent across the two levels, one could safely deduce that a certain moderator had a strong influence on the results of the meta-analysis. Considering the rule of thumb that the relationship between cases and variables in a regression should be roughly 1 to 10, the reviewers decided to test only 1 or 2 moderators at a time to explore if there was a significant and substantive correlation of the variable to the effect sizes. Meta-regression functions were available in Comprehensive Meta-Analysis version 3 (CMA3), which facilitated the exploration.

When doing the analysis with the 110 *non-independent* effect sizes, the study variables named ‘test instrument’, ‘units of analysis’, ‘matched cohorts’, ‘counterfactual IE method’ and ‘statistical significance’ link to Hedge’s *g*. This, however, would not be clear when the analysis was conducted at the level of studies (possibly due to low statistical power). The only variable that yielded statistically significant correlation to the effect size in both levels of analysis was ‘grade of assessment’. The results were almost identical in both models. Table S3 (online only) presents the results from the regression conducted on the 110 effect sizes using ‘grade of assessment’ as the predictor variable.

From the regression results, we can observe that the grade in which the learning assessments are conducted has a clear negative correlation with the effect size and this is also statistically significant. This explains at least some of the variances between the studies. This is also clearly illustrated in the scatter plot of Hedge’s *g* by grade of assessment (Figure 2.3), which shows that all studies regress along the same slope (black line) and are all contained within the same prediction interval (blue lines). The only outliers are the effect sizes from the Schollar (2015) study, which had previously been flagged to produce exceptionally high results.

Figure 2.2 *Graphical illustration of SA education Multi-Treatment Meta-analysis (MTMA)*

Figure 2.3 *Regression of effect size on school grade*

The finding from our meta-regression gives further evidence to the general education theories and the meta-analytical consideration, raised by Bloom et al. (2008), that the higher the school grades in which the intervention occurs (and thus also where the learning assessments are conducted), the lower the effect sizes are to be expected. For the education policy debates, these findings further strengthen the call for more interventions to be conducted in the earlier years of schooling for more impact to be expected on learner achievement.

The meta-regression technique was useful to assess the correlation of continuous numerical variables on the outcome variable of effect size. However, there are also

Chapter 2

some moderators which are dichotomous or categorical, which are worth exploring using other techniques analogous to the analysis of variance (ANOVA). This is commonly conducted in meta-analysis as part of the analysis of heterogeneity.

We have seen earlier that certain parameters from the evaluation design could have potentially played a role in influencing the magnitude of the impact. The researcher decided, therefore, to test if there is a significant difference in the effect sizes of the various studies, depending on the methodological framework that was used in the evaluation. As discussed earlier, there was an indication that 'counterfactual IE method', 'assessment instrument', 'units of analysis' and 'matched cohorts' could have influenced the magnitude of the impact results. A variance test was conducted of the methodological parameters used throughout the various studies. In a similar way to an ANOVA test, a heterogeneity test was performed using Q-value statistics to analyse variance of different sub-groups of studies within the meta-analysis.

What is evident from Table 2.3 is that all of the moderators of methodology (impact evaluation technique, units of analysis, test instrument, the use of pre- and post-matched cohorts), produced significantly different results within the meta-analysis, therefore a portion of the heterogeneity in the effect sizes is derived by the above variables and not just by the intervention itself. These results further confirm some of the findings of Lipsey et al. (2012): that impact studies generally report larger effects when researcher-developed test instruments (versus state assessments) are used to measure outcomes and analysis is conducted at learner level, rather than school level. As expected, larger effect sizes were also linked to statistical significance, though this was often dependent on the standard error and sampling process. To further reduce the amount of effect sizes under observation, the reviewers decided to focus only on the *55 statistically significant results* in the meta-analysis, where the confidence intervals of Hedge's g didn't include zero.

As demonstrated before in the meta-regression, the grade in which the learning assessments are conducted has a substantial and significant influence on the effect size. With the same logic, interventions conducted in lower phases of the South African school system should also report a larger impact than those occurring at the later stages of school grades. This assumption was tested again using the same ANOVA-like heterogeneity test on the 55 statistically significant Hedge's g , with the results illustrated in Table 2.4.

Just as before, with regression on grade of assessment, the pattern is unmistakable. The lower the phases of the school system in which interventions occur, the higher the impact observed on learning outcomes. Even with a limited number of studies and effect sizes, the analysis above proves that the school phase is statistically and substantially significant in predicting the impact of the interventions.

As much as these meta-regressions provide some useful indications about the various factors that influence effect size, the results above need to be interpreted

with caution, as they are based on a fairly limited number of cases which violate some of the key regression assumptions. As discussed earlier, the 110 effect sizes under analysis are not completely independent, as they are correlated with one another within each study group. Nonetheless, it is clear that grade and school level in which the evaluation is conducted does clearly have an effect on the results of education impact studies. From the coefficient estimate, one can also see that there is a negative correlation – as the school phase increases the Hedge's g reported in the impact study tends to decrease.

Sub-group analysis

Having established that heterogeneity in effect sizes, the second phase involved a sub-group analysis. Rather than asking which is the most effective intervention to improve learning outcomes in South African schools, it is more appropriate to continue the investigation asking when, where and how are education interventions effective?

The forest plots (Figures 2.4-2.5) break down the results of the meta-analysis according to the different learning areas, intervention types and implementing agencies. This provides useful insights for policy-makers and investors who need to make decisions about which interventions to design, invest and support in different contexts of South Africa's education sector.⁴

From the subject sub-group analysis (Figures 2.4-2.5), one can observe that literacy and numeracy interventions have a greater impact on learner achievement when implemented in earlier grades in primary schools, such as in the case of Back to the Basics (B2B), COUNT, Gauteng Primary Language and Maths Strategy (GPLMS), Integrated Education Program (IEP), and sJsK Literacy Project. Learning interventions on sciences and other subjects taught at higher stages of the schooling system register a much lower impact than learning occurring in language and mathematics in lower grades. The Dinaledi (Figure 2.6) and the Epoch & Optima Programme (Figure 2.5), however, emerge as exceptional cases in which interventions at high school level produce substantial effects on science and maths learning – even when utilising grade 12 NSC assessments, which are known to normally yield relatively small effect sizes.

What seems to have a significant impact on high school pass rates and test scores is the introduction of additional teachers, assistants and human resources in order to improve educator-learner ratios, reduce class sizes, offer extra remedial

⁴ Depending on the range of effects registered in each sub-group, forest plot scales have been adjusted from $(-1/+1)$ to $(-2/+2)$ accordingly, for example, with the inclusion of the Schollar (2015) results which include effect sizes above $+1,0$. As seen earlier, grade and school phase are very prominent factors influencing effect size; therefore from this point forward they will be included among the descriptive identifiers in the next sub-group analyses.

Chapter 2

Figure 2.4 *Impact of interventions to improve language abilities*

classes and generally provide more time for, and attention to, the individual learners, as illustrated by the findings emerging from the Besharati (2014) study on secondary school interventions in North West and Limpopo provinces, as well as from the evaluations of the Dinaledi schools (Blum et al., 2010), Plus Time (Prinsloo, 2009) and the Epoch & Optima Trusts Maths Challenge Programme (Schollar & Mouton, 2014).

Forest plots, in Figures 2.7-2.8, analyse the effect sizes of some of the most popular types of interventions frequently implemented in South Africa, such as the provision of LTSMs, teacher-development programmes and integrated school development. It needs to be noted that most programmes of LTSM provision in South Africa (i.e. READ, GPLMS, B2B, EGRS) are structured pedagogic programmes (Sniltstveit et al., 2016), with multiple components – including teacher training and coaching.

The most striking finding from the forest plot on LTSMs (Figure 2.7) is that Back to the Basics (B2B) workbooks appear as the most effective, as well as one of the least effective, educational support materials for grade 6 pupils in South Africa. Two well-performed RCTs by two different authors (Fleisch et al., 2010; Schollar, 2015) evaluating the same intervention, which supposedly was implemented in a similar manner, led one study to conclude that the workbooks have a phenomenal effect size of +1,21 and the other to report a disappointing effect size of -0,17. At first, a reader will interpret this drastic difference in the results as possibly caused by the different provincial contexts (rural Limpopo versus urban Gauteng), but

Systematic Reviews of Research in Basic Education in South Africa

Figure 2.5 *Impact of interventions to improve mathematics abilities*

Figure 2.6 *Impact of interventions to improve science and other FET subjects*

Chapter 2

Figure 2.7 *Impact of LTSM provision*

a closer analysis of the two studies will highlight that the variation lies in the methodological approach used in implementing the two experiments. Schollar used this as his counterfactual 'null intervention', while Fleisch and Taylor utilised an alternative textbook (*enhanced standardised practice*), which arguably could also have contained some of the pedagogical elements of the B2B workbook. Eric Schollar and Associates (ESA) were very involved in the design and delivery of the B2B workbooks and thus had a different set of interest and passion for the intervention than possibly the JET and Wits University researchers in the 2010 evaluation, who also ended up using a different set of testing instruments from ESA.

Another interesting insight emerging from the forest plots is that contrary to some of the education impact evaluation literature (Glewwe et al., 2004; Hanushek & Woessmann, 2012; Kremer, Brannen & Glennerster, 2013), but of support to other international reviews (Conn, 2017; Krishnaratne et al., 2013; McEwan, 2014), evidence from South Africa shows that simple provision of effective learning and teaching material (i.e., Botes & Mji, 2010; Schollar, 2015; Taylor & Watson, 2015) can yield a similar impact on learning outcomes as more complex and expensive whole school development programmes – shotgun approach (Bloch, 2009; Mouton et al., 2014). It needs to be highlighted, however, that different intervention modalities are not necessarily mutually exclusive. Often the impact of

Systematic Reviews of Research in Basic Education in South Africa

Figure 2.8 *Impact of whole school development programmes*

one intervention is confounded by the effects of other treatment components (Bold et al., 2013; Friedman et al., 2011). For instance, many of the LTSMs provided by educational agencies (workbooks, lesson plans, science equipment, etc.) are often accompanied by systemic training of teachers and management on how to use them. The provision of extra teachers is also often linked to class and school restructuring.

The last forest plot analyses the difference in effect sizes among the education interventions developed by government agencies and academic researchers (Figure 2.9). Even from an eye-ball glance, it is clear (and somewhat expected) that interventions designed and implemented by academics in a semi-controlled environment, for the purpose of research, experimentation and piloting, yield bigger effect sizes and more precise estimates than the on-going education policies and programmes run in large-scale policy environments by government or major private-public partnerships.

Although various studies conducted in South Africa's education sector (i.e., Blum et al., 2010; Fleisch & Schoer, 2014; Gustaffson & Taylor, 2013; Van

Chapter 2

Figure 2.9 *Impact of government programmes versus academic experiments*

der Berg, 2008) suggest that programmes have a different impact in different geographic areas and socio-economic contexts, the results from the above meta-analysis illustrate that geographic locations within South Africa provide relatively little heterogeneity, with regard to the effect sizes of interventions, compared to other influencing factors, such as education phase and evaluation design. Though the original review (2016) provided a geographic sub-group analysis for the provincial education authorities and private investors working in the specific regions of the country, the forest plots on the effects of interventions conducted in the six provinces of South Africa have been omitted from this paper, as this is less relevant for an international audience.

Discussion

Despite twenty-five years of reform efforts, South Africa continues to face serious weakness in the education system. Government, non-profit organisations and the private sector have tried to compensate for these weaknesses through large investments, innovative education models, complex school development programmes, bold education policies and numerous other initiatives implemented by a diversity of agencies. Notwithstanding the financial, technical and political resources invested in the education sector, this study has confirmed that very few interventions have had a substantial impact on learning outcomes in South Africa's public school system (Bloch, 2009; Fleisch, 2008; Sayed et al., 2013; Taylor et al., 2003). This meta-analysis does not take a formal position with regard to the merits of any particular pedagogical approach, school development programme or learning theory. It rather introduces empirical and systematic methods of evaluation and comparative analysis to the field of education programming.

This meta-analysis is a contribution to evidence-based education research and policy in South Africa. It has captured many of the South African impact studies that international reviews had not previously picked up (Evans & Popova, 2015) and has advanced the work of Mouton et al. (2014) by providing the first quantitative meta-analysis of interventions to improve learning outcomes in South Africa's schooling sector. Numerous impact evaluations have been conducted on education programmes in South Africa, with different methods, degrees of quality and scientific rigour. A number of reviews have also been undertaken (Mouton et al., 2014; Roberts & Schollar, 2011; Sayed et al., 2013) to reflect on the experiences and the learning that have emerged in South Africa's education sector post-1994. Nonetheless, education policy-makers and investors require standardised impact measures to allow for more robust empirical comparison of the effectiveness of different options available to them.

This review has presented the application of a methodological framework for the comparison of intervention effectiveness within the education policy arena of South Africa. Through the use of multiple treatment meta-analysis, the study has analysed the impact of some of the major interventions implemented in public schools over the past twenty years. It has identified some of the programmes that have been most successful in improving learning outcomes and the context and circumstances in which these effects have occurred. While not being conclusive on what are the best education programmes, it has shed some light on when, where, how and why some interventions have shown a significant impact on learner achievement.

What has emerged strongly is that context, methodology and study design have a substantial influence on the effect sizes that are reported in education impact evaluation. Assessment tools, evaluation methods, units of analysis and other methodological choices all appear to influence the type of results that emerge from

Chapter 2

these impact studies. The focus of the intervention (Slavin & Smith, 2009), the timing in which evaluations are conducted (Clemens et al., 2004; Harris, 2008) and potential Hawthorne effects (Schollar, 2015) are all factors that influence impact evaluations. The findings from this South African meta-analysis resonate with similar findings emerging from the education meta-analysis conducted by Lipsey et al. (2012). It also confirms several of the findings by Conn (2017) and Evans and Popova (2015), where diverging results of different reviews are dependent on the sample of studies included and the intervention categorisation utilised by the different reviewers. The South African meta-analysis has also shown how results of impact evaluations cannot be taken at face value, as context and study design have a significant influence on the effect sizes reported in the various programme evaluations.

A vivid illustration of the main story emerging out of the South African meta-analysis is the case of the evaluation of the Back to the Basics (B2) workbooks, where the methodological approach undertaken by two different evaluators (Fleisch et al., 2011; Schollar, 2015) led to diametrically opposite results, though they both implemented rigorous RCTs of the same intervention on the same grade level of learners in different provinces of South Africa. Both experiments, though well implemented, had methodological problems and limitations. A third RCT of the B2B workbooks by a different researcher in a new context would definitely be a project worth pursuing in order to reach more definitive conclusions on the impact of guided workbooks in South African primary schools.

The strongest predictors of effect size in this systematic review have been the schooling phase and grade in which the interventions were implemented, school subjects and the implementation of multi-component models that combine resources and professional development. Programmes implemented at lower levels of the school system and in core subjects, such as literacy and numeracy, (i.e., Fleisch & Schoer, 2014) report much larger learning impact than programmes implemented in higher grades (i.e., Taylor & Watson, 2015). This is also consistent with the general child development literature and some of the international meta-analytical work done in the field of education (Bloom et al., 2008; Lipsey et al., 2012). Several evaluations of high school interventions in South Africa (Besharati, 2014; Blum et al., 2010; Schollar & Mouton, 2014) have discovered that major learning gains occur when there is an injection of additional human resources and teachers into the schools, that help reduce class size and teacher-learner ratios. Numerous studies conducted in primary schools have highlighted the importance of interventions in the area of language (i.e., Early Grade Reading Study, Learning for Living, READ, Reading Catch Up Programme, sJsK Literacy Project), which produce important gains not only on language outcomes, but also on other learning areas (Bohmer, 2014; Fleisch et al., 2015; Schollar, 2005; Taylor et al., 2017). While not included in this systematic review, the very latest studies (Cilliers et

al., 2019; Kotze et al., 2018) largely confirm these findings. Finally, well-designed and comprehensively-delivered learning and teaching resources (LTSMs) appear impactful, as well as cost-effective, in improving learning outcomes relatively quickly – compared to more complex and expensive integrated school development programmes (Taylor & Watson, 2015).

The only two exceptions to the above finding have been the Epoch & Optima programme and the Dinaledi initiative, which have both produced remarkable results on mathematics and science learning outcomes in the FET band (senior high school), which is an education phase traditionally known to report very small gains in learner achievement. Schollar and Mouton's (2014) evaluation of the Epoch & Optima programme, however, suffered from potential selection bias and the World Bank (Blum et al., 2010) evaluation of Dinaledi had some serious data limitation. A follow-up replication of the Dinaledi study is well warranted, as the government flagship programme has meanwhile evolved and the Department of Education data has also significantly improved.

This meta-analysis has also shown that experimental evaluations conducted by researchers in carefully controlled environments, such as the testing of innovative teaching approaches or special learning aids (Bohmer, 2014; Hattingh, 2003; Louw et al., 2008; Padayachee et al., 2011), produce high effect sizes. However, there is no guarantee that the same impact results are to be expected when the interventions are transferred to the broader policy environment, where beneficiaries often receive partial or incorrect exposure to the treatment. High effect sizes in academic studies can also be due to publication bias towards studies with positive results and to some degree to Hawthorne effects.

Finally, this review has underlined the need for more empirical evidence in South Africa's education sector and the use of rigorous counterfactual evaluations, such as RCTs and strong quasi-experimental designs, which address selection bias and allow for rigorous analysis of causality. Attention needs to be given to sampling processes to provide more power and confidence in the results of education impact studies (Fleisch et al., 2015; McEwan, 2014; Taylor et al., 2017). It is hoped that this meta-analysis inspires both public and private agencies to improve the quality of their evaluations and to utilise the best available scientific methods to systematically plan future programmes, make smarter investments and improve their on-going efforts in South Africa's education sector.

References

References marked with an asterisk () indicate studies included in this meta-analysis.*

- Bamberger, M. (2009). Strengthening the evaluation of program effectiveness through reconstructing baseline data. *Journal of Development Effectiveness* 1(1), 37-59. <https://doi.org/10.1080/19439340902727610>
- Bamberger, M., Rao, V., & Woolcock, M. (2010). Using mixed methods in monitoring and evaluation: experiences from international development evaluation. In A. Tashakkori, & C. Teddlie (Eds.), *SAGE handbook of mixed methods in social and behavioral research* (2nd ed., pp. 613-641). Thousand Oaks, CA: Sage Publications. <https://doi.org/10.4135/9781506335193.n24>
- *Baumgartner, J., Smuts, C., Malan, L., Kvalsvig, J., Van Stuijvenberg, M., Hurrell, R., Zimmermann, M. (2012). Effects of iron and n3 fatty acid supplementation, alone and in combination, on cognition in school children: a randomized, double-blind, placebo-controlled intervention in South Africa. *American Journal of Clinical Nutrition*, 96, 1327-38. <https://doi.org/10.3945/ajcn.112.041004>
- *Besharati, N. (2014). *Platinum & passes: the impact of mining investments on education outcomes in South Africa* (Research Report No. 16). Johannesburg: SAIIA & Wits School of Governance.
- Besharati, N. (2015). *Private sector investment in South Africa's education sector: nature, extent and incentives* (Research Brief, 2015). Johannesburg: Jet Education Services.
- Besharati, N.A. (2016). *Evaluating development effectiveness assessing and comparing the impact of education intervention in South Africa* (University of the Witwatersrand, Johannesburg. Doctoral Dissertation).
- Bloch, G. (2009). *The toxic mix: what's wrong with South Africa's schools and how to fix it*. Cape Town: Tafelberg.
- Bloom, H., Hill, C., Black, A.R., & Lipsey, M. (2008). Performance trajectories and performance gaps as achievement effect-size benchmarks for education interventions. *Journal of Research on Educational Effectiveness*, 1, 289-328. <https://doi.org/10.1080/19345740802400072>
- Blum, J., Krishnan, N., & Legovini, A. (2010). *Expanding opportunities for South African youth through math & science: the impact of the Dinaledi program*. Washington DC: World Bank Development Impact Evaluation Initiative (DIME).
- Bohmer, B. (2014). *Testing Numeric: Evidence from a randomized controlled trial of a computer based mathematics intervention in Cape Town High Schools*. Unpublished Masters Thesis. University of Cape Town.

- Bold, T., Kimenyi, M., Mwabu, G., Ng'ang'a, A., & Sandefur, J. (2013). *Scaling up what works: experimental evidence on external validity in Kenyan education*. Working Paper, (321). Washington DC: Center for Global Development. <https://doi.org/10.2139/ssrn.2241240>
- Botes, H., & Mji, A. (2010). Language diversity in the mathematics classroom: does a learner companion make a difference? *South African Journal of Education*, 30(1), 123–138. <https://doi.org/10.15700/saje.v30n1a318>
- Caldwell, D.M. (2005). Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *British Medical Journal*, 7521, 897. <https://doi.org/10.1136/bmj.331.7521.897>
- Campbell Collaboration (2015). *Campbell Collaboration systematic reviews: policies and guidelines* (Monograph Series, Supplement 1). Oslo: Campbell Collaboration.
- Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2019). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 0618-9538R1. https://doi.org/10.35489/BSG-RISE-WP_2018/024
- Centre for Development and Enterprise (CDE, 2011). *Schooling Reform is Possible: Lessons for South Africa from International Experience*. (Round Table No. 18). Johannesburg: Centre for Development Enterprise.
- Chisholm, L., Volmink, J., Ndhlovu, T., Potenza, E., Mahomed, H., Muller, J., & Malan, B. (2005). *A South African curriculum for the twenty-first century*. Report of the Review Committee on Curriculum. Pretoria: Human Science Research Council.
- Christie, P., Butler, D., & Potterton, M. (2007). *Schools that work* (Report to the Minister of Education). Pretoria: Department of Education, Republic of South Africa.
- Clemens, M.A., Radelet, S., & Bhavnani, R.R. (2004). *Counting chickens when they hatch: the short term effect of aid on growth* (Center for Global Development Working Paper No. 44). Washington DC: Center for Global Development. <https://doi.org/10.2139/ssrn.567241>
- Coe, R. (2002). *It's the effect size, stupid: what effect size is and why it is important*. Retrieved from <http://www.leeds.ac.uk/educol/documents/00002182.htm>
- Coetzee, M., & Van der Berg, S. (2012). *Report on a scoping study to evaluate the feasibility of undertaking an impact evaluation of the national school feeding programme and Grade R, based on existing data sources*. Study for the Department of Basic Education and the Department of Policy Evaluation and Monitoring in the Presidency. Stellenbosch: Stellenbosch University.

- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd edition). Hillsdale: Lawrence Erlbaum.
- Conn, K.M. (2017). Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Impact Evaluations. *Review of Educational Research*, DOI: 10.3102/0034654317712025. <https://doi.org/10.3102/0034654317712025>
- Cranney, A., Guyatt, G., Griffith, L., Wells, G., Tugwell, P., Rosen, C., et al. (2002). Summary of meta-analyses of therapies for post-menopausal osteoporosis. *Endocrine Review*, 23, 570. <https://doi.org/10.1210/er.2001-9002>
- De Chaisemartin, T. (2010). *Evaluation of the Cape teaching and leadership institute*. Johannesburg: JET Education Services.
- Deacon R., & Simkins, C. (2011). *Value in the classroom: The quantity and quality of South Africa's teachers*. Johannesburg: Centre for Development and Enterprise (CDE).
- Duflo, E., Glennerster, R., & Kremer, M. (2008). Using randomization in development economics research: A toolkit. In T. Schultz & J. Strauss (Eds.), *Handbook of Development Economics*, Vol. 4. Amsterdam and New York: North Holland. [https://doi.org/10.1016/S1573-4471\(07\)04061-2](https://doi.org/10.1016/S1573-4471(07)04061-2)
- Elmore, R. (2008). Leadership as the practice of improvement. *Improving School Leadership*, 2, 40-41. <https://doi.org/10.1787/9789264039551-4-en>
- Evans, D., & Popova, A. (2015). *What really works to improve learning in developing countries? An analysis of divergent findings in systematic reviews*. (February 26, 2015). World Bank Policy Research Working Paper, (7203). Washington, DC: World Bank. <https://doi.org/10.1596/1813-9450-7203>
- Fleisch, B. (2006). Bureaucratic accountability in the education action zones of South Africa. *South African Journal of Education*, 26(3), 369-382.
- Fleisch, B. (2008). *Primary education in crisis: Why South African school children underachieve in reading and mathematics*. Johannesburg: Juta & Co. Ltd.
- *Fleisch, B., Taylor, N., Herholdt, R., & Sapire, I. (2011). Evaluation of the Back to Basics Mathematics Workbooks: A randomized Controlled Trial of the primary mathematics research project. *South African Journal of Education*, 31(4), 488-501. <https://doi.org/10.15700/saje.v31n4a466>
- *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), 01-12. <https://doi.org/10.15700/201409161040>

- *Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2015). *Assessing the impact of the RCUP: The findings of the impact evaluation of the Reading Catch-Up Programme*. Report, Johannesburg.
- 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*, 49, 157-174. <https://doi.org/10.1016/j.ijedudev.2016.02.006>
- Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2017). Failing to catch up in reading in the middle years: The findings of the impact evaluation of the Reading Catch-Up Programme in South Africa. *International Journal of Educational Development*, 53, 36-47. <https://doi.org/10.1016/j.ijedudev.2016.11.008>
- Fleisch, B. (2018). *The Education Triple Cocktail: System-wide Instructional Reform in South Africa*. UCT Press/Juta and Company (Pty) Ltd.
- Friedman, W., Kremer, M., Miguel, E., & Thornton, R. (2011). *Education as liberation?* (NBER Working Paper No. 16939). Cambridge, MA: NBER (National Bureau of Economic Research). <https://doi.org/10.3386/w16939>
- Glewwe, P., Kremer, M., Moulin, S., & Zitzewitz, E. (2004). Retrospective vs. prospective analyses of school inputs: The case of flip charts in Kenya. *Journal of Development Economics*, 74(1), 251-268. <https://doi.org/10.1016/j.jdeveco.2003.12.010>
- Glewwe, P.W., Hanushek, E.A., Humpage, S.D., & Ravina, R. (2014). School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010 in *Education Policy in Developing Countries*, Ed. P. Glewwe, Chicago and London: University of Chicago Press.
- Gustafsson, M. (2007). Using the hierarchical linear model to understand school production in South Africa. *South African Journal of Economics*, 75(1), 84-98. <https://doi.org/10.1111/j.1813-6982.2007.00107.x>
- Gustafsson, M., & Taylor, S. (2013). *Treating schools to a new administration. The impact of South Africa's 2005 provincial boundary changes on school performance*. Stellenbosch: Stellenbosch University.
- Hansen, L.O., Young, R.S., Hinami, K., Leung, A., & Williams, M.V. (2011). Interventions to reduce 30-day rehospitalization: a systematic review. *Annals of Internal Medicine*, 155(8), 520-528. <https://doi.org/10.7326/0003-4819-155-8-201110180-00008>
- Hanushek, E., & Woessmann, L. (2012). Do better schools lead to more growth? Cognitive skills, economic outcomes, and causation. *Journal of Economic Growth*, 17(4), 267-321. <https://doi.org/10.1007/s10887-012-9081-x>

- Harris, D. (2009). Toward policy-relevant benchmarks for interpreting effect sizes combining effects with costs. *Educational Evaluation and Policy Analysis*, 31(1), 3-29. <https://doi.org/10.3102/0162373708327524>
- *Harvey, J., & Prinsloo, C. (2015). *Hope for rebuilding language foundations: part 2 – replication in a second cohort*. Evaluation Report. Pretoria: Human Science Research Council.
- Hattie, J. (2009). *Visible learning: A synthesis of 800+ meta-analyses on achievement*. Abingdon: Routledge.
- Hattingh, A. (2003). Examining learning achievement and experiences of science learners in a problem-based learning environment. *South African Journal of Education*, 23(1), 52.
- Hobden, P., & Hobden, S. (2009). *RUMEP Project Summative Evaluation*. Johannesburg: Zenex Foundation.
- Hombrados, J.G., & Waddington, H. (2012). *Internal validity in social experiments and quasi-experiments: An assessment tool for reviewers*. Mimeo, London: International Initiative for Impact Evaluation (3ie).
- Hunt, K., Schoer, V., Ntuli, M., Rankin, N., & Sebastiao, C. (2011). A blurred signal? Comparability of NSC mathematics scores and former SC mathematics scores: how consistent is the signal across time? *Education as Change*, 15(1), 3-16. <https://doi.org/10.1080/16823206.2011.574097>
- Joint Education Trust (JET, 2006). *READ mother tongue literacy programme*. Evaluation Report. Johannesburg: Zenex Foundation.
- Joint Education Trust (JET, 2007). *Towards optimal usage of mindset resources in South African Schools: Five Case Studies in Western Cape and Gauteng*. Johannesburg: JET Education Services.
- Kanjee A. (2007). *Large-scale assessments in South Africa: supporting formative assessments in schools*. Paper presented at the 33rd International Association for Educational Assessment Conference, Baku, Azerbaijan, 16 to 21 September.
- Kim, Y.S.G., Lee, H., & Zuilkowski, S.S. (2019). Impact of Literacy Interventions on Reading Skills in Low-and Middle-Income Countries: A Meta-Analysis. *Child Development*. <https://doi.org/10.1111/cdev.13204>
- Kotze, J., Fleisch, B., & Taylor, S. (2018). Alternative forms of early grade instructional coaching: Emerging evidence from field experiments in South Africa. *International Journal of Educational Development*. <https://doi.org/10.1016/j.ijedudev.2018.09.004>
- Kremer, M., Brannen, C., & Glennerster, R. (2013). The challenge of education and learning in the developing world. *Science*, 340(6130), 297-300. <https://doi.org/10.1126/science.1235350>

- Krishnaratne, S., White, H., & Carpenter, E. (2013). *Quality education for all children? What works in education in developing countries?* Working Paper (20). New Delhi: International Initiative for Impact Evaluation (3ie) <https://doi.org/10.23846/WP0020>
- Lipsey, M.W., & Wilson, D.B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications, Inc.
- Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., & Busick, M.D. (2012). *Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms*. (No. ED-IES-09-C-0021). Washington DC.
- Louw, J., Muller, J., & Tredoux, M. (2008). Time-on-task, technology and mathematics achievement, *Evaluation and Program Planning*, 31(1), 41-50. <https://doi.org/10.1016/j.evalprogplan.2007.11.001>
- Lu, G., & Ades, A.E. (2006). Assessing evidence inconsistency in mixed treatment comparisons. *Journal of the American Statistical Association*, 101 (474), 447-459. <https://doi.org/10.1198/016214505000001302>
- Marzano, R.J. (2003). *What works in schools: Translating research into action*. ASCD.
- McEwan, P.J. (2015). Improving learning in primary schools of developing countries: a meta-analysis of randomized experiments. *Review of Educational Research*, 85(3), 353-394. <https://doi.org/10.3102/0034654314553127>
- Motala, S., Morrow, S., & Sayed, Y. (2014). Gauteng Department of Education: A policy review In F. Maringe & M. Prew (Eds.), *Twenty years of education transformation in Gauteng 1994 to 2014: an independent review* (Chapter 2). Johannesburg: Gauteng Department of Education.
- Mouton, J., Wildschut, L., Richter, T., & Pocock, R. (2013). *Zenex Review Project* (Final Revised Report). Johannesburg: Evaluation Research Agency (ERA).
- National Planning Commission (NPC, 2012). *National development plan vision 2030*. Pretoria: Presidency.
- National Planning Commission (NPC, 2011). *Diagnostic overview*. Pretoria: Presidency.
- Organisation for Economic Co-operation and Development (OECD, 2008). *Review of national policies for education: South Africa*. Paris: OECD.
- Padayachee, P., Boshoff, H., Olivier, W., & Harding, A. (2011). A blended learning Grade 12 intervention using DVD technology to enhance the teaching and learning of mathematics: original research. *Pythagoras*, 32(1), 1-8. <https://doi.org/10.4102/pythagoras.v32i1.24>

- Piper, B. (2009). *Integrated education program: Impact study of SMRS using early grade reading assessment in three provinces in South Africa*. Park, NC: RTI International, Research Triangle.
- Prinsloo, C., & Kanjee, A. (2005). *Improving learning in South African schools: the Quality Learning Project (QLP) summative evaluation (2000-2004)*. Pretoria: Human Sciences Research Council.
- Prinsloo, C.H. (2009). *Extra classes, extra marks? Report on the Plus Time Project*, Pretoria: Human Science Research Council.
- Research Institute for Education Planning (RIEP) (2008). *Report on the impact study in the Eastern Cape and KwaZulu-Natal on literacy and numeracy grade 4*. Bloemfontein: University of Free State.
- Roberts, J. (2006). Evaluation of the numeracy programme operated by COUNT in Uitenhage schools. Final Report. Johannesburg: Zennex Foundation.
- Roberts, J., & Schollar, E. (2006). Meta-evaluation of interventions in math, science and language 1998-2006. Johannesburg: ESA.
- Roberts, J., & Schollar, E. (2011). Meta-Evaluation of programmes and projects supported by the Zenex Foundation 2006-2011. Johannesburg: ESA.
- Roche, C., & Kelly, L. (2005). Evaluating the performance of development agencies. In G. Pitman, O. Feinstein, & G. Ingram, Vol. 7. *Evaluating development effectiveness. World Bank Series on Evaluation and Development*. Edison, NJ: Transaction Publishers.
- *Sailors, M., Hoffman, J.V., Pearson, P.D., Beretvas, S.N., & Matthee, B. (2010). The effects of first- and second-language instruction in rural South African schools. *Bilingual Research Journal*, 33(1), 21-41. <https://doi.org/10.1080/15235881003733241>
- Salanti, G. (2012). Indirect and mixed-treatment comparison, network, or multiple-treatments meta-analysis: many names, many benefits, many concerns for the next generation evidence synthesis tool. *Research Synthesis Methods*, 3(2), 80-97. <https://doi.org/10.1002/jrsm.1037>
- Sayed, Y., Kanjee, A., & Nkomo, M. (2013). *The search for quality education in post-apartheid South Africa: Interventions to improve learning and teaching*. Cape Town: HSRC Press.
- Schollar, E. (1999). *Final Report of the Evaluation of the Impact of the INSET Component of the Mpumalanga Primary Schools Initiative (MPSI)*. Pretoria: DFID.
- Schollar, E. (2001). A Review of Two Evaluations of the application of The READ Primary Schools Programme in the Eastern Cape province of South Africa. *The International Journal of Educational Research (IJER)*, 35(2). [https://doi.org/10.1016/S0883-0355\(01\)00017-9](https://doi.org/10.1016/S0883-0355(01)00017-9)

- Schollar, E. (2002). *The Evaluation of the Pilot EQUIP project of the National Business Initiative* (Final Report). Johannesburg: ESA.
- Schollar, E. (2005). *The evaluation of the Learning for Living Project*. Johannesburg: Eric Schollar and Associates.
- Schollar, E. (2008). *The Primary Mathematics Research Project 2004–2007: towards evidence-based educational development in South Africa* (Final Report). Johannesburg: Eric Schollar & Associates.
- Schollar, E. (2013). The Learning for Living Project: a book-based approach to learning of language in South African primary schools. In Y. Sayed, A. Kanjee, & M. Nkomo (Eds.), *The search for quality education in post-apartheid South Africa* (pp. 237–256). Cape Town, South Africa: HSRC Press.
- Schollar, E. (2015). *The Primary Mathematics Research Project: 2004–2012. An evidence based programme of research into understanding and improving the outcomes of mathematical education in South African primary schools*. PhD Thesis. Cape Town: University of Cape Town.
- Schollar, E., & Mouton, J. (2014). *Epoch and Optima: Public Schools Math Challenge Programme. An impact evaluation* (Final Report). Johannesburg: ESA & ERA.
- Simkins, C. (2010). *The maths and science performance of South Africa's public schools: some lessons from the past decade* (Executive Summary No. 1). Johannesburg: Centre for Development Enterprise.
- Simkins, C., & Paterson, A. (2005). *Learner performance in South Africa: social and economic determinants of success in language and mathematics*. Pretoria: HSRC Press.
- Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500–506. <https://doi.org/10.3102/0162373709352369>
- Snilstveit, B., Stevenson, J., Menon, R., Phillips, D., Gallagher, E., Geleen, M., ... & Jimenez, E. (2016). The impact of education programmes on learning and school participation in low- and middle-income countries. <https://doi.org/10.23846/SRS007>
- South African Institute for Distance Education (SAIDE, 2007). *Evaluation of COUNT Family Math School Pilot Project*. Main Report. Johannesburg: SAIDE.
- South African Schools Act (SASA), 1996 (Act No. 84 of 1996). Retrieved from <http://www.acts.co.za/south-african-schools-act-1996/index.html>
- Spaull, N. (2011). *A preliminary analysis of SACMEQ III South Africa* (Stellenbosch Economic Working Papers 11/11). Stellenbosch: University of Stellenbosch.

- Spaull, N. (2013). *South Africa's education crisis: the quality of education in South Africa 1994-2011* (CDE Report). Johannesburg: Centre for Development and Enterprise.
- *Summers, H.C. (2008). *Maths Centre/Zenex Foundation Project: upgrading of Teacher Qualifications at Foundation Phase KwaZulu-Natal and the Eastern Cape (2005-2008)*. Summative Evaluation Report. Johannesburg: Zenex Foundation.
- Taylor, N., Fleisch, B., & Shindler, J. (2008). *Changes in education since 1994*. Pretoria: South Africa Presidency.
- Taylor, N. (2009). Standards-based accountability in South Africa. *School Effectiveness and School Improvement*, 20(3), 341-356. <https://doi.org/10.1080/09243450902916704>
- Taylor N., (2011) *A five-year plan for South African schooling: a report commissioned by the National Planning Commission*. Johannesburg: JET Education Services.
- Taylor, N., Muller, J., & Vinjevoold, P. (2003). *Getting schools working*. Cape Town: Pearson Education South Africa.
- Taylor, S., & Watson, P. (2015). *The impact of study guides on matric performance: evidence from a randomised experiment*. Stellenbosch Economic Working Paper. University of Stellenbosch.
- Taylor, S., Fleisch, B., Cilliers, J., Prinsloo, C., & Reddy, V. (2017). *Early Grade Reading Study Summary Report: Results of Year 2 Impact Evaluation*. Pretoria: Department of Basic Education.
- Trialogue. (2015). *The Trialogue 2015 CSI handbook*. Cape Town: Trialogue.
- Van der Berg, S. (2008). How effective are poor schools? Poverty and educational outcomes in South Africa. *Studies in Educational Evaluation*, 34(3), 145-154. <https://doi.org/10.1016/j.stueduc.2008.07.005>
- Van Stuijvenberg, M.E., Kvalsvig, J.D., Faber, M., Kruger, M., Kenoyer, D.G., & Benadé, A.S. (1999). Effect of iron-, iodine-, and β -carotene-fortified biscuits on the micronutrient status of primary school children: a randomized controlled trial. *The American Journal of Clinical Nutrition*, 69(3), 497-503. <https://doi.org/10.1093/ajcn/69.3.497>
- Wilson, S.J., Tanner-Smith, E.E., Lipsey, M.W., Steinka-Fry, & Morrison, J. (2011). Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Systematic Reviews* 2011 (8). <https://doi.org/10.4073/csr.2011.8>

Systematic Reviews of Research in Basic Education in South Africa

Analysis of heterogeneity by evaluation design parameters

Group	Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity		
	Num. Studies	Point estimate	Stand. error	Lower limit	Upper limit	Z-value	P-value	Q-value	df Q	P-value
Counterfactual Method										
PSM	20	0,00	0,06	-0,10	0,11	0,08	0,94			
RCT	35	0,11	0,03	0,05	0,16	3,60	0,00			
RDD	2	0,37	0,30	-0,21	0,95	1,26	0,21			
Simple match	53	0,28	0,03	0,22	0,34	9,60	0,00			
Total between								28,40	3	0,000
Overall	110	0,17	0,02	0,13	0,21	8,85	0,00			
Assessment Instrument										
Own	71	0,22	0,03	0,17	0,27	8,15	0,00			
State	39	0,12	0,04	0,04	0,20	2,96	0,00			
Total between								4,38	1	0,036
Overall	110	0,19	0,02	0,14	0,23	8,42	0,00			
Units of Analysis										
Learners	84	0,21	0,02	0,16	0,26	8,59	0,00			
Schools	26	0,10	0,06	-0,02	0,21	1,63	0,10			
Total between								2,95	1	0,086
Overall	110	0,19	0,02	0,15	0,24	8,58	0,00			
Matched Cohorts										
No	50	0,12	0,03	0,06	0,19	3,66	0,00			
Yes	60	0,23	0,03	0,17	0,29	7,66	0,00			
Total between								6,16	1	0,013
Overall	110	0,18	0,02	0,14	0,23	8,12	0,00			
Statistical significance										
Yes	55	0,34	0,03	0,28	0,40	11,17	0,00			
No	55	0,00	0,01	-0,02	0,02	0,02	0,98			
Total between								109,7	1	0,000
Overall	110	0,04	0,01	0,02	0,06	3,89	0,00			

Chapter 2

Analysis of heterogeneity by school phase

Group	Effect size and 95% confidence interval					Test of null (2-Tail)		Heterogeneity		
	Num. Studies	Point estimate	Stand. error	Lower limit	Upper limit	Z-value	P-value	Q-value	df Q	P-value
Prim - 1 Foun.	9	0,53	0,08	0,36	0,69	6,34	0,00			
Prim - 2 Inter.	27	0,36	0,04	0,29	0,43	10,19	0,00			
Sec - 3 Senior	7	0,34	0,07	0,21	0,47	5,10	0,00			
Sec - 4 FET	12	0,14	0,07	0,00	0,28	2,02	0,04			
Total between								12,84	3	0,00
Overall	55	0,34	0,03	0,29	0,40	12,70	0,00			