



Improving Grade R mathematics teaching in South Africa: Evidence from an impact evaluation of a province-wide intervention

Eleanor Hazell

Executive Manager, Monitoring and Evaluation, JET Education Services, Cape Town, South Africa
ehazell@jet.org.za
<https://orcid.org/0000-0003-0243-0082>

Garth Spencer-Smith

Centre for Education Practice Research, Faculty of Education, University of Johannesburg, Johannesburg, South Africa
garth@kelello.org
<https://orcid.org/0000-0002-8950-1500>

Nicky Roberts

Centre for Education Practice Research, Faculty of Education, University of Johannesburg, Johannesburg, South Africa
nroberts@uj.ac.za
<https://orcid.org/0000-0002-1910-0162>

(Received: 17 February 2019; accepted: 24 April 2019)

Abstract

We present the impact on learner outcomes of a province-wide Grade R mathematics intervention (termed R-Maths) in relation to theoretical frameworks established from a meta-evaluation of evaluations of education interventions in South Africa and a review of other meta-evaluation and synthesis studies. We compare the changes in Mathematics performance from baseline to end line, of learners in the intervention group (taught by R-Maths-trained teachers/practitioners) to the comparison group (learners in schools in the same districts, but whose teachers/practitioners had not yet received the R-Maths intervention). The intervention group performed 2.9 percentage points better than the comparison group over the whole Marko-D test of mathematical competencies, with a small effect size. The greatest effects on performance were from language of learning and teaching, and district. The R-Maths case indicates that a modified cascade model which includes some elements of Fleisch's "educational triple cocktail" (structured learning materials, teacher training, and support) may be successful by working with, and through, department of education structures. Whether the effects are retained over time and if these effects can be replicated in different contexts is not yet known.

Keywords: mathematics, early grade, at scale, Marko-D, Grade R, evaluation

Introduction

In a context of poor education outcomes, substantial spending on education, and limited documented evidence regarding what works, there is a pressing need for evaluation research to be subjected to academic scrutiny and published in the public domain. In this paper, we contribute to addressing this gap.

We report on the findings of a quasi-experimental impact evaluation of a province-wide early grade mathematics intervention (termed R-Maths). The R-Maths project was implemented across the Western Cape province in South Africa in Grade R and led by the Western Cape Education Department (WCED). The intervention aimed to strengthen the teaching and learning of mathematics in Grade R and, ultimately, to improve the conceptual understanding and mathematical skills of Grade R learners, such that they would enjoy mathematics and be (academically) successful in the Foundation Phase. The project targeted Grade R and a specific learning area (mathematics). Its impact on learner achievement is therefore pertinent and we discuss this later in this article.

R-Maths made use of elements of the “educational triple cocktail” to which Fleisch (2018) refers in the title of his book, in that it includes teacher training, learning, teaching and support materials (LTSM), and follow-up support. However, its implementation differed in that rather than training and supporting teachers/practitioners¹ directly, subject advisors were trained and supported and they, in turn, trained and supported the teachers/practitioners. R-Maths therefore offers another implementation model for a large-scale (province-wide) intervention in mathematics delivered via existing department of education structures.

The findings of the R-Maths evaluation are situated in context by our presenting them in relation to the headline findings of review and synthesis studies of evaluations of education interventions that are relevant to the South African context and were recently reviewed by Hazell (2019).

Theoretical orientation

There are two aspects that inform our theoretical orientation to this paper. First, we draw on education and, specifically, school intervention literature to present various frameworks for describing interventions and we use one to describe R-Maths. Second, we draw on impact evaluations of education interventions that are primarily experimental or quasi-experimental in design and that focus specifically on impact on learner outcomes. We use these to identify promising levers for change and indicate which were included or excluded from the R-Maths project.

¹ Teachers/practitioners is the terminology used for consistency purposes to refer to the individuals teaching Grade R learners. The majority of the individuals teaching Grade R in the Western Cape (and even more so in most other provinces in South Africa) are practitioners with various levels of early childhood development (ECD) training; very few are qualified educators.

Frameworks for describing school interventions

Interventions that aim to improve learner education outcomes can be categorised in a variety of ways. The level of the intervention, primary target group, intervention type, and conceptual/theoretical basis for the intervention can all be considered. Typologies include that developed by Snilstveit et al., (2015) who developed a typology of school outcome interventions based on the level of intervention and focusing on the primary beneficiary of the inputs:

- Child level: school feeding, school-based health, merit-based scholarships, providing information (about education) to children;
- Household level: eliminating user-fees, cash-transfers, scholarships and allowances, providing information (about education) to parents;
- School level: structured pedagogy, computer assisted learning, remedial education, grouping students by ability, providing materials, new schools and infrastructure;
- Teacher level: teacher hiring incentives, teacher performance incentives, teacher training, diagnostic feedback (providing teachers with information about learners);
- System-level interventions: school-based management, community-based monitoring, private-public partnerships and private provision of schooling; and
- Multi-level interventions: with interventions that may include any combination of those outlined above.

In South Africa, Besharati and Tsotsotso (2015) conducted a systematic review and meta-analysis of interventions implemented since 1994 that aim to improve learner performance outcomes. They identified the following broad types of interventions:

- Learner-targeted support;
- Teacher-centred initiatives;
- Provision of LTSM;
- Management and governance;
- Infrastructure and facilities;
- Structural reforms, policies and incentives;
- Community/family involvement; and
- Whole school development.

Their classification mixes the levels of the intervention with the type of support (or what is done for the primary beneficiaries).

Taking a very broad view of types of school interventions, McEwan (2015), who conducted a meta-analysis of randomised control trials of primary school interventions in low- and middle-income countries, identified three broad types:

- Instructional: information and communications technology, teacher training, class size or composition, instructional materials, and grants.

- Health and/or nutrition-based: food, beverage and/or micronutrients, and deworming.
- Incentive-based: contract or volunteer teachers, student and/or teacher incentives, school management or supervision, and informational.

This type of classification focuses on what is being done, rather than on which levels of the system or which mechanisms are used.

In reflecting on how to describe and classify types of school interventions, Mouton, Wildschut, Richter, & Pocock (2013) distinguished between interventions targeting different levels, stages, or phases of schooling, learning areas (school subjects) and non-learning areas (governance, school leadership and management, and curriculum management), and intervention types.

Homing in on Sub-Saharan Africa, Conn (2017) conducted a meta-analysis of randomised control trials and quasi-experimental impact evaluations, and identified five broad types of interventions:

- Quality of instruction: class size and composition, instructional time, pedagogical interventions (including technology-assisted learning) and school supplies;
- School or community financial limitations: cash transfers and infrastructure;
- School or system accountability: information provision and management interventions;
- Cognitive processing: school meals and health treatments;
- Motivation: student incentives and teacher incentives; and
- School type: particular types of schools.

Here we see a greater focus on levers for change or, in other words, how change occurs.

The above are illustrative examples of the plethora of ways in which school interventions are categorised. It appears that most meta-evaluators create their own typologies for classification and that there is little agreement in the field on a systematic way to describe interventions.

We found the heuristic and decision framework developed by Mouton et al. (2013) to be helpful in distinguishing different components of the descriptive typology of interventions. This conceptualises school interventions according to:

- Component 1: the target group (which brings in consideration for level of the system; target phases at the school, the domains of learning and the scale in terms of the population of target schools);
- Component 2: intervention type or mode (which identifies particular levers for change); and
- Component 3: implementation theory (which includes consideration for how the intervention will work, location, duration, dosage, cost and so on).

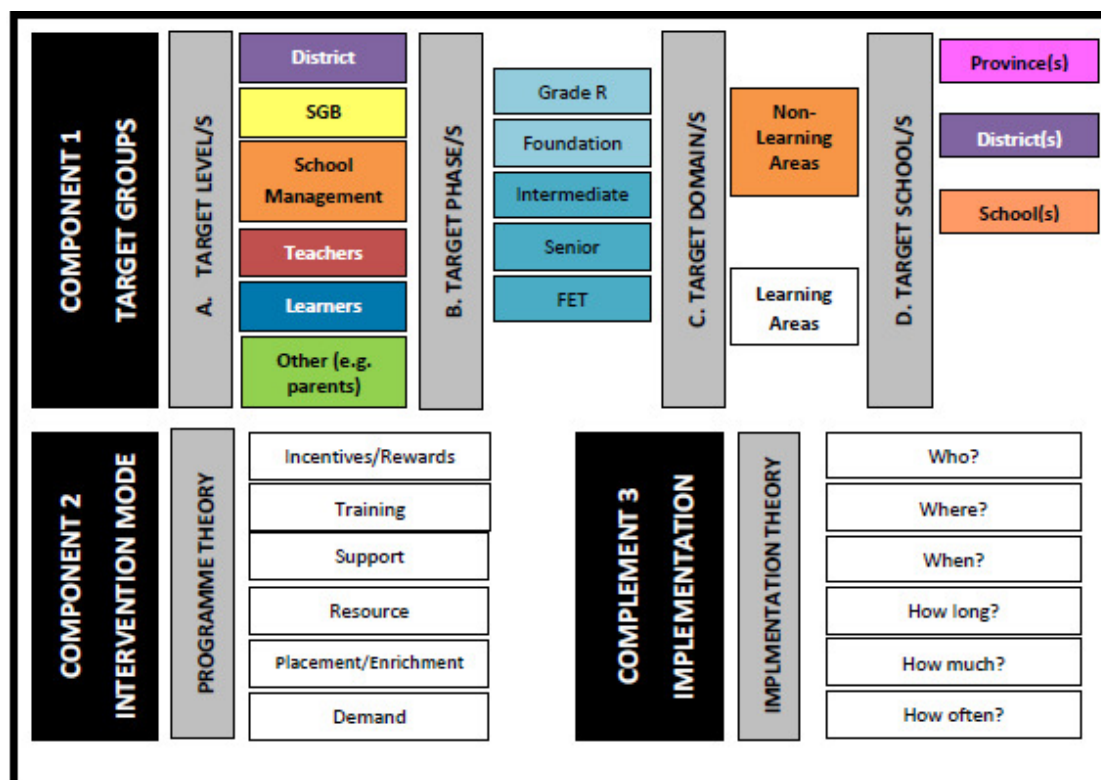


Figure 1: Intervention design framework (Source: Mouton et al., 2013)

Guidance on how to classify and describe each sub-component is, however, absent. This is understandable given that it was intended as an intervention design framework to guide decision-making. Nonetheless, we argue that this framework provides a coherent and detailed way of describing interventions and can therefore be used as a typology for intervention type. This framework has also been applied elsewhere in relation to mathematics interventions in South African schools (Roberts, Mostert, & Takane, 2015). We therefore apply this to our description of the R-Maths intervention.

Successful interventions and promising levers for change identified from a scan of other studies

In this section we present examples, drawn from the literature, that made use of distinct typologies for classifying types of education interventions. First, we provide findings from meta-analyses on what types of education interventions were the most effective in improving learner performance. Second, we provide some findings from specific South African education interventions.

In our scan of the literature, we were struck by the lack of detail provided in most meta-analyses; their application of typologies varied and focused usually only on one (or, at the most, two) components of the Mouton et al. (2013) intervention design framework.

Across a number of recent meta-review and synthesis studies, interventions that target teachers and aim to enhance the quality of instruction, via the introduction of specific teaching methods and/or capacity building, alongside the provision of LTSM, are identified

as promising. For example, Snilstveit et al. (2015) found that the largest and most consistently positive effects in terms of learner performance outcomes were for teacher-level structured pedagogy interventions. These are described as interventions aimed at improving the content and quality of teaching by introducing new and improved content and teaching methods (for example, the provision of lesson plans and training and support for teachers to use them). Structured pedagogy interventions were found to have an average effect² of 0.23 on language test scores and 0.14 on mathematics test scores.

Similarly, McEwan (2015) found that teacher training interventions had the second largest mean effect size (0.12) while Conn (2017) found that interventions that aim to improve teaching quality were most successful: pedagogical and instructional time interventions had the greatest average effect size (0.918 and 0.464 respectively). Further analysis was conducted by Conn (2017) to investigate *what types* of pedagogical interventions were most effective. Evidence was found that (both teacher and computer-led) interventions that assess and adapt to a learner's level were particularly effective; teacher-training interventions that included mentoring and/or in-school coaching had consistently positive effects; and the provision of materials in local languages featured commonly in successful interventions. Remedial education (average effect of 0.19 on mathematics test scores and 0.16 on language test scores) was also identified as a promising approach by Snilstveit et al. (2015). The successful interventions described above have some commonalities with the promising levers for change identified by Hazell (2019) in the South African context.

Other types of interventions that the reviews and synthesis studies have identified as promising with respect to improving education outcomes are quite diverse and include for McEwan (2015), interventions that use computers or technology (average effect of 0.15); class size and/or composition (0.12); contract or volunteer teachers (0.10); student and/or teacher performance incentives (0.09); and LTSM (0.08). For Snilstveit et al. (2015) these include extra time in school (average effect of 0.09 on mathematics test scores and 0.19 on language test scores), school feeding (average effect of 0.10 on mathematics test scores and 0.09 on language test scores), and merit-based scholarships (average effect of 0.11 on mathematics test scores and 0.04 on language test scores).

In a unique study, Besharati and Tsotsotso (2015) investigated the influence of target phase and found that interventions implemented in lower grades and phases of the South African schooling system have a greater effect on learner performance. Further, they investigated differences in terms of intervention scale and found that interventions designed and implemented by academics/researchers for the purposes of research and piloting had greater effects than interventions implemented at scale by the private sector and government. However, they suggested that it may not be the interventions per se that are better or worse, but, rather, that it is more challenging to attain a large effect on learner performance when implementing at scale.

² Average standardised mean difference.

A challenge in comparing the findings across studies is that researchers, for example, Snilstveit et al. (2015) and Conn (2017) used different typologies for classifying interventions, thus making it difficult to compare like with like, and few considered the level of granular detail like target phase/grades and intervention scale that the more detailed South African study by Besharati and Tsotsotso (2015) did. Nevertheless, looking across these studies one can find a common thread in that typically the most effective interventions focus on teachers in their classrooms with appropriate pedagogical mentoring/support, and LTSM which are appropriate and tailored to the cognitive level of learners. Other interventions were found to be effective by one or two authors, like for example, computers or technology, LTSM (only), incentives and changing class size or composition, but not across contexts.

A recent meta-evaluation conducted by Hazell (2019) reveals that there have been promising interventions implemented on a variety of scales³ over the past five years. She found that the two types of interventions that have been evaluated rigorously and that have demonstrated promising results are 1) ones that offer Fleisch's (2018) "education triple cocktail" of LTSM, lesson plans, and individual coaching (Fleisch, 2018) and 2) those that commence with diagnostic testing and target LTSM and teaching to learners' current ability level.

An example of the former type is the Gauteng Primary Literacy and Mathematics Strategy (GPLMS), which was undertaken in 1,040 under-performing schools in Gauteng province, South Africa, targeting Grade 1 to 7 literacy/language and numeracy/maths teachers, who were provided with just-in-time training, lessons plans and other LTSM, and individual coaching. Phase 1 of the GPLMS was implemented from 2010–2014. The programme was evaluated via a quasi-experimental study that exploited a so-called natural experiment that occurred when some under-performing schools, which should have received the intervention, were left out mistakenly. Schools that received the intervention recorded improved learner performance in early grade maths test scores as compared to similar schools which had not benefitted from the programme. Significant differences were found between treatment and comparison schools Grade 1 and 3 mathematics Annual National Assessment (ANA) test scores after one year, with an effect of between 0.35-0.61 standard deviations (sd), increasing after two years. Weakly statistically significant differences were found for Grade 2 and non-statistically significant differences were found for Grade 4. A cohort analysis following a sample of learners from Grades 1 to 3 found a difference (effect size) of 0.7 after two years, and following a cohort of learners from Grades 2 to 4 showed an effect size of 0.5 sd after one year, which disappeared by Grade 4 (Fleisch, Schöer, Roberts, & Thornton, 2016).

The Department of Basic Education followed these encouraging results, by conducting randomised control trials with multiple treatment arms to test the efficacy of different variations of this model including: 1) the relative efficacy of training with LTSM or training with LTSM and individual coaching in improving literacy in home language in the Foundation Phase; and 2) the relative efficacy and cost-effectiveness of providing in-person or virtual coaching in improving English literacy in the Foundation Phase. In the first study, the groups were equivalent at baseline and, after two years, learners whose teachers received

³ From small-scale, (i.e. in 5-18 schools), to programmes rolled out at a provincial scale.

LTSM, training, and coaching were 0.252 sd ahead of the control group, as compared to learners whose teachers received LTSM and training only who were 0.12 sd ahead of the control group (Kotzé, Fleisch, & Taylor, 2018).

The second type of promising interventions to which Hazell (2019) referred were implemented at smaller scale and have not been assessed as rigorously, but also indicate promising results in terms of helping to address learning gaps. The Primary Maths Research Project, a research-based pedagogical intervention developed and evaluated by Schollar (2015), was found to be successful in improving learner performance in a custom-designed mathematics test, with Grade 4 learners whose teachers received training in the project methodology and LTSM experiencing significant gains (as compared to a control group) after completing the 14-week intervention. A follow-up study conducted three years later found that the learning gains had been sustained. The Learner Regeneration Project trained teachers and community facilitators to support learning and provided LTSM at the appropriate level. Substantial improvements were found in literacy and language performance in the Foundation and Intermediate phases. Both these interventions commenced with diagnostic testing and the provision of LTSM and teaching tailored to their level (Prinsloo, Harvey, Thaba, & Moodley, 2015). The evaluation studies that reported on the second group of promising interventions were reports, not published papers, and did not report effect size in the same way as did the first group of studies.

In conclusion, the effect size of interventions identified as promising in international synthesis studies is often quite small, upwards of around 0.1 sd. Interventions identified as promising in the South African context have slightly larger effect sizes (upwards of around 0.2 sd) and the effect is often measured after a period of two years. In some instances, effects found after a period of one year were found to have tapered off a year later.

R-Maths project description

Using the Mouton et al. (2013) intervention design framework, we describe the R-Maths intervention in terms of component 1 (the target group), component 2 (the intervention type or mode) and component 3 (implementation theory).

Component 1: The R-Maths target group

The Grade R Early Mathematics (R-Maths) project was an initiative led by the WCED. Mathematics content development, training and support was provided by the Schools Development Unit of the University of Cape Town (UCT). The R-Maths project worked with and targeted Subject Advisors directly, who, in turn, provided support to Grade R teachers/practitioners while learners were the group that the project aimed to benefit ultimately.

The overarching goal identified for the project was to improve the conceptual understanding and mathematical skills of Grade R learners in Western Cape, such that they would enjoy mathematics and be (academically) successful in the Foundation Phase. It targeted Grade R

(which is now considered part of the Foundation Phase) and a specific learning area—mathematics).

Component 2: R-Maths intervention type or mode

The project fits a number of the intervention modes identified by Mouton et al. (2013): appropriate LTSM developed specifically for the project (in the form of facilitators' guides and participants' materials for the training and cluster workshops, a Mathematics concept guide, termly guides that provide a curriculum framework, lesson ideas, and teaching aids); training in early grade mathematics content and teaching methodologies, and follow-up support were provided to Foundation Phase Subject Advisors. This included what were called "principles of teaching" that promote learning opportunities for mathematics in Grade R settings, appropriate practical classroom methodology that encourages thinking and reasoning, as well as play-based teaching that demonstrates how to teach the mathematics concepts, all within, and mindful of, the South African context. The materials were translated and made available in English, Afrikaans, and isiXhosa. In turn, the Subject Advisors provided LTSM, workshops, training and (limited) follow-up support to Grade R teachers/practitioners in early grade mathematics content and teaching methodologies. Subject Advisors' interaction with teachers/practitioners was mainly through cluster-based workshops and training sessions, and via WhatsApp chat groups but not individually nor at their schools.

Although not the focus of this paper, it should be noted that the nature of the learning materials differed from the daily structured lessons offered by the GPLMS and subsequent studies conducted by the Department of Basic Education. The R-Maths materials included structured support on how to train and support teachers/practitioners as well as a weekly rhythm of classroom activities (which, while addressing the core mathematical ideas in the national Curriculum and Assessment Policy Statement (CAPS), did not make use of the suggested detailed learning programme offered in CAPS). So, the materials were not daily lesson plans, but offered a core teaching focus for each week, and a related set of workstations for group activity that were then repeated over each five-day cycle.

The key levers for change that the programme was expected to operate through were: Subject Advisors would be upskilled in content and pedagogy and resourced to train and support Grade R teachers/practitioners; and teachers/practitioners would be upskilled in content and methodology to teach Grade R mathematics and resourced to deliver interactive, interesting, age and grade appropriate mathematics lessons.

Component 3: R-Maths implementation theory

Implementation for this province-wide initiative was through a two-phase, modified cascade model. The first level of training conducted by the UCT Schools Development Unit was of all Foundation Phase Subject Advisors across the Western Cape Province. The second level was the training conducted by the Foundation Phase Subject Advisors of all Grade R

teachers/practitioners across the province. The second level training was carried out in two phases in 2017 and 2018; each phase targeted roughly half the Grade R teachers/practitioners.

The R-Maths Project was delivered by the UCT Schools Development Unit to Foundation Phase Subject Advisors who received initial training over a period of five days (28 hours). The intervention was not a pure cascade model in that the Schools Development Unit team continued to provide support to the Subject Advisors in all districts after the initial training. This support took various forms, but, most notably, included helping the Subject Advisors prepare for training the teachers/practitioners through a two- to three-hour dry run held before each cluster workshop and a four-day dry run prior to the block training that Subject Advisors facilitated for teachers/practitioners. Subject Advisors received up to 42 hours of additional support following the initial training.

Grade R teachers/practitioners received seven two-hour cluster workshops (14 hours), and a five-day (28-hour) block training; these were delivered by Subject Advisors in the language of teaching and learning that the teachers/practitioners used in the classroom. The cluster training sessions served as regular mentoring and support for the pedagogic use of the LTSM. Additional support was provided via WhatsApp chat groups and some teachers/practitioners received school-based support visits.⁴ Subsequent to the training, the teachers/practitioners participated in one two-hour reflection meeting and one two-hour Professional Learning Community meeting focused on R-Maths. The trained teachers/practitioners were then meant to implement the R-Maths content, concepts, and pedagogical ideas in their classrooms.

Methodology

The evaluation had an outcome and impact evaluation component that determined the extent to which the expected/intended short-, medium-, and long-term outcomes (outlined in the programme theory) that were expected to occur at the level of Foundation Phase Subject Advisors, teachers/practitioners, teaching and learning, and learners did occur. The focus of this paper is the impact on learner outcomes. To that end, we focus on one of the evaluation questions: Did the R-Maths Project have an impact on Grade R learners' Mathematics knowledge and skills (as measured using the Marko-D assessment)? Changes in Grade R learners' mathematical knowledge and skills were assessed via a quasi-experimental, difference-in-difference design (Gertler, Martinez, Premand, Rawlings, & Vermeersch, 2010).

The evaluation focussed on collecting data in two of the eight WCED districts. The two chosen districts—one of the four urban districts, and one of the four more-rural districts—were purposively sampled to ensure that the schools whose teachers/practitioners who were to be trained in phase 1 and phase 2 were as similar as possible in relation to certain key criteria which may influence learner performance in Grade R, such as learner performance at

⁴ The ratio of Grade R teacher/practitioners (and schools) to Subject Advisor is very high and it is a challenge for Subject Advisor to get to all of the schools/teachers/practitioners they are responsible for supporting to provide individual support.

Grade 3 level in systemic mathematics assessments, language of learning and teaching, school quintile, and enrolment.

Learners attending schools participating in Phase 1 (2017) of the project comprised the intervention group, and learners attending schools participating in Phase 2 (2018) made up the comparison group.

A simple random sample of all Grade R learners in the two case study districts was made. This sampling method was used rather than cluster sampling because of evaluation cost constraints. The final sample was 168 learners attending schools participating in Phase 1 of the Project and 168 learners in schools to be included in Phase 2, in each of the two chosen districts (thus, 672 learners in total). These learners were tested at baseline (February/March 2017) and 8 months later at end line (October/November 2017).

The instrument used to assess the learners was the 47-item demonstration (demo) version of the Marko-D test of mathematical competencies.⁵

The Marko-D is an individual oral test of early number concept development normed for children at the beginning of Grade 1, but which can be “administered to younger pre-school or Grade R children and older children (second graders)” (Henning et al., 2019, p. 12). It is an oral one-on-one test administered in the language of instruction of the school. The choice of the Marko-D was because it is the only Mathematics test for this age group that has been recently developed and which has been validated with South African learners in Sesotho, English, Afrikaans, and isiZulu (Henning et al., 2019). One of the ways in which the Marko-D can be employed is to “assess the effects of an intervention, by administering the test before and after the intervention (Henning et al., 2019).

The Marko-D test was administered by 17 trained test administrators (who had engaged with the underlying theoretical model of the test, and the demonstration video, as advised by Henning et al. (2019), overseen by one assessment team leader in each district who observed and gave feedback to each test administrator after they administered a test to a learner.

The test was marked by the test administrators directly since they completed the answer sheet during the test so there was no need for further coding. Since this was not a written test, no moderation of the marking of scripts was possible. Data capturers captured the data during the two-week period following the tests using a restricted input system. Ten percent of the data captured was checked by a project manager as part of the quality assurance process, and error rates were calculated. At both baseline and end line, the overall error rate was less than 0.5 errors per test captured. All the data was then cleaned by the project manager by checking for blank cells and invalid codes or responses and verifying the number of participant responses captured.

⁵ The full Marko-D has 48-items and was published for use only in South Africa in 2019.

The raw scores obtained from the 47-item (demo) version, were converted into percentages to ensure comparability with the full Marko-D for the assignment to the five Marko-D conceptual levels within the specified upper and lower bounds, as shown in Table 1 below.

Table 1: Marko-D raw score boundaries and levels

Marko-D level	Lower bound raw score (%)	Upper bound raw score (%)	Level descriptor	Norm for beginning of Grade 1
I	≤ 7 (15%)	14 (29%)	Counting	M – 1SD ≈ 30%
II	15 (31%)	24 (50%)	Ordinal number line	M ≈ 47%
III	25 (52%)	31 (65%)	Cardinality	M + 1SD ≈ 63%
IV	32 (67%)	35 (73%)	Part-part-whole relations	
V	36 (75%)	≥41 (85%)	Equidistant number line intervals	

Source: Henning et al., 2019, p. 87.

The Marko-D norms situate the mean at Marko-D level II (M ≈ 47%), with M - 1SD being at the top of level I (≈ 30%) and M + 1SD being at level III (≈ 63%) for children at the beginning of Grade 1. It is to be expected that children towards the end of Grade R would approach the beginning of Grade 1 norms (with an expected mean at Marko-D level II).

In terms of analysis, descriptive statistics were calculated per group, and represented as graphs and tables. After the end line was completed, independent sample t-tests to establish the significance of the observed differences in the two districts were conducted. T-tests do not have the power to control for differences between participants such as age, gender, and other factors relevant to testing. Nor do they have the ability to hold baseline scores constant or investigate differences between group, district, and language simultaneously. A general linear model (GLM) was thus created to satisfy this purpose. It was constructed for each level of the Marko-D and the total Marko-D so as to determine which factors had an influence on learner performance in the end line test. All learner test statistics were computed using ICT software: SPSS version 24 or Excel 2016.

Ethical clearance was obtained from the University of Johannesburg for conducting the evaluation and permission was obtained from the WCED to conduct research in schools in the province. Ethical guidelines regarding confidentiality of the data (Republic of South Africa, 2006) were followed. This was a province-wide intervention in which involvement of all WCED schools was mandatory. For learners, their experience of the R-Maths intervention was through their normal Grade R teacher/practitioner, as part of usual school activities. Special precautions were taken to protect children involved in the learner test. The testing took place during school time and on the school premises. Consent was obtained from the school principal and Grade R teachers/practitioners for each learner who participated. Since the children were too young to give legal consent, they assented to participate in the study. A courtesy letter was provided to each school, and the school principal was requested to distribute this to inform the parents about the test and provide them with the option to opt out.

Limitations

There are several limitations worth noting. The choice of which districts to select for in-depth research was limited by logistical and budgetary constraints. Given cost constraints, the agreed sampling approach for the learner test was simple—random sampling rather than cluster sampling. Because of the tight timeframes involved, the learners were first tested when implementation of the R-Maths Project had already been underway for between five and eight weeks. Although efforts were made to use the same test administrators at both baseline and end line, this was not possible in all cases. The Marko-D test used to assess the Grade R learners included questions on number concepts only, whereas the R-Maths project covered all five Mathematics content areas in the Grade R national curriculum. Our rationale for its use, despite this limitation, has been argued above. The Marko-D was the only assessment found to be suitable for the Grade R and Grade 1 levels. The Marko-D had already been validated in South Africa for English, Afrikaans, Sesotho, and isiZulu. In the Western Cape, administration was required in English, Afrikaans, and isiXhosa (the translation of which was conducted making use of the isiZulu and English versions). The isiXhosa version of the test had not been validated at the time of its use; the data collected through R-Maths was subsequently used for this purpose.

Findings

Here we provide the evaluation findings on the changes in the learners' scores from baseline to end line. These findings will help us answer the research question, "Did the R-Maths Project have an impact on Grade R learners' Mathematics knowledge and skills?"

We look first just at the mean learner scores and standard deviations for the Marko-D test at baseline and end line by group (intervention or comparison). However, this analysis is limited since it does not control for other variables that may explain the changes such as age and language of learning and teaching. We therefore introduce the findings from a GLM conducted on the whole sample since a GLM is able to control for other factors.

Difference-in-difference analysis in each district

A total of 622 learners completed the Marko-D test at both baseline and end line, divided between districts and groups as shown in Table 2. These learners came from 101 different schools in the Urban District 1 and 47 different schools in the Rural District 2.

Table 2: Number of matched tests, by district and group

District	Number of completed tests (intervention group)	Number of completed tests (comparison group)
Urban District 1	160	157
Rural District 2	150	155

We focus first on the urban district.

Most learners ($\approx 70\text{-}80\%$) were assessed in English. Only a small minority of those assessed in English did not have English as their home language. Almost all remaining learners were learning mathematics in their home language of isiXhosa.

The test results of the learners in the urban district are summarised in Table 3.

Table 3: Urban district 1: learner test results

Group	Test phase	N	Minimum	Maximum	Mean	Standard error of the mean	Standard deviation
Intervention	Baseline	160	8.5%	87.2%	37.8%	1.2%	15.3%
	Endline	160	19.1%	95.7%	57.5%	1.3%	16.7%
Comparison	Baseline	157	2.1%	83.0%	35.0%	1.3%	16.5%
	Endline	157	14.9%	95.7%	55.3%	1.5%	18.9%

Table 3 shows that the Urban District 1 intervention group’s baseline mean was 37.8% (Marko-D level II) and rose by 19.6 percentage points to 57.5% (Marko-D level III) at end line. The Urban District 1 comparison group’s baseline mean was 35.0% and rose by 20.3 percentage points to 55.3% at end line. The intervention group thus improved by 0.7 percentage points fewer than the comparison group.

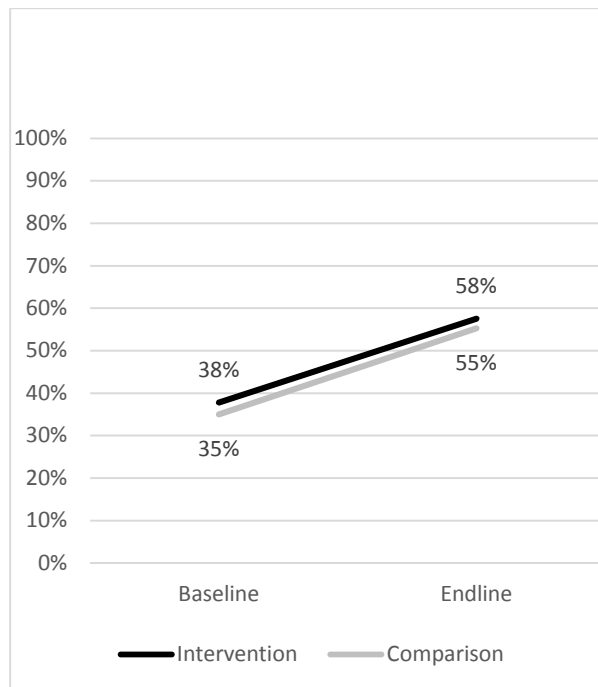


Figure 2: Changes in mean Marko D scores for the intervention and comparison groups in Urban District 1

From these values it can be seen that over the 8-month period between the pre-test and the post-test, average learner performance in the Marko-D increased considerably in this district for both the intervention and the comparison groups. This was also the case in the rural

district 2. This substantial increase indicates that the learners in the sample learnt a great deal of mathematics in their Grade R year (but does not, of course, mean that the R-Maths project led to this, since these changes occurred for learners in both intervention and comparison groups).

The intervention group thus improved less (from a higher baseline) than the comparison group did. Independent samples t-tests showed that the net shifts were not significantly different for the intervention group ($M = 19.6$; $SD = 17.0$) compared with the comparison group ($M = 20.3$; $SD = 17.9$); $t(314) = -0.33$, $p = 0.371$; but, nonetheless, this is a negative finding for the R-Maths intervention.

We turn now to the rural district. Most learners ($\approx 75\%$ in the intervention group, and 100% in the comparison group) were assessed in Afrikaans. All remaining learners were learning mathematics in their home language of isiXhosa.

The test results of the learners in this district are summarised in Table 4.

Table 4: Rural district 2: learner test results

Group	Test phase	N	Minimum	Maximum	Mean	Standard error of the mean	Standard deviation
Intervention	Baseline	150	2.1%	91.5%	42.0%	1.5%	17.9%
	Endline	150	12.8%	97.9%	58.4%	1.5%	18.6%
Comparison	Baseline	155	0.0%	89.4%	40.7%	1.3%	16.2%
	Endline	155	17.0%	87.2%	54.0%	1.3%	15.8%

The Rural District 2 intervention group's baseline mean was 42.0% (Marko-D level II) and rose 16.3 percentage points to 58.4% (Marko-D level III) at end line. The Rural District 2 comparison group's baseline mean was 40.7% and rose by 13.3 percentage points to 54.0% at end line. The intervention group thus improved by 3.0 percentage points more than the comparison group.

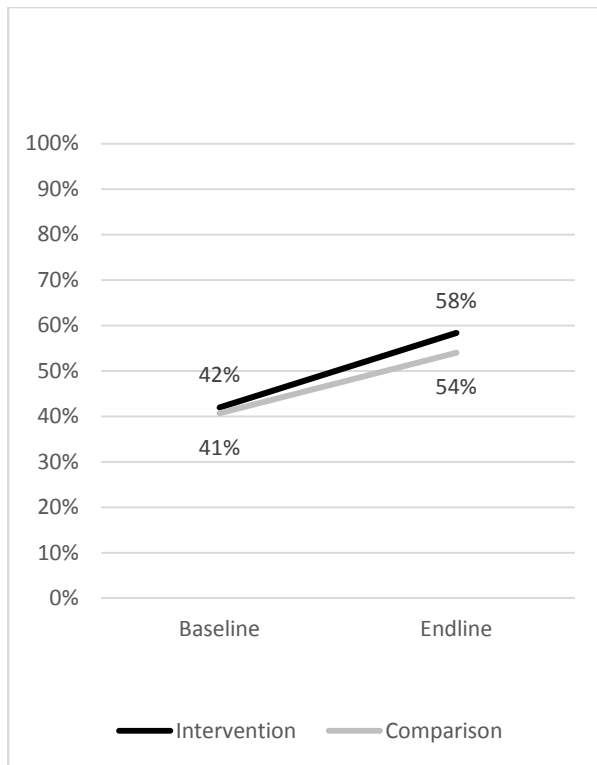


Figure 3: Changes in mean Marko D scores for the intervention and comparison groups in Rural District 2

The intervention group improved more over time (from a higher baseline) than the comparison group. Independent samples t-tests showed that the net shifts were significantly higher for the intervention group learners ($M = 16.3$; $SD = 15.1$) compared with the comparison group learners ($M = 13.3$; $SD = 14.0$); $t(300) = 1.81$, $p = 0.035$ (with a *small* effect size of $d = 0.21$). This can be considered a positive finding for the impact of the R-Maths intervention.

General linear model on the whole sample

The two districts were combined and other factors/covariates were taken into account: Gender (Male; Female), Quintile (1; 2; 3; 4; 5), Home Language (English; Afrikaans; Xhosa; Other), Language of Learning and Teaching, Language of Testing, and Age were included in the model as factors (Nominal or Ordinal Variables) and covariates (Scale/Ratio Variables) as appropriate; all these were included as fixed effects. In addition, consideration was given to each of the Marko-D levels (I to V). The following findings arose from the GLM.

The intervention group performed better than the comparison group: 2.9 percentage points better over the whole Marko-D test and approximately 5 percentage points better in the three levels where the difference in improvement between the groups was significant. The relative improvement varied between 0.17 and 0.24 sd. In all cases, the Cohen's d effect size was *small* ($d \approx 0.19$ to 0.24).

However, the greatest effects on Marko-D performance were from language of learning and teaching, and, by proxy, language of testing, and district. It was found that isiXhosa speakers

and urban learners improved the most from baseline to end line. The effects of language of learning and teaching were of a *small to medium* size ($d \approx 0.30$ to 0.50) and the effects of district of *medium* size ($d \approx 0.50$).

Older learners performed better than younger learners at three levels of the test, but this difference was also *small* ($d \approx 0.20$). None of the other factors/covariates in the GLM were found to be significant.

For the R-Maths intervention, the biggest effects were evident at Marko-D level II and level III. Thus, the gains evident were mainly at the lower levels of the Marko-D scale. Greatest improvement at the lower levels was to be expected, since the Grade R learners assessed in this evaluation were on the younger end of the spectrum of learners that could be assessed using the Marko-D (normed for beginning of Grade 1). That the biggest effects are evidenced at Marko-D level II and level III is also to be expected considering the normed mean is Marko-D level II.

When comparing the average effect of the intervention and the average effect of age on Marko-D performance at level II and level III, receiving the intervention was found to be equivalent to approximately six additional months of age. It should be noted, however, that this refers to six months of general child development and not six months of schooling, since all learners in this study were in Grade R.

The GLM accounts for confounds between language and district. We must, thus, conclude that the effect of the intervention was similar across our case study districts and language groups.

Discussion and conclusion

It is unrealistic to expect large differences in improvements in learner scores (for the intervention and comparison groups) in the eight months between pre- and post-tests for the R-Maths intervention. Overall, therefore, the fact that the R-Maths intervention had a generally small but positive effect on the mathematics results of children whose teachers/practitioners had been exposed to the intervention —about 2.9 percentage points over the whole test, with a small effect size ($d = 0.2$; equivalent to 0.17 sd)—is encouraging. Grade R children in the intervention group were performing similarly to those Grade R learners in the comparison schools who were six months older (when assessed at levels II and III of Marko-D). Put another way, this means that younger intervention group children performed roughly as well as children in the comparison group who were six months older in the same year of schooling.

To put the R-Maths findings into context, the effect size of interventions identified as promising in international synthesis studies was often quite small: Snilstveit et al. (2015) reported an average effect of 0.23 on language test scores and 0.14 on mathematics test scores for teacher-level structured pedagogy interventions, which had the largest and most consistently positive effects on learner performance. In the South African context, promising

interventions that offered Fleisch's (2018) "education triple cocktail" of LTSM, lesson plans and individual coaching were found to have effects upwards of 0.252 sd after two years. However, the effects of some interventions which initially seemed promising had disappeared after two years, highlighting the importance of assessing the sustainability of learner performance gains.

To further understand the relative strength of the benefits obtained through the R-Maths intervention, it is worth reflecting on the evaluation of other programmes that have been implemented on the same scale. An analysis of the GPLMS intervention's impact on Mathematics performance, following cohorts of learners for two years (Grade 1 to Grade 3) found that learners in treatment schools performed 0.7 sd better. However, whilst learners in treatment schools followed from Grade 2 to Grade 3 performed 0.5 sd better, this disappeared by Grade 4 (Fleisch et al., 2016).

A key aspect of R-Maths that distinguishes it from the typical "educational triple cocktail" model was that in the case of R-Maths the teachers/practitioners were not trained and coached directly. Rather, the Foundation Phase Subject Advisors in the province were trained, and then they offered training and ongoing support to Grade R teachers/practitioners through cluster meetings and via WhatsApp groups, while at the same time they continued to receive support from the UCT Schools Development Unit trainers.

While this has the advantage of increased sustainability and enables capacity-development within the provincial education department structures, this approach does create a long chain of effect from service provider to subject advisor to Grade R teachers/practitioner to Grade R learner. Perhaps dilution at each stage of transfer is one of the factors contributing to the relatively small effect size (of 0.2 sd).

Two further possible explanatory factors for this are worth reflecting on. First, the measurement of learner outcomes took place in the first year of R-Maths implementation and one may expect there to be greater capacity and implementation stability in the second year from Subject Advisors who had already supported the first cohort. Second, the pre- and post-tests were conducted within a short (8-month) period. It would be better to have had pre- and post-tests at least an academic year apart.

The R-Maths evaluation generated some promising evidence of a small but positive effect on learners' mathematics performance. It does not yet offer evidence of a sustained impact on learner outcomes; for this to be established, a delayed post-test would be necessary. Further, to enhance external validity, it would be ideal for additional evidence to be generated (positive or otherwise) of the effectiveness of the R-Maths over time and/or the implementation of the R-Maths model in other contexts.

Further research should investigate how best to increase the benefits that children receive from education interventions and also uncover factors that may improve or reduce the effect of the intervention. It is likely that both the home and classroom, as well as the background and characteristics of both the teachers/practitioners and the Subject Advisors will have some

influence on both the initial size of the positive effect of the intervention, and on the endurance of the effects of the intervention over time.

In terms of context, the Western Cape is one of the most urban provinces, and the WCED has a generally well-capacitated human resource infrastructure at both provincial and district level capacity (including Subject Advisors), which has been pivotal to this intervention. In many other South African provinces this kind of district level capacity is lacking and implementing R-Maths, using the same implementation model, may not be viable there. Nonetheless, such an approach – where attention is paid to multiple levels of the state schooling system and strengthening the internal capacity of the system via a modified cascade model which provides on-going support rather than once-off training may have value for the implementation of other interventions at scale in similarly-resourced regions and countries.

Acknowledgements

The R-Maths project evaluation, on which this article draws, was commissioned and funded by the Zenex Foundation, the ELMA Foundation, and Maitri Trust.

The authors would like to thank the R-Maths Project Steering Committee (PSC) for their support, guidance, and robust engagement with the evaluation. The PSC members were:

- Luke Aspinall, Maitri Trust;
- Gail Campbell, Zenex Foundation;
- Jonathan Clark, UCT Schools Development Unit;
- Jo Davies, Maitri Trust;
- Karen Dudley, WCED;
- Lauren Fok, Zenex Foundation;
- Genevieve Koopman, WCED;
- Cally Kuhne, UCT Schools Development Unit;
- Bernadette Moffat, The ELMA Philanthropies
- Kirstin O’Sullivan, The ELMA Philanthropies; and
- Gillian van Wyk, WCED.

Thanks also to Matthew Snelling for running the GLM, and to Hogrefe and the University of Johannesburg who gave permission for the Marko-D learner test to be used in the evaluation and translated into isiXhosa. We are also very grateful to the many Grade R teachers/practitioners and learners who participated willingly in the research activities.

References

- Besharati, N. A., & Tsotsotso, K. (2015). *In search for the education panacea: A systematic review and comparative meta-analysis of interventions to improve learner achievement in South Africa*. Johannesburg, RSA: University of the Witwatersrand.

- Conn, K. M. (2017). Identifying effective education interventions in Sub-Saharan Africa: A meta-analysis of impact evaluations. *Review of Educational Research*, 87(5), 863–898.
- Fleisch, B. (2018). *The education triple cocktail: System-wide instructional reform in South Africa*. Cape Town: University of Cape Town Press.
- 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.
- Gertler, P. J., Martinez, S., Premand, P., Rawlings, L. B., & Vermeersch, C. M. J. (2010). *Impact evaluation in practice*. Washington DC: The International Bank for Reconstruction and Development/The World Bank. Retrieved from http://siteresources.worldbank.org/EXTHDOFFICE/Resources/5485726-1295455628620/Impact_Evaluation_in_Practice.pdf
- Hazell, E. (2019). *A meta-evaluation and synthesis of evaluations of South African education programmes: 2013–2018* (Unpublished master's report). Stellenbosch University, Stellenbosch, RSA.
- Henning, E., Ehlert, A., Balzer, L., Ragpot, L., Herholdt, R., & Fritz, A. (2019). *Marko-D SA: Assessment of number concept development*. Johannesburg, RSA: University of Johannesburg.
- Kotzé, 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*, 66, 203–213. <https://doi.org/10.1016/j.ijedudev.2018.09.004>
- McEwan, P. (2015). Improving learning outcomes in primary schools of developing countries: A meta-analysis of randomized experiments. *Review of Educational Research*, 85(3), 353–394.
- Mouton, J., Wildschut, L., Richter, T., & Pocock, R. (2013). *Review project: Final report* (Unpublished report). Johannesburg, RSA: Zenex Foundation.
- Prinsloo, C. H., Harvey, J., Thaba W., & Moodley, M. (2015). *Human Sciences Research Council evaluation report: siyaJabula siyaKhula's Learner Regeneration Project Vhumbedzi and Malamulele North-East, Vhembe, Limpopo Evidence after two years (2013–2014/15)*. Retrieved from <http://www.nstf.org.za/wp-content/uploads/2017/10/Part-2-Report-PDF-1.pdf>

- Republic of South Africa. (2006). Ethical rules of conduct for practitioners registered under the Health Professions Act, 1974. *Government Gazette*, No 29079. Pretoria, RSA: Government Printer.
- Roberts, N., Mostert, I., & Takane, T. (2015). *Zenex Foundation landscape review of mathematics interventions in South African schools*. Johannesburg: Zenex Foundation.
- 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* (Unpublished doctoral dissertation). University of Cape Town, Cape Town, RSA.
- Snilstveit, B., Stevenson, J., Phillips, D., Vojtkova, M., Gallagher, E., Schmidt, T., Jobse, H., Geelen, M., Pastorello, M., & Eyers, J. (2015). *Interventions for improving learning outcomes and access to education in low- and middle-income countries*. Retrieved from http://www.3ieimpact.org/media/filer_public/2016/07/12/sr24-education-review.pdf