SA-TIED
Southern Africa – Towards Inclusive Economic Development

SA-TIED YOUNG SCHOLARS PROGRAMME

# Impact evaluation of Funda Wande in-service teacher coaching intervention

## Findings from the first year

Tiaan Meiring

SA-TIED Working Paper #156 | January 2021

UNITED NATIONS UNIVERSITY UNU-WIDER
national treasury Department: National Treasury REPUBLIC OF SOUTH AFRICA
INTERNATIONAL FOOD POLICY RESEARCH INSTITUTE IFPRI
planning, monitoring & evaluation Department: Planning, Monitoring and Evaluation REPUBLIC OF SOUTH AFRICA
the dtic Department: Trade, Industry and Competition REPUBLIC OF SOUTH AFRICA
SARS
TIPS

## Young Scholars

This paper was produced as a part of the SA-TIED Young Scholars' programme. The programme is a part of SA-TIED's capacity building initiatives, designed to support the development of skills and capabilities in the research and policymaking aspects of economic development. Every year, the programme recruits Masters' level college students from Southern Africa to become SA-TIED Young Scholars.

SA-TIED Young Scholars work with top academics and officials in their research fields to complete original research projects as a part of the programme and as a part of the research component of their Masters' degree.

## About the programme

### Southern Africa –Towards Inclusive Economic Development (SA-TIED)

SA-TIED is a unique collaboration between local and international research institutes and the government of South Africa. Its primary goal is to improve the interface between research and policy by producing cutting-edge research for inclusive growth and economic transformation in the southern African region. It is hoped that the SA-TIED programme will lead to greater institutional and individual capacities, improve database management and data analysis, and provide research outputs that assist in the formulation of evidence-based economic policy.

The collaboration is between the United Nations University World Institute for Development Economics Research (UNU-WIDER), the National Treasury of South Africa, the International Food Policy Research Institute (IFPRI), the Department of Monitoring, Planning, and Evaluation, the Department of Trade and Industry, South African Revenue Services, Trade and Industrial Policy Strategies, and other universities and institutes. It is funded by the National Treasury of South Africa, the Department of Trade and Industry of South Africa, the Delegation of the European Union to South Africa, IFPRI, and UNU-WIDER through the Institute's contributions from Finland, Sweden, and the United Kingdom to its research programme.



Corresponding author: tiaanmeiring7@gmail.com

# Impact evaluation of Funda Wande in-service teacher coaching intervention

Findings from the first year

Tiaan Meiring

November 2020

**Abstract:** Low-resource public schools in South Africa suffer from a lack of (i) access to home-language reading materials and (ii) teacher capacity on how to teach early-grade reading in African languages. In this context, the Funda Wande structured pedagogical intervention provides a complementary bouquet of learner and teacher resources and teacher professional development (in the form of in-class teacher coaching). After one year of implementation, results from a randomized control trial evaluation shows that the programme improves student reading proficiency by 0.17 standard deviation. Learning gains translate to between 20 and 27 per cent of a year's worth of learning for Grade 2 learners and between 33 and 58 per cent for Grade 1 learners. Encouraging from a policy perspective is that the effects are positive across all sub-tasks that were reliably measured, and that the intervention has positive impacts across the distribution of learners by initial reading proficiency.

**Key words:** early-grade reading, teacher professional development, coaching, structured pedagogical intervention, randomized control trial

**JEL classification:** C93, I20, I21, I28

**Acknowledgements:** This research has benefitted from the mentoring, guidance, and input of Professor Cally Ardington (South African Labour and Development Research Unit [SALDRU], University of Cape Town), who supervised the master's research dissertation from which this paper was adapted. I would also like to thank Professor Murray Leibbrandt (SALDRU, University of Cape Town) for his invaluable comments, suggestions, and support. Financial support from the NRF National Research Chair of Poverty and Inequality Research, a research assistant role within SALDRU, and the UNU-WIDER's SA-TIED Young Scholars programme made this research possible and is also gratefully acknowledged.

# 1 Introduction

South Africa faces a particularly persistent challenge in providing quality education to the majority of learners from low-income settings. The 2016 Progress in International Reading Literacy Study (PIRLS) international benchmark test shows that South Africa is unique among upper-middle income countries in that 78 per cent of Grade 4 children cannot read for meaning (i.e. at the PIRLS low international benchmark) in any language (Howie et al. 2017). South Africa's comparatively poor performance in early-grade reading (and later learning outcomes) persists despite an almost universal primary enrolment rate (Howie et al. 2017), government policies ensuring that the majority of learners have access to mother tongue education for the first three years of primary schooling,[1] and the country's comparatively high expenditure on education by international standards (Motala and Carel 2019; World Bank 2018).

In turn, poor learning outcomes in aggregate are disproportionately driven by the majority of learners who find themselves in the poorest 75–90 per cent of schools in the country's bimodal education system (Spaull and Pretorius 2019; Spaull 2013). With large disparities in terms of access to quality of schooling inputs (both physical and human) and functioning management, support, and accountability structures, inequalities in downstream learning outcomes represent a de facto two-tiered schooling system that sees children with the largest educational deficits attend schools with disproportionately less capacity (Spaull 2015; NEEDU 2013).

In terms of priorities of where to intervene, the largest potential for long-run returns arguably lies in targeting learners early in their schooling careers. Core skills (like basic literacy and numeracy) are hierarchical, forming the foundations on which subsequent learning and skills development take place (Heckman 2006). When a learner falls behind, the deficit is compounded over time as there is a mismatch between the level of classroom instruction and learner's actual learning levels (Glewwe and Muralidharan 2016). If learners do not master reading as a core skill, specifically, they cannot acquire further subject-specific knowledge that relies on the foundation of reading comprehension (Spaull and Pretorius 2019; Spaull and Kotze 2015).

Teacher capacity has proved to be a promising avenue, or policy lever, for early intervention. The quality of teachers in a child's early years of education seems to have large and persistent effects on both schooling and other later life welfare outcomes, across both developed (Hanushek and Rivkin, 2010; Chetty et al. 2014) and developing countries (Bau and Das 2020; Bold et al. 2017; Bruns and Luque 2014).[2]

[1] Mother tongue education is a constitutionally mandated right for learners in Grades 1–3. Therefore, more than 70 per cent of South African children learn to read in an African language before switching to either English or Afrikaans in Grade 4 (Pretorius and Spaull 2016). Approximately 90 per cent of learners transition to English as language of instruction in Grade 4, with the remaining 10 per cent generally receiving instruction in Afrikaans.

[2] For the United States, Chetty et al. (2014) estimate that learners who attend classes from higher-quality teachers in primary school (measured in terms of the teachers' impact on raising their students' test scores) are more likely to attend college, earn higher salaries, and are less likely to have children as teenagers. According to Chetty et al. (2014), 'replacing a teacher whose (value-added) is in the bottom 5% with an average teacher would increase the present value of students' lifetime income by approximately $250,000 per classroom.'

The focus on improving teacher quality has seen an increase in programmes aiming to strengthen teacher capacity through professional development-, or structured pedagogical programmes.[3] These programmes characteristically consist of:

(i) an integrated package approach that includes the provision of curriculum-aligned learner resources[4] (like graded readers and other forms of print materials);
(ii) teacher guidelines (generally in the form of lesson plans); and
(iii) some form of teacher professional development (often consisting of initial teacher training, implementation support, feedback, and/or mentoring) (Cilliers et al. 2019; Fleisch et al. 2016).

Partnering with the Eastern Cape Department of Education (ECDoE), the not-for-profit organization Funda Wande is implementing a novel structured pedagogical intervention in the Eastern Cape of South Africa. This paper evaluates the effect of the intervention on early reading outcomes of students after the first academic year of implementation.

The paper proceeds as follows. Section 2 provides the background to the study and describes the context that teachers face in low-resource public schools across South Africa. Section 3 provides a description of the Funda Wande intervention and its theoretical underpinnings. Section 4 describes the evaluation design, results are reported in Section 5,[5] and Section 6 concludes with a discussion of how findings here relate to the larger literature and future research.

## 2 Background and context

The intervention specifically targets no-fee public schools, where the foundation-phase language of learning and teaching (LOLT) is an African language. These schools represent a significant share of the South African schooling system: about 93 per cent of primary school learners attend public schools, with the no-fee-paying segment of public schools accounting for more than 70 per cent of learners (Howie et al. 2017). The socioeconomic status (fee-paying status) of the school that a learner attends is the greatest predictor of their academic performance (Taylor 2011). However, given the country's historical legacy of racial segregation, a school's socioeconomic status is also highly correlated with characteristics such as its geographic location and the language of instruction in the early grades (Spaull 2019).

For learners to learn how to read, a necessary condition is that they have access to reading materials in the language that they are learning to read in. However, only 30 per cent of schools have access to some form of library (DBE 2018). The libraries that do exist are often poorly stocked (especially with regard to African-language reading materials). Consistent with a general theme in the South African education system, learners in formerly white-only, fee-paying primary schools are more likely to have access to libraries at school (87 per cent of learners) than learners in schools formerly

[3] Notably the range of pilot- and at-scale intervention studies in Kenya by Piper et al. (2014, 2015, 2018a) and Zuilkowski and Piper (2017), as well as similar programmes in South Africa (Cilliers et al. 2019; Kotze et al. 2019), Liberia (Piper and Korda 2011), and Uganda (Kerwin and Thornton 2019; Lucas et al. 2014).

[4] Often referred to as learning and teaching support material (LTSM).

[5] Results are also published in Ardington and Meiring (2020), whose report has been cited in full, data used as applicable, as well as relevant images/tables reproduced in the present paper with permission from the South African Labour and Development Research Unit (SALDRU).

classified as rural homelands or urban African schools (where only 35–36 per cent of learners today have access to libraries at school, in any form) (DBE 2014).

In addition to resource constraints, the lack of content and pedagogical knowledge of the majority of South African teachers who work in no-fee public schools serves a binding constraint to student learning (NEEDU 2013; Venkat and Spaull 2015).[6] More specifically, most teachers have historically not received any meaningful training in how to teach reading (Taylor and Taylor 2013; Van der Berg et al. 2016).[7]

In practice, common findings from classroom observations on norms in teacher instructional practices suggest that these do not align with best practice teaching methods, lesson pacing or curriculum coverage (Hoadley 2012, 2016, 2018; NEEDU 2013; Taylor and Taylor 2013). Illustratively, often the principal method for teaching reading is to read to the class and have learners collectively chorus back what was read (e.g., Prinsloo 2008). Conversely, there is often a lack of individualized learner attention (NEEDU 2013) and a neglect of formal instruction of other skills foundational to learning to read (such as phonics, spelling, and vocabulary; see Spaull and Pretorius 2019).

In an environment where both teacher absenteeism (Reddy et al. 2010; Spaull 2011) and curriculum coverage are low,[8] current evidence suggests that it is the lack of productive teaching activity taking places *despite* the presence of teachers that is a major binding constraint to learning (Carnoy et al. 2012; Hoadley and Gallant 2016; Van der Berg et al. 2016).

## 3 Programme description and theoretical underpinnings

The Funda Wande intervention builds upon the lessons learnt from the international literature and promising insights from previous iterations of similar approaches to improving teacher instruction and learning outcome of learners in the South African context,[9] the most notable among which is the Department of Basic Education-led early-grade reading study (EGRS) (Cilliers et al. 2019; Kotze et al. 2019). For these packaged interventions (often categorized as structured pedagogical interventions; Snilstveit et al. 2016) to be successful, indications are that (a) some degree of teacher support, monitoring, and feedback are required, and (b) that the programme should focus on improving specific shortcomings in teacher capacity, whether in instructional techniques, lesson planning, or the effective use of complementary materials provided.

---

[6] This despite South African teachers generally being adequately qualified with respect to academic qualifications (Taylor et al. 2013).

[7] Teachers in low-resource South African schools have in recent years been on the receiving end of a plethora of (often competing) supporting strategies and models (Van der Berg et al. 2016). However, very recently there has been little rigorous evaluation of these competing teacher support and professional development programmes, leaving policy makers with little sense of which programmes are working and why they are (not) working (Van der Berg et al. 2016).

[8] Evidenced in various systemic evaluations: the National School Effectiveness Study, National Education Evaluation and Development Unit case studies, and the School Monitoring Survey (see Dechaisemartin 2013; NEEDU 2013; DBE 2015).

[9] Examples include the Gauteng Primary Language and Mathematics Strategy programme (Fleisch et al. 2016; Fleisch and Schöer 2014) the Reading Catch-Up Study (Fleisch et al. 2017), the Systematic Method for Reading Success study (Piper 2009), and the Learning for Living project (Sailors et al. 2010).

Drawing on these insights, the Funda Wande programme makes use of a carefully designed, multimedia course to train foundation-phase (Grades R to 3) teachers.[10] The course teaches the major components of reading and writing in isiXhosa (the pilot language), with subtitles in English. The essential components of the intervention are:

- *Learning and teaching support material (LTSM) box*: Each teacher is provided with a set of Funda Wande materials, readers, and additional graded reading aides (such as posters and phonics flashcards) all integrated into lesson plans provided. Materials for teachers include structured lesson plans, handwriting booklets, baseline assessment booklets, group-guided reading booklets, online resources for teachers, and a pre-loaded flash drive with the full set of Funda Wande videos and multimedia resources. All materials are aligned to the national curriculum.
- *Teacher coaching and professional development*: This consists of six expert coaches who are experienced foundation-phase literacy educators with a coach-to-school ratio of 1:5. The coaches observe teachers in their classrooms, provide targeted advice on how to improve their practice, and provide lesson role modelling. Coaches visit each school an average of three times a month. Coaches provide further group training in the form of on-site phase meetings once per week and occasional off-site workshops. The latter allows teachers to work on common issues and the Funda Wande course materials collectively, to gain a stronger theoretical understanding of teaching literacy, and to plan for upcoming terms. Training consists of both whole-phase meetings after school (three per term), and one-on-one in-classroom visits with each teacher in the foundation phase (at least once per term).
- *Head of department (HOD) training*: HODs are trained to take over the role of coach and literacy specialist after the intervention finishes. To that end, all foundation-phase HODs were given a bursary by the programme to enrol (part-time) in the nationally accredited two-year Funda Wande course at a local university.

### 3.1 The importance of complementary input provision and coaching

Within the education production function framework, the earlier experimental literature focused on providing different kinds of seemingly lacking resources to schools, teachers, and households in resource-constrained environments (for an overview, see Kremer et al. 2013; Glewwe et al. 2014). Findings suggest that simply supplying inputs on their own, without complementary guidance, incentives, or accountability structures is not enough to shift learning outcomes (Kremer et al. 2013).

The main lesson from earlier studies is that business-as-usual input provision is rarely effective and often expensive (Glewwe and Muralidharan 2016). Research on South African schools similarly suggests that the provision of additional school resources often had no impact on learning outcomes, because they were not well managed by the schools (Van der Berg 2008; Taylor 2011). Subsequent programmes have focused on providing a combination of physical inputs integrated alongside complementary intervention components such as teacher professional development, coaching, community interventions, and personalized computer-assisted learning programmes (Piper et al. 2018b).

[10] The Funda Wande literacy course and materials were developed over two years with input from over 15 South African academics from five universities. The course is nationally accredited (by the South African Qualifications Authority) and has strong support from the national Department of Basic Education, the Eastern Cape Department of Education, and Rhodes University. Materials were developed using professionally filmed in-classroom videos, info-graphics, and other multimedia.

In an attempt to synthesize the meta-analyses and systematic reviews of high-quality[11] empirical evidence on the interventions[12] that have improved learning outcomes in developing countries (Conn 2017; Glewwe and Muralidharan 2016; Ganimian and Murnane 2016; McEwan 2015; Snilstveit et al. 2016), Evans and Popova (2016) highlight two classes of programmes that have shown positive effects with some consistency:

(i) pedagogical intervention that tailors teaching to actual learning levels of a learner (instead of the rigid expected levels of curriculums), either by means of teacher methodology or adaptive learning software; and
(ii) individualized, repeated teacher training/coaching interventions that promote a specific task or tool.

Within structured pedagogy programmes, the mode of delivery has played an important role in programme effectiveness (Popova et al. 2018). On-site teacher coaching (as opposed to centralized training workshops) has proved to be an especially important component (Kraft et al. 2018; Cilliers et al. 2019).

The idea behind in-classroom coaches is to provide teachers with the support, monitoring, and feedback required to integrate new materials and novel pedagogical techniques into their daily classroom practices. Although professional development programmes characteristically share a focus on imparting knowledge, teaching is a skill that should arguably be developed through iterative practice and learning-by-doing (Kennedy 2016; City et al. 2009).

Alongside instructional coaching, structured lesson plans specifically target teachers' ability to integrate resource use into classroom practices. Lesson plans are assumed necessary to guide teachers who have historically lacked even basic classroom resources, but are now flooded with an array of LTSM resources. Lesson plans are also intended to reduce the daily cost of adopting new teaching practices (especially when coaches are not around) and reduce teachers' planning and administrative workloads, allowing them to focus on actual teaching and implementing new pedagogical skills (Cilliers et al. 2019).

### 3.2 Understanding how children learn to read in African languages

International evidence suggests that school-entering children are best equipped to learn to read in the language that they are immersed in on a daily basis—their mother tongue (Ball 2010). In turn, these learners are able to use foundational literacy skills acquired in their home language, such as how to decode words, to better position them to learn subsequent languages (Cummins 2001, 2007).[13] The importance of home-language-first reading instruction is echoed in the limited evidence from the South African context. Taylor and Von Fintel (2016) find that learners who receive home-language instruction in the first three years of schooling fare better in English acquisition in subsequent grades.

Although the effectiveness of structured pedagogical programmes is well understood, the role of the specific LOLT used is not, neither in South Africa nor elsewhere on the continent (Brunette

[11] High-quality evidence generally refers to research that attempts to identify causal effects by establishing a well-defined counterfactual (with somewhat different cut-off points in what is considered convincing evidence on the spectrum between quasi- and experimental research).

[12] Note that 'interventions' and 'programmes' are used interchangeably here.

[13] Piper et al. (2016) provide a review of the evidence.

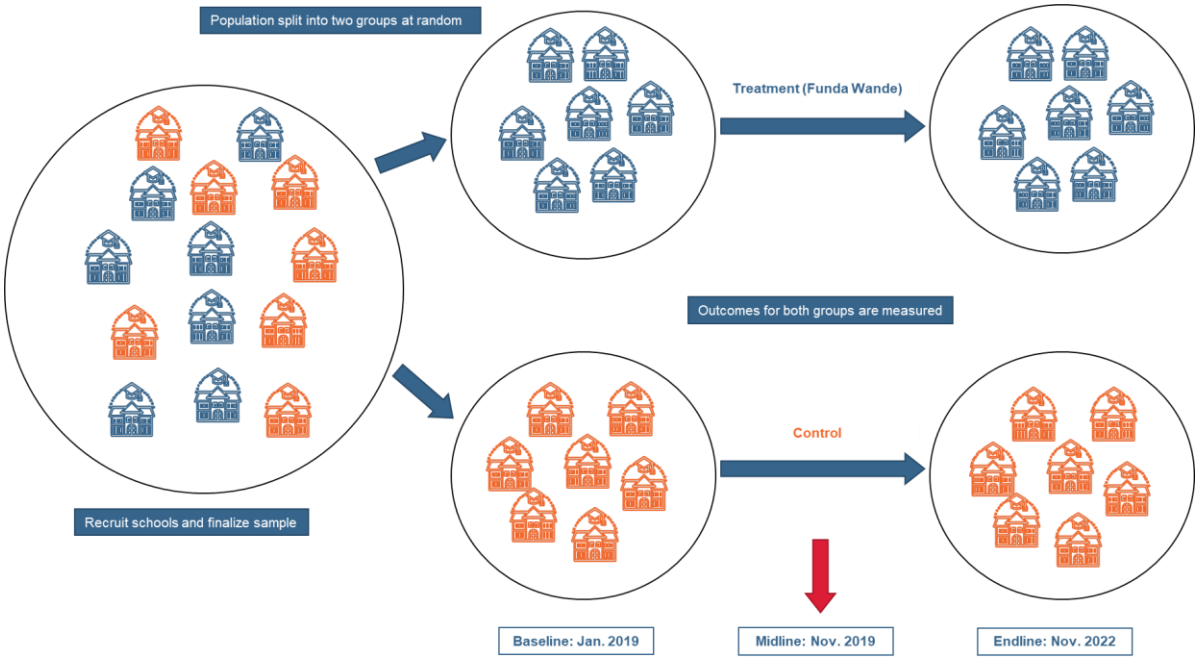
et al. 2019).[14] Just as the effects of structured pedagogy interventions are generally larger on language subjects than on mathematics (Snilstveit et al. 2016), one might expect the effectiveness of these programmes to also differ by the characteristics and complexity of the specific language itself (Spaull et al. 2020). More concretely, early reading pedagogical practices in South African schools are generally derived from teaching English (Funda Wande 2018) and, therefore, not adapted to language-specific characteristics. This differentiates the Funda Wande programme from previous structured pedagogic interventions: its focus on the 'linguistic and orthographic underpinnings of early reading instruction' (Funda Wande 2018: 11).

# 4 Evaluation design

## 4.1 Research questions and methodology

The primary aim of the evaluation is to assess the impact of the intervention on both foundational reading skills and reading comprehension in the learners' home language (see Figure 1). The results here are based on the midline assessment completed at the end of the first academic year (November/December 2019), following on baseline reading assessments conducted with learners in all schools before the programme started (in January 2019). The midline results form part of an impact evaluation in which schools are randomly assigned into one of two arms—Funda Wande and control—for a four-year period (2019–22).

Figure 1: Funda Wande impact evaluation design



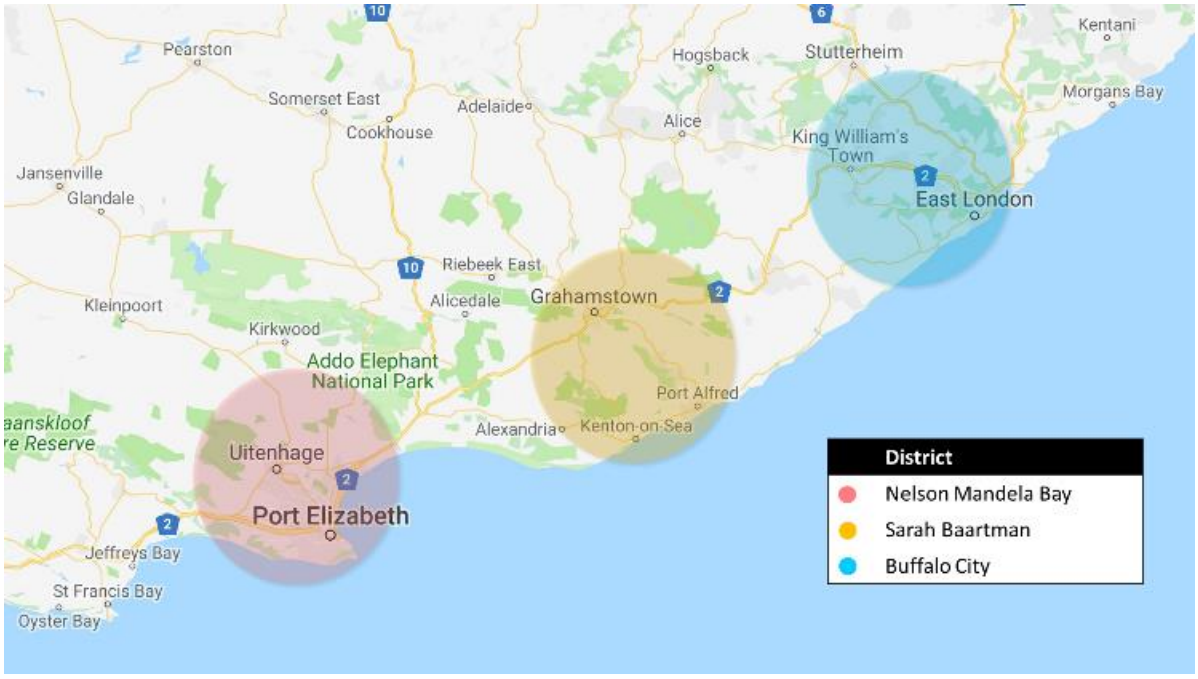

Source: author's illustration based on Funda Wande data (see SALDRU 2019).

[14] Brunette et al. (2019) find suggestive evidence from Uganda on the heterogeneity of a reading programme's effects by language characteristics. Their results indicate large variations in the same programme's effectiveness depending on the complexity of the language in which it was implemented.

## 4.2 Recruitment and random assignment

All the schools in the evaluation sample are no-fee, quintile three public primary schools with an isiXhosa LOLT. Together with the ECDoE, these schools were selected from three urban and peri-urban districts in the Eastern Cape (Nelson Mandela Bay, Sarah Baartman, and Buffalo City; Figure 2). The final baseline sample consists of 29 control schools and 30 treatment schools, with schools randomly assigned to treatment or control within each of the three strata (the three educational districts—see Appendix I). Within each school, one Grade 1 class and one Grade 2 class were randomly selected; and within each of these selected classes, ten learners were randomly selected. Appendix I provides further details on the school selection process and how representative the evaluation sample is of the underlying population of schools in the Eastern Cape.

Figure 2: Location of Funda Wande impact evaluation schools




Source: author's illustration based on Funda Wande data (see SALDRU 2019).

## 4.3 Instruments

An extended early-grade reading assessment (EGRA) that included a range of pre-literacy and literacy tasks was administered to each of the randomly selected learners at baseline and midline. The assessments used the standard isiXhosa EGRA with adaptations developed by Funda Wande and the evaluation's principal investigator, drawing on other similar assessment instruments (see Ardington and Meiring 2020: 20).

Appendix Table C1 shows the full range of sub-tasks and indicates the grade(s) and data collection rounds in which they were administered. The inclusion of a range of EGRA sub-tasks in the baseline assessment was motivated by two key concerns. First, learning to read depends on a complex set of interconnected skills, including oral language and literacy-related skills (Snow 2017). An important consideration in the evaluation is examining the relationship among these various skills both concurrently and longitudinally as they develop. Understanding where the greatest deficits lie and which skills the intervention most effectively affects is essential for ongoing programme design.

Second, as the vast majority of learners are not reading at the appropriate level for their age, one would expect floor effects (i.e. many learners scoring zero) in many of the core EGRA sub-tasks, particularly for Grade 1 learners. A range of sub-tasks are thus employed, including pre-literacy measures, to ensure that there is good discrimination between learners at baseline and those at midline. Appendix II further outlines the rationale for how, why, and when certain tasks were assessed.

Learners assessments were supplemented with learner, teacher and principal interviews, as well as student weight and height measurements, in both rounds of data collection. All assessments and interviews were conducted entirely in isiXhosa by isiXhosa home-language enumerators.

### 4.4 Baseline sample description

Overall, 1,187 learners were assessed at baseline.[15] Table 1 provides a brief description of the sample at baseline (full baseline descriptions are available in Appendix Tables A1–A3). The sample is roughly evenly split between boys and girls. The average age of Grade 1 learners is 6 years and 5 months, while that of Grade 2 learners is 7 years and 6 months. Almost 10 per cent of Grade 1 learners are repeating their grade, with less than 5 per cent of Grade 2 learners repeating.

Table 1: Learner sex, age, and grade repetition by grade

| | Grade 1 | Grade 2 | Total |
|---|---|---|---|
| Girl | 49% | 51% | 50% |
| Age | 6 years, 5 months | 7 years, 6 months | 6 years, 11 months |
| Repeating grade | 9.7% | 4.6% | 7.1% |
| Observations | 595 | 592 | 1,187 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

A comprehensive discussion on learners' access to print resources, both at home and in the classroom, is provided in the baseline report (Ardington 2019). Suffice to summarize here that the evaluation schools follow the wider trend in low-resource schools, with a lack of access to reading materials and limited opportunities to engage with text in a meaningful way (both at school and at home). More than half the learners did not have access to a library at school (either on site or mobile), the number of readers available per class are generally insufficient for the number of learners in the class, and learners face a similar lack of access to reading material at home, as almost two-thirds of learners (65 per cent) report that they have no books other than schoolbooks to read at home.

Random assignment of schools to treatment and control groups eliminates any possible selection bias (Angrist and Pischke 2009; Athey and Imbens 2016). It ensures that any subsequent differences in learners' midline reading proficiency by treatment status can be attributed to the Funda Wande intervention (i.e. the treatment effect). Random assignment ensures that schools are balanced on observable and unobservable characteristics in expectation, with any incidental differences between the treatment and control groups at baseline occurring by chance. Although

[15] The original sample design and power calculations were based on assessing ten Grade 1 and ten Grade 2 learners at each school. At the start of baseline fieldwork, the research team decided to explore whether it would be possible to complete twelve learners in each grade. Within the first week, it became clear that this was an unrealistic target and we reverted back to the original plan of ten learners per grade. Baseline data include eleven schools with twelve Grade 1 learners, two schools with twelve Grade 2 learners, and three schools with eleven Grade 2 learners. For these schools, the additional learners assessed at baseline were used as replacements for unavailable (absent/transferred/refused) learners.

it is not strictly necessary to conduct an array of individual tests to assess whether baseline imbalances are statistically significant (see Athey and Imbens 2016; Bruhn and McKenzie 2009), sample balance checks are common practice in the evaluation of social programmers and are reported in Appendix Tables A1–A3 for the full array of observable learner characteristics at baseline. As one would expect, the balance tables demonstrate the similarity between the treatment and control groups in terms of learner and household characteristics and reading skills, with only one of the 47 significance tests (2 per cent) indicating a statistically significant difference at the 5 per cent significance level (no more than expected to occur by chance).[16]

### 4.5 Midline sample, attrition, and balance

Attrition has two potential impacts on the evaluation design. The first is a small reduction in statistical power with a slightly smaller sample. However, the power calculations behind the sample design are based on fairly conservative assumptions, implying that the small reduction in sample size here is not of great concern.[17] The second is the potential for selection bias to be introduced into the sample, thereby threatening a key strength of the randomized control trial methodology: the internal validity of the estimated programme effects.

Column (1) in Table 2 shows the regression of treatment status on whether or not a learner attrited, taking into account the experimental design[18] (Athey and Imbens 2016). The overall attrition rate for the learner sample was 6 per cent.[19] There is no statistically significant differential attrition between learners or teachers in the two groups. As expected from previous evaluations (e.g., Cilliers et al. 2019), teacher attrition was slightly higher at 8 per cent than learner attrition at 6 per cent.[20]

Table 2: Tests for learner and teacher differential attrition

| | Learner attrite | Teacher attrite |
|---|---|---|
| Treatment | 0.030 | 0.035 |
| | (0.032) | (0.053) |
| Control attrition | 0.046 | 0.069 |
| Observations | 1,187 | 1,187 |
| *R*-squared | 0.054 | 0.162 |
| Strata fixed effects | Yes | Yes |

Note: standard errors reported in brackets.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

[16] Refer to the baseline report by Ardington (2019) for a more extensive discussion on baseline characteristics.

[17] See the baseline report for the original power calculations (Ardington 2019). These power calculations were particularly conservative given (i) the high levels of correlation between baseline and midline measurements of learner reading proficiency (especially for Grade 2 learners) and (ii) the relative homogeneous nature of schools in the sample (implying that most of the variation in outcomes is between learners within schools and not between schools themselves). Viewed in conjunction with the fact that there is very little attrition at this stage of the evaluation, this implies that treatment effects can be fairly precisely estimated.

[18] In other words, taking into account the sub-district (or strata) within which schools were randomly assigned and clustering standard errors at the school level.

[19] The attrition rate takes into account the seven learners who were replaced (i.e. attrition is calculated for the sample of 1,187 learners).

[20] This proportion includes substitute teachers as non-attritors as the information collected from teachers is about the availability and use of reading materials in the class rather than information about the individual teacher. Excluding the replacement teachers, the attrition rate is 16 per cent.

Appendix Table A5 reports the standardized mean differences in baseline outcomes and characteristics for the midline sample (Athey and Imbens 2016; Imbens and Rubin 2015: Chapter 14). Rather than reporting statistical significance, the focus is on the size of the differences between the groups (and whether any imbalances affect the outcomes of interest).[21] Standardized mean differences (of effect sizes[22]) provide a scale invariant measure to assess whether the two groups are equivalent at midline. As expected, the differences between treatment and control learners on all variables are within the limits to satisfy equivalence (Imbens and Rubin 2015: Chapter 14; WWC 2020).[23] Variables for which there are slight imbalances are included as covariates in the subsequent analyses.

In sum, the randomization process faced no challenges and the midline analysis sample remained balanced; therefore, the programme effects can be reliably estimated.

# 5 Midline results

## 5.1 Development of reading skills in status quo schooling environments

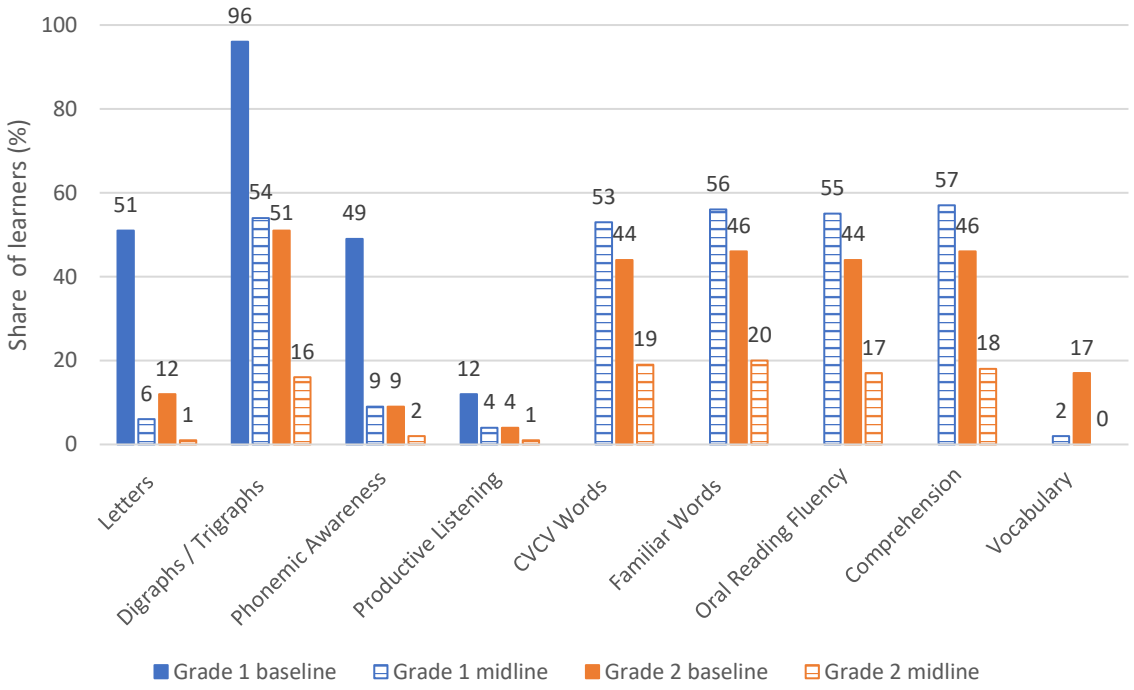
An understanding of teacher support and learning conditions in control schools are required to get a sense of what any intervention effects translate to in practice. For all the tasks that were conducted at both baseline and midline, the percentage of control group learners scoring zero and the change in learners' average scores are also summarized in Figures 3 and 4, respectively. Table 3 is a summary of Appendix Table A5, which presents the distribution of midline scores for each of the EGRA sub-tasks assessed for Grade 1 and Grade 2 learners by treatment status.

[21] The variables reported in Appendix Table A5 are all included as covariates in the subsequent analysis because (i) they are predictive of midline reading proficiency (improving precision), (ii) they were incidentally imbalanced at baseline, and/or (iii) they display slight imbalances for the midline analytical sample (based on the effect sizes).

[22] Effect sizes are calculated as the difference in means between the treatment and control groups, divided by the pooled standard deviation for the variable.

[23] If the effect sizes are 0.05 or less in absolute value, the two groups are considered equivalent on that dimension. When effect sizes are in the range between 0.05 and 0.25, the baseline measures are included as controls in the model estimating programme effects to satisfy equivalence. Variables for which such adjustments are required include 11 of the 18 sub-tasks, learner age, whether learners have non-academic books at home, whether their household owns a computer or some form of motor vehicle, or has a toilet inside their home. Variables with effect sizes in this range are displayed in red font in Appendix Table A4. These variables are added as controls to satisfy equivalence between the two groups (and not only to improve the precision of the estimates of programme impact). No effect size is greater than 0.25 in absolute value—the level at which the samples are not considered to be equivalent any more (WWC 2020).

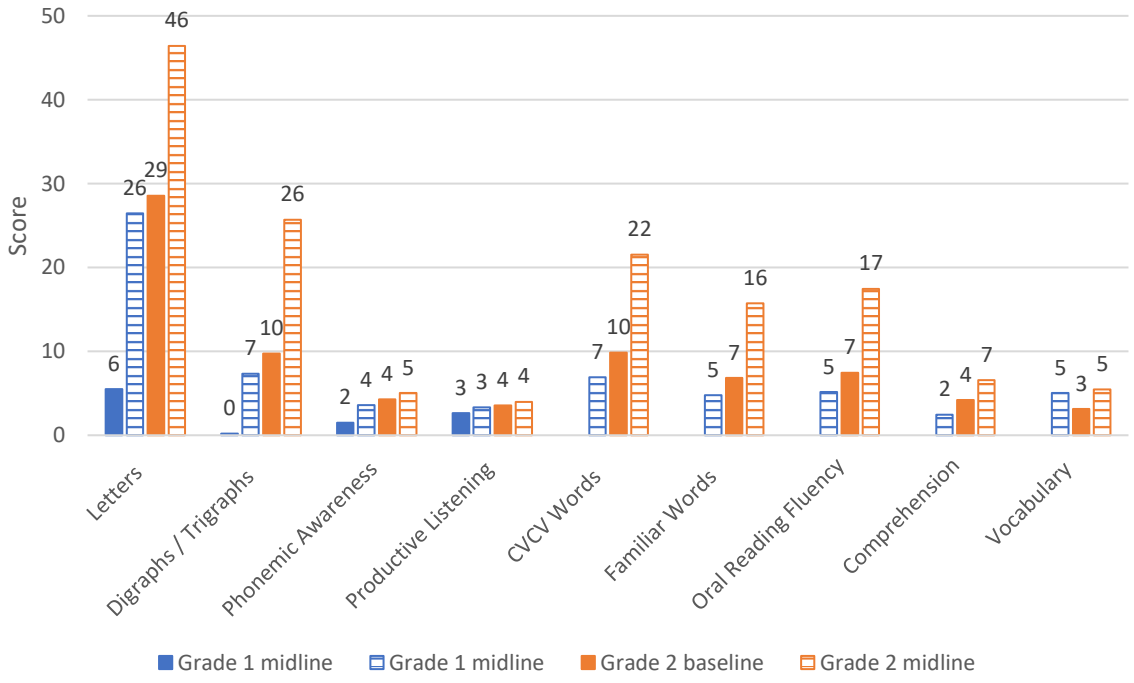
Figure 3: Percentage of control group learners scoring zero at baseline and midline by grade




Note: CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Figure 4: Average scores for control group learners at baseline and midline by grade




Note: CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table 3: Midline EGRA score for control group learners, by grade

| | Grade 1 | | | Grade 2 | | |
|---|---|---|---|---|---|---|
| | **Mean** | **SD** | **% zero** | **Mean** | **SD** | **% zero** |
| Correct letter-sounds per minute | 24.3 | 18.5 | 8 | 44.8 | 20.3 | 1 |
| Correct digraphs/trigraphs per minute | 6.2 | 10.6 | 58 | 24.3 | 19.6 | 15 |
| Phonemic awareness | 3.3 | 2.1 | 10 | 5.0 | 1.9 | 2 |
| Productive listening comprehension | 3.2 | 1.4 | 4 | 3.9 | 1.2 | 1 |
| Correct CVCV words per minute | 6.3 | 9.5 | 50 | 20.4 | 16.4 | 20 |
| Correct words per minute | 4.2 | 6.7 | 50 | 14.7 | 11.9 | 22 |
| Oral reading fluency (passage 1) | 4.5 | 7.6 | 50 | 16.7 | 14.0 | 16 |
| Reading comprehension (passage 1) | 2.3 | 3.3 | 58 | 6.4 | 4.0 | 18 |
| Expressive vocabulary | 9.1 | 3.5 | 0 | | | |
| Vocabulary | | | | 5.5 | 0.8 | 0 |
| Sentence choice | | | | 5.4 | 3.8 | 27 |
| Oral reading fluency (passage 2) | | | | 15.3 | 13.5 | 24 |
| Reading comprehension (passage 2) | | | | 4.0 | 3.1 | 27 |
| Observations | 279 | | | 278 | | |

Note: SD, standard deviation; CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Looking at selected foundational (decoding) skills by grade level suggests that despite some progress on average, the majority of learners are not on track to read with comprehension by the end of Grade 3. For example, the percentage of Grade 1's scoring zero for letter-sounds decreased from 46 per cent to 8 per cent over the year; but more than half of learners (58 per cent) could still not sound out a single digraph or trigraph at the end of the year.[24] The ability to recognize different digraphs and trigraphs is a key foundational skill for learners to progress to decoding words in isiXhosa. This is consistent with half of the learners not being able to read a single word from a short paragraph by the end of Grade 1 (with reference to the oral reading fluency sub-task).

Indicative of similar foundational skill-gaps developing among a large share of Grade 2 learners, 15 per cent of learners could not correctly identify a single digraph or trigraph after two full academic years. By the end of Grade 2, the percentage of learners who are unable to correctly read a word from two short reading passages range from 18 to 24 per cent.

## 5.2 Estimating treatment effects

*Estimation equations*

The main equation to estimate is:

$$Y_{igsd1} = \beta_0 + \beta_1(Treatment\ 1)_s + \boldsymbol{X}'_{i0}T + \mu_d + \varepsilon_{igsd1} \quad (1)$$

The outcome measure, $Y_{igsd1}$, is a midline measure of reading proficiency for learner *i* in grade *g* of school *s*; (*Treatment*) is the dummy variable indicating the treatment status for school *s*; $\boldsymbol{X}'_{is0}$ is a vector of baseline controls; $\mu_d$ is the specific school district (or strata) fixed effects;[25] and $\varepsilon_{igsd1}$ is

[24] Digraphs and trigraphs are a combination of two or three letters (like 'ph' or 'ndl') representing a single sound.

[25] As discussed in Section 4.2, random assignment of schools occurred within each of the three educational districts.

the error term clustered at the school level. The parameter of interest ($\beta_1$) is the average treatment effect on learner reading outcomes.

Random assignment and that the two groups were still balanced at midline ensures that a simple comparison of means across learners in the intervention and control schools provides unbiased estimates of the programme effects. However, regression analysis of the programme effects allows one to (i) control for any incidental pre-randomization differences between the two groups, (ii) account for non-random attrition, and (iii) increases the precision of the estimates by including variables that explain a large share of the variation in outcomes (but which are independent of treatment assignment). All results reported, therefore, control separately for each relevant measure of reading proficiency collected at baseline,[26] learner level characteristics, household assets, as well as strata fixed effects.[27]

In the cases where learners had missing data on a specific dimension of the baseline control variables, a missing value was imputed and a separate dummy variable was included to indicate missingness as a control.[28] Since schools were randomly assigned to intervention arms, standard errors are clustered at the school level (Abadie et al. 2017; Glennerster and Takavarasha 2013).

Heterogeneous treatment effects are estimated as follows:

$$Y_{igsd1} = \beta_0 + \beta_1(Treatment)_s + \beta_2(Treatment \times x)_{is} + \boldsymbol{X}'_{i0}T + \mu_d + \varepsilon_{igsd1} \qquad (2)$$

where $x_i$ is the moderating variable of interest at the learner level (here, the learner's baseline reading proficiency and/or their sex). The moderating variable is also included in the vector of baseline controls.[29]

*Intervention effects measured in standard deviations*

In a similar fashion to Cilliers et al. (2019), the main outcome of interest is a composite score of the isiXhosa reading proficiency, constructed by means of principal components analysis on the range of EGRA sub-tasks reliably assessed.[30] The results from the main estimation Equation (1) are reported in Figures 5–7. Estimated effect sizes are reported for standardized versions of the various sub-tasks to provide a sense of the relative size of the effects on each sub-task. The exercise is repeated for each grade, including grade-specific controls.[31] The darker shaded areas of the bars

[26] In cases where the analysis is done on the full sample, this includes the common tasks assessed for both Grade 1 and Grade 2 at baseline. For analyses on one grade only, the relevant grade-specific assessments at baseline were included as additional controls. See Table 2 for the layout of common and grade-specific tasks at baseline.

[27] Full lists of covariates are reported in Appendix Table A4.

[28] Missing values were assigned a value of zero if the variable is categorical, while missing observations on continuous variables were set equal to the sample mean (in a similar fashion to Cilliers et al. 2019).

[29] In the case of baseline reading proficiency, each constituent component of the composite score is controlled for separately—similar to the main estimation model.

[30] Refer to Appendix II for details on the construction of the composite reading proficiency scores.

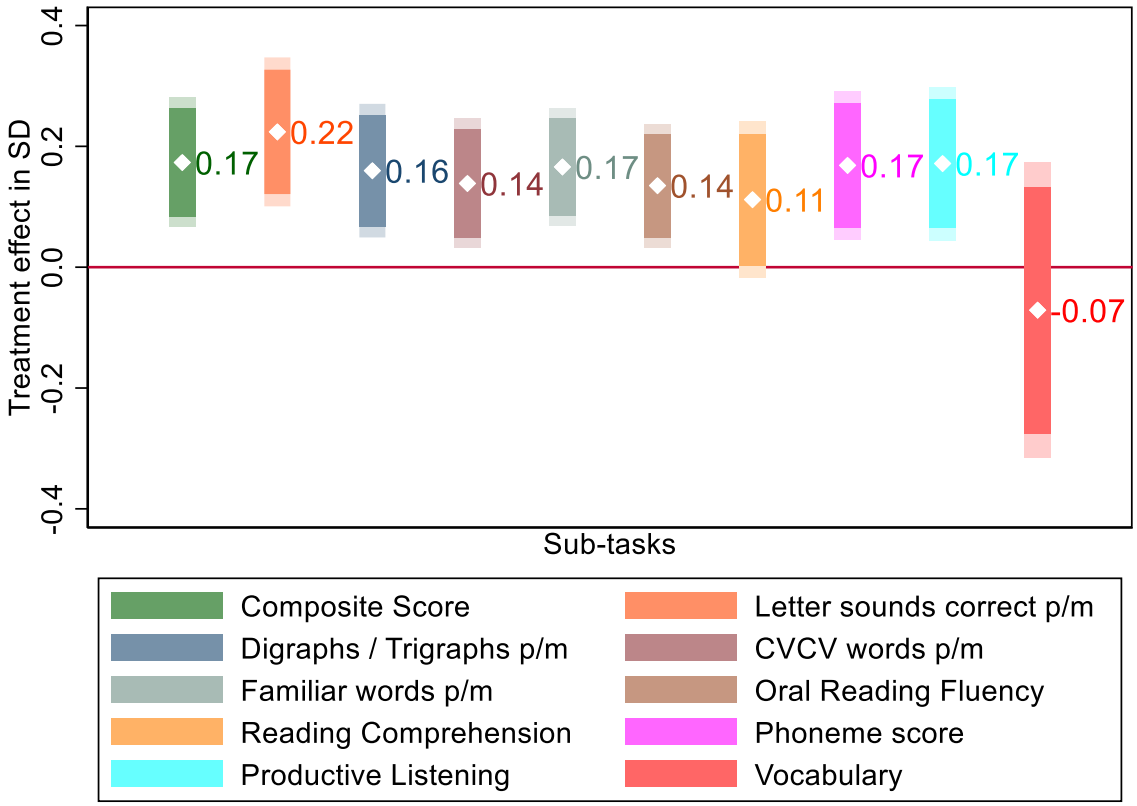
[31] A large degree of the variation in midline composite reading proficiency is explained by the baseline covariates included in the models, driven by relatively high correlations between baseline and midline assessments. For the full sample, 67 per cent of midline learner reading proficiency outcomes are explained by the baseline covariates (excluding treatment status). For Grade 2–specific models, baseline covariates explain an even higher 75 per cent of midline variation.

in Figures 5–7 display the 90 per cent confidence interval; lighter shaded fringes end at the 95 per cent confidence intervals.

Figure 5 shows the estimated effects for the full sample. The key result is that the impact of the programme on reading proficiency for the full cohort of learners over one year of exposure is 0.17 standard deviation (SD). Appendix Table A3 reports the effect size point estimates in SDs, standard errors of the estimates, the regression estimated *p*-values, as well as the randomization-based inference constructed *p*-values (as recommended by Athey and Imbens 2016) for all the results that follow.

Composite reading proficiency gains are driven by relatively large and consistent effects across the range of emergent and early literacy tasks. The effect on learner vocabulary is very noisily estimated. This is to be expected, given that learners generally scored at or close to the maximum on this task and, therefore, the task was unable to discriminate between learners. The vocabulary task is therefore excluded from subsequent analyses (included here only for transparency purposes). Reading comprehension had a positive point estimate (0.11 SD) that was not statistically significant for the full sample.

Figure 5: Treatment effects for common tasks assessed on both grades, overall and by sub-task



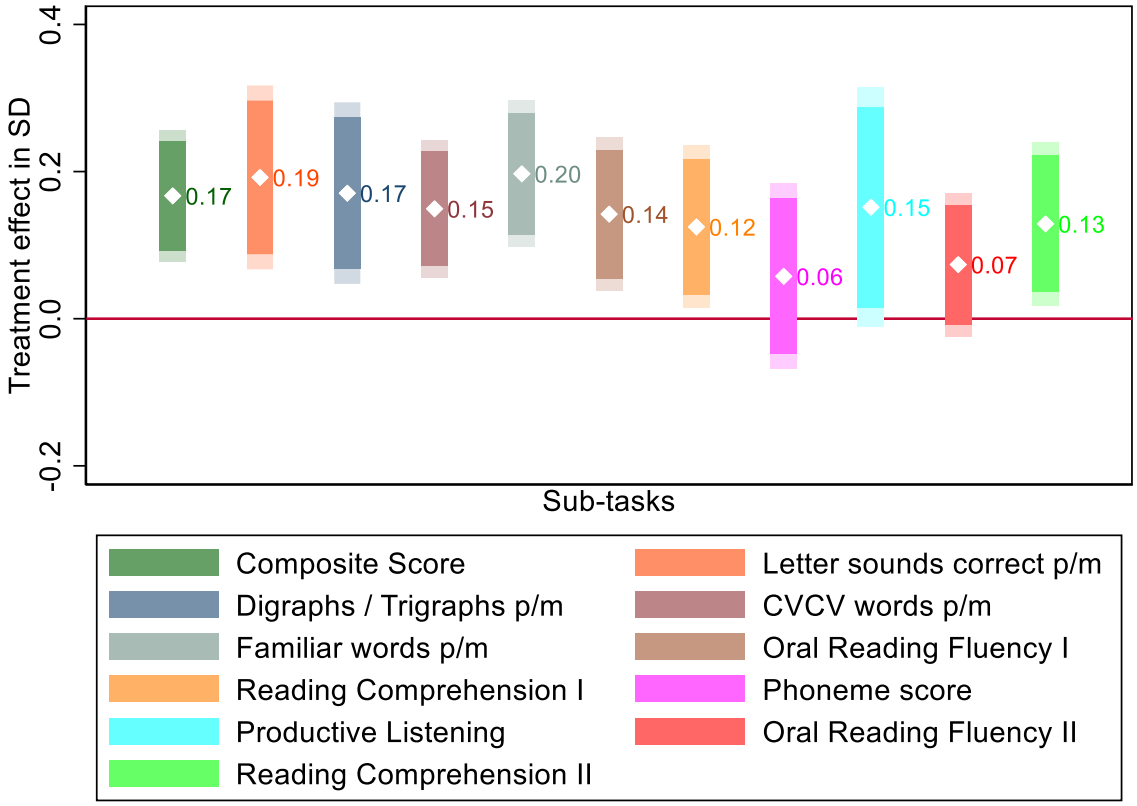

Note: SD, standard deviation; p/m, per minute; CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Figures 6 and 7 display the estimates of programme impact on Grade 1 and Grade 2 reading proficiency, respectively. The effect on Grade 1 reading proficiency is estimated to be 0.21 SD (relatively less precisely estimated[32]) and 0.16 SD for Grade 2 with no significant difference in effect by grade level.

Programme effects for Grade 1 learners are largest on emergent- and pre-literacy skills: letter-sound recognition (0.27 SD for single letters and 0.19 SD for more complex digraphs and trigraphs), phonemic awareness (0.28 SD), and productive listening comprehension (0.20 SD) (Figure 6).

Figure 6: Treatment effects for Grade 1 learners, overall and by sub-task

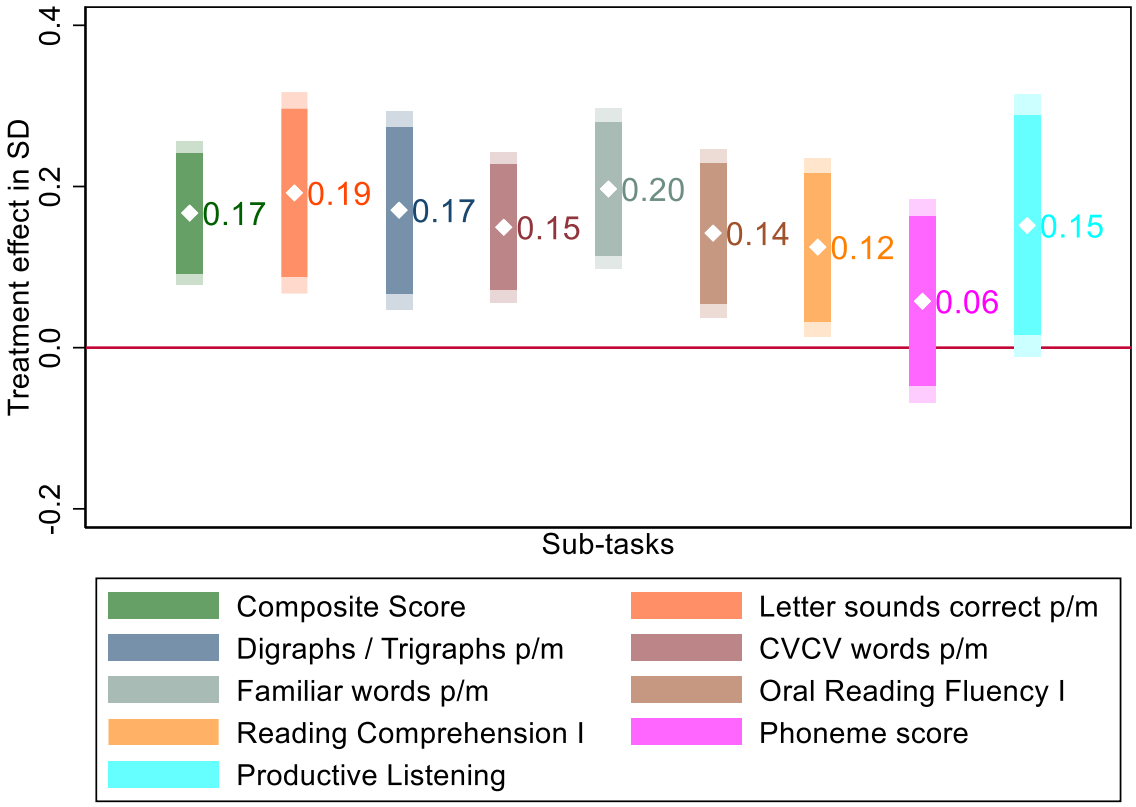



Note: SD, standard deviation; p/m, per minute; CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

[32] Effects are more noisily estimated for Grade 1 learners because (i) they could not be assessed on higher-order literacy skills right at the start of their schooling career, (ii) there were floor effects on the basic decoding tasks (such as the letter-sound recognition tasks) assessed at baseline that would have predicted downstream word and paragraph reading outcomes at midline, and (iii) that more than half the Grade 1 learners assessed scored zero on these respective basic decoding tasks at midline.

Figure 7: Treatment effects for Grade 2 learners only, overall and by sub-task




Note: SD, standard deviation; p/m, per minute; CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Programme effects for Grade 2 learners are broadly in line with those estimated for the Grade 1 cohort on foundational decoding skills, except for the phonemic awareness and productive listening sub-tasks. The latter effects are more imprecisely estimated and not statistically distinguishable from zero for Grade 2 learners. There is a relatively large difference in point estimates on phonemic awareness for Grade 1 and Grade 2 learners (0.22 SD, although we do not have enough power to say that the difference is significant). Viewed alongside the relatively larger point estimate for Grade 1 learners on another first-order literacy component, simple letter-sound recognition, these results would be consistent with the idea that letter-sound knowledge and the ability to manipulate phonemes are important foundational skills required for learners to sound out novel words and better progress towards word reading (Spaull et al. 2020: 5–6). Declining effect sizes on these two tasks over time would be in line with broader findings that alphabetic awareness has a narrow developmental window (Ouellette and Haley 2013).

Grade 2 learners' reading fluency was assessed on two separate passages at midline. The first passage was the same passage used to assess reading fluency and comprehension for Grade 2 learners at baseline, while the newly introduced second passage was slightly longer and more challenging.[33] The effect of the programme on the reading fluency for Grade 2 learners on the

[33] The original passage consisted of 41 words and had a descriptive picture. The newly introduced passage consisted of 55 words and had no descriptive picture accompanying it.

original, simpler passage was 0.14 SD, in contrast to the smaller and not significant effect of 0.05 SD on the new, more challenging passage.

Effects on reading comprehension are consistent with the notion that a certain level of emergent and pre-literacy skills must be acquired before seeing shifts among these higher-order outcomes. In contrast to the full sample, Grade 2 learners experience a statistically significant effect on reading comprehension (for both passages assessed at the Grade 2 level). Learners in treatment schools score 0.12 to 0.13 SD higher than peers in control schools.[34] Considered overall, the programme significantly shifts reading comprehension outcomes at the Grade 2 level.

*Intervention effects measured in terms of a year of learning*

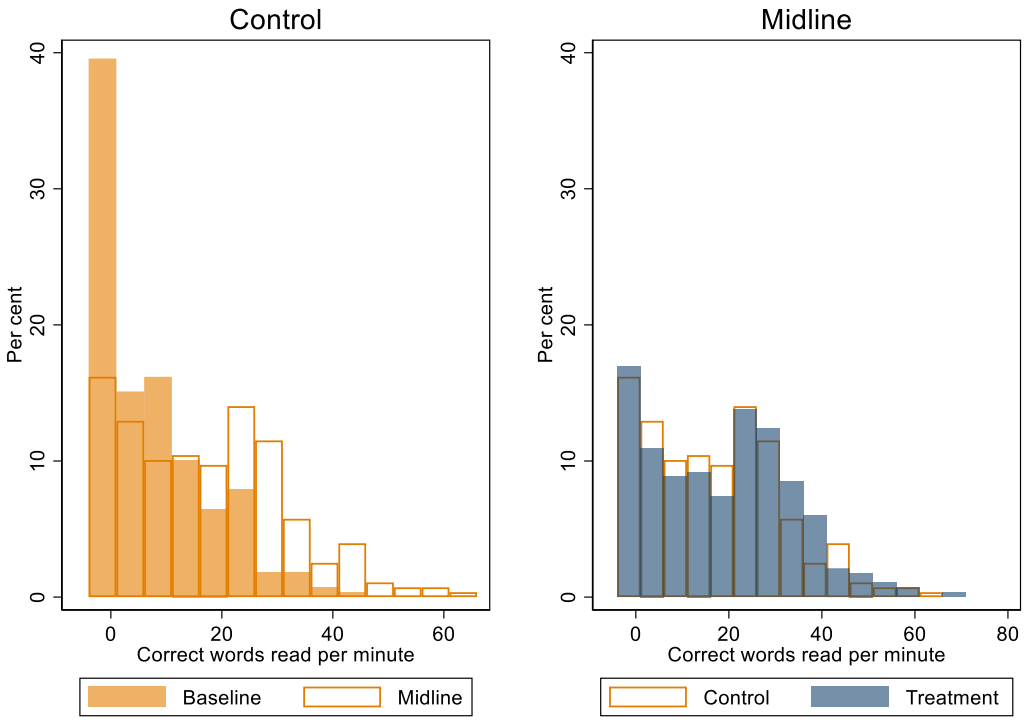
Although measures in SDs are useful in providing a relative sense of the size on programme effects, they are not very intuitive and provide less of a sense of what learning gains translate to in practice. One way to better gauge the practical significance of learning effects is to interpret point estimates relative to a year of learning in the control group. This provides an estimate of how large the additional learning gains in the intervention group are relative to the 'business-as-usual' learning gains that accrued over the academic year status quo schooling environments learners. This measurement requires a sub-task to have been assessed on the relevant grade at both baseline and midline assessments.[35]

The focus of the analysis here is on point estimates of effects relative to *mean* outcomes in the control group. However, accompanying histograms of selected sub-tasks for the respective grade levels show how these effects translate into shifts in the *distribution* of the outcomes. In Figures 8 and 9, (i) left panels indicate the growth (rightward shift) in the distribution of outcomes for learners in status quo schooling environments over the year, and (ii) right panels indicate to what extent treatment further shifted rightward (or improved) the distribution of learner outcomes after one year.

[34] The estimate on the impact on the additional sentence comprehension task was smaller and less precisely estimated. This smaller estimated effect size and the noise in the estimate might well result from the nature of the task itself. As discussed in Appendix II, the task is limited in its ability to discriminate among learners' underlying reading comprehension ability.

[35] As noted earlier, the vocabulary task is not discussed due to extreme ceiling effects. In effect, therefore, results reported are only for those sub-tasks that could be reliably measured.

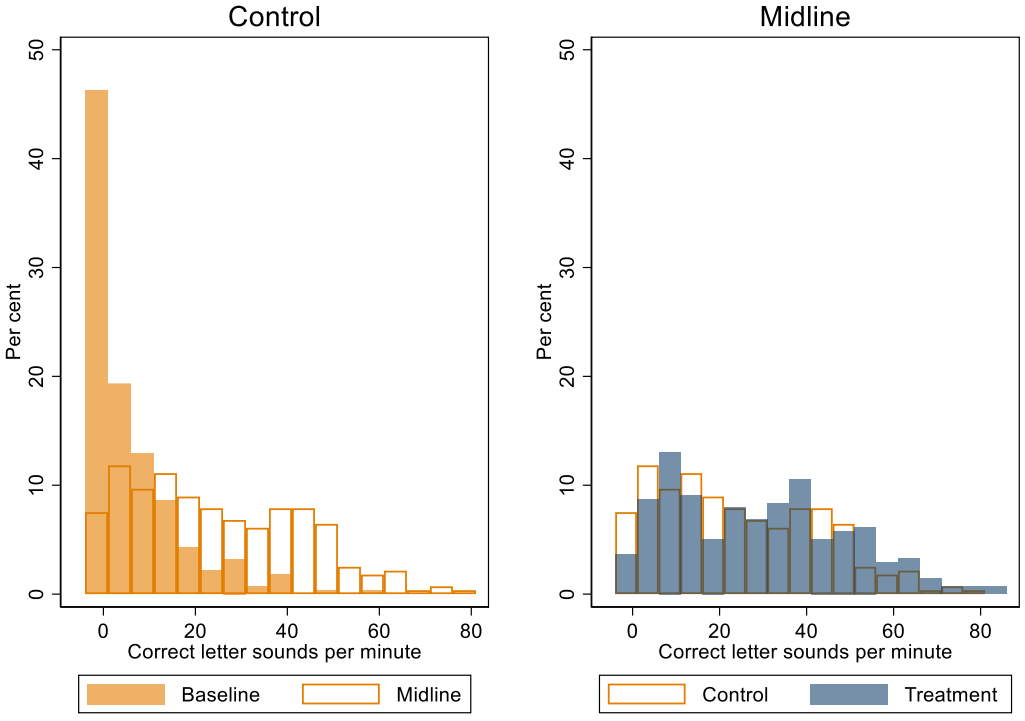
Figure 8: Histograms of shifts in Grade 2 oral reading fluency ability




Note: bin-width equals five-letter words (with a separate bin for zero scores only).

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Figure 9: Histograms of shifts in Grade 1 letter-sound recognition ability




Source: author's calculation based on Funda Wande data (see SALDRU 2019).

For each sub-task conducted on Grade 2 learners, Table 4 displays for the control group the mean outcome at baseline and the growth (or difference in means) between baseline and midline, while the second part of the table provides the estimate of the effect size (also in task-specific units, such as words read correctly per minute) and reinterprets this as a percentage of the learning that took place in control schools.

Table 4: Treatment effects in terms of a year of learning, Grade 2

| | Control (year of learning) | | Treatment effect | |
|---|---|---|---|---|
| | Baseline mean (units) | Growth in mean | Effect size (units) | % of year of learning |
| Letter-sounds | 29.0 | 15.8 | 4.2 | 27 |
| Digraphs and trigraphs | 9.3 | 15.0 | 3.0 | 20 |
| CVCV words | 9.6 | 10.8 | 2.2 | 20 |
| Familiar words | 6.7 | 8.0 | 2.1 | 26 |
| Oral reading fluency | 7.5 | 9.2 | 1.8 | 19 |
| Reading comprehension I | 4.3 | 2.1 | 0.5 | 24 |

Note: CVCV, consonant–vowel–consonant–vowel.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Effect sizes for Grade 2 learners translate to a range of between 19 and 27 per cent of a year of learning on the various sub-tasks (or roughly a 'school term's worth' of learning on a task such as reading comprehension). To demonstrate the logic for the oral reading fluency: in control schools, Grade 2 learners read fewer than 8 words correctly per minute on average from a short passage at baseline, increasing to almost 17 correct words per minute at midline. This implies that the two additional words per minute treatment effect represents a fifth of a year's worth of learning for reading fluency. The effect on the distribution of oral reading fluency scores is graphically illustrated by the greater mass at higher fluency scores for treatment school learners.

For Grade 1 learners, effect sizes on the four emergent and pre-literacy tasks reliably measured in both waves of assessment ranged between a third and almost two-thirds of a year's worth of learning (Table 5). By the end of Grade 1, control school learners more than quadrupled the correct letter-sounds identified per minute (from 6 to 24). The estimated six additional correct letter-sounds identified by intervention school learners is thus large, both relative to what was gained under the business-as-usual Grade 1 schooling environment, and with respect to the four letter-sounds-per-minute treatment effect for Grade 2 learners. Again, Figure 9 shows graphically what this effect translates to for learners also in treatment schools in terms of a greater share of learners at higher letter-sound scores.

Table 5: Treatment effects in terms of a year of learning, Grade 1

| | Control (year of learning) | | Treatment effect | |
|---|---|---|---|---|
| | Baseline mean (units) | Growth in mean | Effect size (units) | % of year of learning |
| Letter-sounds | 6.0 | 18.3 | 6.0 | 33 |
| Digraphs and trigraphs | 0.2 | 6.0 | 3.5 | 58 |
| Phonemic awareness | 1.6 | 1.7 | 0.6 | 36 |
| Productive listening | 2.6 | 0.6 | 0.3 | 46 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

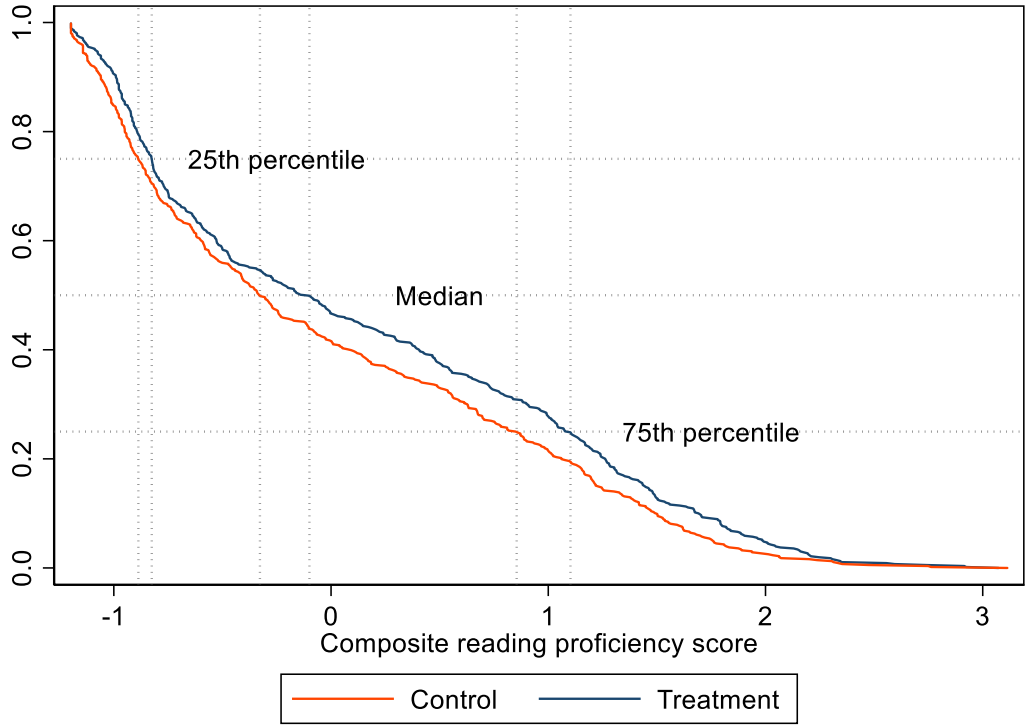
### 5.3 Programme effects across the distribution of reading proficiency

This section investigates whether the intervention had any differential effects based on learner's reading proficiency levels [i.e. whether there were any heterogeneous treatment effects Equation (2)]. Results from other structured pedagogical programmes similar to the Funda Wande intervention (like the EGRS and Reading Catch-Up Study in South Africa and the Tusome studies in Kenya), and educational interventions more broadly, often find differential effects on certain sub-groups. A general theme from the literature. is that many programmes have the greatest impact on the already better performing learners, those who are better equipped to take advantage of the programme (Cilliers et al. 2019, Fleisch et al. 2017). Alternatively, some programmes seem to have the greatest impact on the weakest learners, those who often lag behind curriculum prescribed levels of learning and still need development in certain foundational skills. Given the data and sample size available, the analysis of heterogeneous treatment effects at this stage is limited to arguably the most pertinent question from a programme design and policy point of view.[36]

There is more than one way to see whether intervention effects differ across the distribution of reading proficiency. One method is to compare the entire midline distributions of literacy outcomes for the intervention and control groups to see whether the 'gap' in outcomes varies substantially at different points across the distribution. The distribution of the midline composite score is shown separately by treatment status in Figure 10. For each level of the composite score, the lines indicate the proportion of learners with scores at that level or greater. Intervention school learners outperform those in the control schools across the distribution. Comparing learners at the same point in the two groups' respective distributions of midline reading proficiency scores, learners in the intervention group score 0.06 SD higher at the 25th percentile, 0.23 SD higher at the median, and 0.25 SD higher at the 75th percentile. The same result generally holds across both grades for sub-tasks reliably measured (see Appendix Figure E1 for letter-sounds in Grade 1 and Appendix Figure E2 for familiar word reading in Grade 2).

[36] The risk of investigating multiple possible sources of differential treatment effects is called data mining. In other words, if one tests for differential treatment effects by a whole range of characteristics and sub-combination of them, the probability of finding a statistically significant result just by chance is increased.

Figure 10: Distribution of midline composite scores by treatment status (full sample of both grades combined)



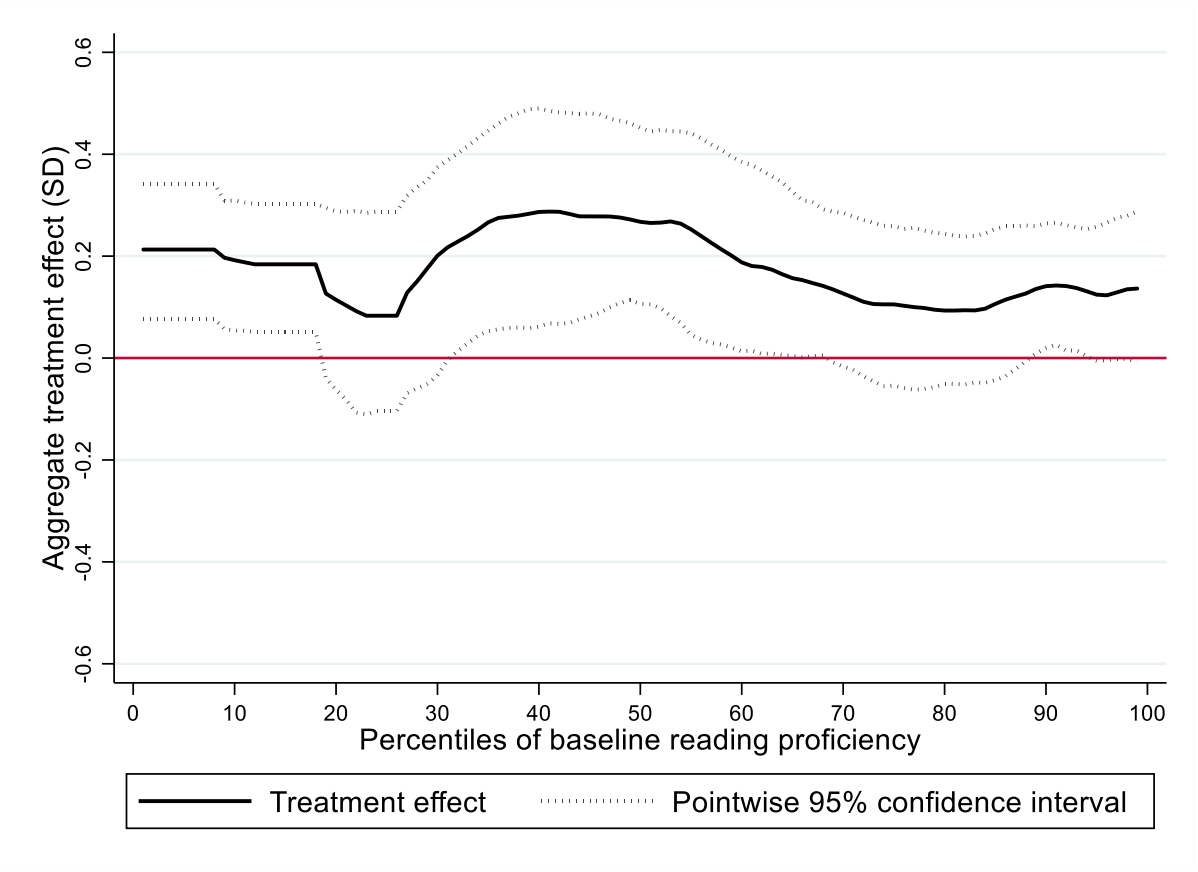

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

A second approach is to investigate whether the impact of the intervention differs depending on learner baseline reading proficiency. Figure 11 displays the estimated programme impact at each percentile of learner baseline reading proficiency, indicating that programme effects are positive and fairly constant across the distribution of where learners' rank in terms of baseline reading proficiency.[37] This result is consistent with the comparisons of the scores distributions of midline reading proficiency between treatment and control groups.[38] Robustness checks confirming these results are reported in Appendix III.

[37] More precisely, Figure 11 displays the local polynomial regression estimates of the effect size at each percentile of baseline reading proficiency. Estimates are obtained by first creating a value-added measure of reading proficiency, constructed by subtracting each learner's predicted score (based on the range of baseline covariates) from their actual midline reading proficiency score. The value-added measure of the intervention is therefore equal to the residual, which is assumed to be attributable to the learner's intervention status and other learner-level idiosyncrasies. Second, value-added measures are then estimated for all learners based on their percentile rank of baseline reading proficiency, separately for intervention and control groups. The intervention effect estimates at each point in the distribution the difference between fitted values of each respective control percentile from the corresponding intervention percentile of student baseline reading proficiency. Finally, a pointwise 95 per cent confidence interval is created using a bootstrap resampling of baseline percentiles (500 iterations), stratifying by sub-districts and clustering at the school level.

[38] This estimate of effects across the distribution of baseline scores should yield similar results to those above, at least to the extent that treatment impacts are non-negative and fairly consistent across the distribution of baseline reading proficiency.

Figure 11: Non-parametric intervention effects by baseline reading proficiency




Source: author's calculation based on Funda Wande data (see SALDRU 2019).

## 6 Discussion

The Funda Wande intervention had a 0.17 SD impact on learner reading proficiency after one year of implementation. The programme impacts are positive across all sub-tasks that were reliably measured. In practical terms, learning gains on the sub-tasks on which the intervention had a positive effect translated to between 20 and 27 per cent of a year's worth of learning for Grade 2 learners. Grade 1 learners in intervention schools gained even more over their peers for the emergent- and pre-literacy skills on which the programme had positive effects. For letter recognition tasks, phonemic awareness, and productive listening comprehension skills, these Grade 1 learner outcomes improved between 33 and 58 per cent more than the 'business-as-usual' development in control schools.

When investigated by grade, certain dynamics of the learning trajectories of learners in the different grades and how these relate to the programme impacts come to the fore. The intervention effects on Grade 1 learners' foundational skills (letter-sound recognition and phonemic awareness) are particularly large, both relative to the impacts on other Grade 1 literacy skills and the impacts on similar skills for Grade 2 learners. For Grade 2 learners, the impacts of the intervention are more consistent across foundational (letter-sound recognition) and higher-order literacy skills (such as word recognition, reading fluency, and reading comprehension outcomes), but not significant on phonemic awareness. These dynamics suggest that certain foundational decoding skills, such as letter-sound knowledge, phonemic awareness, and word recognition are important for learners to master before they can effectively progress towards reading passages fluently. The results support

the idea that learners require a range of foundational literacy abilities before they can read with some level of speed and accuracy (i.e. fluency), and in turn, then need to read with a certain minimum level of fluency to comprehend what they are reading.[39]

Encouraging from a policy perspective is that the intervention seems to have fairly consistent positive effects for learners across the distribution of reading proficiency. This is a significant finding given that benefits from similar interventions (and early education interventions, more broadly) are often concentrated among learners with higher baseline literacy levels (Cilliers et al. 2019; Fleisch et al. 2017; Glewwe et al. 2009). The in-depth qualitative classroom observations planned for subsequent rounds of evaluation will shed more light on the potential mechanisms driving these findings.

Questions regarding external validity and moving towards implementation at scale for these structured pedagogy programmes are arguably best assessed by replication of similar interventions (i.e. functioning on the same underlying rationale) across different contexts learning and iterating along the way (Banerjee et al. 2017; Piper et al. 2018b). In the South African policy space, this could imply replication across different provinces, time periods, partnering provincial bureaucracies, implementing institutions, and languages of instruction. To date, the state of evidence from EGRS I (Cilliers et al. 2019), EGRS II (Kotze et al. 2019), and Funda Wande coaching interventions suggests that structured pedagogy programmes can be effective in more than one province (the North West, Mpumalanga, and the Eastern Cape) and in more than one language of instruction (Setswana home language, English as first additional language, and isiXhosa home language).

Given the consistent positive impacts found for the structured pedagogy programmes that have been assessed in the South African context, these programmes are arguably past the proof of concept stage and have shown some degree of consistency in shifting learner reading outcomes across these different contexts. However, many questions that remain are related to implementing a version of these programmes at scale: (i) What role do the individual inputs and their combinations play in driving programme impacts (e.g., the provision of home-language resources and instruction)? (ii) How cost effective are different iterations of the class of intervention? (iii) How would the relation of programme costs and benefits change if it were implemented at scale within a national-level public education system in future? These questions provide fruitful avenues for future research.

## References


Abadie, A., S. Athey, G.W. Imbens, and J. Wooldridge (2017). 'When Should You Adjust Standard Errors for Clustering?'. NBER Working Paper 24003. Cambridge, MA: National Bureau of Economic Research (NBER). https://doi.org/10.3386/w24003

Angrist, J., and J.S. Pischke (2009). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press. https://doi.org/10.1515/9781400829828

Ardington, C. (2019). 'Impact Evaluation of the Funda Wande Coaching Intervention: Baseline Findings'. Cape Town: SALDRU, University of Cape Town.


[39] See Spaull et al. (2020: 5–8) for a discussion of the hierarchical nature of language acquisition and its applicability to learning to read African languages in the South African context.

Ardington, C., and C.J. Meiring (2020). 'Impact Evaluation of Funda Wande Coaching Intervention Midline Findings'. Cape Town: SALDRU, University of Cape Town. Available at: https://fundawande.org/news/impact-evaluation-of-funda-wande-coaching-intervention-midline-findings-23 (accessed 7 July 2020).

Athey, S., and G.W. Imbens (2016). 'The Econometrics of Randomized Experiments'. In A.V. Banerjee and E. Duflo (eds), *Handbook of Economic Field Experiments: Volume 1* (Chapter 3, pp. 73–140). Cambridge, MA: Elsevier. https://doi.org/10.1016/bs.hefe.2016.10.003

Ball, J. (2010). *Enhancing Learning of Children from Diverse Language Backgrounds: Mother Tongue-Based Bilingual or Multilingual Education in the Early Years*. Paris: UNESCO. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000212270 (accessed October 2020).

Banerjee, A., R. Banerji, J. Berry, E. Duflo, H. Kannan, S. Mukerji, M. Shortland, and M. Walton (2017). 'From Proof of Concept to Scalable Policies: Challenges and Solutions, With an Application'. *Journal of Economic Perspectives*, 31(4): 73–102. https://doi.org/10.1257/jep.31.4.73

Bau, N., and J. Das (2020). 'Teacher Value-Added in a Low-Income Country'. *American Economic Review: Economic Policy*, 12(1): 62–96. https://doi.org/10.1257/pol.20170243

Bold, T., D. Filmer, G. Martin, E. Molina, C. Rockmore, J. Svensson, and W. Wane (2017). 'What Do Teachers Know and Do? Does It Matter? Evidence from Primary Schools in Africa'. World Bank Policy Research Working Paper 7956. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-7956

Brunette, T., B. Piper, R. Jordan, S. King, and R. Nabacwa (2019). 'The Impact of Mother Tongue Reading Instruction in Twelve Ugandan Languages and the Role of Language Complexity, Socioeconomic Factor, and Program Implementation'. *Comparative Education Review*, 63(4): 591–612. https://doi.org/10.1086/705426

Bruns, B., and J. Luque (2014). 'Great Teachers: How to Raise Student Learning in Latin America and the Caribbean'. Washington, DC: World Bank. https://doi.org/10.1596/978-1-4648-0151-8

Bruhn, M., and D. McKenzie (2009). 'In Pursuit of Balance: Randomization in Practice in Development Field Experiments'. *American Economic Journal: Applied Economics*, 1(4): 200–32. https://doi.org/10.1257/app.1.4.200

Carnoy, M., L. Chisholm, and B. Chilisa (2012). *The Low Achievement Trap: Comparing Schooling in Botswana and South Africa*. Pretoria: HSRC Press.

Chetty, R., J. Friedman, and J. Rockoff (2014). 'Measuring the Impacts of Teachers II: Teacher Value Added and Student Outcomes in Adulthood'. *American Economic Review*, 104(9): 2633–79. https://doi.org/10.1257/aer.104.9.2633

Cilliers, J., B. Fleisch, C. Prinsloo and S. Taylor (2019). 'How to Improve Teaching Practice? An Experimental Comparison of Centralized Training and In-Classroom Coaching'. *Journal of Human Resources*, 55: 926–62. https://doi.org/10.3368/jhr.55.3.0618-9538R1

City, E.A., R.F. Elmore, S.E. Fiarman, and L. Teitel (2009). *Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning*. Cambridge, MA: Harvard Education Press.

Conn, K.M. (2017). 'Identifying Effective Education Interventions in Sub-Saharan Africa: A Meta-Analysis of Impact Evaluations'. *Review of Educational Research*, 87(5): 863–98. https://doi.org/10.3102/0034654317712025

Cummins, J. (2001). *Negotiating Identities: Education for Empowerment in a Diverse Society*. 2nd edn. Los Angeles: California Association for Bilingual Education.

Cummins, J. (2007). 'Rethinking Monolingual Instructional Strategies in Multilingual Classrooms.' *Canadian Journal of Applied Linguistics*, 10(2): 221–40.

DBE (2014). *Second Detailed Indicator Report for Basic Education Sector*. Pretoria: Department of Basic Education (DBE).

DBE (2015). Action Plan to 2019: Towards the Realisation of Schooling 2030. Pretoria: Department of Basic Education (DBE).

DBE (2018). *School Monitoring Survey 2017/18: Quantitative and Qualitative Summary Report*. Pretoria: Department of Basic Education (DBE).

Dechaisemartin, T. (2013). 'Writing Matters: The Neglect of Writing in South African Schools. In N. Taylor, S. Van der Berg, and T. Mabogoane (eds), *Creating Effective Schools: Report of South African's National Effectiveness Study* (pp. 159–180). Cape Town: Pearson.

Evans, D.K., and A. Popova (2016). 'What Really Works to Improve Learning in Developing Countries? An Analysis of Divergent Findings in Systematic Reviews'. *The World Bank Research Observer*, 31(2): 242–70. https://doi.org/10.1093/wbro/lkw004

Fleisch, B., and V. Schöer (2014). 'Large-Scale Instructional Reform in the Global South: Insights from the Mid-Point Evaluation of the Gauteng Primary Language and Mathematics Strategy'. *South African Journal of Education*, 34(3): 1–11. https://doi.org/10.15700/201409161040

Fleisch, B., V. Schöer, G. Roberts, and A. Thornton (2016). 'System-Wide Improvement of Early-Grade Mathematics: New Evidence from the Gauteng Primary Language and Mathematics Strategy'. *International Journal of Educational Development*, 49: 157–74. https://doi.org/10.1016/j.ijedudev.2016.02.006

Fleisch, B., S. Taylor, V. Schöer, and T. Mabogoane (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

Funda Wande (2018). Funda Wande: Reading for Meaning [Project Document]. Cape Town: Funda Wande.

Ganimian, A.J., and R.J. Murnane (2016). 'Improving Educational Outcomes in Developing Countries: Lessons from Rigorous Impact Evaluations'. NBER Working Paper 20284. Cambridge, MA: National Bureau of Economic Research (NBER). Available at: http://www.nber.org/papers/w20284 (accessed 7 February 2019).

Glennerster, R., and K. Takavarasha (2013). *Running Randomized Evaluations: A Practical Guide*. Princeton, NJ/Oxford: Princeton University Press. https://doi.org/10.2307/j.ctt4cgd52

Glewwe, P., E.A. Hanushek, S. Humpage, and R. Ravina (2014). 'School Resources and Educational Outcomes in Developing Countries: A Review of the Literature from 1990 to 2010'. In P. Glewwe (ed.), *Education Policy in Developing Countries* (pp. 13–64). Chicago, IL: University of Chicago Press. https://doi.org/10.7208/chicago/9780226078854.003.0002

Glewwe, P., M. Kremer, and S. Moulin (2009). 'Many Children Left Behind? Textbooks and Test Scores in Kenya'. *American Economic Journal: Applied Economics*, 1(1): 112–35. https://doi.org/10.1257/app.1.1.112

Glewwe, P., and K. Muralidharan (2016). 'Improving Education Outcomes in Developing Countries: Evidence, Knowledge Gaps, and Policy Implications'. In S. Machin, L. Woessmann, and E.A. Hanushek (eds), *Handbook of the Economics of Education: Volume 5* (Chapter 10, pp. 653–743). Cambridge, MA: Elsevier. https://doi.org/10.1016/B978-0-444-63459-7.00010-5

Gove, A., and Wetterberg, A. (eds.) (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2011.bk.0007.1109

Hanushek, E., and S. Rivkin (2010). 'Generalizations about Using Value-Added Measures of Teacher Quality'. *American Economic Review*, 100(2): 267–71. https://doi.org/10.1257/aer.100.2.267

Heckman, J.J. (2006). 'Skill Formation and the Economics of Investing in Disadvantaged Children'. *Science*, 312: 1900–02. https://doi.org/10.1126/science.1128898

Hoadley, U. (2012). 'What Do We Know about Teaching and Learning in South African Primary Schools?'. *Education as Change*, 16(2): 187–202. https://doi.org/10.1080/16823206.2012.745725

Hoadley, U. (2016). 'A Review of the Research Literature on Teaching and Learning in the Foundation Phase in South Africa'. Research on Socioeconomic Policy (ReSEP) Working Papers: 05/16. Stellenbosch University. Available at: https://resep.sun.ac.za/a-review-of-the-research-literature-on-teaching-and-learning-in-the-foundation-phase-in-south-africa/ (accessed 15 April 2020).

Hoadley, U. (2018). *Pedagogy in Poverty: Lessons from Twenty Years of Curriculum Reform in South Africa*. Abingdon/New York: Routledge. https://doi.org/10.4324/9781315680927

Hoadley, U., and J. Gallant (2016). 'An Analysis of the Grade 3 Department of Basic Education Workbooks as Curriculum Tools'. *South African Journal of Childhood Education*, 6(1): 1–28. https://doi.org/10.4102/sajce.v6i1.400

Howie, S., C. Combrink, K. Roux, M. Tshele, G. Mokoena, and N. Palane (2017). *Progress in International Reading Literacy Study 2016: South African Children's Reading Literacy Achievement*. Pretoria, South Africa: Centre for Evaluation and Assessment, University of Pretoria.

Imbens, G.W., and D. Rubin (2015). *Causal Inference in Statistics, Social, and Biomedical Sciences: An Introduction* (Chapter 14). Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781139025751

Kennedy, M. (2016). 'How Does Professional Development Improve Teaching?' *Review of Educational Research*, 86(4): 945–80. https://doi.org/10.3102/0034654315626800

Kerwin, J.T., and R.L. Thornton (2019). 'Making the Grade: The Sensitivity of Education Program Effectiveness to Input Choices and Outcome Measures'. https://doi.org/10.2139/ssrn.3002723

Kremer, M., C. Brannen, and R. Glennerster (2013). 'The Challenge of Education and Learning in the Developing World'. *Science*, 340: 297–300. https://doi.org/10.1126/science.1235350

Kotze, J., B. Fleisch, and S. Taylor (2019). 'Alternative Forms of Early Grade Instructional Coaching: Emerging Evidence from Field Experiments in South Africa'. *International Journal of Educational Development*, 66(1): 203–13. https://doi.org/10.1016/j.ijedudev.2018.09.004

Kraft, M.A., D. Blazar, and D. Hogan (2018). 'The Effect of Teacher Coaching on Instruction and Achievement: A Meta-Analysis of the Causal Evidence'. *Review of Educational Research*, 88(4): 547–88. https://doi.org/10.3102/0034654318759268

Lucas, A.M., P.J. McEwan, M. Ngware, and M. Oketch (2014). 'Improving Early-Grade Literacy in East Africa: Experimental Evidence from Kenya and Uganda'. *Journal of Policy Analysis and Management*, 33(4): 950–76. https://doi.org/10.1002/pam.21782

McEwan, P.J. (2015). 'Improving Learning in Primary Schools of Developing Countries: A MetaAnalysis of Randomized Experiments'. *Review of Educational Research*, 85(3): 353–94. https://doi.org/10.3102/0034654314553127

Menendez, A., and C. Ardington (2018). *Impact Evaluation of USAID/South Africa Story Powered School Program—Baseline. Technical Report*. Available at: https://nalibali.org/research (accessed 5 October 2019).

Motala, S., and D. Carel (2019). 'Education Funding and Equity in South African Schools'. In N. Spaull and J.D. Jansen (eds), *South African Schooling: The Enigma of Inequality* (Chapter 4, pp. 67–86). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-18811-5_4

NEEDU (2013). *NEEDU National Report 2012: The State of Literacy Teaching and Learning in the Foundation Phase*. Technical report. Pretoria: National Education and Evaluation Development Unit (NEEDU), Department of Basic Education.

Ouellette, G.P., and A. Haley (2013). 'One Complicated Extended Family: The Influence of Alphabetic Knowledge and Vocabulary on Phonemic Awareness'. *Journal of Research in Reading*, 36(1): 29–41. https://doi.org/10.1111/j.1467-9817.2010.01486.x

Piper, B. (2009). *Integrated Education Program: Impact Study of SMRS Using Early Grade Reading Assessment in Three Provinces in South Africa*. Pretoria: United States Agency for International Development (USAID).

Piper, B., J. Destefano, E.M. Kinyanjui, and S. Ong'Ele (2018a). 'Scaling up successfully: Lessons from Kenya's Tusome National Literacy Program'. *Journal of Educational Change*, 19: 293– 321. https://doi.org/10.1007/s10833-018-9325-4

Piper, B., and M. Korda (2011). *Early Grade Reading Assessment (EGRA) Plus: Liberia: Program Evaluation Report*. Research Triangle Park, NC: RTI International.

Piper, B., D. Kwayumba, A. Oyanga, and E. Jepkemei (2015). *The Primary Math and Reading (PRIMR) Initiative Endline Impact Evaluation on the DFID Kenya Rural Expansion Programme*. Prepared for DFID Kenya under contract 202657-108. RTI International, Research Triangle Park, NC.

Piper, B., S.S. Zuilkowski, M. Dubeck, E. Jepkemei, and S.J. King (2018b). 'Identifying the Essential Ingredients to Literacy and Numeracy Improvement: Teacher Professional Development and Coaching, Student Textbooks, and Structured Teachers' Guides'. *World Development*, 106: 324–36. https://doi.org/10.1016/j.worlddev.2018.01.018

Piper, B., S.S. Zuilkowski, and A. Mugenda (2014). 'Improving Reading Outcomes in Kenya: First-Year Effects of the PRIMR Initiative'. *International Journal of Educational Development*, 37: 11–21. https://doi.org/10.1016/j.ijedudev.2014.02.006

Piper, B., S.S. Zuilkowski, and S. Ong'Ele (2016). 'Implementing Mother Tongue Instruction in the Real World: Results from a Medium-Scale Randomised Control Trial in Kenya'. *Comparative Education Review*, 60(4): 776–807. https://doi.org/10.1086/688493

Popova, A., D.K. Evans, M.E. Breeding, and V. Arancibia (2018). 'Teacher Professional Development around the World: The Gap between Evidence and Practice'. Policy Research Working Paper 8572. Washington, DC: World Bank. https://doi.org/10.1596/1813-9450-8572

Pretorius, E.J., and N. Spaull (2016). 'Exploring Relationships between Oral Reading Fluency and Reading Comprehension amongst English Second Language Readers in South Africa'. *Reading and Writing*, 29: 1449–71. https://doi.org/10.1007/s11145-016-9645-9

Prinsloo, C. (2008). 'Practice Makes Perfect—Limpopo Study Shows the Value of Exercising Literary Muscles'. *HSRC Review*, 6(4): 5–6.

Reddy, V., C. Prinsloo, T. Netshitangani, R. Moletsane, and J. Van Rensberg (2010). *An Investigation into Education Leave in the South African Ordinary Public Schooling System*. Pretoria: Human Sciences Research Council.

SALDRU (2019). Funda Wande Baseline and Midline Evaluations [Data file]. Cape Town: SALDRU, University of Cape Town.

Sailors, M., J.V. Hoffman, P.D. Pearson, S.N. Beretvas, and B. Matthee (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

Snilstveit, B., J. Stevenson, R. Menon, D. Philips, E. Gallagher, M. Geleen, H. Jobse, T. Schmidt, and E. Jimenez (2016). *The Impact of Education Programmes on Learning and School Participation in Low- and Middle-Income Countries*. Systematic Review Summary 7. London: International Initiative for Impact Evaluation (3ie). https://doi.org/10.23846/SRS007

Snow, C. (2017). 'Early Literacy Development and Instruction: An Overview'. In N. Kucirkova, C. Snow, V. Grøver, and C. McBride-Chang (eds), *The Routledge International Handbook of Early Literacy Education: A Contemporary Guide to Literacy Teaching and Interventions in a Global Context* (Chapter 1, pp. 9). New York, NY: Routledge. https://doi.org/10.4324/9781315766027-2

Spaull, N. (2011). *Primary School Performance in Botswana, Mozambique, Namibia and South Africa: A Comparative Analysis of SACMEQ III*. South African Consortium for Monitoring Educational Quality.

Spaull, N. (2013). 'Poverty and Privilege: Primary School Inequality in South Africa'. *International Journal of Educational Development*, 33(5): 436–47. https://doi.org/10.1016/j.ijedudev.2012.09.009

Spaull, N. (2015). 'Accountability and Capacity in South African Education'. *Education as Change*, 19(3): 113–42. https://doi.org/10.1080/16823206.2015.1056199

Spaull, N. (2019). 'Equity: A Price Too High to Pay'. In N. Spaull and J.D. Jansen (eds), *South African Schooling: The Enigma of Inequality* (Chapter 1, pp. 1–24). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-18811-5_1

Spaull, N., and U. Hoadley (2017). 'Getting Reading Right: Building Firm Foundations'. In L. Jamieson, L. Berry, and L. Lake (eds), *South African Child Gauge 2017* (pp. 77–83). Cape Town: Children's Institute: University of Cape Town.

Spaull, N., and J. Kotze (2015). 'Starting Behind and Staying Behind in South Africa. The Case of Insurmountable Learning Deficits in Mathematics'. *International Journal of Educational Development*, 41: 13–24. https://doi.org/10.1016/j.ijedudev.2015.01.002

Spaull, N., and E. Pretorius (2019). 'Still Falling at the First Hurdle: Early Grade Reading in South Africa'. In N. Spaull and J.D. Jansen (eds), *South African Schooling: The Enigma of Inequality* (Chapter 8, pp. 147–68). Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-18811-5_8

Spaull, N., E. Pretorius, and N. Mohohlwane (2020). 'Investigating the Comprehension Iceberg: Developing Empirical Benchmarks for Early-Grade Reading in Agglutinating African Languages'. *South African Journal of Childhood Education*, 10(1): 1–14. https://doi.org/10.4102/sajce.v10i1.773

Taylor, N., and S. Taylor (2013). 'Teacher Knowledge and Professional Habitus'. In N. Taylor, S. Van den Berg, and T. Mabogoane (eds), *Creating Effective Schools: Report of South Africa's National Schools Effectiveness Study* (pp. 31–60). Cape Town: Pearson.

Taylor, N., S. Van der Berg, and T. Mabogoane (2013). *Creating Effective Schools*. Cape Town: Pearson.

Taylor, S. (2011). 'Uncovering Indicators of Effective School Management in South Africa Using the National School Effectiveness Study'. Stellenbosch Economic Working Papers 10/11. Stellenbosch University. Available at: https://resep.sun.ac.za/wp-content/uploads/2017/10/wp-10-2011.pdf (accessed 5 October 2020).

Taylor, S., and M. Von Fintel (2016). 'Estimating the Impact of Language of Instruction in South African Primary Schools: A Fixed Effects Approach'. *Economics of Education Review*, 50: 75–89. https://doi.org/10.1016/j.econedurev.2016.01.003

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 der Berg, S., N. Spaull, G. Wills, M. Gustafsson, and J. Kotze (2016). *Identifying Binding Constraints in Education*. Report commissioned by the South African Presidency and funded by the European Union's Programme to Support Pro-Poor Policy Development (PSPPD) initiative. Pretoria: PSPPD. https://doi.org/10.2139/ssrn.2906945

Venkat, H., and N. Spaull (2015) 'What Do We Know about Primary Teachers' Mathematical Content Knowledge in South Africa? An Analysis of SACMEQ 2007'. *International Journal of Educational Development*, 41: 121–130. https://doi.org/10.1016/j.ijedudev.2015.02.002

WWC (2020). *Standards Handbook*. Version 4.1. Washington, D.C.: Institute of Education Sciences. Available at: https://ies.ed.gov/ncee/wwc/Handbooks (accessed 15 October 2020).

World Bank (2018). *An Incomplete Transition: Overcoming the Legacy of Exclusion in South Africa*. Washington, DC: Systematic Country Diagnostic.

Zuilkowski, S.S., and B. Piper (2017). 'Instructional Coaching in Kenya: Supporting Teachers to Improve Literacy Outcomes'. In M. Akiba and G.K. LeTendre (eds), *International Handbook of Teacher Quality and Policy* (pp. 505–16). New York, NY: Routledge. https://doi.org/10.4324/9781315710068-33

# 8 Appendix

## 8.1 Appendix tables and figures

Table A1: Baseline balance in Grade 1 EGRA scores

| | Treatment | | | Control | | | p-value |
|---|---|---|---|---|---|---|---|
| | **Mean** | **s.d.** | **N** | **Mean** | **s.d.** | **N** | |
| Receptive listening comprehension | 9,55 | 0,85 | 301 | 9,35 | 1,19 | 294 | 0,14 |
| Productive listening comprehension | 2,73 | 1,54 | 301 | 2,58 | 1,53 | 294 | 0,36 |
| Expressive vocabulary | 10,35 | 3,91 | 301 | 10,95 | 4,18 | 294 | 0,22 |
| Letter sounds per minute | 5,00 | 9,07 | 301 | 6,01 | 9,60 | 294 | 0,31 |
| Digraphs and trigraphs per minute | 0,15 | 1,21 | 301 | 0,24 | 1,37 | 294 | 0,55 |
| Phonemic awareness | 1,41 | 1,85 | 300 | 1,61 | 1,80 | 294 | 0,30 |
| Word choice | 1,68 | 1,71 | 301 | 1,90 | 1,58 | 293 | 0,19 |
| Rapid automatized naming | 12,30 | 3,92 | 301 | 12,50 | 3,89 | 294 | 0,61 |
| Write your name | 4,49 | 1,05 | 298 | 4,59 | 0,82 | 294 | 0,37 |
| Copy a word | 4,33 | 1,41 | 298 | 4,32 | 1,37 | 294 | 0,92 |
| Write letters | 1,36 | 1,46 | 298 | 1,73 | 1,51 | 294 | 0,08 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table A2: Baseline balance in Grade 2 EGRA scores

| | Treatment | | | Control | | | p-value |
|---|---|---|---|---|---|---|---|
| | **Mean** | **s.d.** | **N** | **Mean** | **s.d.** | **N** | |
| Receptive listening comprehension | 9,71 | 0,61 | 302 | 9,61 | 0,83 | 290 | 0,19 |
| Productive listening comprehension | 3,47 | 1,36 | 302 | 3,61 | 1,33 | 290 | 0,31 |
| Expressive vocabulary | 12,15 | 3,45 | 302 | 12,36 | 3,84 | 290 | 0,63 |
| Letter sounds per minute | 28,42 | 20,27 | 302 | 28,48 | 19,53 | 290 | 0,98 |
| Digraphs and trigraphs per minute | 8,88 | 11,44 | 302 | 8,10 | 11,25 | 290 | 0,62 |
| Phonemic awareness | 4,21 | 2,31 | 302 | 4,32 | 2,39 | 290 | 0,72 |
| CVCV words per minute | 10,17 | 12,29 | 302 | 9,41 | 11,80 | 290 | 0,65 |
| Familiar words per minute | 7,11 | 8,93 | 302 | 6,54 | 8,59 | 290 | 0,64 |
| Oral reading fluency | 7,53 | 9,71 | 302 | 7,33 | 9,07 | 290 | 0,88 |
| Reading comprehension | 4,13 | 4,82 | 302 | 4,27 | 4,59 | 290 | 0,83 |
| Vocabulary | 3,08 | 2,40 | 301 | 3,19 | 2,25 | 290 | 0,74 |
| Sentence comprehension | 4,44 | 4,46 | 302 | 4,52 | 4,34 | 290 | 0,90 |
| Write letters | 3,93 | 1,42 | 292 | 4,01 | 1,26 | 290 | 0,65 |
| Write words | 13,79 | 7,11 | 292 | 14,47 | 6,53 | 290 | 0,52 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table A3: Baseline balance in learner characteristics and home assets (both grades)

| | Treatment | | | Control | | | p-value |
|---|---|---|---|---|---|---|---|
| | **Mean** | **s.d.** | **N** | **Mean** | **s.d.** | **N** | |
| Female | 0,51 | 0,50 | 603 | 0,49 | 0,50 | 584 | 0,58 |
| Age in months | 84,16 | 10,66 | 603 | 83,37 | 10,21 | 584 | 0,31 |
| Wearing spectacles | 0,02 | 0,14 | 603 | 0,02 | 0,14 | 584 | 0,92 |
| Height (cm) | 119,20 | 6,82 | 603 | 118,93 | 6,64 | 584 | 0,59 |
| Absent >= 1-day past week | 0,27 | 0,44 | 596 | 0,29 | 0,46 | 577 | 0,42 |
| Teacher absent >= 1-day past week | 0,23 | 0,42 | 563 | 0,24 | 0,43 | 545 | 0,70 |
| Library in the school | 0,36 | 0,48 | 489 | 0,37 | 0,48 | 481 | 0,89 |
| Readers or storybooks in classroom | 0,87 | 0,34 | 597 | 0,92 | 0,27 | 580 | 0,04 |
| Lives with mother | 0,89 | 0,31 | 603 | 0,91 | 0,29 | 584 | 0,44 |
| Lives with father | 0,62 | 0,49 | 603 | 0,65 | 0,48 | 584 | 0,28 |
| Lives with both parents | 0,59 | 0,49 | 603 | 0,62 | 0,48 | 584 | 0,33 |
| Lives with neither parents | 0,08 | 0,27 | 603 | 0,06 | 0,24 | 584 | 0,26 |
| Books (not schoolbooks) at home | 0,32 | 0,47 | 602 | 0,38 | 0,48 | 583 | 0,12 |
| Radio | 0,70 | 0,46 | 602 | 0,72 | 0,45 | 584 | 0,66 |
| Mobile | 0,99 | 0,10 | 602 | 0,99 | 0,11 | 584 | 0,82 |
| Electricity | 0,97 | 0,17 | 602 | 0,96 | 0,19 | 583 | 0,61 |
| Television | 0,95 | 0,23 | 602 | 0,94 | 0,24 | 583 | 0,70 |
| Computer | 0,30 | 0,46 | 603 | 0,33 | 0,47 | 578 | 0,34 |
| Fridge | 0,91 | 0,28 | 602 | 0,94 | 0,23 | 583 | 0,12 |
| Toilet | 0,59 | 0,49 | 601 | 0,61 | 0,49 | 583 | 0,54 |
| Bicycle | 0,45 | 0,50 | 518 | 0,40 | 0,49 | 534 | 0,30 |
| Vehicle | 0,50 | 0,50 | 601 | 0,53 | 0,50 | 584 | 0,32 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table A4: Midline equivalence in learner baseline test score, characteristics, and home assets

| | Treatment | | Control | | p -value | Pooled | Effect |
|---|---|---|---|---|---|---|---|
| | Mean | s.d. | Mean | s.d. | | s.d. | size |
| **Common tasks** | | | | | | | |
| Letter sounds per minute | 17,31 | 19,78 | 17,49 | 19,08 | 0,91 | 19,42 | 0,01 |
| Digraphs and trigraphs per minute | 5,32 | 11,49 | 4,76 | 10,76 | 0,58 | 11,13 | 0,05 |
| Productive listening comprehension | 3,12 | 1,49 | 3,10 | 1,52 | 0,86 | 1,50 | 0,01 |
| Receptive listening comprehension | 9,63 | 0,74 | 9,50 | 1,01 | 0,13 | 0,88 | 0,15 |
| Phonemic awareness | 2,87 | 2,51 | 2,98 | 2,52 | 0,62 | 2,51 | 0,04 |
| Expressive vocabulary | 11,28 | 3,80 | 11,69 | 4,04 | 0,31 | 3,93 | 0,11 |
| Write letters | 2,70 | 1,93 | 2,89 | 1,80 | 0,25 | 1,86 | 0,10 |
| **Grade 1-only tasks** | | | | | | | |
| Word choice | 1,73 | 1,74 | 1,90 | 1,60 | 0,36 | 1,67 | 0,10 |
| Rapid automatized naming | 36,99 | 12,00 | 37,63 | 11,59 | 0,59 | 11,79 | 0,05 |
| Write your name | 4,55 | 1,00 | 4,59 | 0,80 | 0,65 | 0,90 | 0,05 |
| Copy a word | 4,41 | 1,31 | 4,33 | 1,36 | 0,49 | 1,34 | 0,07 |
| **Grade 2-only tasks** | | | | | | | |
| CVCV words per minute | 10,40 | 12,65 | 9,61 | 11,86 | 0,65 | 12,26 | 0,06 |
| Familiar words per minute | 7,16 | 9,00 | 6,67 | 8,62 | 0,69 | 8,81 | 0,05 |
| Oral reading fluency | 7,74 | 9,92 | 7,47 | 9,14 | 0,84 | 9,53 | 0,03 |
| Reading comprehension | 4,20 | 4,83 | 4,34 | 4,59 | 0,83 | 4,71 | 0,03 |
| Vocabulary | 3,14 | 2,39 | 3,23 | 2,24 | 0,78 | 2,32 | 0,04 |
| Sentence comprehension | 4,54 | 4,47 | 4,63 | 4,34 | 0,88 | 4,40 | 0,02 |
| Write words | 14,03 | 7,00 | 14,63 | 6,46 | 0,57 | 6,73 | 0,09 |
| **Learner characteristics** | | | | | | | |
| Grade 1 | 0,49 | 0,50 | 0,50 | 0,50 | 0,28 | 0,50 | 0,01 |
| Grade 2 | 0,51 | 0,50 | 0,50 | 0,50 | 0,28 | 0,50 | 0,01 |
| Female | 0,51 | 0,50 | 0,49 | 0,50 | 0,37 | 0,50 | 0,04 |
| Age in months | 84,16 | 10,63 | 83,36 | 10,21 | 0,32 | 10,42 | 0,08 |
| Height for age z-score | -0,37 | 1,04 | -0,38 | 0,98 | 0,90 | 1,01 | 0,01 |
| **Household assets** | | | | | | | |
| Books other than schoolbooks to read at home | 0,33 | 0,47 | 0,38 | 0,49 | 0,10 | 0,48 | 0,12 |
| Radio | 0,71 | 0,46 | 0,72 | 0,45 | 0,73 | 0,45 | 0,02 |
| Television | 0,94 | 0,23 | 0,94 | 0,24 | 0,69 | 0,23 | 0,02 |
| Computer | 0,30 | 0,46 | 0,34 | 0,47 | 0,34 | 0,47 | 0,07 |
| Toilet | 0,59 | 0,49 | 0,62 | 0,49 | 0,54 | 0,49 | 0,05 |
| Vehicle | 0,33 | 0,47 | 0,38 | 0,49 | 0,10 | 0,48 | 0,12 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table A5: Sub-task midline raw score distributions, by grade and treatment status

| | Control | | | | | | | | | Treatment | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Task** | **N** | **% zero** | **Mean** | **s.d.** | **Mean (excl. 0)** | **Percentiles** | | | | **N** | **% zero** | **Mean** | **s.d.** | **Mean (excl. 0)** | **Percentiles** | | | |
| | | | | | | **25th** | **50th** | **75th** | **max** | | | | | | **25th** | **50th** | **75th** | **max** |
| **Grade 1** | | | | | | | | | | | | | | | | | | |
| Letters p/m | 279 | 8% | 24,3 | (18,5) | 26,3 | 8 | 22 | 39 | 76 | 276 | 4% | 28,6 | (20,0) | 29,7 | 10 | 27 | 43 | 84 |
| Di-/ trigraphs p/m | 279 | 58% | 6,2 | (10,6) | 14,7 | 0 | 0 | 9 | 55 | 276 | 50% | 8,5 | (13,7) | 17,0 | 0 | 1 | 12 | 71 |
| CVCV words p/m | 279 | 52% | 6,3 | (9,5) | 13,2 | 0 | 0 | 10 | 40 | 276 | 54% | 7,5 | (11,3) | 16,1 | 0 | 0 | 14 | 45 |
| Familiar words p/m | 279 | 55% | 4,2 | (6,7) | 9,4 | 0 | 0 | 6 | 27 | 276 | 57% | 5,4 | (8,2) | 12,4 | 0 | 0 | 9 | 32 |
| Oral reading fluency | 279 | 56% | 4,5 | (7,6) | 10,3 | 0 | 0 | 6 | 29 | 276 | 53% | 5,8 | (9,1) | 12,4 | 0 | 0 | 10 | 38 |
| Reading comp. I | 279 | 58% | 2,3 | (3,3) | 5,5 | 0 | 0 | 5 | 13 | 276 | 57% | 2,6 | (3,6) | 6,0 | 0 | 0 | 6 | 12 |
| Productive listening | 279 | 4% | 3,2 | (1,4) | 3,3 | 2 | 3 | 4 | 6 | 276 | 4% | 3,5 | (1,4) | 3,6 | 3 | 4 | 4 | 6 |
| Vocabulary | 279 | 2% | 5,1 | (1,2) | 5,2 | 5 | 5 | 6 | 6 | 276 | 2% | 5,0 | (1,4) | 5,1 | 5 | 5 | 6 | 6 |
| Phonemic awareness | 279 | 10% | 3,3 | (2,1) | 3,7 | 2 | 3 | 5 | 10 | 276 | 7% | 3,9 | (2,0) | 4,2 | 3 | 4 | 5 | 9 |
| Expressive vocabulary | 279 | 0% | 9,1 | (3,5) | 9,2 | 7 | 9 | 11 | 20 | 276 | 0% | 9,6 | (3,6) | 9,6 | 7 | 9 | 12 | 20 |
| **Grade 2** | | | | | | | | | | | | | | | | | | |
| Letters p/m | 278 | 1% | 44,8 | (20,3) | 45,4 | 34 | 49 | 59 | 93 | 283 | 1% | 48,1 | (20,1) | 48,6 | 35 | 51 | 63 | 91 |
| Di-/ trigraphs p/m | 278 | 15% | 24,3 | (19,6) | 28,7 | 5 | 26 | 39 | 87 | 283 | 17% | 27,1 | (20,5) | 32,6 | 6 | 30 | 43 | 80 |
| CVCV words p/m | 278 | 20% | 20,4 | (16,4) | 25,6 | 4 | 19 | 33 | 79 | 283 | 17% | 22,6 | (17,1) | 27,2 | 7 | 22 | 35 | 67 |
| Familiar words p/m | 278 | 22% | 14,7 | (11,9) | 18,8 | 3 | 15 | 24 | 45 | 283 | 18% | 16,8 | (12,9) | 20,4 | 5 | 18 | 27 | 54 |
| Oral reading fluency | 278 | 16% | 16,7 | (14,0) | 19,9 | 4 | 16 | 26 | 63 | 283 | 17% | 18,2 | (14,8) | 22,0 | 3 | 19 | 29 | 68 |
| Reading comp. I | 278 | 18% | 6,4 | (4,0) | 7,9 | 3 | 8 | 10 | 14 | 283 | 19% | 6,7 | (4,1) | 8,3 | 3 | 8 | 10 | 14 |
| Productive listening | 278 | 1% | 3,9 | (1,2) | 4,0 | 3 | 4 | 5 | 6 | 283 | 1% | 4,1 | (1,2) | 4,1 | 3 | 4 | 5 | 6 |
| Vocabulary | 278 | 0% | 5,5 | (0,8) | 5,5 | 5 | 6 | 6 | 6 | 283 | 1% | 5,4 | (1,0) | 5,5 | 5 | 6 | 6 | 6 |
| Phonemic awareness | 277 | 2% | 5,0 | (1,9) | 5,1 | 3 | 5 | 6 | 10 | 283 | 1% | 5,1 | (2,0) | 5,2 | 4 | 5 | 6 | 10 |
| Oral reading fluency II | 278 | 24% | 15,3 | (13,5) | 20,3 | 1 | 15 | 25 | 59 | 283 | 22% | 16,3 | (13,3) | 20,9 | 3 | 16 | 26 | 60 |
| Reading comp. II | 278 | 27% | 4,0 | (3,1) | 5,5 | 0 | 5 | 6 | 10 | 283 | 23% | 4,3 | (3,2) | 5,6 | 1 | 5 | 7 | 10 |
| Sentence choice | 278 | 27% | 5,4 | (3,8) | 7,3 | 0 | 7 | 9 | 10 | 283 | 29% | 5,6 | (4,0) | 7,8 | 0 | 7 | 9 | 10 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Table A6: Treatment effects by sub-task

| | Effect size (s.d.) | s.e. | p-values | |
|---|---|---|---|---|
| | | | Regression | Random. inference |
| **Both grades** | | | | |
| Composite score | 0,17 | 0,05 | 0 | 0 |
| Letters p/m | 0,22 | 0,06 | 0,00 | 0 |
| Di-/ trigraphs p/m | 0,16 | 0,06 | 0,01 | 0,02 |
| CVCV words p/m | 0,14 | 0,05 | 0,01 | 0,04 |
| Familiar words p/m | 0,17 | 0,05 | 0,00 | 0 |
| Oral reading fluency | 0,14 | 0,05 | 0,01 | 0,02 |
| Reading comp. I | 0,11 | 0,06 | 0,09 | 0,13 |
| Productive listening | 0,17 | 0,06 | 0,01 | 0 |
| Vocabulary | -0,07 | 0,12 | 0,56 | 0,34 |
| Phonemic awareness | 0,17 | 0,06 | 0,01 | 0,01 |
| **Grade 2** | | | | |
| Composite score | 0,16 | 0,04 | 0 | 0 |
| Letters p/m | 0,19 | 0,06 | 0 | 0 |
| Di-/ trigraphs p/m | 0,17 | 0,06 | 0,01 | 0,01 |
| CVCV words p/m | 0,15 | 0,05 | 0 | 0,01 |
| Familiar words p/m | 0,19 | 0,05 | 0 | 0 |
| Oral reading fluency | 0,14 | 0,05 | 0,01 | 0,03 |
| Reading comp. I | 0,13 | 0,06 | 0,03 | 0,09 |
| Productive listening | 0,05 | 0,06 | 0,39 | 0,26 |
| Phonemic awareness | 0,15 | 0,08 | 0,07 | 0,04 |
| Oral reading fluency II | 0,07 | 0,05 | 0,13 | 0,09 |
| Reading comp. II | 0,12 | 0,06 | 0,03 | 0,02 |
| Grade 2 composite score | 0,15 | 0,04 | 0 | 0 |
| **Grade 1** | | | | |
| Composite score | 0,21 | 0,09 | 0,02 | 0,01 |
| Letters p/m | 0,27 | 0,09 | 0,01 | 0 |
| Di-/ trigraphs p/m | 0,19 | 0,08 | 0,03 | 0,08 |
| CVCV words p/m | 0,14 | 0,08 | 0,1 | 0,08 |
| Familiar words p/m | 0,17 | 0,08 | 0,03 | 0,05 |
| Oral reading fluency | 0,16 | 0,08 | 0,04 | 0,06 |
| Reading comp. I | 0,15 | 0,09 | 0,12 | 0,21 |
| Phonemic awareness | 0,28 | 0,09 | 0 | 0,01 |
| Productive listening | 0,2 | 0,08 | 0,02 | 0,04 |
| Composite score | 0,21 | 0,09 | 0,02 | 0,01 |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

## 8.2 Appendix I: school selection and underlying population

Working with ECDoE, Funda Wande invited schools from the three urban and peri-urban districts in the Eastern Cape (Nelson Mandela Bay, Sarah Baartman, and Buffalo City) to apply to be part of the programme. Funda Wande's approach is to work with schools who do want an intervention to take place in their schools. The main consideration is ethical—given the number of schools requiring support in South Africa, implementing partners chose to start with schools that expressed a willingness to receive support. To be eligible for the Funda Wande intervention, school principals had to write a letter of motivation asking to be included in the intervention. The letter had to be signed by themselves, the Deputy Principal, the HOD, and a School Governing Body (SGB) member.

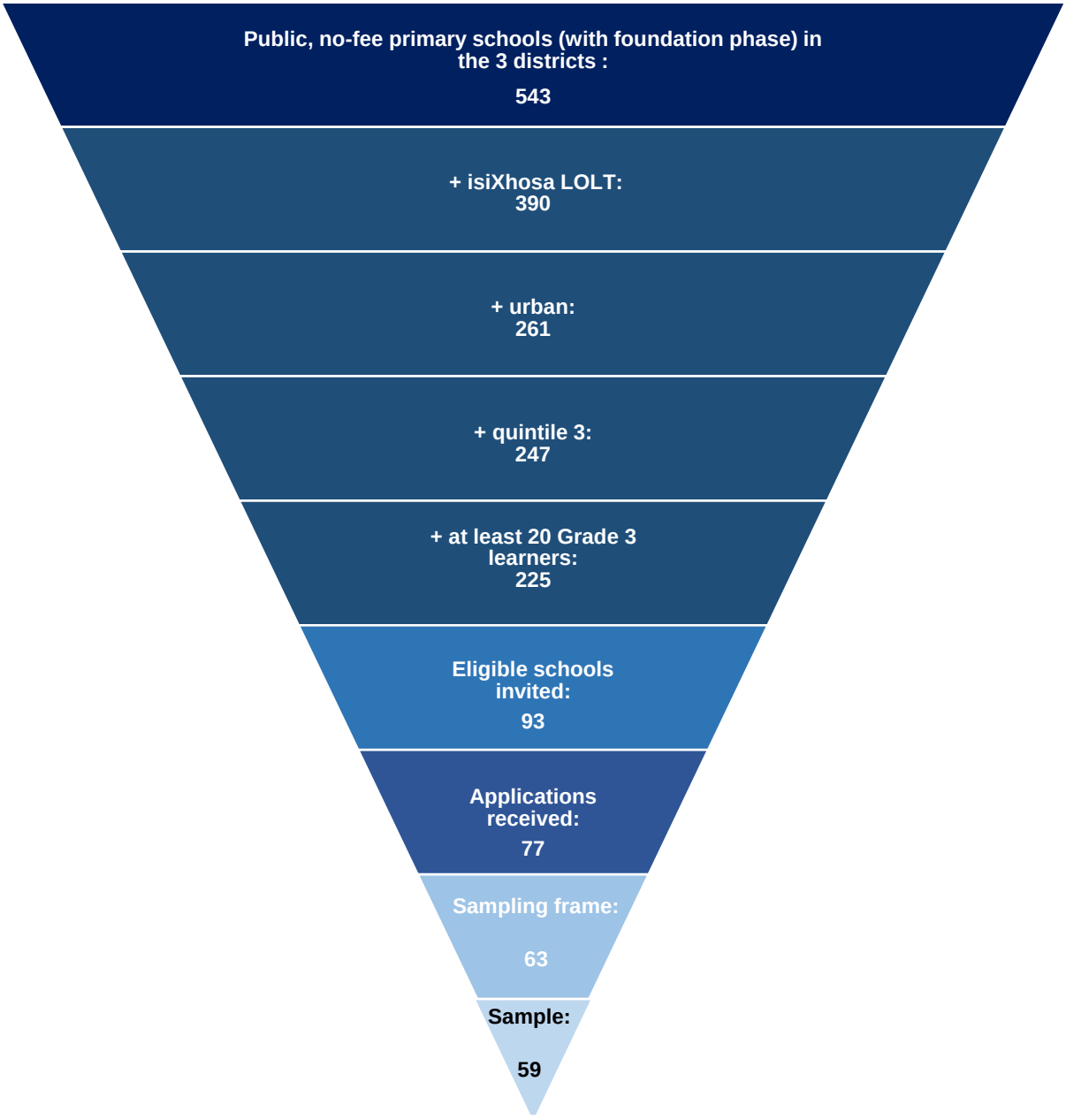
Based on the primary inclusion criteria, Funda Wande received a list of eligible schools from the three respective district managers. The list of 93 schools were based on the explicit criteria that schools should be no-fee, public primary or combined schools (i.e. have Grade 1–3 learners), with no other major literacy intervention ongoing and an isiXhosa language of learning and teaching (LOLT). Invitations were sent to all 93 schools from the district official lists, of which 77 schools both i) returned completed application forms, and ii) were self-described as motivated to take part in the study. Funda Wande further screened the applications to exclude schools with chronic management problems, severe overcrowding (greater than 50 learners per class), or fewer than 20 learners per grade.

Of the returned applications, 63 schools were selected for the programme. From a programme administrative standpoint, Funda Wande also had an informal selection criterion of not including schools that were outside of approximately a one-and-a-half-hour drive from either of the three central locations (East London, Port Elizabeth, or Makana [Grahamstown]).

Figure C1 locates the final sample within the universe of public, ordinary, no-fee schools offering foundation phase in the three districts. The figures are based on Education Management Information Systems (EMIS) administrative data for the period when the school selection process took place (term three of 2018) merged with data from the Data Driven Districts (DDD) dashboards.[1] Of the 543 schools, 78 per cent have isiXhosa as the sole language of learning and teaching (LOLT). Sixty-seven per cent of these schools are urban. Almost all (95 per cent) of these schools are classified as quintile three. Finally, 91 per cent of the remaining schools have at least 20 Grade 3 learners. The total number of schools satisfying these criteria is 225.

[1] This database is a collaborative effort by the DBE and the Michael and Susan Dell Foundation, providing education practitioners, administrators, and researchers with fine-grained learner level data. From this dataset, a school's LOLT was determined to be isiXhosa if all Grade 3 learners had marks for isiXhosa home language (therefore also excluding dual medium schools).

Figure C1: School selection process from plausible populations of schools in the three districts




Source: author's illustration based on Funda Wande data (see SALDRU 2019).

With respect to creating the three strata within which randomization took place, the original aim was to have a sample of ten treatment and ten control schools per district. The total number of schools in the Sarah Baartman district was only 14. These schools were therefore merged with the Uitenhage schools to create a group of 20 schools. There was a total of 22 schools in the Port Elizabeth district, with two of these schools randomly selected as possible replacement schools—randomly assigning one to treatment and the other to control. Within each group (or strata) of 20 schools, half of the schools were randomly assigned to receive the Funda Wande programme, with the other half serving as control schools.

Furthermore, post randomization, it was discovered that the LOLT of two control schools was not isiXhosa throughout the foundation phase. These schools were subsequently dropped, and one school from the replacement group was added to the treatment group.

## 8.3 Appendix II: learner assessment details, rationale, and reading outcome measures

*Range of early- and pre-literacy domains assessed*

Many of the extended Early Grade Reading Assessment (EGRA) task used in the Funda Wande evaluation instrument built on minor adaptations made by NORC, at the University of Chicago, for the Story Powered Schools Impact Evaluation.[2] The instrument development also benefited hugely from the input of the Story Powered Schools and Funda Wande evaluation field teams, particularly on translations and appropriate language.

The range of literacy and pre-literacy assessments conducted at baseline were generally used again for the midline learner evaluations (see Table D1). At baseline, Grade 1 learners were not assessed on certain higher order skills that one would not expect them to have acquired right at the start of their schooling career. However, Grade 1 learners were assessed for most of these skills at midline, including word reading, paragraph reading fluency, and reading comprehension tasks. Of the higher order skills from the baseline assessment, only the sentence choice task was conducted on Grade 2 learners only.

A few sub-tasks from baseline were not included in the midline. The rapid automatized naming (RAN) task was included at baseline to identify learners who had zero, single, or double RAN and phonological awareness deficits (Dubek et al. 2017) at baseline, with the interest in tracking the literacy development of these three groups of learners through the waves of the study. The receptive listening task was excluded from midline due to ceiling effects (i.e. many learners scoring full marks) at baseline. In the interests of avoiding learner fatigue during the assessments, baseline writing tasks were also excluded at midline.[3]

Grade 2 learners' reading fluency was assessed on two separate passages at midline. The first passage was the same passage used to assess reading fluency and comprehension at baseline, whilst the newly introduced second passage was slightly longer and more challenging. The newly introduced passage provides a second measure of reading fluency and subsequent reading comprehension assessment for Grade 2 learners. Having two different texts on which reading fluency and comprehension are assessed allows one to go beyond only measuring learners' progression, but also to distinguish whether changes in scores for Grade 2 learners on these tasks are purely down to skills acquired over the academic year (and not to any extent due to learners recalling the texts).

For both passages, learners were only asked comprehension questions based on the point up to which they had completed the preceding reading passage. Low levels of reading fluency therefore posed a potential hurdle to assessing reading comprehension: even if learners could at least start reading from the passage, the majority of learners could not read far enough for them to complete all the subsequent comprehension questions (discussed further below). Learners were therefore assessed on their reading fluency based on how many words they could read accurately in the first 60 seconds. Learners were allowed an additional two minutes to continue reading from the two

[2] This is a randomized controlled trial impact evaluation of Nal'ibali's Story Powered School programme involving over 9,000 Grade 2–4 learners in 360 rural Eastern Cape and KwaZulu-Natal schools. The evaluation runs from early 2017 to late 2019. See Menendez and Ardington (2018).

[3] Writing tasks are likely to be included in some of the subsequent rounds of data collection, as learners become more proficient readers, and the assessment of higher order abilities (like creative writing) becomes increasingly important to differentiate among the better performing learners.

passages, before they were asked comprehension questions based on the respective passages immediately thereafter.

Table D1. Reading skills and sub-tasks in baseline and midline assessments

| Skill | Sub-task and measurement | Baseline | Midline |
|---|---|---|---|
| Receptive listening comprehension | Performing actions following verbal instruction from the enumerator | Grade 1 & 2 | |
| Productive listening comprehension | Number of questions answered correctly about a passage read aloud by the enumerator | Grade 1 & 2 | Grade 1 & 2 |
| Expressive vocabulary | Learner is asked to name items in shop and animals | Grade 1 & 2 | Grade 1 |
| Letter sound knowledge | Number of letters sounds identified in 60 seconds | Grade 1 & 2 | Grade 1 & 2 |
| Digraph/trigraph sound knowledge | Number of digraphs and trigraphs identified in 60 seconds | Grade 1 & 2 | Grade 1 & 2 |
| Phonemic awareness | Identifying and manipulating phonemes (starting and ending sounds of words, segmenting words) | Grade 1 & 2 | Grade 1 & 2 |
| Word recognition | Selecting the word read by the enumerator from four possible CVCV words | Grade 1 | |
| Rapid Automatized Naming | Number of familiar pictures correctly identified in 60 seconds | Grade 1 | |
| Word recognition | Number of correct CVCV words read in 60 seconds | Grade 2 | Grade 1 & 2 |
| Word recognition | Familiar word reading, number of correct words read in 60 seconds | Grade 2 | Grade 1 & 2 |
| Oral reading fluency | Connected text reading, number of words read correctly from the first reading passage in 60 seconds | Grade 2 | Grade 1 & 2 |
| Reading comprehension | Number of questions answered correctly about the passage read aloud by the learner | Grade 2 | Grade 1 & 2 |
| Oral Reading fluency II | Connected text reading, number of words from a second reading passage read correctly in 60 seconds | | Grade 2 |
| Reading comprehension II | Number of questions answered correctly about the passage read aloud by the learner | | Grade 2 |
| Receptive vocabulary | Identifying correct picture to match word | Grade 1 & 2 | |
| Reading comprehension | Identifying whether each of 20 short sentences make sense | Grade 2 | Grade 2 |
| Writing | Writing name | Grade 1 | |
| Writing | Copying a word | Grade 1 | |
| Writing | Writing letters | Grade 1 & 2 | |
| Writing | Writing words | Grade 2 | |

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

*Primary outcome measures*

**a) Composite reading proficiency**—In a similar fashion to Cilliers et al. (2019), a composite score of the isiXhosa reading proficiency was constructed based on the different EGRA sub-tasks on which learners were assessed. The statistical method used, Principal Components Analysis (PCA), reduces the data from the different sub-tasks to create a single variable that captures the most common variation among them: the first principal component. Intuitively, the principal component is taken to be reflective of a common underlying construct, which we here take to reflect isiXhosa reading proficiency. Only the control group's midline sub-tasks are used to construct the index, as these scores give the 'business as usual' weighting of the respective factors to the composite reading proficiency index. In order to simplify interpretation, the composite index was standardized by subtracting the control group mean and dividing by its standard deviation (allowing for interpretation in terms of standard deviations). A baseline composite reading proficiency score is constructed in a similar fashion but using the EGRA scores for the

full baseline sample (i.e. when test scores in the treatment group were still unaffected by the treatment).

The purpose of the composite score is to create one transparent and clearly defined overarching measure of programme impacts. Statistically, it serves as a reassurance that our overall assessment of programme impact, heterogeneous treatment impacts, and robustness checks are not selectively reported for certain sub-tasks and/or sub-groups. Nevertheless, given that i) the sub-tests do not all necessarily fit together in one coherent whole, and ii) that we are also interested in the impacts of the programme on certain foundational components on the path to reading for meaning, results for the main estimation model are also reported for each sub-task individually.

Given the aim of constructing a reading proficiency index, two tasks were left out of the index: i) the vocabulary task and ii) the productive listening task. The former had had severe ceiling effects (with more than 50 per cent of learners scoring full marks)—which affects its usefulness in the index. An exploratory factor analysis indicated that the productive listening task had a low item-rest correlation and loaded higher on the second underlying factor that seems to be indicative of oral literacy skills (and not of reading proficiency) (Ardington 2019).

**b) Reading comprehension**—Given the stated programme objective that all learners should be reading for meaning by the end of Grade 3, reading comprehension measures are also considered as primary outcomes. There are particular methodological considerations to accurately and reliably measuring reading comprehension, often neglected in applied research, but given particular consideration in the Funda Wande evaluation.

Both the Funda Wande intervention and evaluation therefore place a particular emphasis on the programme's main outcome: reading with comprehension. The measures of comprehension are significantly more extensive than those generally used in Early Grade Reading Assessment (EGRA) type literacy tests (Gove and Wetternberg 2011). For example, recent adaptations of the EGRA assessments like those in Liberia (Piper and Korda, 2011), Kenya (Piper et al. 2014), and South Africa (Cilliers et al. 2019) all ask four to five short questions subsequent to learners' one reading fluency task. Generally, the sole comprehension task consists of four basic literal questions and one more challenging inferential type question, with learners only asked questions related up to the section of paragraph that they managed to read in the one minute.

In contrast, three separate reading comprehension tasks were employed in the Funda Wande evaluation, all of which provide a more extensive assessment of learners' comprehension skills. The first reading comprehension task is based on a short passage of 41 words. It consists of 14 questions (of which 11 are literal and three are interpretive questions), with learners allowed to keep the text in front of them to aid in answering the questions. The second reading passage is longer (55 words in length) and the subsequent comprehension task more challenging. In this case, the passage is no longer available as a reference for the comprehension questions and learners are asked ten questions (split halfway between literal and interpretive questions).

The additional two minutes of reading time enabled learners to answer a far greater share of the comprehension questions. Table D2 below provides an illustrative example based on the first reading comprehension task completed by both grades. For learners to have read far enough to attempt the first inferential comprehension question (the fifth question out of 14), they must have read at least the first 13 words from the passage. Only 20 per cent of Grade 1's and 65 per cent of Grade 2's managed to read that many words in the first 60 seconds of the reading fluency task. The additional two minutes allowed an additional 23 per cent of Grade 1's and 17 per cent of Grade 2's to reach the point in the passage where they could attempt the first inferential question. The additional two minutes also allowed a fifth of Grade 1 learners to complete the whole passage,

where no Grade 1 could achieve this in only one minute. Almost two thirds of Grade 2's could complete the whole passage after three minutes, up from less than ten per cent in one minute.

Table D2: How many comprehension questions learners can attempt based on their reading speed

| | Can attempt 1st inferential question (>= 13 words read)[4] | | Can attempt all questions (= 41 words read) | |
|---|---|---|---|---|
| | **Grade 1** | **Grade 2** | **Grade 1** | **Grade 2** |
| % after 1 minute | 20% | 65% | 0% | 7% |
| % after 3 minutes | 43% | 82% | 21% | 63% |

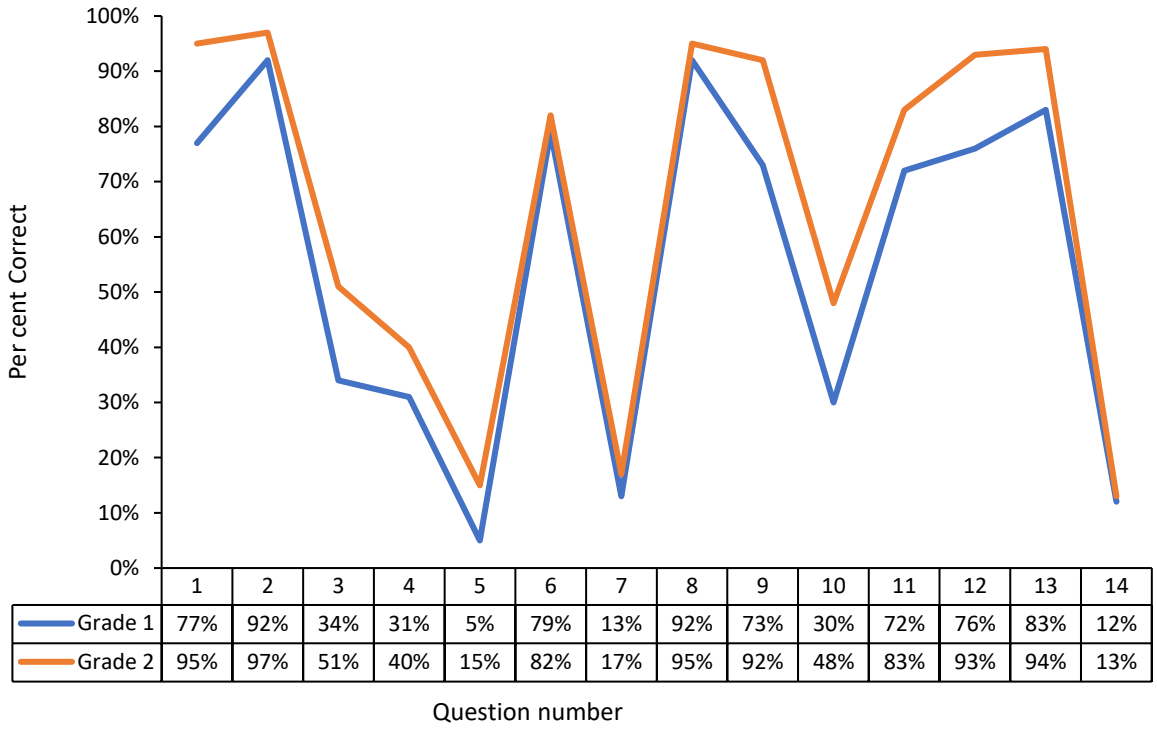
Source: author's calculation based on Funda Wande data (see SALDRU 2019).

The average scores on the comprehension tasks are thus slightly misleading if one is only interested in learner reading comprehension conditional on how fast they read. For example, the average comprehension score for all Grade 2's is 47 per cent on the first comprehension task and 41 per cent on the second. However, only six in ten Grade 2 learners could attempt all 14 comprehension questions on the first comprehension task, whilst only approximately half (47 per cent) of the same group read fast enough to attempt all ten comprehension questions on the second passage. In contrast, for these two subsets of Grade 2's who attempted all the respective comprehension task questions, their average scores were 65 per cent and 67 per cent on the two tasks, respectively.

Figures D1 indicates the percentage of learners who could correctly answer each question of the two paragraph reading comprehension tasks, but only for those learners who finished reading the whole passage in three minutes. There is a large variability in learners' ability to answer the questions across both reading comprehension tasks. In particular, learners fared the worst in the interpretive and inferential questions in the first comprehensions task (questions 5, 7, and 14). For the second comprehension task, learners also scored very low on three of the interpretive questions (questions 2, 9, and 10), whilst the other lowest scoring item (question 7) was a factual detail from the story that learners generally struggled to recall.

[4] A learner was judged able to attempt a comprehension question based on the amount of words that they attempted (i.e. how far into the passage they read, whether or not they read the respective words correctly).

Figure D1: How learners who attempted all questions fared by question (Reading Comprehension 1, by grade)




| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Grade 1 | 77% | 92% | 34% | 31% | 5% | 79% | 13% | 92% | 73% | 30% | 72% | 76% | 83% | 12% |
| Grade 2 | 95% | 97% | 51% | 40% | 15% | 82% | 17% | 95% | 92% | 48% | 83% | 93% | 94% | 13% |

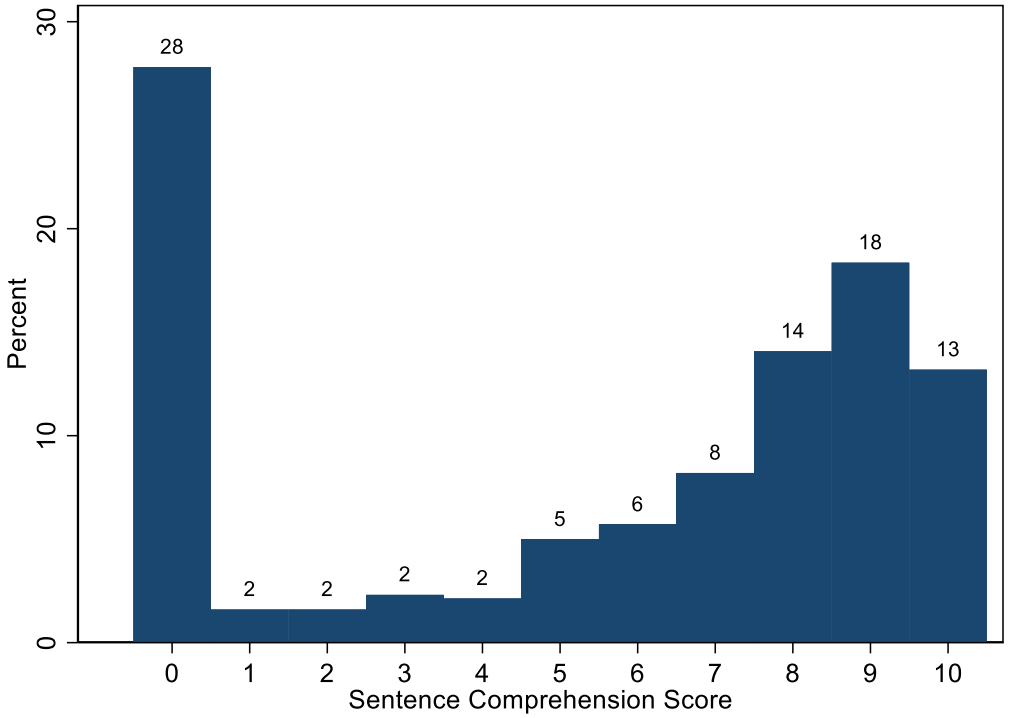
Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Given the challenges in measuring reading comprehension in a context of low levels of reading fluency levels, an additional sentence choice comprehension task was also included. More precisely, the sentence choice comprehension sub-task was untimed and consisted of 20 short sentences (typically two words in isiXhosa), which learners had to read and then indicate whether the sentence makes sense or not. Each sentence had a pair, for example 'Fire is cold' and 'Fire is hot'. Learners scored one point if their responses for both items in the pair were correct and scored zero otherwise.

Figure D2 shows the distribution of how Grade 2 leaners scored on untimed sentence choice comprehension task. The scores for the task are clustered at both the bottom (28 per cent score zero) and top (46 per cent score 80 per cent or more) of the distribution. One in five learners could not read the first three sentences and the task was discontinued. These learners make up the vast majority of the learners scoring zero on the task. Almost all (91 per cent) of the learners who did not attempt a reading comprehension question scored zero on the sentence comprehension task. This suggests that the task is somewhat limited in the extent to which it can discriminate between learners reading comprehension ability at both the lower and upper ends of the distribution.

In sum, the sentence choice task provided only limited discrimination between learners reading comprehension abilities. The former two paragraph reading comprehension tasks are thus the preferred measures of learners reading comprehension abilities.

Figure D2: Histogram of sentence comprehension scores (Grade 2 only)

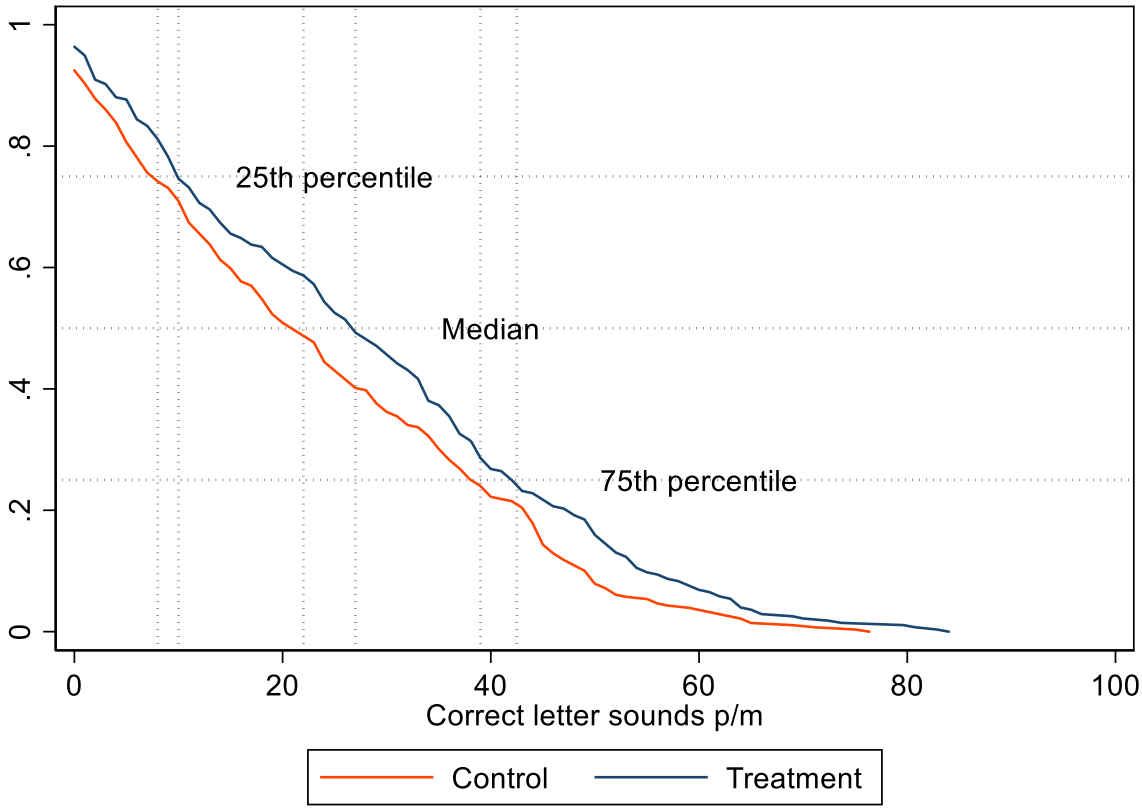



Source: author's calculation based on Funda Wande data (see SALDRU 2019).

## 8.4 Appendix III: heterogeneous treatment effects – robustness checks

Figures E1 and E2 below show the percentage of learners scoring at or below a certain level for selected grade relevant reading proficiency sub-tasks by intervention group status. More specifically, Figure E1 indicates that a greater share of intervention school Grade 1's could identify a certain number of correct letter sounds per minute at every point along the distribution. At the 25th percentile, intervention school Grade 1 learners could read two correct letter sounds per minute more (ten versus eight). At the medians there is a five letter sounds difference (27 versus 22), with a three-and-a-half letter sound difference at the 75th percentiles (42.5 versus 39). The largest consistent difference between the groups was in the middle of the distribution (a seven-letter sound difference from the 55th to the 59th percentiles in the two groups).

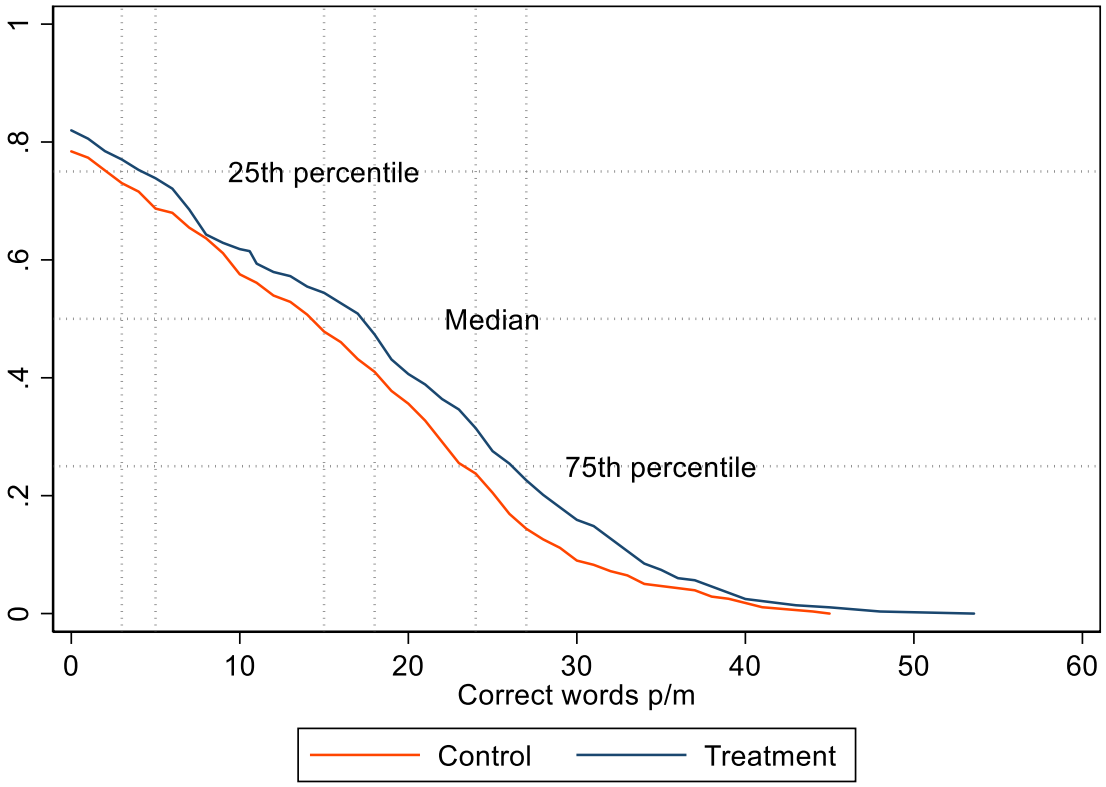
Figure E1: Correct letter sounds per minute by intervention group—Grade 1




Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Similarly, for the distribution of correct familiar words read per minute by Grade 2 learners in the two groups (Figure E2), learners in the intervention group outperform control school Grade 2's across the entire distribution. At the lower end of the distribution, 81 per cent of learners in treatment schools could identify at least one word correctly per minute at midline, compared to a slightly lower 78 per cent of the control group's Grade 2's. At the 25th percentile of the respective distributions, intervention school learners read two words more correctly per minute (five, as opposed to three in the control group). At both the medians and the 75th percentiles, intervention school learners read three words more correctly per minute (18 versus 15 at the medians, and 27 versus 24 at the 75th percentile).

Figure E2: Correct familiar words per minute by intervention group—Grade 2




Source: author's calculation based on Funda Wande data (see SALDRU 2019).

The same models from the main analysis are used, but adding interaction terms for treatment status and the learner's baseline level of reading proficiency. The same set of controls are used, including a separate control for each relevant measure of reading proficiency collected at baseline. Table E1 below reports the coefficients on the school's intervention status, the interaction terms testing for heterogeneous effects based on baseline reading proficiency, the p-value of the test of whether the latter is statistically significant, as well as the sample on which the analysis was conducted.

The first column in table E1 suggests that there is not a linear relationship between learners' baseline reading proficiency and the effect of the programme, given that the interaction term is not statistically different from zero (p-value=0.344). Column two tests for a quadratic relationship between the intervention and learners baseline reading proficiency.[5] It might well be the case, for example, that the programme has no effect on the weakest or best performing learners at baseline, but that the impact of the programme increases and peaks as one moves to the middle of the baseline reading proficiency distribution. However, a test for the joint significance of the two interaction terms in column two suggests that this is also not the case (p-value=0.593). Overall, the results suggest that the impact of the programme is positive and consistent across the distribution of learners' baseline reading proficiency.

[5] The model in column (2) of Table E1 therefore also includes one additional control variable: the squared score of learners' baseline reading proficiency score, alongside the additional interaction term of intervention status with the squared baseline reading proficiency score.

Table E1: Intervention effects by baseline reading proficiency

| | Outcome variable: midline composite score | |
|---|---|---|
| Treatment | 0.174*** | 0.181*** |
| | (0.0538) | (0.0648) |
| Treatment x baseline composite score | -0.0420 | -0.0320 |
| | (0.0440) | (0.0508) |
| Treatment x baseline composite score squared | | -0.00711 |
| | | (0.0257) |
| Sample | FULL | FULL |
| Observations | 1,104 | 1,104 |
| R-squared | 0.678 | 0.684 |
| Heterogeneous treatment effect: p-value | 0.344 | 0.593 |

Note: standard errors reported in brackets.

Source: author's calculation based on Funda Wande data (see SALDRU 2019).

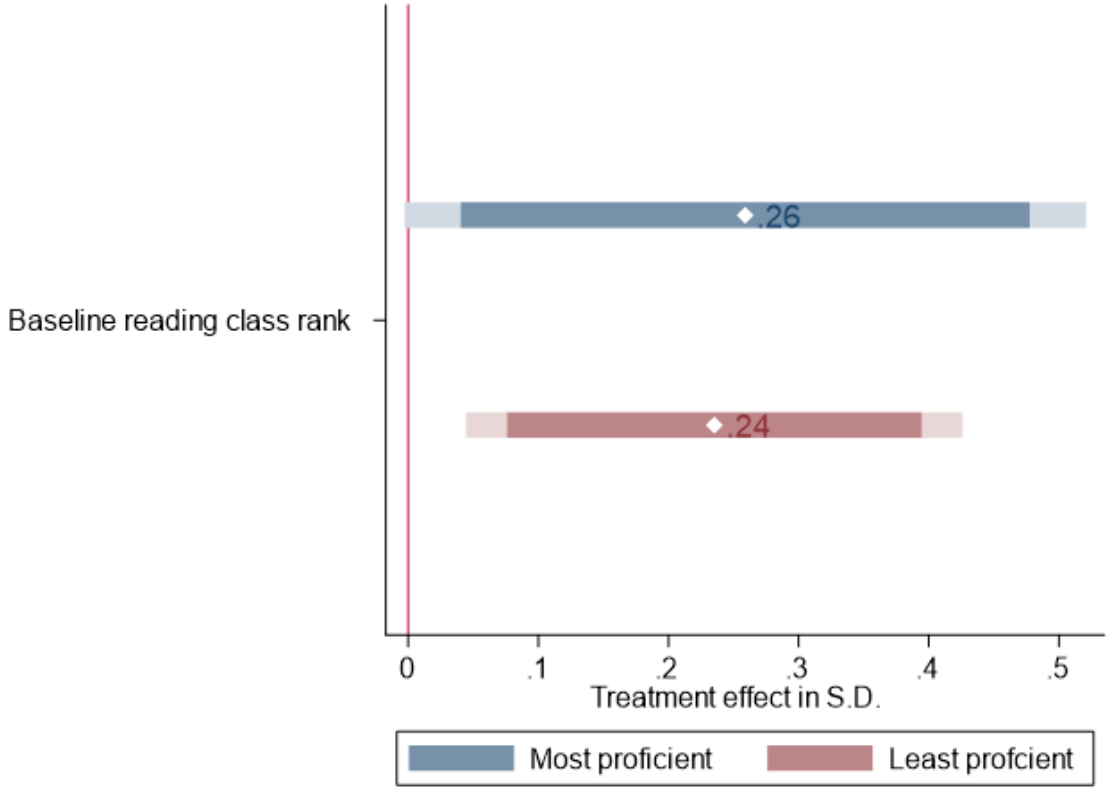
One might also be interested in whether the programme's effects are as consistent based on learners' relative ranks within the same classrooms. Even though this is unlikely to occur in the scenario where programme impacts are consistent across the sample distribution of baseline reading proficiency (as is the case here), the programme could hypothetically still have the greatest effect for learners who tend to lie at a certain end of the spectrum *within* their class. For example, intervention teachers might be both more effective in general and devote their attention disproportionately to helping learners lagging behind to catch up. If what counts as 'lagging behind' varies significantly across different classrooms and schools, and a large share of learners clump at the bottom of their classroom's specific distribution at any one time, we could still see gains across the sample, but which are concentrated at the lower end for specific classrooms.

Figure E3 below displays the estimated impact of the programme from two separate regressions for only those learners who ranked either first or last in their class for reading proficiency at baseline, respectively[6]. The intervention was effective in shifting the reading outcomes of those

[6] The ranking methodology worked as follows. For the 'most proficient reader' (or top ranked learner), the learner(s) who scored the highest mark in the class on the relevant task were assigned a ranking score of one. If learners were tied for first place, both would receive a score of one. All subsequent learners would receive a ranking score of one plus the number of learners between them and first place. In the example where two learners were tied first, the next-best ranking learner would receive a rank score of three, and so on. The ranking methodology for the 'least proficient reader' (or bottom ranked learner) worked in exactly the same manner, but with the learner with lowest score in the class now receiving a rank of one. If, for example, eight learners in the class all scored zero on a task, they would all receive a rank of one, with the next worst learner receiving a rank score of nine. Learners with tied scores at the top and (especially) at the bottom of the class presented challenges to the construction of rank scores based on tasks like learners' correct letter sounds identified or reading fluency. Both Grade 1 and Grade 2 learners were ranked based on baseline reading proficiency scores. For Grade 1 learners, baseline reading proficiency scores were derived using PCA of letter sound recognition, phonemic awareness, and copying letters tasks (i.e. the same index score as the baseline composite score used in Table E1). The Grade 2 rank is based on a reading proficiency score constructed similarly, but including word recognition, reading fluency, and comprehension tasks from baseline assessments (sub-tasks assessed on Grade 2 learners only at baseline). The result is that the Grade 2 reading proficiency rank is based on similar sub-domains of reading proficiency as the midline composite reading proficiency score. Excluding these baseline tasks from the Grade 2 composite score would have limited our ability to differentiate the most proficient reader among those Grade 2 learners in a class, as most learners were already highly proficient in lower order decoding

learners who were the least proficient in their class at baseline (effect size of 0.24 s.d.) and for the most proficient readers (0.26 s.d.). The is not a statistically significant differential impact for a classroom's most or least proficient readers at baseline.

Figure E3: Intervention impacts estimated on most and least proficient readers at baseline, respectively

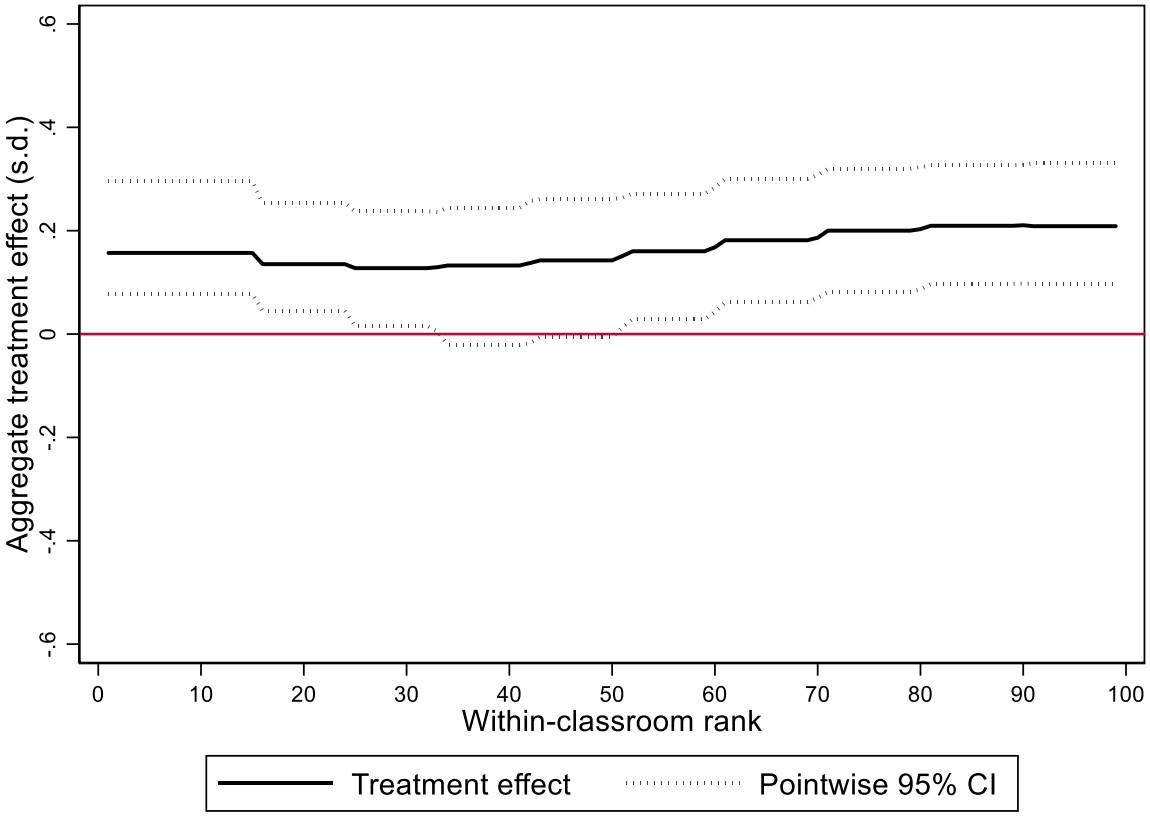



Source: author's calculation based on Funda Wande data (see SALDRU 2019).

Figure E4 repeats the same non-parametric estimation of treatment impacts based on learners' baseline reading proficiency as in figure E3 above, but now based on a learner's baseline rank within the ten learners assessed per class. Treatment effects are consistent and positive, independent of where a learner ranked within their classroom at baseline. Together, these results suggest that neither learners' absolute levels of baseline reading proficiency, nor whether they ranked at the top or the bottom of the class before the programme started, served as a constraint to the programme's effectiveness.

tasks, but vary more in performance on higher order word recognition, paragraph reading, and comprehension tasks. In classrooms where more than one learner was tied as the least proficient reader, all the least proficient readers in the class were used in the analysis but down-weighted in proportion to how many learners a classroom is contributing to the analysis (i.e. inverse probability weighting).

Figure E4: Non-parametric intervention impacts by within-class rank of baseline reading proficiency




Source: author's calculation based on Funda Wande data (see SALDRU 2019).