THE EARLY GRADE READING STUDY:

A report on the baseline data collection and Year 1 programme activities

IMPROVING EARLY GRADE READING IN SOUTH AFRICA [P2.10.SA.IE]

BASELINE REPORT, 29 OCTOBER 2015

RESEARCH TEAM (AUTHORS)

Stephen Taylor

Jacobus Cilliers

Brahm Fleisch

Cas Prinsloo

Servaas van der Berg

Vijay Reddy

The Early Grade Reading Study team gratefully acknowledges the generous support from the donors and partners listed below.

basic education
Department:
Basic Education
REPUBLIC OF SOUTH AFRICA

Responsible Department: Project Design, Management and Governance

unicef
unite for children

planning, monitoring and evaluation
Department:
Planning, Monitoring and Evaluation
REPUBLIC OF SOUTH AFRICA

PROUDLY SUPPORTED BY ANGLO AMERICAN

UNIVERSITY OF THE WITWATERSRAND JOHANNESBURG

Project Administration

ZENEX FOUNDATION

GAUTENG PROVINCE
EDUCATION
REPUBLIC OF SOUTH AFRICA

Adapted from a programme developed and paid for by the Gauteng Department of Education.

class act

Service Provider

HSRC
Human Sciences Research Council

3ie International Initiative for Impact Evaluation

CONTENTS

ABBREVIATIONS AND ACRONYMS

DBE	Department of Basic Education
PIRLS	Progress in International Reading Literacy Study
RCT	Randomized Controlled Trial
ANA	Annual National Assessments
EGRS	Early Grade Reading Study
SGB	School Governing Body
RDD	Regression Discontinuity Design
EGRA	Early Grade Reading Assessment
HSRC	Human Sciences Research Council
CRC	Community Reading Coach
GPLMS	Gauteng Primary Language and Mathematics Strategy

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

This report describes the initial phase of the South African Department of Basic Education's Early Grade Reading Study (EGRS), especially the results from the baseline data collection. The EGRS project involves the implementation and evaluation of three alternative programmes all aimed at improving the acquisition of home language reading and literacy. The project is being implemented in two districts in the North West Province, in which the main home language is Setswana. The EGRS is working with the grade 1 class of 2015 for a two-year period, following the same learners into grade 2 in 2016.

The first intervention (implemented in 50 schools) provides teachers with lesson plans, additional reading support materials and training at centralized workshops twice a year. The second intervention (implemented in a different group of 50 schools) provides teachers with the same set of lesson plans and additional reading support materials but provides ongoing support to teachers through monthly on-site coaching and small cluster training sessions. The third intervention (implemented in a further 50 schools) holds weekly meetings with grade 1 parents to inform them of the importance of learning to read in the early grades and to empower them with knowledge and tools to become involved in their own child's reading acquisition.

Assignment to each of the three intervention or "treatment" groups and to a further group of 80 control schools was done through a computerized lottery. This ensures comparability across the groups. This randomized assignment is the key design feature of the EGRS and will be the basis for making claims about the causal impacts of each intervention on reading outcomes, when measured at the end of grade 1 and again at the end of grade 2.

The baseline data collection was administered by the Human Sciences Research Council (HSRC) in all 230 schools in February 2015. A random sample of 20 grade 1 learners per school participated in oral assessments of reading and pre-reading skills. Questionnaires were also administered to the school principal, to all grade 1 teachers and to parents of the 20 tested learners.

The learner tests were adapted from the well-known Early Grade Reading Assessment (EGRA) tool and covered the following skills in Setswana: expressive vocabulary, letter recognition fluency, short-term memory, phonological awareness, word recognition fluency, sentence reading and sentence comprehension.

The baseline testing confirms the success of the randomization: on all measures of reading ability there is a good balance across the four treatment groups. Some test subtasks yielded ceiling effects (where many learners achieved the maximum score) and other subtasks yielded floor effects (where many learners achieved the minimum score) but across the entire test there was a good variation in scores.

Girls outperformed boys on the reading tests. This advantage for girls is consistent with what is observed in standardized tests for higher grades in South Africa, such as in the Annual National Assessments. It is interesting that this gender gap is evident right at the start of grade 1, which would suggest that the girl advantage may be due to some factor other than school practices, most likely differences in the physiological development of girls and boys at this age.

The parent questionnaires indicate that the majority of parents or guardians in participating schools have low levels of education and they also have high levels of unemployment. Parent education is also predictive of cognitive ability and basic reading skills at the start of school. Similarly, the children of parents/guardians who read with them performed better on the baseline tests.

The grade 1 teachers are almost always female, and are rather old with an average age of 50. About 26% of teachers are 56 years old or older. Less than 10% of teachers are younger than 40. This has implications for the future provisioning of Foundation Phase education in general and for home language learning in particular (since other evidence suggests that low proportions of new Foundation Phase teacher graduates are specialized in the African languages). It may also affect the theory of change for the interventions if there are any differences in the way older teachers versus younger teachers react to support programmes. One teacher characteristic that was positively correlated with child learning outcomes was the teacher's own reading comprehension. It is unlikely that this reflects a causal relationship between teacher quality and learner performance since learners have just joined the school. It is possible that this reflects a selection effect where both learners and teachers select themselves into better schools. If this is indeed the case it represents a striking phenomenon: stronger children tend to be taught by stronger teachers.

Throughout the course of 2015 the three interventions have been running in schools. The lesson plans and reading support materials have been delivered to all schools in Treatment groups 1 and 2. Two training sessions have been held for teachers in Treatment 1 schools to date, in February and in July 2015. These 2-day training sessions were well attended – 100% attendance at the first and 85% attendance at the second. Schools in Treatment 2 have been receiving monthly coaching visits and afternoon clustered coaching sessions – three coaches share the 50 schools more or less equally between them. Teacher attendance rates at the cluster sessions were 100% in Term 2, 82% in Term 3 and 93% in Term 4. Some qualitative analysis based on classroom observation in a handful of Treatment 1 and 2 schools would suggest that the implementation of the prescribed lesson plans has been done with greater fidelity in Treatment 2 schools (coaching) than in Treatment 1 schools (training). However, this is based on a small sample of schools so no conclusions can be drawn with certainty.

One Community Reading Coach (CRC) has been recruited per Treatment 3 school and trained to run weekly afternoon sessions open to all grade 1 parents. A total of 30 sessions is scheduled for each year covering a total of 10 topics per year. Each topic has 3 sessions where the topic is the same but the activities of the session differ. Thus a parent can attend roughly 1

in 3 sessions and still be exposed to all topics, while parents who attend more regularly can still enjoy fresh activities. In a few schools it has proven difficult to recruit a CRC or the CRC has had to be replaced. Parent attendance has also been a challenge in these schools with attendance rates dropping from 35% for the orientation sessions and Topic 1 to 18% for Topic 4. Creative ways to encourage greater attendance in 2016 will need to be considered.

Interventions are scheduled to continue throughout 2016. A midline data collection is taking place between the 26th of October and the 13th of November 2015. The endline data collection is scheduled for October/November 2016.

BACKGROUND TO THE EARLY GRADE READING STUDY

The acquisition of reading is foundational to all subsequent learning; yet the majority of South African children are being left behind in this regard. The PIRLS study of 2006 showed that a striking 80% of South African children were not yet reading with comprehension after five years of schooling. The problem is particularly severe amongst poor children. Consequently, massive inequalities in educational achievement are established early in primary school and there is no evidence of these inequalities being reduced in later years. Therefore, early interventions, such as improving the acquisition of reading amongst poor children, can be expected to have larger effects than interventions later in the school programme.

The recently introduced Annual National Assessments (ANA) have raised public awareness of the weak literacy achievement of children in the primary school grades. Although the DBE and provincial education departments are implementing various strategies to support early grade reading, there is little or no sense of what is working and why. Moreover, there are competing models of support in the system. Some provinces favour the traditional model of teacher training workshops, while the province of Gauteng has provided additional graded readers and clearly scripted lesson plans together with specialist reading coaches who visit teachers on monthly basis to observe lessons and offer assistance. It is important that a national reading strategy be based on scientific evidence regarding what most improves the acquisition of reading.

A randomized controlled trial (RCT) design allows a credible estimation of the true causal impact of interventions, and thus has the potential to inform responsible policy decisions. Through the use of a lottery to allocate schools to intervention and control groups it is possible to construct a credible “counterfactual” scenario – what would have happened to those who received an intervention had they not received that intervention.

Moreover, by directly comparing the impacts on reading outcomes of alternative programmes, each with different cost implications, we can identify the most cost-effective intervention. This project is designed to explicitly compare the impact and cost of a new model of teacher development (on-school support) to the impact and cost of a more traditional model (training at central venues). The third intervention, which aims at improving parent involvement in schools and in home-based reading activities, relies on a rather different theory of change and is less expensive. By measuring the success of each intervention on the same scale, this project will provide a sense of the cost-effectiveness of different policy options.

The primary implementing partner is the South African government, in particular the Department of Basic Education. A key role is also being played by the North West provincial education department, which is contributing financially and is championing the project within the schools.

A service provider has been appointed to run the three interventions on behalf of the DBE for the purposes of this impact evaluation. The service provider is an organisation called “Class Act”, which is highly involved in partnerships with government to run literacy interventions. For

example, “Class Act” was a service provider in the Gauteng Province’s implementation of the Gauteng Primary Literacy and Maths Strategy (GPLMS) over the last few years. Programme interventions are being funded by a coalition of donors, including the ZENEX Foundation, UNICEF, Anglo American and the Department of Planning, Monitoring and Evaluation in the Presidency. These funds are being managed by the University of the Witwatersrand, which ran a tender for the service provider work and subsequently entered into a contract with Class Act.

The evaluation side of the project is being supervised by the Research Team while the data collection and capturing is being managed by South Africa’s Human Sciences Research Council (HSRC). The evaluation is being funded by the International Initiative for Impact Evaluation (3ie).

DESCRIPTION OF INTERVENTIONS

This study evaluates three different interventions, all aimed at improving early-grade reading in the home language, which in the case of the North West province is Setswana. All three interventions work with children entering Grade 1 at the start of 2015 over a two-year period (thus working with grade 2 learners in 2016).

Treatment 1: Training, scripted lessons, graded readers.

Treatments 1 and 2 aim to apply the same set of instructional practices in the teaching of home language literacy in grade 1 and 2 classrooms. Both treatments provide teachers with lesson plans, which are aligned to the curriculum as specified in the Curriculum and Assessment Policy Statements (CAPS) for home language literacy in the Foundation Phase. The lesson plans provide detailed specification for each lesson including information on methodology and content to be taught for each instructional day. The lesson plans incorporate the use of learning support materials including the government-provided workbooks as well as certain additional materials (graded reading booklets, flash cards, posters, etc.), which are provided through the EGRS. The graded reading booklets provide a key resource for the teacher to use in group-guided reading and individual work so as to facilitate reading practice at an appropriate pace and sequence of progression.

Treatment 1 trains the teachers on how to use the lesson plans and accompanying materials through central training sessions, each lasting 2 days, and occurring twice yearly. The first session was conducted in February 2015 and the second occurred in July 2015. Similar sessions are scheduled for 2016.

Treatment 2: Reading Coaches, scripted lessons, graded readers.

Exactly the same set of instructional materials (scripted lesson plans, graded reading booklets and other materials) is provided to Treatment 2 schools. However, instead of central training

sessions, ongoing support to teachers consisting of regular (monthly) on-school coaching from specialist "reading coaches" is provided. In addition to these on-site visits, there are occasional meetings with the coach and a small cluster of nearby Treatment 2 schools. The evaluation of treatments 1 and 2 should thus shed light on whether the fairly prescriptive instructional regime has the ability to improve reading acquisition and whether the mode of teacher support is important in mediating effectiveness.

Treatment 3: Parental involvement

Treatment 3 is designed to promote parental involvement to support their children's reading progress. At each of the 50 schools in this treatment arm a Community Reading Coach (CRC) was recruited. The CRC was identified through communication with the school principal who recommended a suitably qualified but available person in the community. The CRCs attend a 1-day training session facilitated by the service provider (Class Act) at the start of each school term (quarterly). The CRCs are trained to deliver weekly training sessions for grade 1 parents at their respective schools. A total of 30 sessions is scheduled for each year covering a total of 10 topics per year. Each topic has 3 sessions where the topic is the same but the activities of the session differ. Thus a parent can attend roughly 1 in 3 sessions and still be exposed to all topics, while parents who attend more regularly can still enjoy fresh activities. For their services, CRCs are paid a stipend of R400 per month (about $35).

The topics covered in these sessions include the importance of learning to read for later educational and labour market success, training on how to support their child's reading at home and the provision of low-cost materials and reading games to use at home.

THEORY OF CHANGE

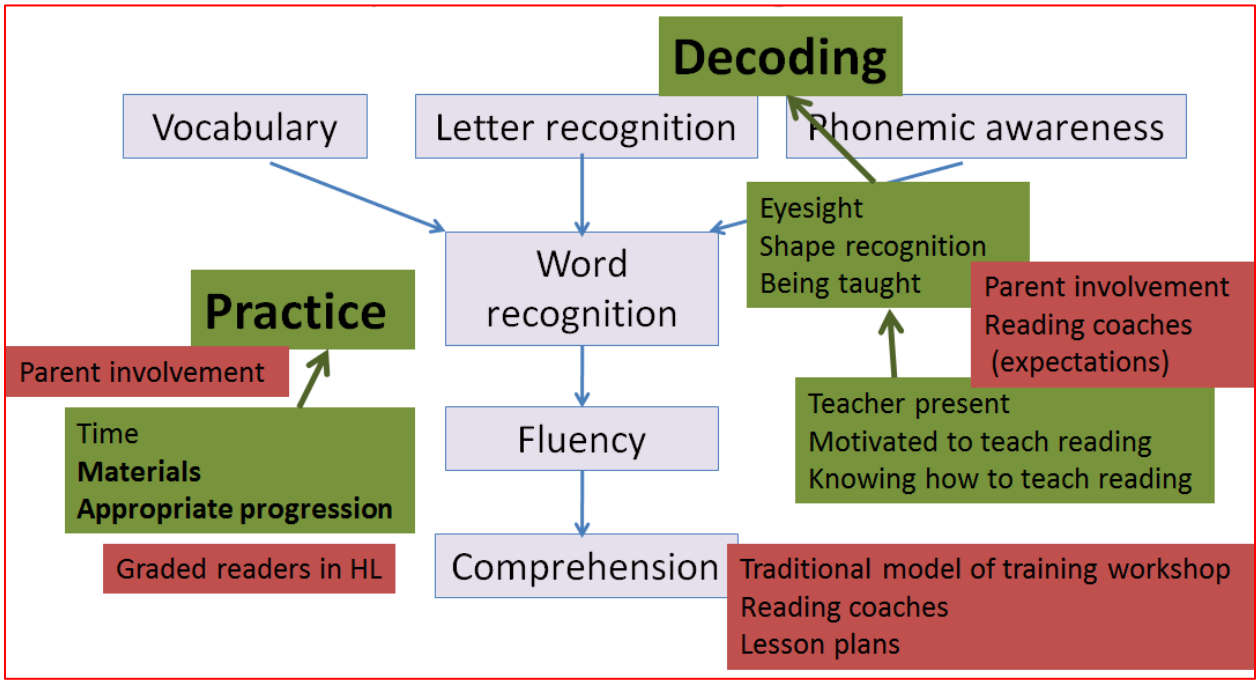
Reading acquisition

All three interventions relate to the educational theory of how reading acquisition occurs. Reading comprehension is the product of two components: vocabulary and decoding. To a great extent vocabulary (and more broadly language acquisition) comes naturally through speaking and hearing others speaking. Through speaking and hearing others speaking, phonological awareness also develops - this involves sound segmentation and recall of sound patterns. This phonological awareness is important for children to learn to decode. Particular written shapes are associated with particular sounds. Decoding thus consists of letter recognition and phonemic awareness. Unlike learning to speak, decoding does not come naturally; it is a method that must be taught systematically. It is important to emphasize that reading is produced by the product of vocabulary and decoding: If one has a perfect vocabulary but has not been taught the method of decoding one will not be able to read at all. Letter recognition and phonemic awareness are mastered through systematic teaching and consistent practice. This leads to the next stage of reading acquisition: word recognition. Through practice

and appropriate progression from simpler sounds and words to more complex ones word recognition becomes established leading to the next phase of reading acquisition: fluency. It is only once decoding and word recognition have become fluent that it is possible to reach the ultimate goal of reading comprehension.

In order to learn the basics of decoding, a child requires a teacher who is present, capable and motivated to deliver systematic reading instruction. In order for decoding to become fluent a child requires suitable graded materials and the discipline (perhaps imposed) to practice a lot. The interventions to be tested in this study address these needs in various ways. Figure 1 presents a theoretical diagram illustrating how reading acquisition occurs, what supportive conditions need to be in place and how each of the interventions being evaluated in the EGRS address key points in the development of reading acquisition.

Figure 1: Theoretical diagram of how reading acquisition occurs



There is a growing body of evidence from developing countries that early grade reading interventions can have a significant impact. The "EGRA Plus" programme administered in Liberia produced substantial gains in reading achievement relative to comparison children who did not receive the programme. Key aspects of this programme included a cascading model of reading coaches, the distribution of scripted lesson plans and reading assessment tools, and the dissemination of report cards to parents (Gove and Wetterberg, 2011). A supplementary reading curriculum administered in India also produced significant improvements in both public schools and pre-schools (He, Linden and MacLeod, 2009).

However, these studies cannot tell us which component of the intervention is responsible for the success of the program. This is important for policy purposes, because we want to find the most cost-effective intervention which could be scaled up by government. For example, the “EGRA plus” programme in Liberia was clearly highly resource-intensive because it required ongoing monitoring from qualified reading coaches, but we do not know if one might be able to reach the same results with a sub-component of the program. Moreover, there is uncertainty about the transferability of the findings given different language and social contexts.

Similar programs have been implemented in South Africa, but since they were not credibly evaluated, we do not know if they truly improved pupils’ reading acquisition. The Department of Basic Education typically holds training programs similar to our intervention 1; and Gauteng has implemented a model of reading coaches, similar to intervention 2. Since it has not been possible to produce a robust empirical impact evaluation of these programmes, we do not know if they truly work or not. Fleisch and Schoer (2014) attempted a Regression Discontinuity Design (RDD) to evaluate the impact of the Gauteng Primary Language and Mathematics Strategy (GPLMS) and findings pointed to a positive impact, though the findings were tentatively made given significant data constraints. Sailors et al (2010) evaluated a reading intervention in South Africa, which followed a similar model to intervention 2, but there are large methodological challenges to the study.

There is also a growing international literature providing information to parents and fostering parental involvement in schools can improve learning outcomes, but there is much we still do not know. In Pakistan, pupils who came from villages where the community was provided with information of school performance performed better in independently administered tests, compared to pupils from villages where no such information was administered. The improvement was particularly large for schools with low initial learning outcomes (Andrabi et al, 2013). In a different programme in India, school communities were informed of their school performance and also educated on their rights, roles and responsibilities in school governance through 8 public meetings. Education performance improved as a result (Pandey et al, 2013). However, in a recent impact evaluation in Kenya, informing parents on their child’s reading progress had zero impact (Lieberman, Posner and Tsai, 2013). The authors hypothesize necessary conditions for an information-intervention to work, all of which we address in our study: (i) information is new; (ii) it highlights under-performance and potential to improve; (iii) it is combined with measures which enable parents to act on this information.

All interventions aim to improve reading acquisition in the home language. Strictly speaking, the targeted outcome is home language literacy more broadly, since this is the Foundation Phase curriculum area being given support through our programmes. The choice to address home language literacy is motivated by research showing long-term benefits to strong home language skills prior to switching to a second language. Taylor and Coetzee (2013), for instance, show that in South Africa using home language as the language of instruction during grades 1, 2 and 3 has been associated with better English acquisition in grades 4, 5 and 6.

Intervention 1:

This programme is intended to impart the capacity to ensure that it is possible for the teacher to provide effective and systematic reading instruction in the classroom. Scripted lessons provide a structure to assure systematic practice and learning based on sound pedagogical theory. It can act as a substitute to low teacher capability or low motivation to prepare lesson plans. The accompanying reading materials aim to ensure that all the necessary instructional infrastructure is in place for a systematic reading programme to be effectively implemented.

Intervention 2:

The reading coach intervention provides more intensive training to improve teacher capacity. The assumption is that, just like learning to read, the ability to teach is a skill that needs to be developed over time and might not be accomplished in one-off training. Furthermore, the reading coaches could also improve teacher motivation as they are frequently monitored, provided with much-needed additional support, and can also find inspiration from watching an excellent example provided occasionally by coaches. This programme thus addresses both teacher capacity and teacher motivation. Another way to describe the difference between Treatments 1 and 2 is that while they share an underlying pedagogical theory of change (centered around instructional alignment and coherence using prescriptiveness as a vehicle), they differ in their theory of action (where Treatment 2 has a stronger component focused on changing behavior using accountability and motivation).

Intervention 3:

Parents pay a critical component to learning to read, as it requires continuous practice, both at school and at home. For parents to be *willing* to play this role they need to appreciate (i) the importance of reading; and (ii) that their child is most likely not learning enough at school and requires additional support. This is the purpose of the information. For parents to be *able* to play this role, they need to understand the necessary steps in learning to read and also have appropriate material to practice reading with their child. This is the purpose of the training and additional practice material.

Each of these three interventions has a different theory of change and also has different cost implications. Treatment 3 has the lowest cost amounting to approximately R16 000 per school per year (i.e. about $1200). Treatment 1 costs approximately R34 000 per school per year (i.e. about $2600). Treatment 2 is the most costly, amounting to approximately R63 000 per school per year (i.e. about $4800).

RESEARCH SITE

The EGRS is being implemented in the North West province, in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema. The North West province was chosen on the basis of 1) it being a relatively poor province, thus making it relevant to the majority of the underperforming South African school system; 2) it is relatively homogenous in terms of home language (Setswana) making it more affordable to develop learning support materials in a single language; 3) it is within driving distance from the Gauteng province where the national DBE is located; and 4) the senior management of the North West provincial education department were eager to partner with the DBE on this project. The district of Bojanala was excluded because another special targeted intervention was taking place in that district at the same time. The district of Dr Ruth Segomotsi Mompati was excluded since it is particularly far West of Gauteng and since enough schools existed in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema. Figure 2 shows a map of South Africa divided into the 83 education districts.

Figure 2: Map of South Africa showing education districts

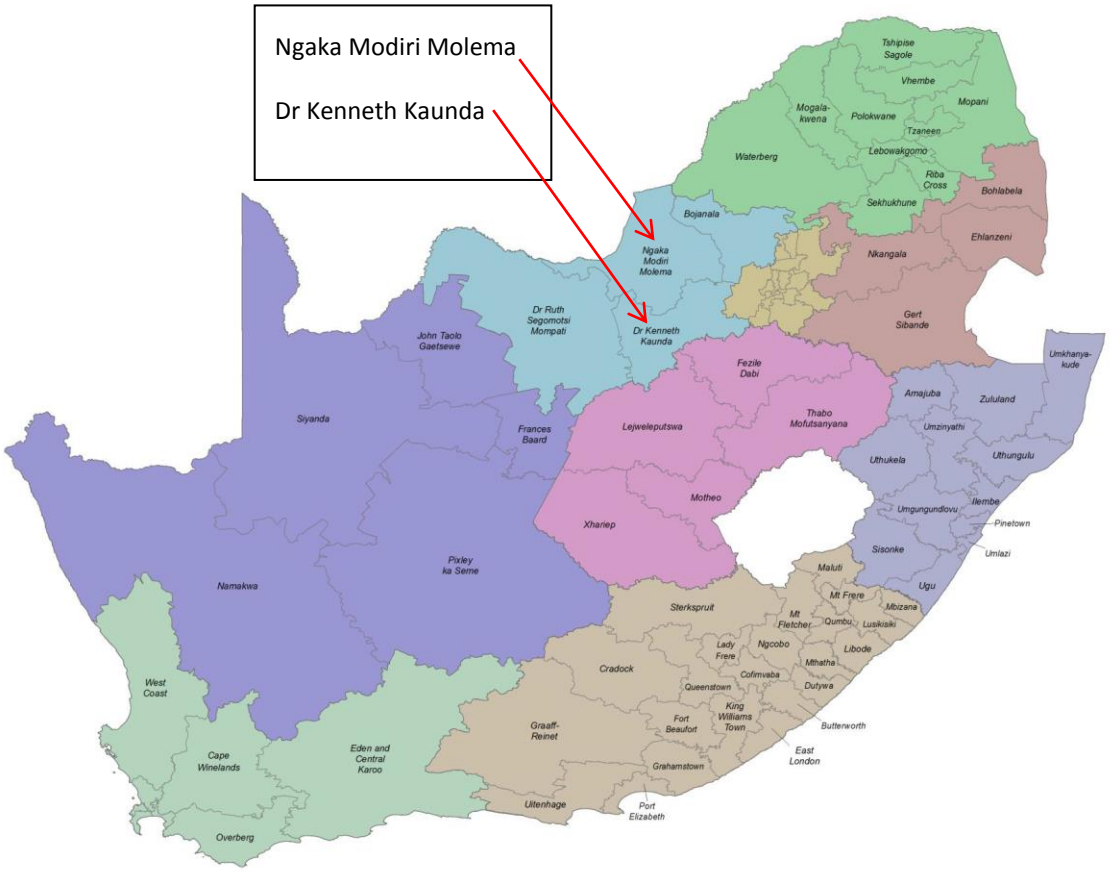


Table 1 below shows the total number of ordinary schools by phase for both Dr Kenneth Kaunda and Ngaka Modiri Molema districts in 2014. We see that Ngaka Modiri Molema district has the highest number of schools across all categories. Of the 248 schools in Dr Kenneth

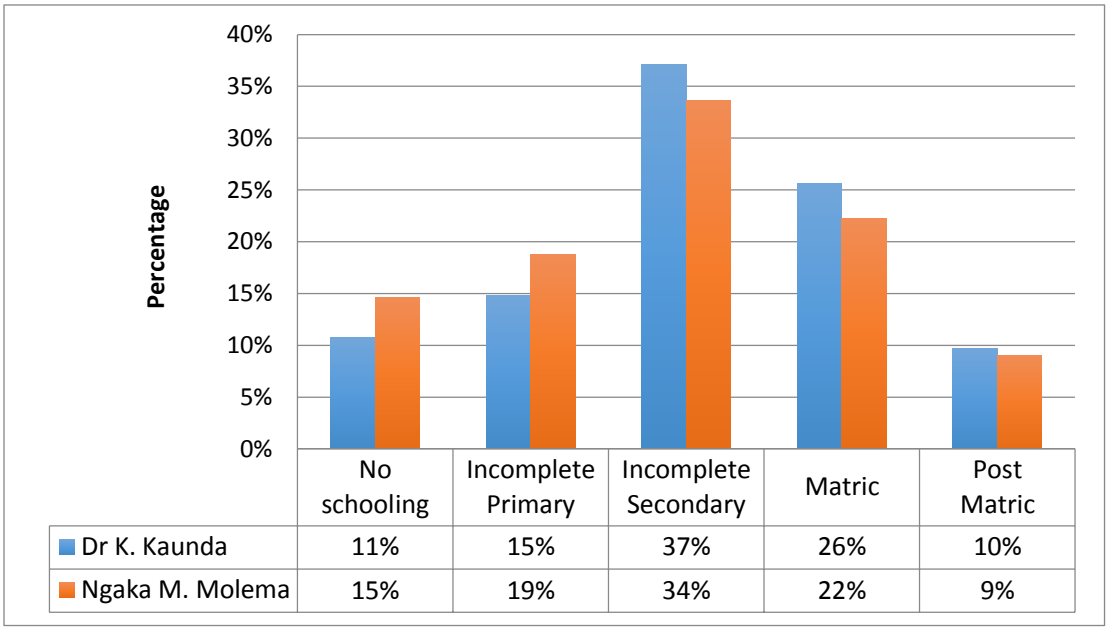
Kaunda district, 14 are independent schools while 11 of the 404 schools in Ngaka Modiri Molema district are independent schools. In Dr Kenneth Kaunda, 81% of schools are no-fee schools (classified as Quintile 1, 2, and 3 according to the official school poverty classification) while the equivalent figure was 91% of schools in Ngaka Modiri Molema district. This confirms that these two districts are largely poor and rural parts of South Africa. The choice of these areas for the EGRS project was deliberate so as to optimize the relevance of the study's findings to the large, underperforming and poor sections of South Africa's school system.

Table 1: Number of schools by phase in Dr Kenneth Kaunda and Ngaka Modiri Molema

	Dr Kenneth Kaunda		Ngaka Modiri Molema	
	Number	%	Number	%
Primary	149	60%	247	61%
Secondary	54	22%	76	19%
Combined	42	17%	67	17%
Intermediate	3	1%	14	3%
Total	248	100%	404	100%

In the 2011 Census, people were asked to indicate the highest level of education that they had completed. It referred to the highest level completed, not the level currently in, if the person was still studying. Figure 3 shows the education levels of adults aged 20 and older by district. The category 'Matric' refers to the secondary school leaving examination. This figure shows that Dr Kenneth Kaunda district had higher proportions of people who had a matric and post matric qualifications compared to those in Ngaka Modiri Molema district. Overall, this figure implies that the majority of people who would be parents to Grade 1 pupils would have relatively low levels of education.

Figure 3: Highest Education level for adults aged 20 and older



The Annual National Assessment (ANA) results provide an indication of school performance at the primary school level. It should be noted, however, that results are not comparable across time or across subjects or grades, since the tests cannot be equated to each other. In 2012 Dr Kenneth Kaunda performed better than Ngaka Modiri Molema. However, the opposite was true in 2013. This seems strange, and may reflect differential test administration and marking practices across time and district. The broad point to note is that language and mathematics performance in both of these districts is at a low level, allowing much room for improvement.

Table 2: Grade 3 learners achieving 50% and above by subject

Subject	Year	Dr Kenneth Kaunda	Ngaka Modiri Molema
Mathematics	2012	30%	18%
	2013	49%	48%
Language	2012	53%	41%
	2013	44%	49%

Table 3: Grade 6 learners achieving 50% and above by subject

Subject	Year	Dr Kenneth Kaunda	Ngaka Modiri Molema
Mathematics	2012	9%	7%
	2013	15%	23%
Language	2012	25%	19%
	2013	40%	45%

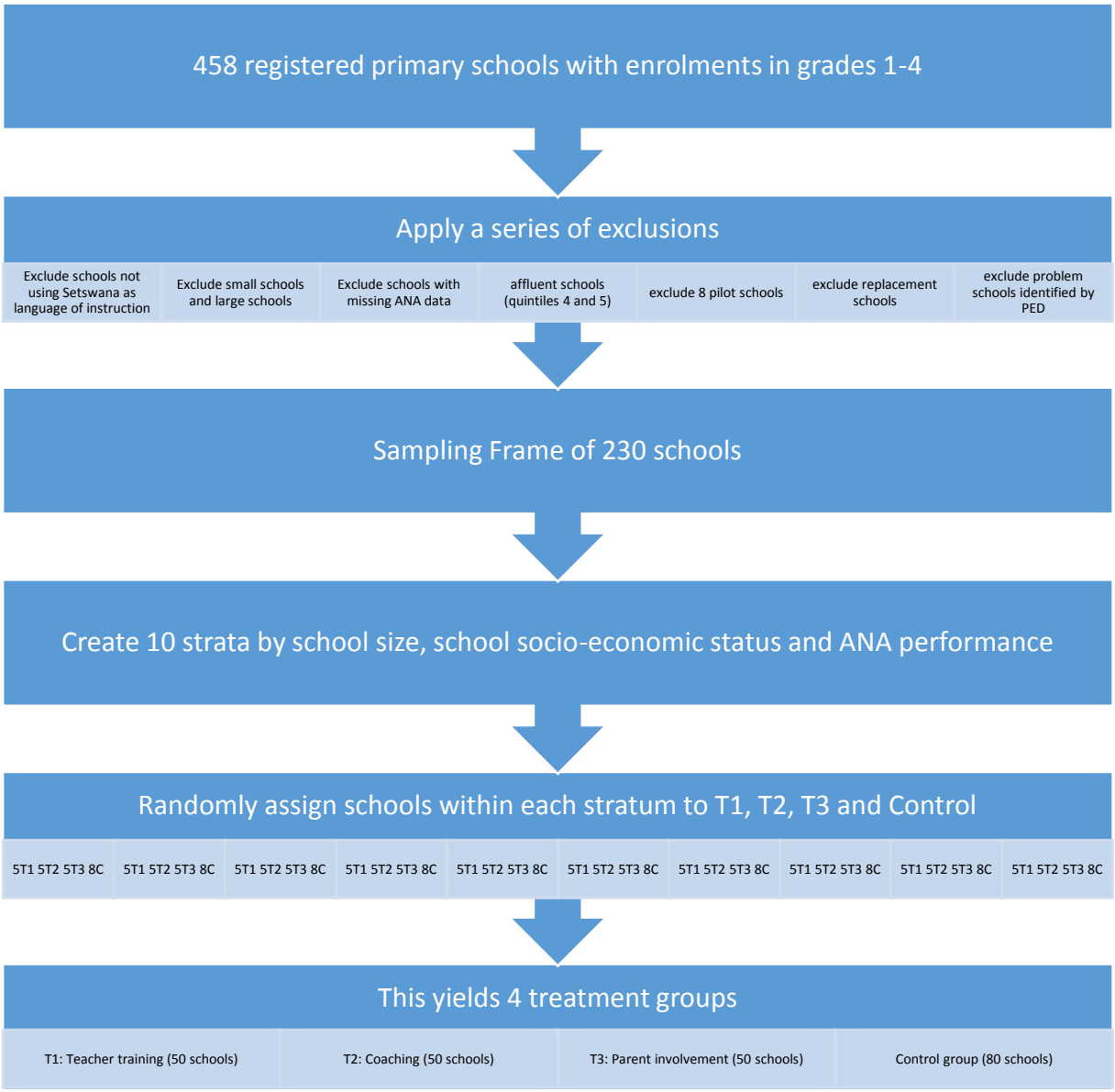
EVALUATION DESIGN

TREATMENT ASSIGNMENT AND SAMPLE SELECTION

Through a process of elimination we developed a sampling frame of 230 eligible schools. Beginning with 458 primary schools registered in 2014 administrative data in the districts of Dr Kenneth Kaunda and Ngaka Modiri Molema we started by excluding relatively affluent schools (those in quintiles 4 and 5). Next, we excluded schools in which the language of instruction in the Foundation Phase was not Setswana. We excluded schools which were missing in the 2014 ANA dataset. We also excluded 8 schools that had already been selected for the purposes of piloting of instruments through the course of this project. We further excluded particularly small schools (fewer than 20 grade 1 enrolments) since many of these schools would practice multi-grade teaching rendering the scripted lesson plans less appropriate. We also excluded particularly large schools (more than 180 grade 1 enrolments) to limit intervention costs. Three more schools were excluded after the North West PED checked our list of schools and found specific problems with these schools (e.g. the school had been closed down, or a particular conflict around school management was occurring in a school). After all of these exclusions 235 eligible schools remained. Using a random number generator, we then excluded 5 schools, which we retained as possible replacement schools. Thus we obtained the sampling frame of 230 schools.

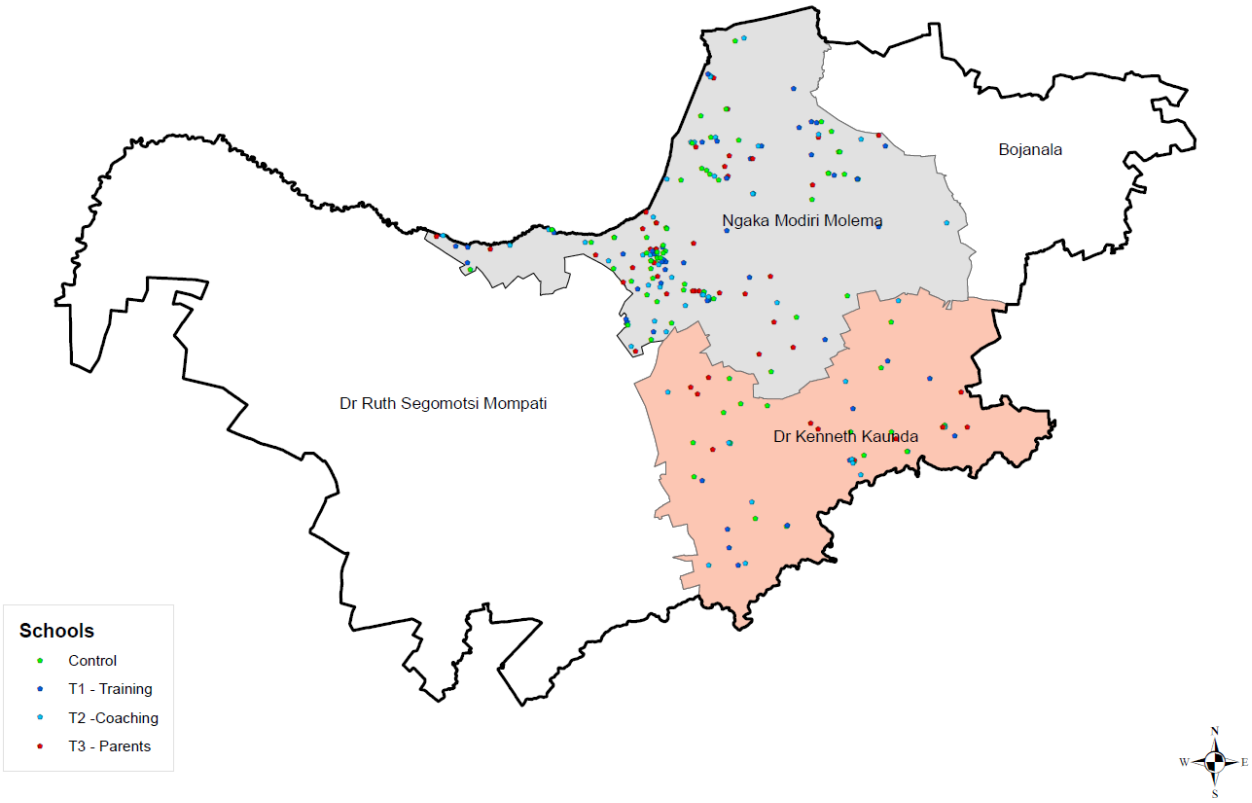
To increase power and assure balance between treatment arms, we performed stratified randomization. We created 10 strata of 23 similar schools based on school size, socio-economic status, and previous performance in the Annual National Assessments. Within each stratum, we then randomly assigned 5 schools to each treatment group and 8 to the control group. Thus we randomly assigned 50 schools to each treatment and 80 to the control. Given that we collect data on 20 grade 1 learners per school, this sample should be sufficient to identify a minimum effect size of 0.21 standard deviations when comparing a treatment group with the control group and a minimum effect size of 0.23 standard deviations when comparing two treatment groups. These calculations assume a 95% confidence interval, an alpha value of 0.8, an intra-class correlation coefficient (rho) of 0.3 and a correlation between pre- and post-test scores of 0.7. Figure 4 presents a schematic diagram to describe the sampling procedure that was followed.

Figure 4: Diagram showing sampling procedure



The following map shows the schools participating in the EGRS and indicates the treatment status of each school. Note that a few schools are not shown on the map due to missing or inaccurate GIS codes.

Figure 5: Map of North West province showing schools by treatment assignment



INSTRUMENT DEVELOPMENT AND PILOTING

The Research Team worked closely with the HSRC to develop four survey instruments for the baseline data collection: a learner test, a school principal questionnaire, a teacher questionnaire and a parent/guardian questionnaire. The learner test was designed in the spirit of the Early Grade Reading Assessment (EGRA) to be administered orally by a fieldworker to one child at a time. The test instrument used parts of the EGRA for Setswana, which had already been developed in South Africa. The letter recognition fluency, word recognition fluency and sentence reading components of the test were based on the Setswana EGRA instrument. A picture comprehension test (or expressive vocabulary test) was included since this was expected to be an easier pre-literacy skill testing vocabulary, and thus useful for avoiding a floor effect at the start of grade 1 when many children are not expected to read at all. A phonemic awareness test component was also added. Similarly, a digit span memory test was included – this involved repeating by memory first two numbers, then three, and so forth up to six numbers, and the

same 5 items for sequences of words. The logic of including this test of working memory is that it is known to be a strong predictor of learning to read. Thus, when estimating the impact of the three interventions after endline testing we can include as a control variable a measure of the child's working memory at baseline and in this way improve the precision of treatment effect estimates.

The school principal, teacher and parent questionnaires were designed in order to collect information to be used in the measurement of heterogeneous treatment effects (i.e. differential impact across relevant sub-groups of schools or learners) and to measure changes in intermediate outcomes along the hypothesized causal chain for each intervention. The parent questionnaire was sent home with those learners who were tested and then brought back to the school on a later day, to be collected by the fieldworkers on a return visit. In addition, a data linkage form was developed upon which all learner names and unique identifier numbers were linked to the appropriate teacher unique identifier and teacher name. All these instruments and the entire data collection process were piloted in 5 schools on the 3rd and 4th of September 2014. Following lessons learnt from the piloting, revisions were made to the instruments.

ETHICAL CLEARANCE

The methodology, with the intended instruments related to the baseline data collection, was formally submitted to the HSRC's Research Ethics Committee in February 2014. The project was approved in principle (i.e., provisionally) on 24 March 2014, pending submission of the final field-test and baseline instruments and site permissions. The relevant field-test documents were submitted and approved on 29 August 2014. Subsequently, after final revisions to the procedures and instruments for the baseline data collection, and submission of final site permissions along with an application for recertification for another year, ethics clearance was provided on 21 January 2015 for the baseline data collection.

DATA COLLECTION

Baseline data collection comprised visits to all 230 EGRS schools (150 treatment and 80 control schools) in order to assess the Setswana language proficiency of 20 Grade 1 learners per school (4 600 learners in total). The HSRC hired a fieldwork agency to recruit fieldworkers and manage their transport to schools during the fieldwork. A total of 60 fieldworkers were recruited, comprising 30 former teachers (to conduct the learner testing) and 30 other fieldworkers who were not necessarily education-specific. The plan was for fieldwork to be conducted in teams of 2, with one fieldworker conducting the learner testing and the other administering the school principal and teacher questionnaires, all in the course of a 1-day visit to each school. The HSRC were directly responsible for the printing and packaging of all instruments and passed these on to the fieldwork agency. The HSRC facilitated the training of the fieldworkers. This was initially a 1-day session. However, after some problems were evident on the first day of data collection a decision was taken to recall all fieldworkers for an additional 1-day re-training.

Data collection occurred between the 4th and the 24th of February. Monitoring of fieldwork occurred at two levels. Firstly, the HSRC sent monitors to observe fieldwork in a randomly selected (by the Research Team) 10% of schools, i.e. 23 schools. Secondly, the DBE made telephone calls to school principals to find out about how fieldwork had occurred at the school. Reports on both levels of monitoring were compiled.

A number of challenges were experienced during the data collection. Firstly, there were some problems with respect to the logistics of school visits. Although all schools should have known about their participation in the EGRS through a set of meetings with all principals at the end of 2014 and through letters from the NW PED, the telephone numbers for schools obtained through the DBE's EMIS data were in some cases incorrect or outdated. To add to this problem, the fieldwork schedule of which fieldworkers should attend which schools on which days, as arranged by the subcontracted fieldwork agency was regularly updated resulting in appointments with schools either not being set up or set up rather late. Fortunately, this did not lead to any outright refusals from schools to participating, and those few schools where initial refusal occurred were re-visited on a later day. Another challenge was that on some occasions fieldworker transport was not efficient so that a team of fieldworkers arrived late at school. This would have compromised the quality of data collection at such a school due to time constraints. A further challenge experienced is of incomplete return of instruments by fieldworkers, possibly partly due to late arrival at schools.

The intention was for the fieldworkers to randomly sample 20 learners per school, using a specified procedure. The fieldworker was to obtain from the teachers the full list of children enrolled in grade 1, putting one class list below the next if a single grade list was not provided. The fieldworker was to tally the total number of children and divide this number by 20. The answer was then to be rounded up to the nearest whole number, n. The fieldworker was then to start with the third learner and select every n^{th} learner for inclusion in the sample. Upon reaching the end of the list the fieldworker was to go back to the top of the list and continue selecting every n^{th} learner, not counting previously selected learners, until 20 learners have been selected. Monitoring of fieldwork indicated that in a few cases, the fieldworkers may not have followed the procedure perfectly. However, there was no evidence of systemic sampling of learners through anybody's recommendation. If a fieldworker attempted the procedure but misunderstood it the resulting sample should still be effectively random. Therefore, there is no reason to expect a systematically stronger or weaker sample to have been selected, and there is certainly no reason to expect any differences in sampling across treatment groups.

The following tables provide a sense of the data completeness as far as instrument returns is concerned. Learner testing occurred in all 230 schools, providing a realized sample of 4539 learners. Table 4 shows that in the majority of schools (204 out of 230 schools) exactly 20 learners were tested and the data successfully captured. In 4 schools there were 21 learners tested. It is not clear why this occurred. It may have been a counting error by the fieldworker or perhaps a small school only had 21 learners and it was felt that a single learner should not be left out. The few cases of 15, 16, 17 and 19 learners tested is not unexpected since there are

known to be some small schools in the sample. Schools with fewer than 20 grade 1 enrolments in 2014 were excluded from the sampling frame; but we know from administering the interventions that some of the schools have lower enrolments in 2015. Although not impossible, it does seem unlikely that schools would only have had 9 or 10 grade 1 enrolments. To some extent, therefore, incomplete fieldwork may have led to fewer than 20 learners were tested.

Table 4: Number of learners successfully tested per school

Number of learners	No of schools with this number of learners
9	1
10	1
15	2
16	4
17	2
19	12
20	204
21	4
Total	230

As Table 5 indicates, the return of parent questionnaires was rather erratic. The parent questionnaire was sent home with tested children and was meant to be brought back to the school and then collected on a later day by the fieldwork agency. The weakness of this method is that children may not always bring the questionnaire back. However, it is more reliable than asking children themselves about hoe characteristics. It is concerning, however, that no parent questionnaires were returned in 49 schools. This is most likely a reflection of poor fieldwork or of a lack of cooperation from school staff. Importantly, there was no significant pattern of instrument return across treatment group, not that one would expect that given that fieldworkers were blind to treatment allocation and that interventions had not yet commenced. For those schools where parent questionnaires were returned the return rates were not too bad, as described in Table 6. About 60% of schools had return rates of greater than 50% (i.e. 10 learners or more). If one excludes, the schools where no parent questionnaires were returned (not shown in Table 6), then about 75% of schools had return rates of 50% or more, and about 60% of schools had return rates of at least 75% (i.e. 15 learners).

Table 5: Numbers of returned learner tests and parent questionnaires

	Learner tests		Parent Questionnaires	
	No Students	No schools	No Students	No schools
Control	1575	80	856	62
Treatment 1	983	50	559	42
Treatment 2	982	50	569	41
Treatment 3	999	50	500	36
Total	4539	230	2484	181

Table 6: Number of parent questionnaires returned per school

Number of parent questionnaires	No of schools with this number of learners	Cumulative percentage
0	49	21.3
2	2	22.17
3	1	22.61
4	7	25.65
5	4	27.39
6	3	28.7
7	3	30
8	7	33.04
9	9	36.96
10	9	40.87
11	10	45.22
12	9	49.13
13	11	53.91
14	15	60.43
15	18	68.26
16	15	74.78
17	8	78.26
18	18	86.09
19	18	93.91
20	14	100

Two separate instruments were supposed to be administered to all teachers in grade 1. The first instrument is the teacher questionnaire, which collected a variety of information about teacher demographics, attitudes and practices. The second instrument was a short reading fluency test for teachers (to be described in more detail in a later section of this report). A questionnaire was also given to school principals to complete. Table 7 shows the numbers of teacher and principal instruments returned. The principal return rate is straightforward since one expects one questionnaire per school. In 14 schools no principal questionnaire was completed and returned

by the fieldwork agency. There were 326 teacher questionnaires returned that could be linked to learners.[1] There were also 320 teacher fluency tests returned. However, in some of these cases the teacher questionnaire data could not be linked to the teacher fluency test data. This may partly reflect inaccurate personal details and incorrect application of unique identifiers by the fieldworkers. However, manual investigation of these unmatchable cases would suggest that there may have been some teachers who only completed one of the instruments. 286 teachers were successfully matched across the two instruments. As Table 7 indicates, although over 300 teacher questionnaires and fluency tests were returned, since more than one teacher could be interviewed per school, the number of schools in which at least one teacher was surveyed was unfortunately less than the intended 230 schools. In only 198 schools was at least one teacher questionnaire returned. The fluency test was successfully administered and captured in 194 schools. It is possible that teacher refusal to be tested could have contributed somewhat to the non-return of teacher fluency data. However, the appropriate procedure for the fieldworker to follow in the case of refusal to participate was to return the test instrument with a field indicating whether the teacher was willing to participate – 40 teachers were not willing according to this variable and thus had missing data on the test score variables.

Table 7: Numbers of returned teacher and principal questionnaires

	Teacher Questionnaire		Teacher Fluency test		Principal Questionnaire
	No Teachers	No schools	No Teachers	No schools	No schools
Control	112	70	107	65	71
Treatment 1	72	43	73	44	49
Treatment 2	77	46	80	45	48
Treatment 3	65	39	60	40	48
Total	326	198	320	194	216

In summary, it would appear that imperfect fieldwork contributed to a lower than intended return rate of survey instruments. Fortunately, non-return is not systematically related to treatment assignment. It is also fortunate that the main priority of learner testing was generally fairly complete. Rather a lot of non-return occurred for the parent, teacher and principal instruments. Moreover, even when instruments were returned there was rather a lot of item non-response. This will limit the evaluation analysis once midline and endline data are collected in several ways. Firstly, the main impact estimation model will not include many parent, teacher and school covariates as controls. The value of such controls is to slightly improve statistical power when estimating the treatment effects. However, this power gain is rather marginal so the loss is not too bad. Moreover, in an RCT setting where the source of variation in treatment assignment is strictly exogenous by design one would expect no bias to have to control for through the

[1] In fact a few more teacher instruments were returned and captured but due to incomplete identification information these could not be linked to learners at schools and were therefore excluded from the merged dataset and this analysis.

inclusion of covariates. Therefore, the inclusion of many covariates in an RCT regression model is in any case not always favoured by analysts. A more worrying limitation is that missing information on baseline characteristics will mean a reduced effective sample size when estimating certain heterogeneous treatment effects and when estimating impacts on intermediate outcomes, such as teacher attitudes and practices. One way to mitigate these problems will be through collecting much of the same information in the midline and endline surveys (November 2015 and November 2016). Certain information, such as teacher age, is not expected to change in response to treatment and can therefore be used in the estimation of heterogeneous treatment effects even if the information was collected after interventions commenced. Treatment effects on intermediate outcomes can be estimated without controlling for baseline characteristics since there is no reason to expect any differences between treatment groups other than because of the causal effect of the interventions. The disadvantage is that power is reduced through the lack of controlling for baseline variation.

A number of steps are being taken in the midline data collection (scheduled for 26 October – 13 November 2015) to improve the data collection. The Terms of Reference for the subcontracting of a fieldwork agency is now much more detailed with respect to fieldworker selection criteria, conditions around approval of and payment for deliverables, and overall functionality criteria for the fieldwork organization. The entire procurement process of the fieldwork agency for midline data collection is happening in good time to ensure adequate lead up time to the data collection. Instead of a single day of fieldworker training, there will be a three-day training programme for fieldworkers including a practice round of data collection (with monitoring and feedback) at five schools not included in the project. The Terms of Reference has specified that exactly 40 fieldworkers should be recruited, 20 of whom will administer the learner tests and must have expertise in early grade teaching. The fieldwork schedule also needs to be submitted well in advance to the HSRC with schools already having been contacted and appointments fixed for specific days made. This process of communicating with schools is also likely to be smoother since we now have an updated database of contact information, which the DBE compiled using information collected in baseline questionnaires and by the implementing agent for interventions. Finally, extensive revisions have been made to the midline instruments, especially the shortening of the school principal and teacher questionnaires.

DATA CAPTURING AND CLEANING

Questionnaires were unpacked and data was cleaned within the HSRC by in-house data capturers. Six separate datasets were thus captured, corresponding to the different instruments. The six datasets were the data linkage file (linking learner, teacher and school unique identifiers), the learner test data file, the parent questionnaire data file, the school principal data file, the teacher questionnaire data file and the teacher reading fluency test data file. A preliminary version of these datasets was provided by the HSRC to the Research Team. Initial analysis of this data identified several data issues. These included one school that was missing from the data, some obvious mistakes in unique identifiers of learners and teachers, one data file that had mixed up the school identifier numbers, etc. The Research Team then sent a set of queries to the HSRC, who in turn investigated these issues. After some re-capturing and

cleaning, the HSRC then provided the final baseline datasets to the Research Team. Some of the data queries had been satisfactorily resolved (e.g. the “missing” school was found – it had initially been confused by the data capturers with another school with a very similar name), while other issues could not be fully resolved as they stemmed from fieldworker errors in capturing information.

Even after receiving the final datasets from the HSRC, the Research Team needed to do additional data cleaning, which was clearly needed once attempting to merge the various datasets. For example, there were a few duplicate learner IDs that needed to be adjusted by manually looking at learner names and surnames and comparing with the linkage data file. Similarly, a number of EMIS numbers (official school unique identifier) were incorrect in the parent questionnaire dataset. These were easily identified and corrected. The data cleaning done by the Research Team is recorded in STATA do-files, which will be made publicly available at the end of the project.

BASELINE RESULTS

LEARNER TEST SCORES

The baseline learner test instrument, which will be made publicly available at the end of the project, was adapted from the Setswana Early Grade Reading Assessment (EGRA). The tests were therefore administered to one child at a time. The average time per test was about 15 minutes. In view of the fact that the baseline assessment took place at the very start of grade 1, one modification was the inclusion of some items which could be described as assessing pre-literacy skills so as to be sure to avoid a floor effect (where a substantial proportion of learners score zero or close to zero on the test as a whole). Section A thus consisted of 10 picture comprehension items, which test expressive vocabulary – a skill which should be fairly well developed by the start of primary school. Six of these items were pictures of well-known objects, such as a car and a spoon. The remaining four pictures displayed some sort of action, such as a bird flying or a child sleeping. In each case, the learner was asked to say the Setswana word for the object or action.

One problem encountered in the scoring for Section A (as for Sections C and D), was that the fieldworker was supposed to mark each of the ten items correct or incorrect and also to indicate the total score out of ten. However, in some cases the sum of the item scores did not tally to the total score recorded by the fieldworker. This occurred for 200 learners out of the total of 4540 learners. In cases where the fieldworker left all individual items blank but entered a valid total score, we used that total score. In cases where the total score was missing we imputed the sum of the individual item scores as the total score. In cases where the difference between the calculated sum of individual scores was 1 or 2 points away from the recorded total score we decided to use the calculated sum of scores under the assumption that this was probably a fieldworker counting error. In cases where the difference was greater than 2 points it is unlikely

that this could be a counting error and therefore we used the recorded total score under the assumption that scoring the individual items was erratically done.

The summary statistics for all items in sub-tests A and C are presented in Table 8, and for sub-test D in Table 9. The summary statistics for the total scores per subtask as well as an overall composite test score are shown in Table 10. For Section A it can be seen that most learners did rather well in this section. Items 4 and 9 were the hardest items in Section A with 61% and 62% of learners getting the answers correct, respectively. The average score out of 10 was 8.58. The inclusion of these easier items was deliberate since it was expected that the majority of learners would struggle with the traditional EGRA items, which require some reading ability. Figure 6 confirms that the majority of students achieved scores of 8, 9 and 10 out of 10 and that there was a ceiling effect on this subtask. The figure also shows the distributions of scores for each of the four treatment arms. The distributions are virtually identical for each treatment group. This confirms the success of the randomization to ensure a well-balanced treatment assignment. One concern with Section A is the low Cronbach's alpha (0.52) that was obtained, indicating that the items are not combining to present a very reliable measure of an underlying construct. There were no individual items that were so problematic that if removed would increase Cronbach's alpha. This analysis of Cronbach's alpha is presented in Table 11.

Table 8: Summary statistics for items in sub-tests A and C

	Observations	Mean	Std. Dev.	Min.	Max.
Score for Item A1	4509	0.92	0.28	0	1
Score for Item A2	4487	0.96	0.19	0	1
Score for Item A3	4503	0.99	0.10	0	1
Score for Item A4	4384	0.61	0.49	0	1
Score for Item A5	4495	0.97	0.17	0	1
Score for Item A6	4441	0.85	0.36	0	1
Score for Item A7	4459	0.84	0.36	0	1
Score for Item A8	4493	0.96	0.20	0	1
Score for Item A9	4391	0.62	0.49	0	1
Score for Item CW1	4404	0.89	0.31	0	1
Score for Item CW2	4404	0.82	0.38	0	1
Score for Item CW3	4296	0.52	0.50	0	1
Score for Item CW4	4216	0.20	0.40	0	1
Score for Item CW5	4174	0.12	0.33	0	1
Score for Item CN1	4399	0.93	0.26	0	1
Score for Item CN2	4383	0.85	0.36	0	1
Score for Item CN3	4276	0.48	0.50	0	1
Score for Item CN4	4206	0.24	0.43	0	1
Score for Item CN5	4132	0.14	0.34	0	1

Table 9: Summary statistics for items in sub-test D

	Observations	Mean	Std. Dev.	Min.	Max.
Score for Item D1_1	4163	0.36	0.48	0	1
Score for Item D1_2	4131	0.26	0.44	0	1
Score for Item D1_3	4112	0.21	0.41	0	1
Score for Item D1_4	4094	0.19	0.40	0	1
Score for Item D1_5	4090	0.16	0.37	0	1
Score for Item D1_6	4070	0.13	0.34	0	1
Score for Item D1_7	4144	0.25	0.43	0	1
Score for Item D1_8	4158	0.28	0.45	0	1
Score for Item D1_9	4119	0.22	0.42	0	1
Score for Item D1_10	3999	0.13	0.33	0	1
Score for Item D1_11	3990	0.10	0.29	0	1
Score for Item D1_12	3978	0.08	0.28	0	1

Table 10: Summary statistics – Aggregate test scores

	Observations	Mean	Std. Dev.	Min.	Max.
Picture comprehension /10	4539	8.58	1.40	0	10
letters correct	4452	5.08	9.86	0	99
digit span words /5	4539	2.45	1.31	0	5
digit span numbers /5	4539	2.54	1.34	0	5
digit span total /10	4539	4.99	2.43	0	10
phonemic awareness /12	4539	2.17	3.11	0	12
words correct	4447	1.91	5.28	0	50
sentence reading comprehension /3	4539	0.73	1.24	0	3
number of sentence words correct /15	4539	1.22	3.38	0	15
combined score mean 0 SD 1 pca	4385	0.00	1.00	-2	5

Figure 6: Kernel Density Curves for Section A (expressive vocabulary) by treatment arm

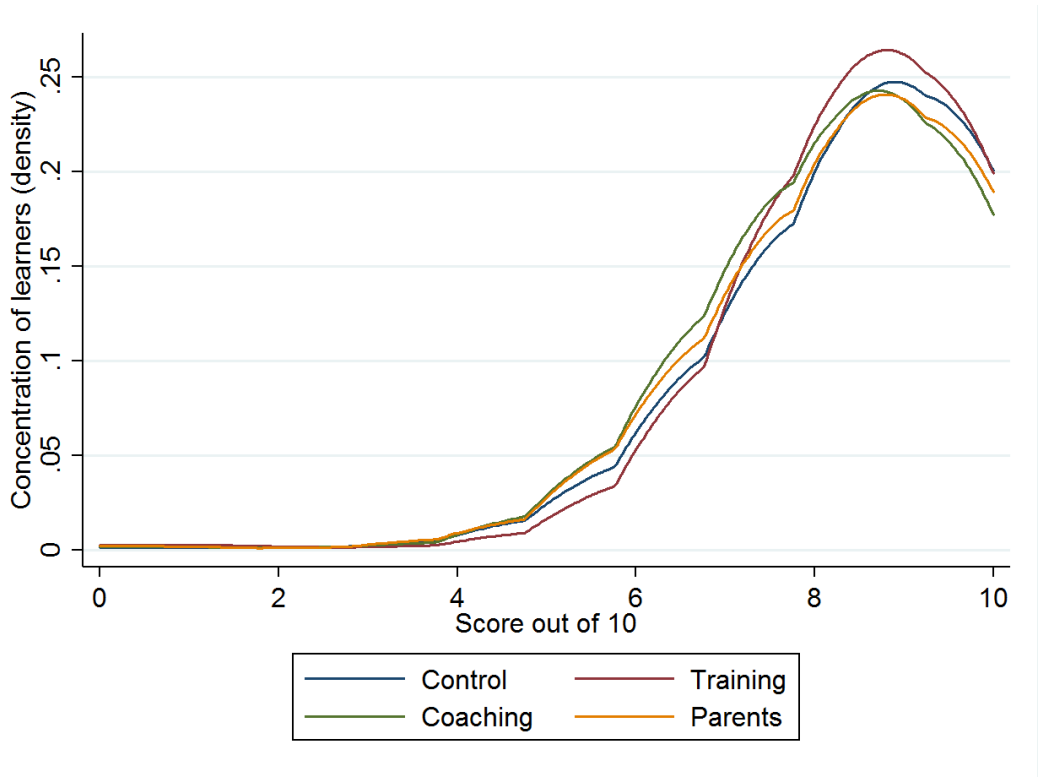
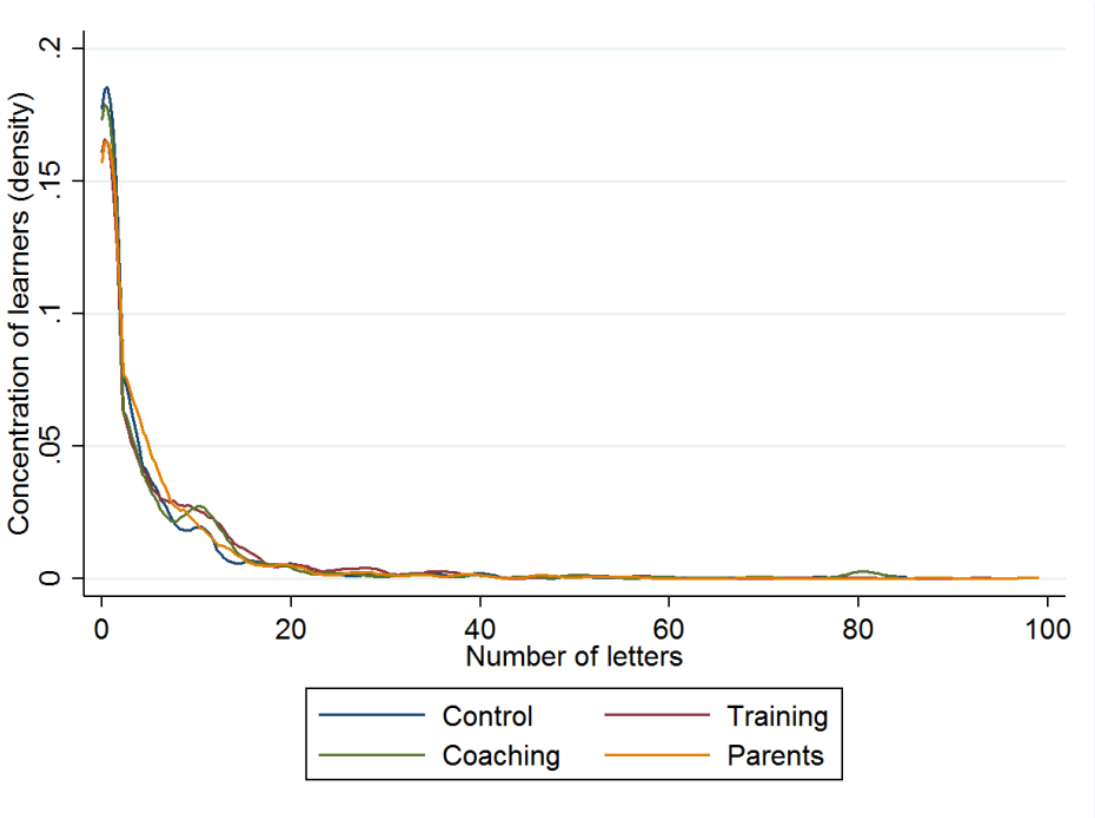


Table 11: Cronbach's alpha for Section A (expressive vocabulary)

Item	Observations	Sign	Item-test correlation	item-rest correlation	Alpha
A1	4509	+	0.43	0.25	0.49
A2	4487	+	0.33	0.19	0.50
A3	4503	+	0.31	0.24	0.51
A4	4384	+	0.54	0.21	0.51
A5	4495	+	0.34	0.22	0.50
A6	4441	+	0.43	0.18	0.51
A7	4459	+	0.51	0.27	0.48
A8	4493	+	0.40	0.26	0.49
A9	4391	+	0.63	0.34	0.45
A10	4488	+	0.43	0.29	0.48
Test					0.52

Section B was the letter recognition test, which is a conventional EGRA task. Learners were given 60 seconds to read as many letter sounds as possible. At the end of the 60 seconds the fieldworker captures the number of letters reached as well as the number of letters correct. As Table 10 indicates, the number of letters reached was typically far higher (averaging 29) than the number of correct letter sounds read (averaging 5). This is to be expected since the fieldworker moves the pointer along to the next letter if the learner has not provided an answer after three seconds. There was quite a substantial floor effect on this subtask, as indicated by Figure 7. About 42% of learners could not read and pronounce any letter sounds. As with Section A, the kernel density curves were virtually identical across the treatment arms. This confirms that the groups are well balanced on baseline.

Figure 7: Kernel Density Curves for Section B (letters correct) by treatment arm



Section C of the test was a digit span memory test designed to provide a measure of a child's working memory, which is known to be a strong predictor of learning to read. The item involved the fieldworker saying two unrelated Setswana words and the learner needed to repeat them back to the fieldworker from memory. The second item involved three unrelated words, the third had four words, the fourth had five words and item five involved six words. The next 5 items followed the same pattern but using numbers (spoken in Setswana). The rationale for including this sub-task is not because it is a reading outcome but because it is predictive of learning to read. This will be important for our final impact evaluation analysis where including baseline measures of cognitive ability, if these are well correlated with reading outcomes at the endline, can be expected to account for some of the variation in reading outcomes and thus increase the

precision with which we can measure the impact of the interventions. There was a good spread of achievement on these items. For the word span test, 89% of learners could successfully repeat the two-word sequence, with smaller proportions being able to repeat more words, down to only 12% who could repeat six words. With the numbers section, 93% of learners could successfully repeat the two-number sequence while only 14% could repeat the six digit sequence. The reliability of Section C is somewhat better than that observed for Section A, as the analysis of Cronbach's alpha in Tables 12 and 13 demonstrate. Figure 8 demonstrates that neither a floor effect nor a ceiling effect exists for Section C. Rather, the distribution of scores approximates a normal distribution, which is encouraging for the purposes of providing a good baseline measure of learner cognitive ability. The figure also confirms good balance across treatment groups.

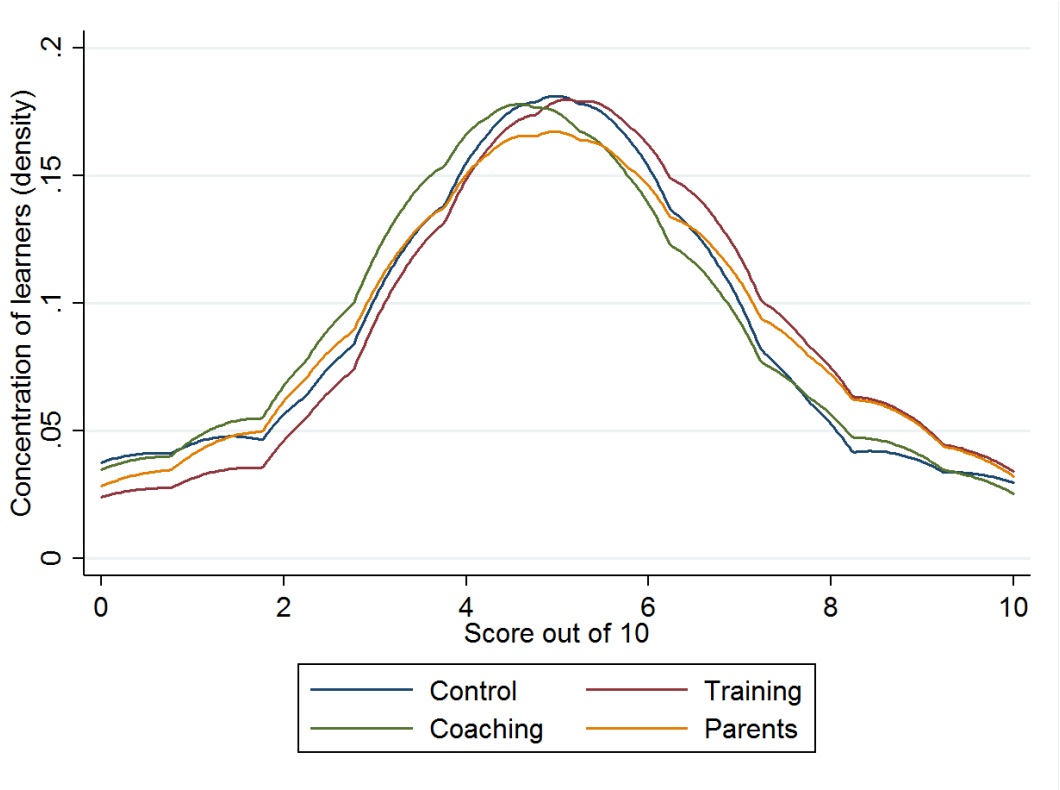
Table 12: Cronbach's alpha for Section C.1 (short-term memory – words)

Item	Observations	Sign	Item-test correlation	item-rest correlation	Alpha
CW1	4404	+	0.58	0.41	0.66
CW2	4404	+	0.66	0.46	0.64
CW3	4296	+	0.78	0.55	0.61
CW4	4216	+	0.70	0.48	0.64
CW5	4174	+	0.59	0.40	0.67
Test					0.70

Table 13: Cronbach's alpha for Section C.2 (short-term memory – numbers)

Item	Observations	Sign	Item-test correlation	item-rest correlation	Alpha
CN1	4399	+	0.52	0.36	0.70
CN2	4383	+	0.63	0.43	0.68
CN3	4276	+	0.79	0.56	0.64
CN4	4206	+	0.76	0.56	0.63
CN5	4132	+	0.67	0.49	0.66
Test					0.71

Figure 8: Kernel Density Curves for Section C (short-term memory) by treatment arm



Section D tested phonological awareness, in three different ways. For the first six items the fieldworker read a Setswana word (e.g. “pitsa”) out loud and the learner had to break the word down into its smallest sound components or phonemes. For items D7 to D9 the fieldworker read a word out loud (e.g. “sega”) and the learner then had to suggest another word beginning with the same two letter sounds (e.g. “seba”). For items D10 to D12 the fieldworker read a word out loud (e.g. “yona”) and the learner then had to suggest another word ending with the same two letter sounds (e.g. “bana”). Table 9 shows that most learners struggled with this subtask, especially with items D10 to D12. The average score out of 12 on Section D was 2.17. There was also a floor effect, as can be seen in Figure 9. Again, the scores appear balanced across treatment groups. Members of the Research Team have observed this subtask being administered and have been somewhat concerned because it is rather difficult for fieldworkers to implement and confusing for learners to understand. While it is designed to test phonological awareness, which is an important component in learning to read, the “rules of the game” are difficult for children to grasp. For example, sometimes children break the word down into syllables rather than the smallest sound components. In such cases the child’s actual phonological awareness may be underestimated due to not understanding what is being requested of them. Despite these concerns the test produced a high degree of reliability as measured by Cronbach’s alpha, which was 0.90 (Table 14).

Figure 9: Kernel Density Curves for Section D (Phonological Awareness) by treatment arm

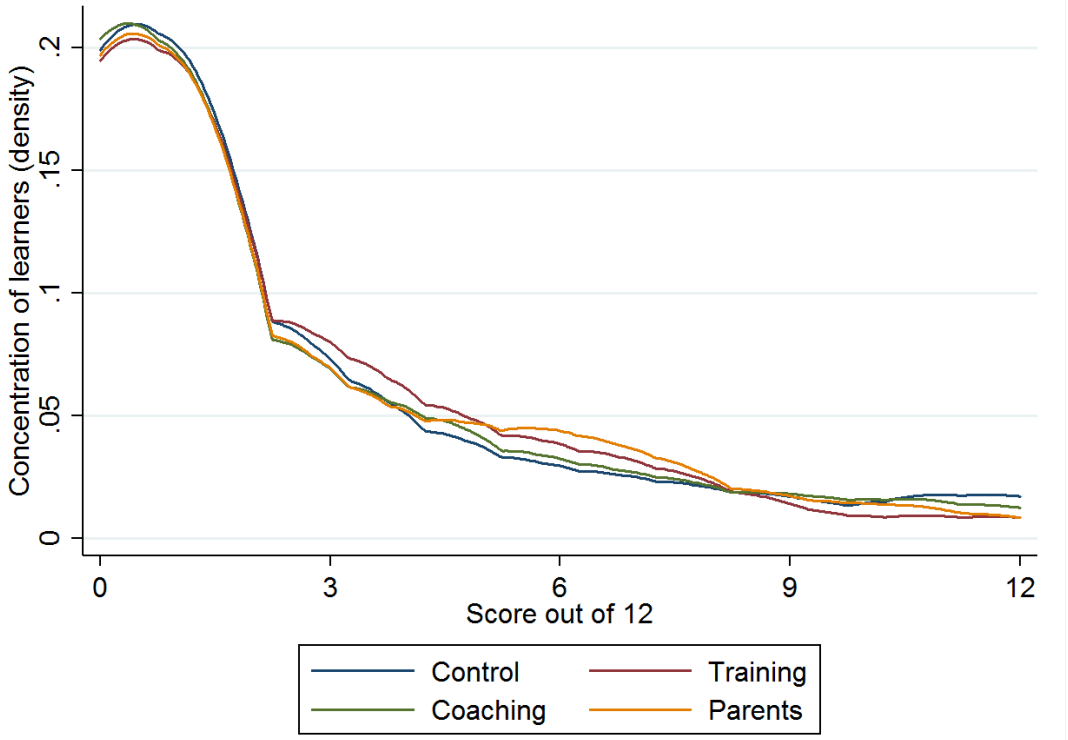
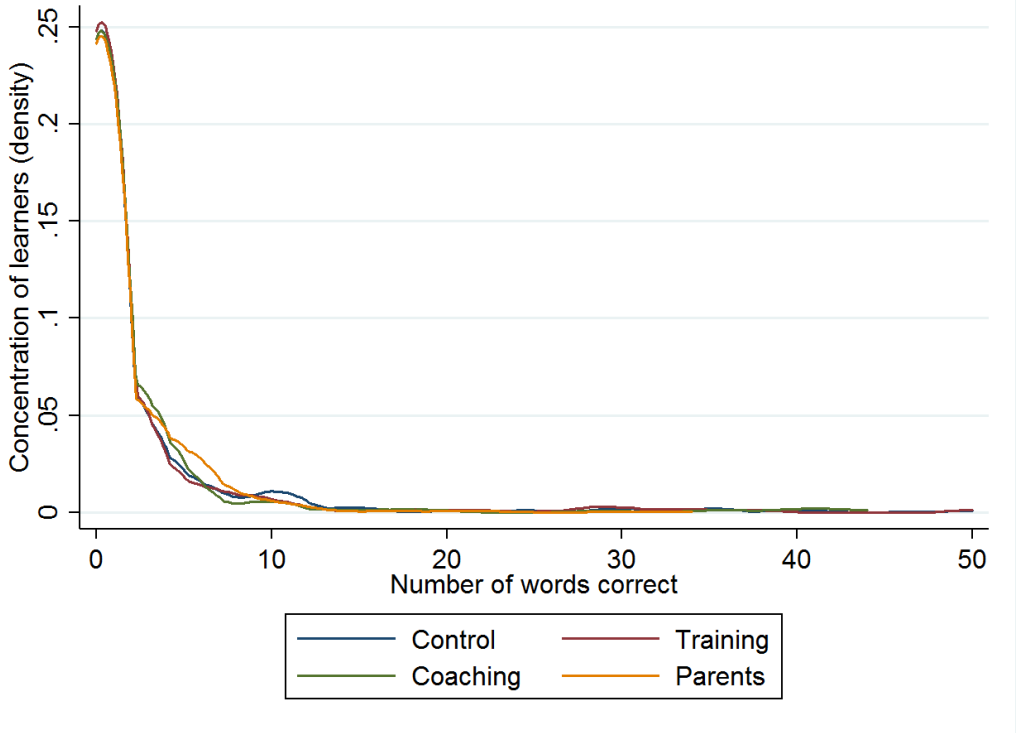


Table 14: Cronbach's alpha for Section D (Phonological Awareness)

Item	Observations	Sign	Item-test correlation	item-rest correlation	Alpha
D1_1	4163	+	0.71	0.62	0.90
D1_2	4131	+	0.79	0.74	0.89
D1_3	4112	+	0.82	0.77	0.89
D1_4	4094	+	0.78	0.73	0.89
D1_5	4090	+	0.79	0.74	0.89
D1_6	4070	+	0.74	0.69	0.89
D2_7	4144	+	0.70	0.62	0.90
D2_8	4158	+	0.65	0.56	0.90
D2_9	4119	+	0.64	0.55	0.90
D3_10	3999	+	0.67	0.60	0.90
D3_11	3990	+	0.69	0.63	0.90
D3_12	3978	+	0.67	0.61	0.90
Test					0.90

Section E was the word recognition test. As with Section B, the learner was given 60 seconds to read as many words as possible out of a maximum of 50 words. As would be expected at the start of grade 1, performance was low on this subtask and there was a substantial floor effect, as seen in Figure 10. The average score on this subtask was 1.91 correct words read. Balance was again good.

Figure 10: Kernel Density Curves for Section E (word recognition) by treatment arm



Section F consisted of three short sentences to be read by the learner. The learner was awarded a mark for every word that was correctly read. Altogether, there were 15 words across the three sentences. As Figure 11 indicates, approximately 80% of children were not able to read any of the words. A small proportion of children (about 4%) were able to read all 15 words. As before, balance was good. After reading each sentence, the learner was asked a comprehension question about that sentence. All answers were one-word answers. The average score out of 3 on the comprehension questions was 0.73 with about 73% of learners scoring zero. Interestingly, 21% of learners scored 3 out of 3 with very few learners scoring 1 or 2 out of 3. It would appear that learners can either read a sentence with comprehension or not and that including all three items did not add much value over and above the first item. For this reason, the midline test instruments will have only two sentences with the second being more complex than the first. Figure 13 indicates the positive association between word recognition and comprehension, as is expected. Similarly, a positive correlation was observed between letter recognition (section B) and comprehension and between word recognition (section E) and comprehension.

Figure 11: Number of words correct in Section F (sentence reading)

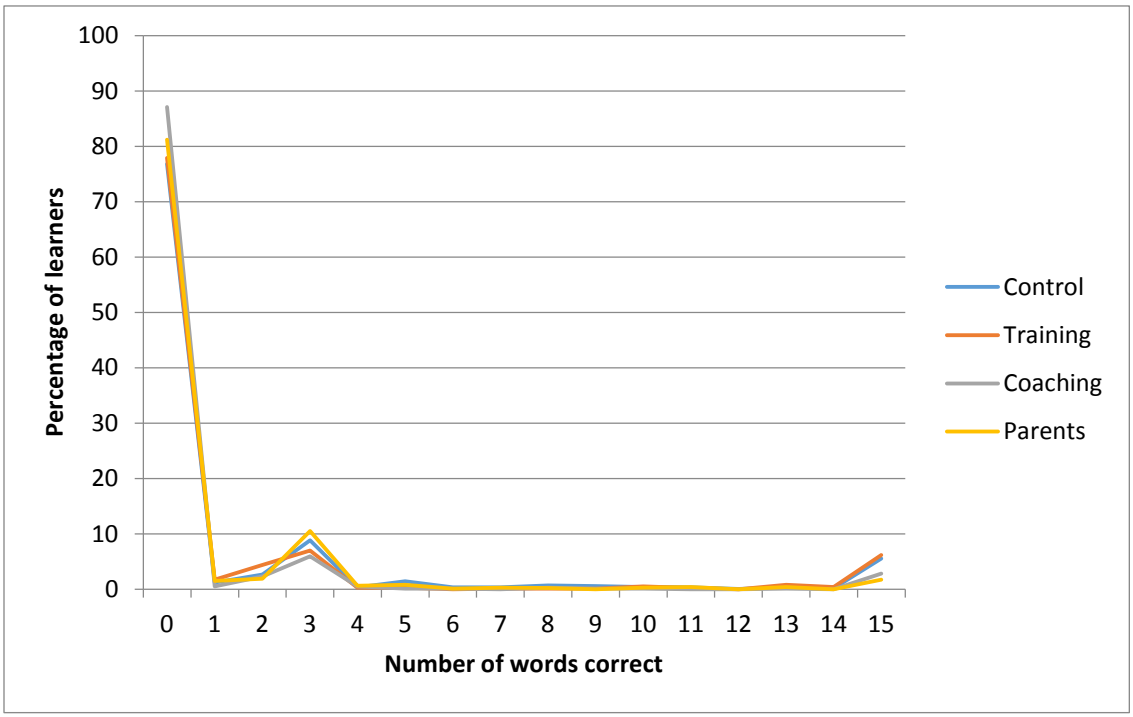


Figure 12: Percentage of learners scoring 0, 1, 2 and 3 for Section F comprehension

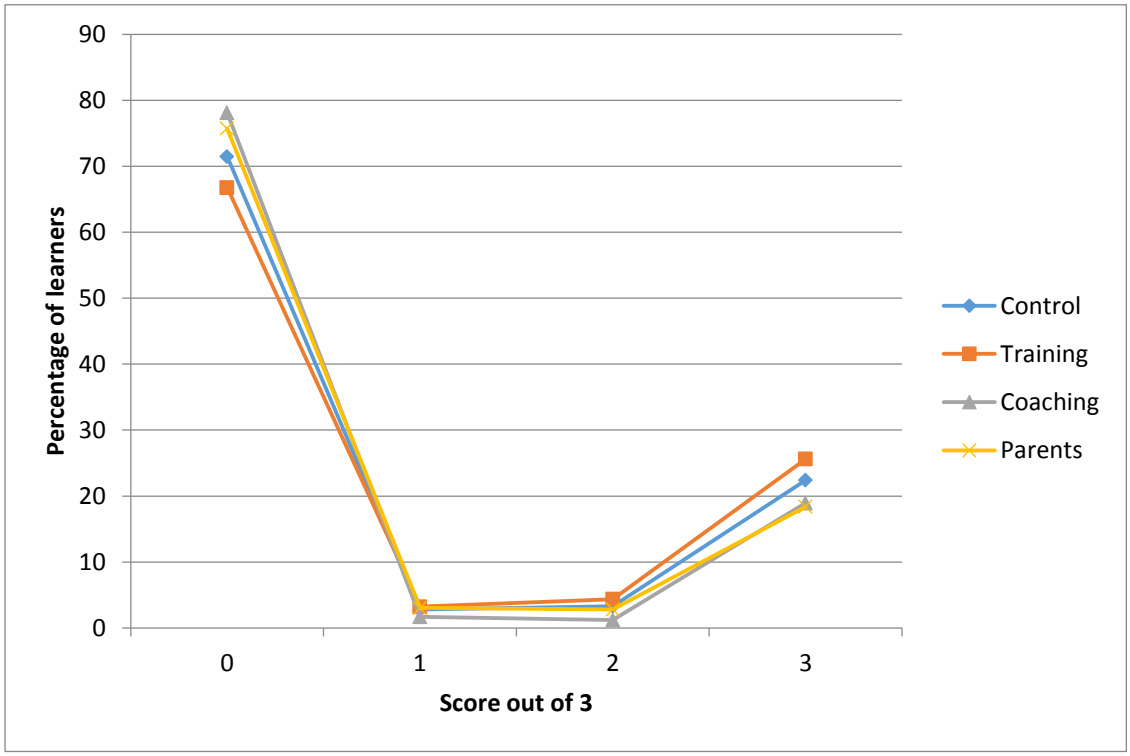
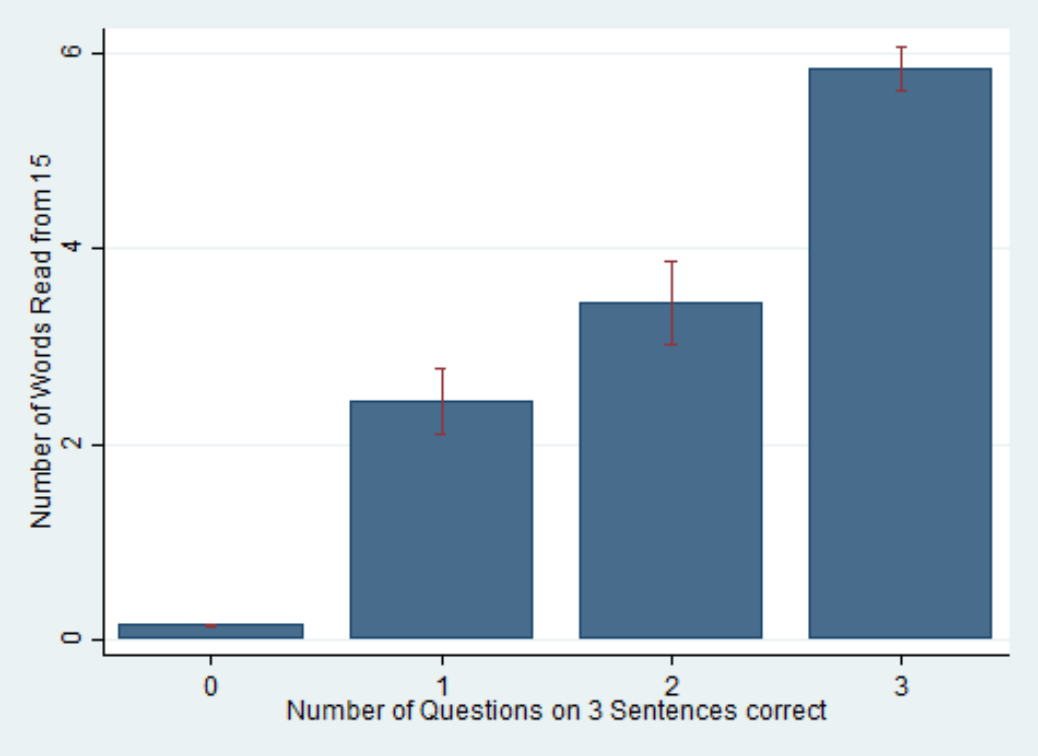


Figure 13: Relationship between words read and comprehension



In order to gain a sense of which sections of the test provide the best information we fitted a 1-parameter Item Response Theory (IRT) model treating each subtask as an individual item. Figure 14 shows item information functions for each subtask. Section A, where a ceiling effect was observed, provides some weak information about the lower part of the ability distribution and little information to distinguish amongst higher ability students. In contrast, Section E provides a lot of information at the high end of the ability distribution but little information about weaker learners. Sections C, D and F provide good information about the upper middle parts of the distribution but little information about the very bottom or very top of the distribution. Figure 15 aggregates all this into a single test information function. This confirms that the test does provide some information to distinguish between students at all parts of the ability distribution, but the information is best amongst the upper middle part of the distribution.

Figure 14: Item Information Functions from a 1-parameter IRT model

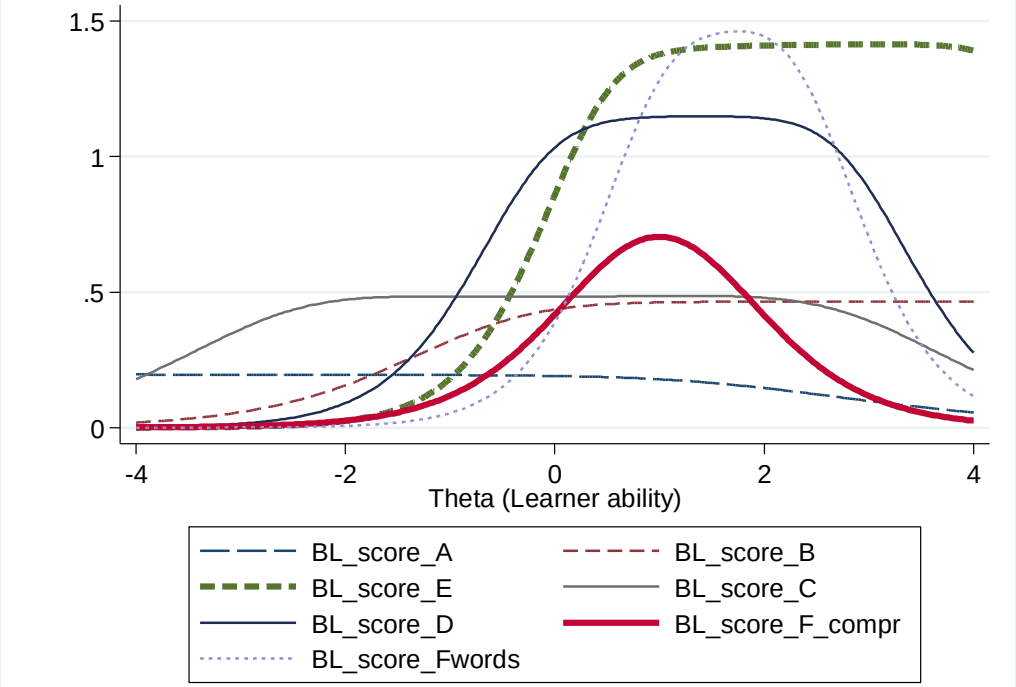
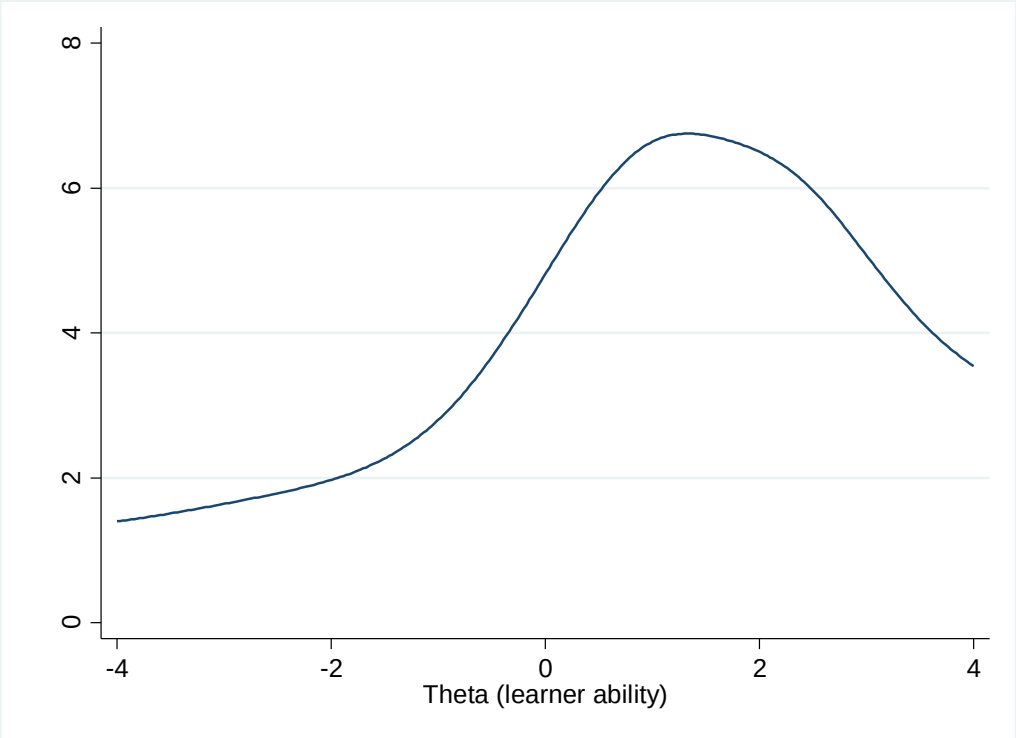
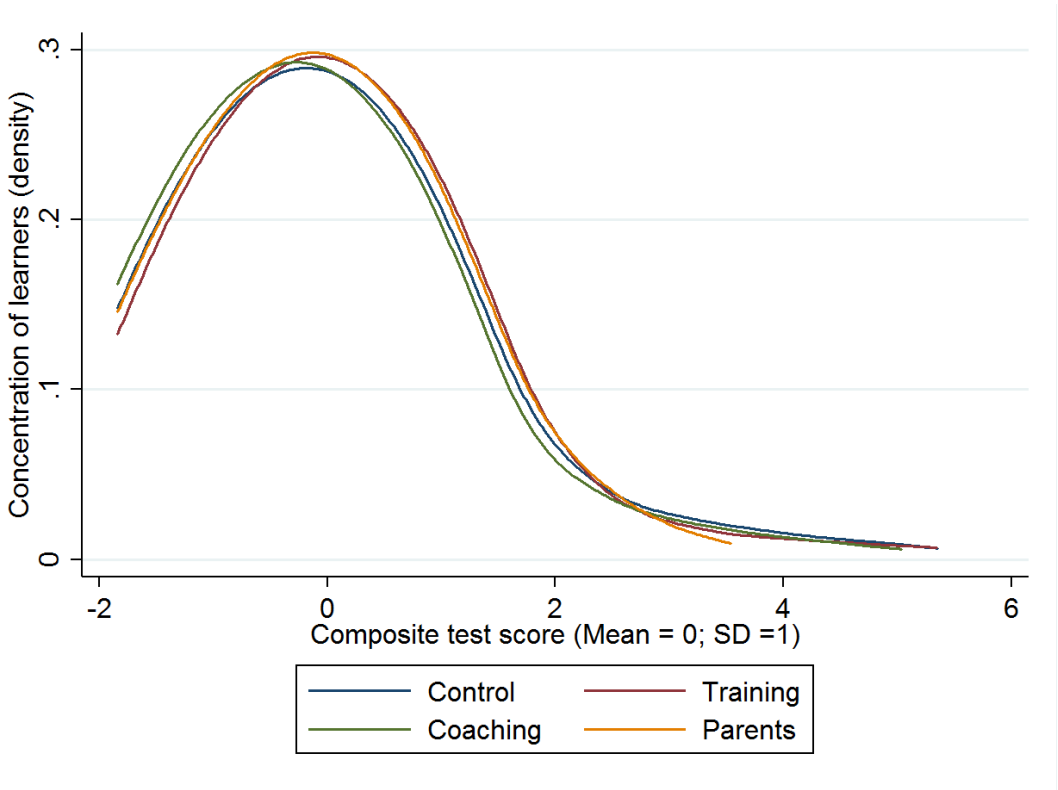


Figure 15: Test Information Function from a 1-parameter IRT model



Throughout the various subtasks there has been some evidence of "floor" and "ceiling" effects for particular subtasks. This is an important consideration, because we want to measure a change in learning outcomes across the whole distribution of pupils. Our statistical power is reduced if there is limited variation in baseline. For example, there is a "ceiling effect" in the vocabulary test where a large proportion got all the answers correct. Similarly, there is a "floor effect" on the number of letters that a pupil correctly read – the majority of the pupils did not get a single letter correct. Nonetheless, when we combine all the different learning measures into one composite score, using principal component analysis, we find a good normal distribution of learning outcomes (shown in Figure 16).[2] This is encouraging, because it means we will be able to detect a change in learning outcomes for *all* pupils across the distribution, and not only the best or worse-performers.

Figure 16: Kernel Density Curves for composite test score by treatment arm



[2] In calculating a composite score one needs to decide how much weight to attach to each subtask in the test. One cannot calculate simply add each subtask's score together, since one subtask may have had more items but should not necessarily carry more significance than another subtask. Therefore, we ran Principal Components Analysis (PCA) on the subtotals for each subtask, treating Section F comprehension as a separate score from Section 5 words correct. In PCA the variation within all variables included is analysed and those linear combinations capturing the most common variation amongst variables are identified. It is assumed that the linear combination, referred to as a principal component, which captures the most common variation amongst the variables included represents the underlying construct of interest. In this case we might think of the primary underlying construct being measured as reading ability or pre-reading ability. The weight given to each variable when calculating the total composite score is then determined by the extent of that variable's correlation with the first principal component. The intuition is that a subtask that is not well correlated with the other subtasks may be measuring something different from the intended underlying construct – this subtask should therefore carry less weight in a composite index.

The preceding analysis has suggested that learning performance is balanced across treatment regimes. Table 15 shows results based on regression analysis to test for balance – to test if the differences in average scores in learning outcomes between treatment groups are statistically significantly different from zero. Each column shows a separate regression on treatment indicators after controlling for district and strata fixed effects. The standard errors are clustered at the school level. One star indicates that the difference in means between one of the treatments and the control is statistically significant at the 10% level. The bottom three rows show the p value for the equality tests on the treatment coefficients. In other words, it shows the pair-wise tests comparing the means between treatment groups. A p value less than .05 would indicate imbalance between the respective treatment groups for the relevant learning outcome. The samples are clearly balanced. Out of the 42 possible comparisons, there is slight imbalance in only 2 cases.

Table 15: Balance tests

	(1) Vocab	(2) Letters correct	(3) Digit span	(4) Phon. awareness	(5) Words correct	(6) Compr. test	(7) Sentence	(8) Combined
Training	-0.0367	1.220	0.440	-0.162	-0.0349	0.121	-0.0164	0.0468
	(0.13)	(1.03)	(0.30)	(0.39)	(0.78)	(0.20)	(0.58)	(0.15)
Coaching	-0.203	1.211	-0.120	-0.100	-0.250	-0.156	-0.801*	-0.109
	(0.15)	(1.51)	(0.30)	(0.40)	(0.72)	(0.18)	(0.45)	(0.15)
Parents	-0.119	0.338	0.230	-0.0772	-0.655	-0.128	-0.681	-0.105
	(0.15)	(0.87)	(0.29)	(0.39)	(0.54)	(0.18)	(0.43)	(0.13)
Control Mean	8.658	4.486	4.873	2.246	2.114	0.766	1.546	0.038
Obs	4539	4452	4539	4539	4447	4539	4539	4385
Training=Coaching: p-value	0.258	0.996	0.079	0.878	0.799	0.191	0.157	0.328
Training=Parents: p-value	0.565	0.383	0.490	0.832	0.377	0.236	0.223	0.295
Coaching=Parents: p-value	0.609	0.558	0.256	0.955	0.516	0.887	0.766	0.979

Note: Each column represents a separate regression on treatment dummies and stratification dummies. Standard errors are clustered at the school level. Bottom three lines show p value of equality of coefficient tests between each treatment dummy

PUPIL CHARACTERISTICS

Table 16 plots results for basic pupil-level characteristics. The average age for pupils is 6.37 years and 47% of the sample is female. The median age is also 6, although a sizable proportion (13%) is 5 years old and roughly 9 percent are older than 7.

Table 16: Descriptive statistics – learner age and gender

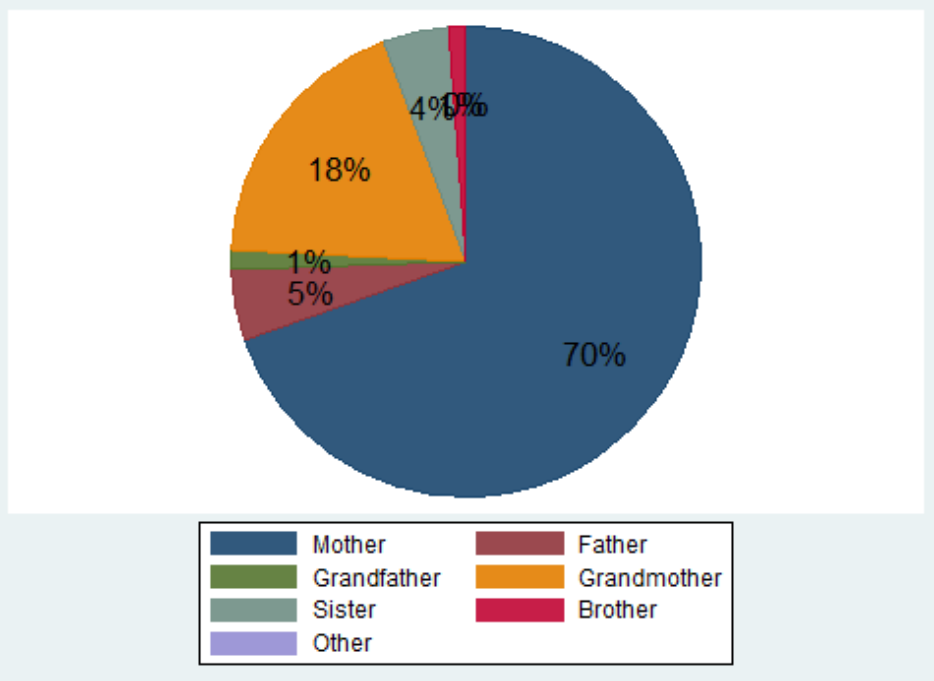
	Observations	Mean	Std. Dev.	Min.	Max.
Pupil age	3881	6.47	0.70	4	10
Girl	4198	0.47	0.50	0	1

PARENT CHARACTERISTICS

Next, we turn to parent characteristics. We sent a survey home with the pupils for the parents/guardians to fill in and return to the school. We only received the forms from 2,484 parents (out of a total of 4,539 pupils who were tested), from only 181 schools. The high response rates in some schools suggest that it is possible to require parents to complete the form. The fact that data is completely missing in 49 schools means that this was a problem of data collection and enumerator training, which should be improved on in the midline and endline rounds of data collection.

Figures 17 to 23 show the main results in a bar graphs and pie charts. First, we discuss parent characteristics. In 97% of cases the primary caregiver filled in the form (not shown). The median age is very young, roughly 25 (note that many gave an answer of 6 or 7 for each. They clearly answered the child's age and we excluded that from the age sample). It is mostly the mother who fills in the form, but note that in a sizable portion (19%) of cases it is the grandmother or grandfather that fills in the form. In only 5% of cases did the child's father complete the form. Even when we restrict the sample to those who claim to be the primary care-givers, roughly 20% are grandparents and over 5% are siblings. However, the mean age of the siblings is 22 years (median is 26), so these are mostly adults.

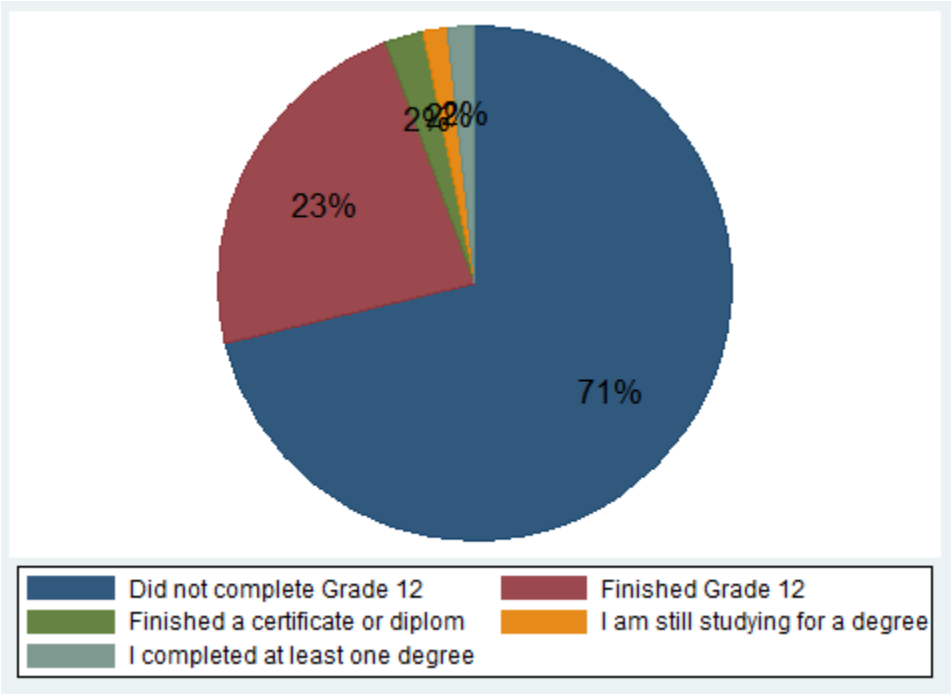
Figure 17: Relationship to pupil



We can see from Figure 18 that most parents have only low levels of education – 71% did not finish matric. Less than 10% completed have a post-matric degree. This size is slightly larger for

grandparents (85% did not finish matric). Unsurprising, given the levels of education, the caregivers rarely read for their own pleasure (not shown). Over a third read less than an hour per week; 41% read 1 to 2 hours.

Figure 18: Education of Guardian



Next we discuss reading activities at home and parents' beliefs and aspirations. Only 10% acknowledge that they never read to their child, yet 27% don't have any books at home and over a quarter read less than one hour a week for their own pleasure. The majority of parents claim that they check if the child is doing his/her homework daily. A third of parents had not spoken to teachers (but since fieldwork occurred in February we cannot place too much weight on this). Only 57% know when the most recent School Governing Body (SGB) meeting was held. This all suggests that there was no parent-teacher meeting at the beginning of the year in many schools or that it was not well communicated to parents.

Figure 19: How often do you read to your child?

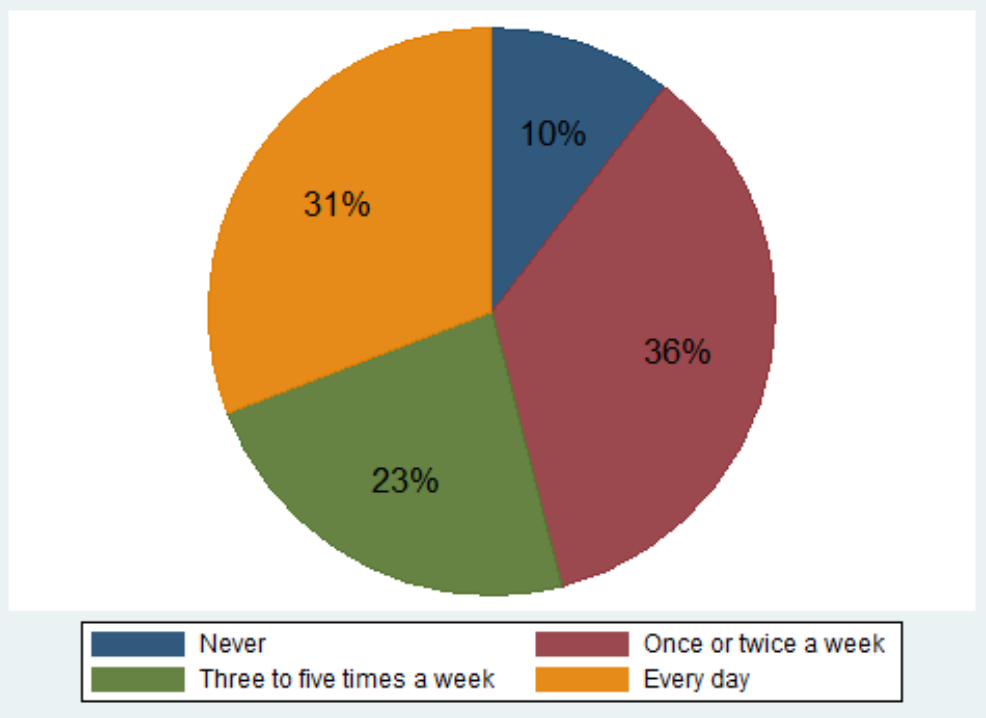


Figure 20: Number of books at home

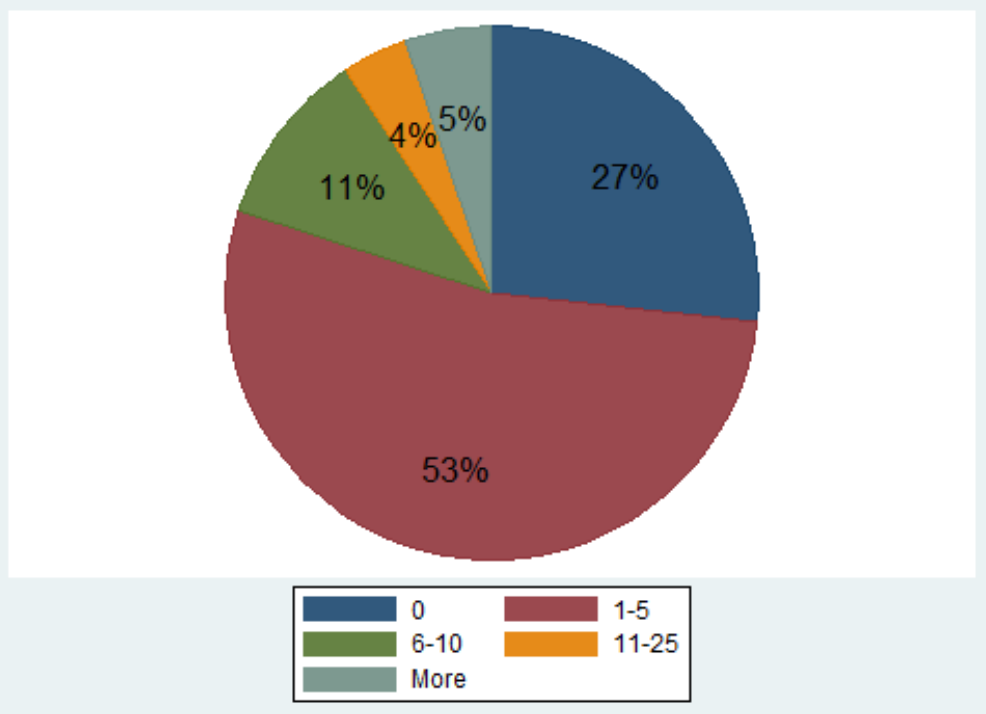


Figure 21: Number of hours read for own pleasure

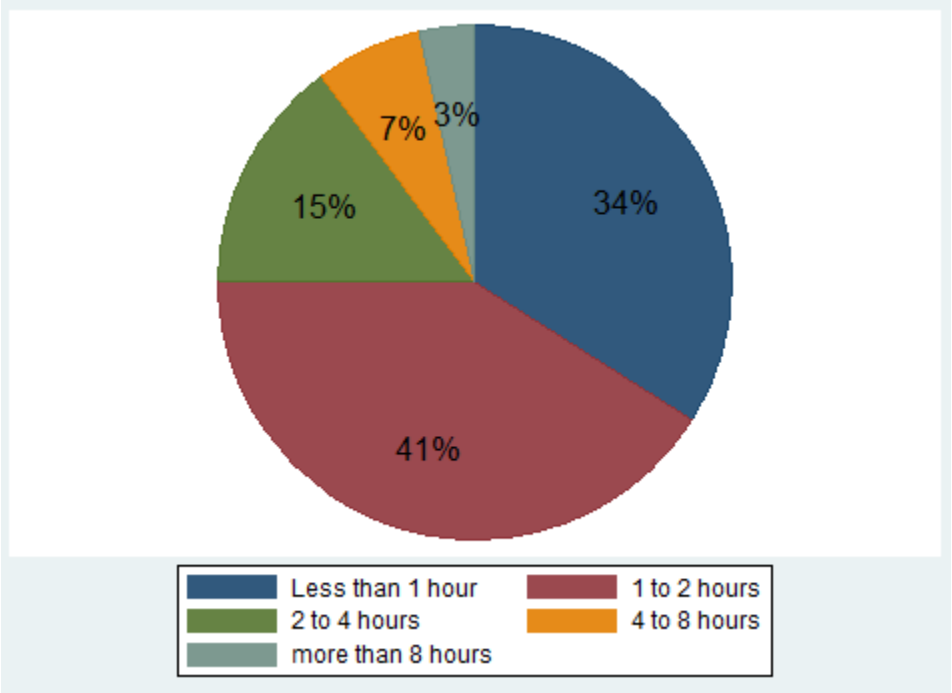
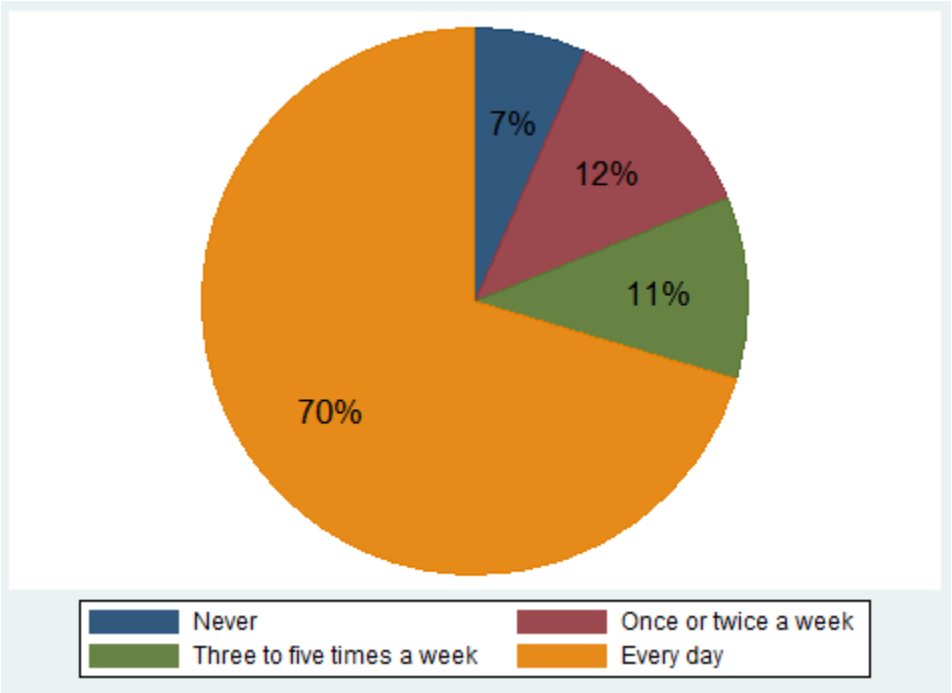


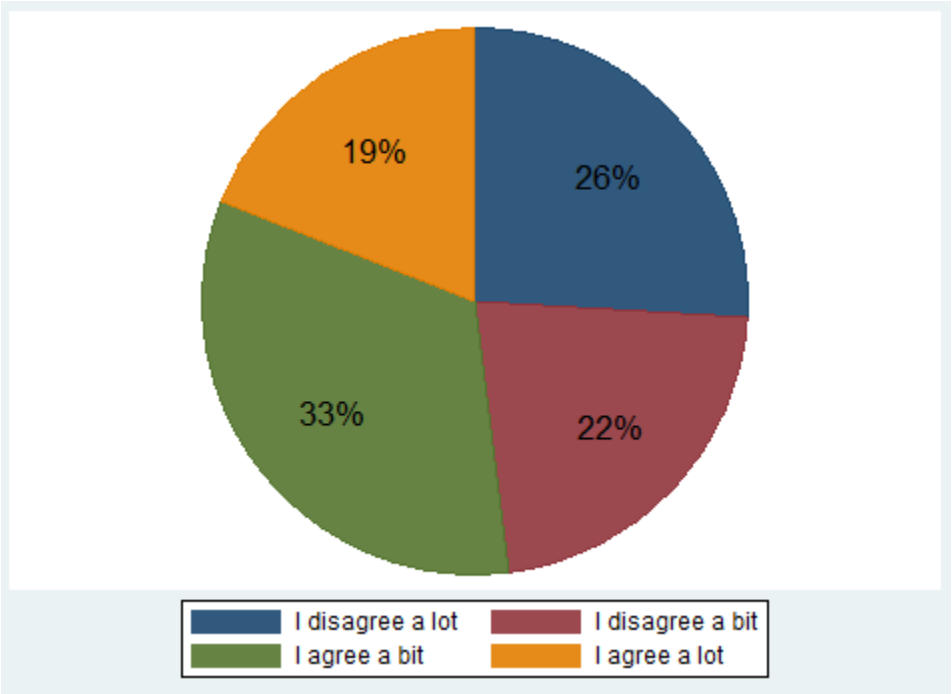
Figure 22: Check that child is doing homework



Just more than half of parents believe that learners at their child's school read poorly. This shows that many parents are critical of the quality of education on offer but a substantial

proportion are probably underestimating the extent of the problem of low learning in schools. It will be interesting to monitor parent beliefs about this at midline and endline assessments, especially in Treatment 3 schools, which are receiving the parent involvement intervention.

Figure 23: Agreement with the following statement: "Our school's learners read very poorly."

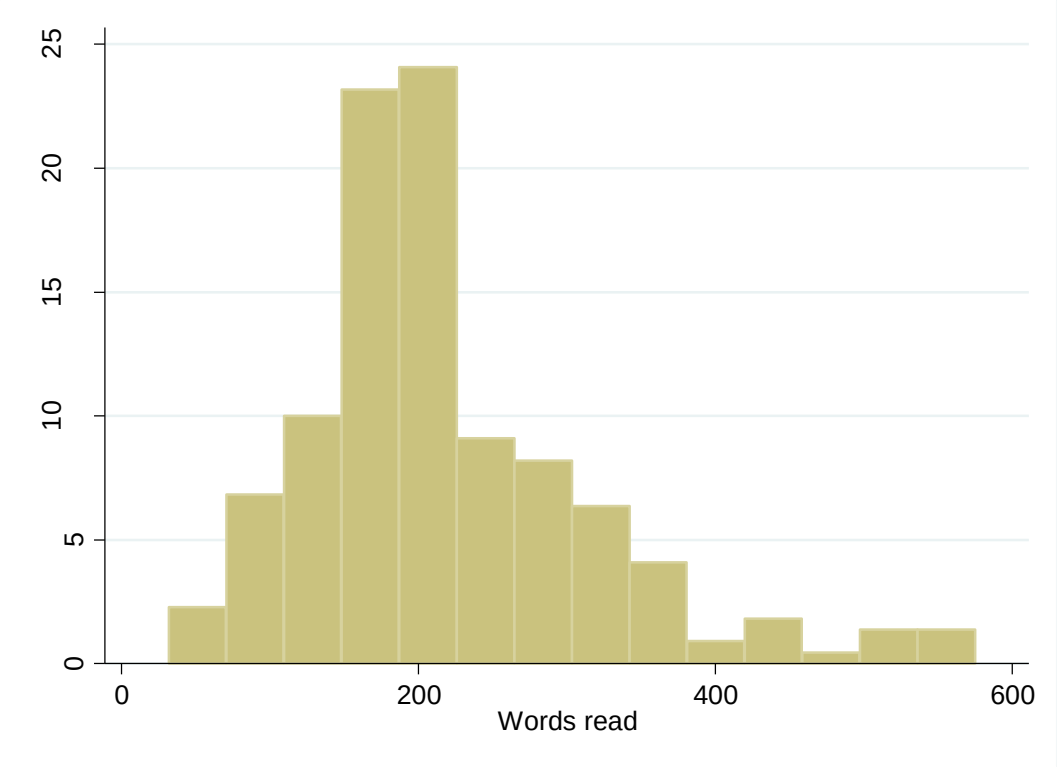


TEACHERS

TEST PERFORMANCE

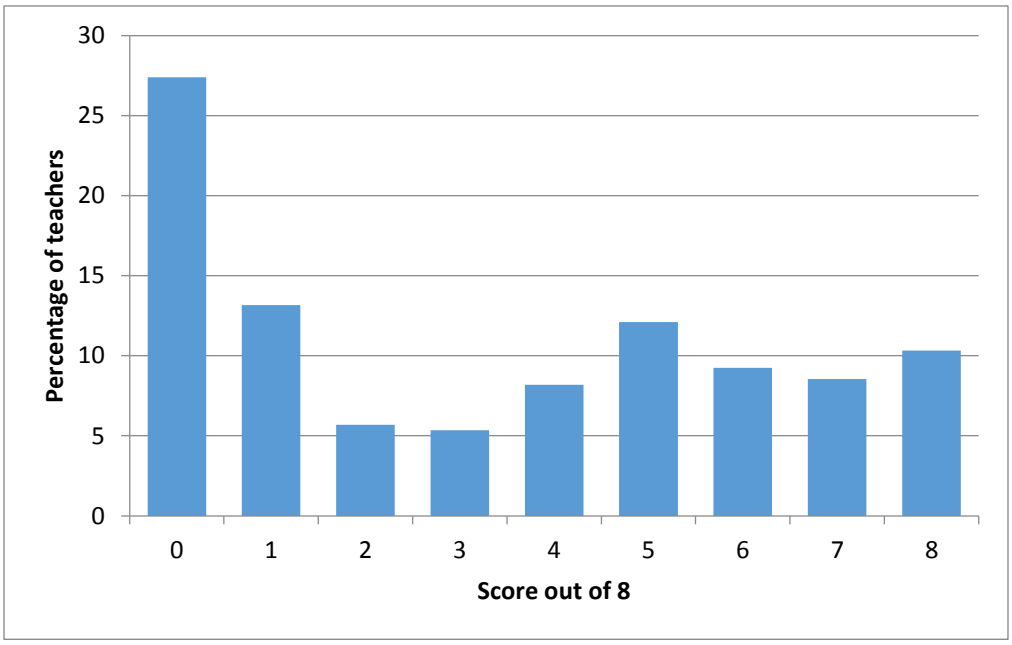
We asked teachers to fill in a questionnaire by themselves whilst the fieldworkers tested the pupils. The teachers were also asked to participate in a short Setswana reading fluency and comprehension assessment. As discussed earlier, there is quite a bit of missing data on the teacher assessment due to non-return of forms in some schools and refusal to participate by 40 teachers. We are left with teacher test data corresponding to about 70% of learners. In the first component of the teacher assessment teachers were given 60 seconds to silently read through a Setswana text consisting of 575 words. The teacher was then asked to indicate how far he/she had gotten. Figure 24 indicates that there is a relatively normal distribution in the number of words read in 60 seconds. The majority of teachers claimed to read a between 150 and 250 words. These estimates of fluency are probably slightly generous but at least they are not wildly unrealistic.

Figure 24: Words read by teachers in 60 seconds



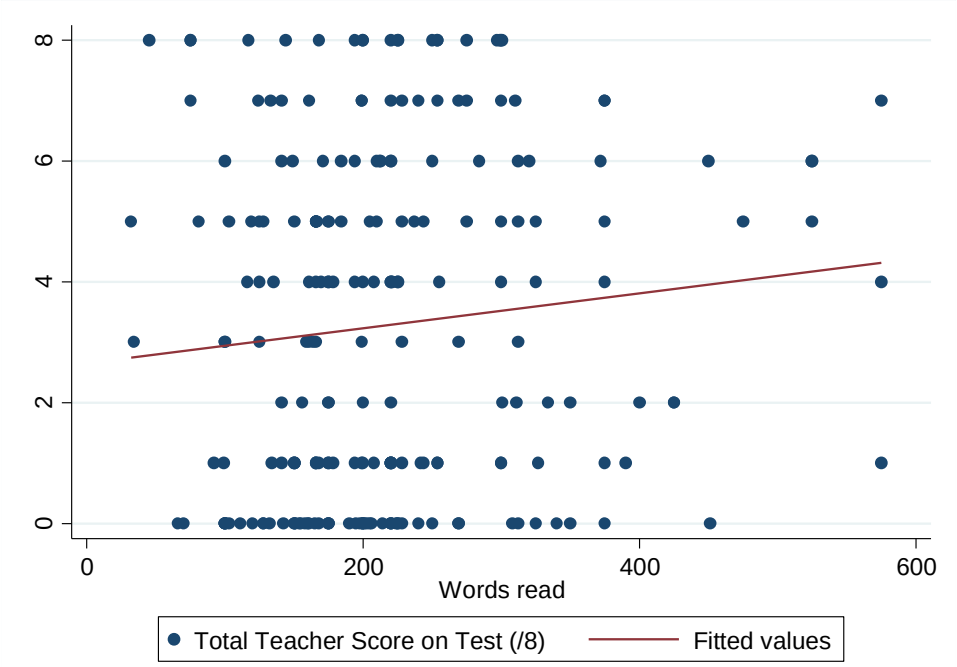
The second section of the teacher assessment was a comprehension test. After spending 60 seconds reading through the text, the teacher was then handed a set of eight multiple choice comprehension questions based on the text. The teacher was then given another three minutes to complete the comprehension questions. The time limit was imposed so as to test the fluency with which teachers are able to read through the text and retrieve answers. The results are disappointing. About 27% of teachers did not get a single question right (Figure 25). It is also possible that these results are biased upwards, if the more competent teachers were more likely to agree to take the test.

Figure 25: Scores on the teacher comprehension test



Furthermore, we can see from Figure 26 that there is a positive correlation between the number of words read and performance in the comprehension test, but this relationship is not nearly as strong as in the case of the pupil test. The weaker validity of the results suggested that administration of the teacher tests may not have been very consistent.

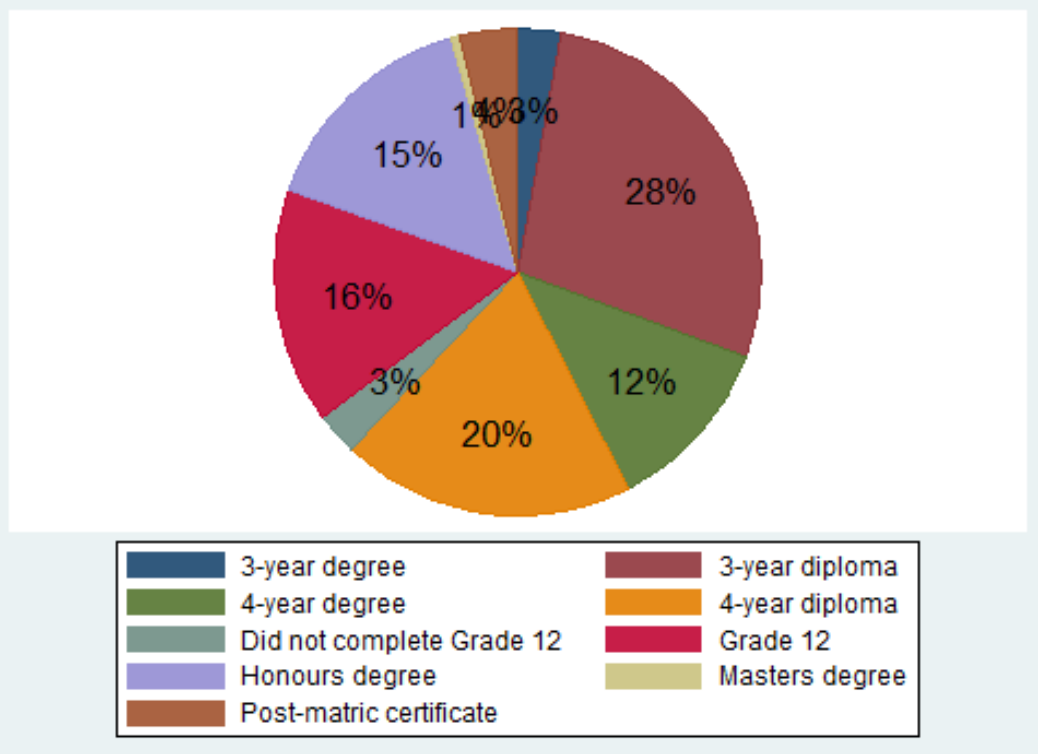
Figure 26: Relationship between words read and questions answered correctly



TEACHER CHARACTERISTICS

Next, we present basic teacher characteristics, shown in Figures 27 to 29 and Table 17. Figure 27 shows the distribution of teachers' education level (note that 45% of teachers did not answer this question. So it is hard to have any confidence in this question). Most have at least a 3-year diploma. 15% only have matric qualifications; 3% have not completed matric.

Figure 27: Teacher qualifications



Figures 28 and 29 reveal an interesting discrepancy in teachers' beliefs. The majority of teachers believe that children should be able to read Setswana fluently (a passage of 50 words in a minute with comprehension) by the end of grade 3 or earlier (84%), yet a slim group actually believe that all children in their school could read by the end of grade 3. Almost half of teachers estimated that only 50-75% of children in their school are able to read by the end of grade 3; roughly a third expect less than half will be able to read! This result is exactly the same if you restrict the sample to teacher who stated they expect that pupils *should* be able to read by grade 3. So, in general teachers don't believe that pupils will reach their expectations.

Figure 28: At what grade should pupils be able to read fluently?

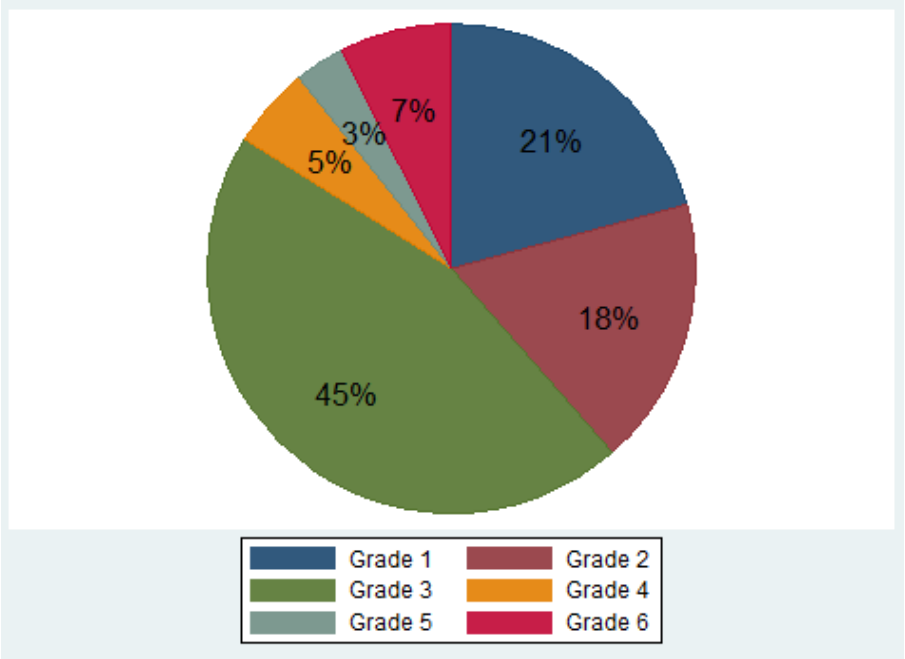


Figure 29: What proportion of children will be able to read by grade 3?

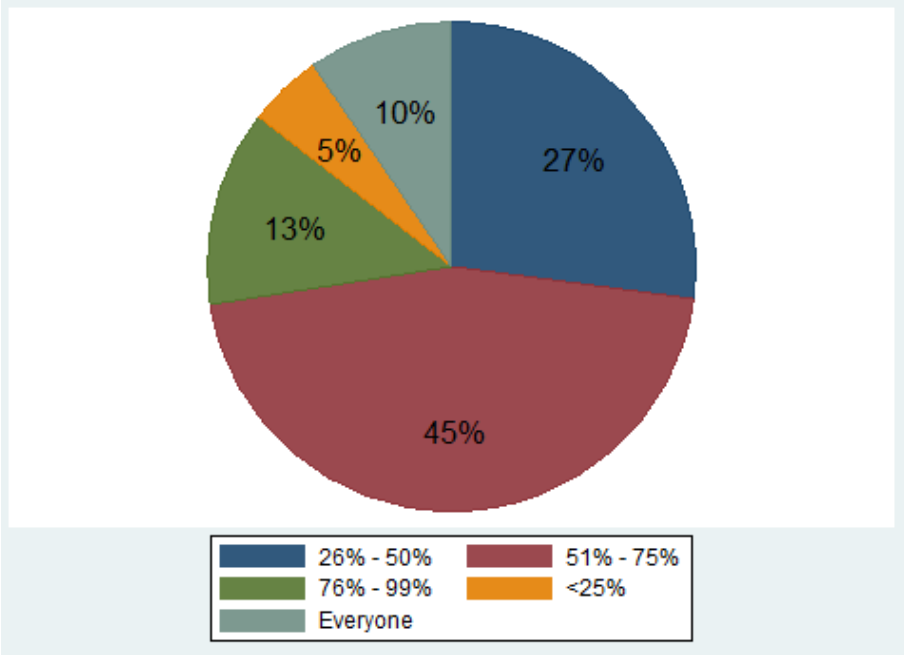


Table 17 shows basic teacher characteristics. Teachers are almost always female, and are rather old with an average age of 50. About 26% of teachers are 56 years old or older. Less than 10% of teachers are younger than 40. 92% speak Setswana most often at home. The average days missed (absent) over the last 10 school days is 1.3; only 49% of teachers claimed not to have been absent at all over the preceding 10 school days. The average class size

("pupils enrolled") is 41 learners. In the majority of classrooms, all the pupils have workbooks and CAPS books. In 60% of classrooms, all the pupils have graded readers (but note the number of missing values for this question). It is unfortunate that on many items there is a lot of missing data due to both non-return of instruments and item non-response.

Table 17: Selected descriptive statistics – teacher characteristics

	Observations	Mean	Std. Dev.	Min.	Max.
Time spent per week teaching reading in Setswana (minutes)	262	171.53	114.50	2	480
days missed last 10 days	203	1.30	2.25	0	10
Age	310	49.78	7.87	24	63
Class size	315	40.62	10.03	3	80
Female teacher	324	0.99	0.10	0	1
Speak Setswana most often at home	311	0.92	0.27	0	1
All learners have CAPS workbooks	241	0.71	0.46	0	1
All learners have workbooks	201	0.78	0.41	0	1
All learners have graded readers	178	0.60	0.49	0	1
At least 3-year diploma	179	0.78	0.42	0	1
At least 3-year degree	179	0.30	0.46	0	1

SCHOOL AND SCHOOL PRINCIPAL

Next, we discuss results from the school principal survey, as shown in Table 18 and Figures 30 to 33.

The majority of school principals have an honours degree, with an average age of 51. The schools are mostly remote rural, with only 1% in a formal suburban area. The schools also come from areas with low levels of socio-economic status: over half of school principals estimated that less than 20% of households have both parents employed. More than half of school principals estimate that the majority of parents have not completed secondary school. The pupil-teacher-ratio in the foundation phase is 38. Almost all the schools (93%) have Setswana as the first language and have a formal language policy.

Table 18: Selected descriptive statistics – school principal questionnaire

	Observations	Mean	Std. Dev.	Min.	Max.
2(a) Age of principal at last birthday:	206	50.85	6.45	28	65
2(c)(vi): Number of filled FP teacher posts	205	5.64	3.14	1	18
Number of foundation phase pupils	230	186.85	134.02	0	545
PTR in Foundation Phase	205	37.84	18.30	0	136
Setswana offered as LoLT	230	0.93	0.26	0	1
School has Formal Language Policy	208	0.91	0.29	0	1
4(a): Learners receive daily meal	230	0.90	0.29	0	1

Figure 30: School principal highest level of education

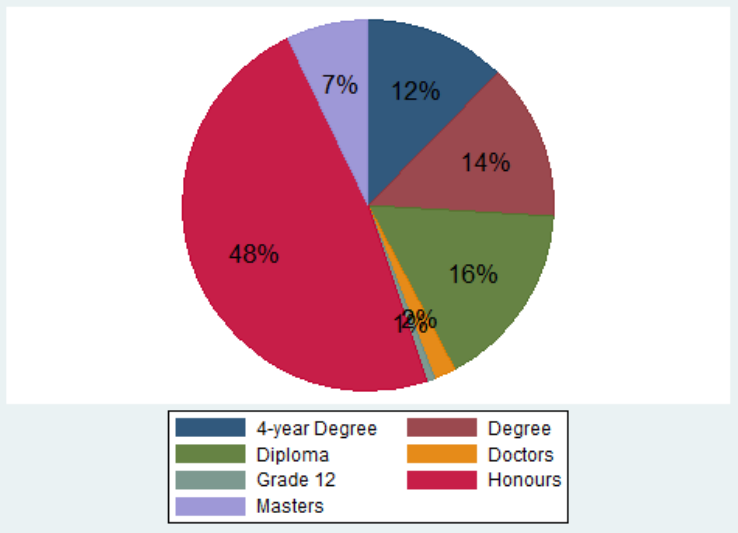


Figure 31: School principal's estimate of average level of parent education

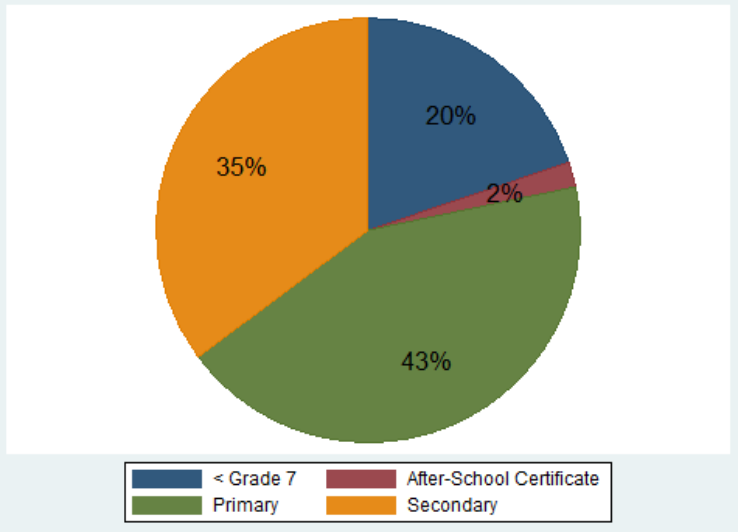


Figure 32: Proportion of learners with both parents employed according to school principal

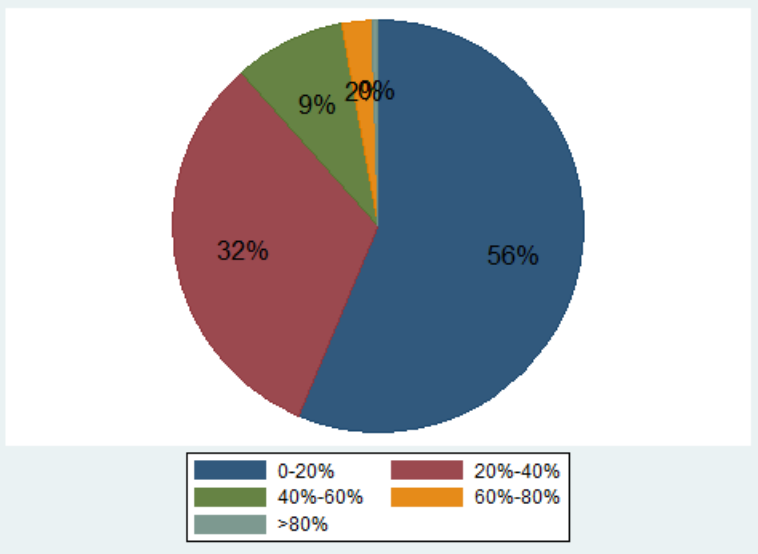
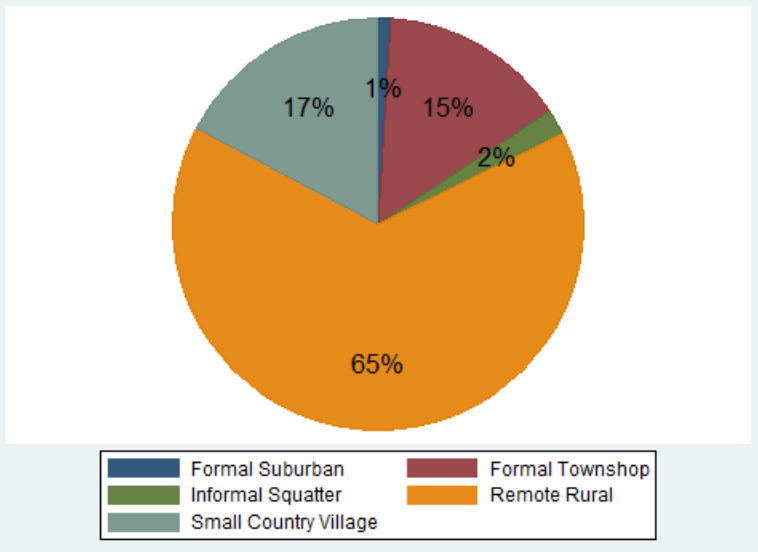


Figure 33: Location of the school according to school principal



WHAT PREDICTS PUPIL TEST SCORES?

Next we examine some pupil, teacher and school characteristics that predict pupil learning outcomes. Note that we can make no causal claims with these regressions – this is merely descriptive analysis, which is informative of possible trends and again tests the validity of our test instruments.

PUPIL CHARACTERISTICS

Table 19 shows results from simple OLS regressions predicting overall composite test score (in terms of standard deviations) (columns 1-3) and letters correct (columns 4-6). The regressions indicate that age does not significantly predict performance. Girls, however, performed significantly better than boys. Girls were able to read about 1 letter more than boys on average, and performed just less than 0.09 SD higher than boys on the composite score. This advantage for girls is consistent and of a similar magnitude with what is observed in standardized tests for higher grades in South Africa, such as in the Annual National Assessments (DBE, 2014), in grade 4 (Howie *et al,* 2012; Fleisch *et al*, 2015) and in grade 6 (Spaull and Taylor, 2015). It is interesting that this gap is evident right at the start of grade 1, which would suggest that the disadvantage may be due to some factor other than school practices that favour girls, most likely differences in the physiological development of girls and boys at this age. Table 19 also indicates that learner performance was significantly better in the district of Ngaka Modiri Molema than in the district of Dr Kenneth Kaunda. This was not anticipated since neither district has shown consistently higher performance in the Annual National Assessments since 2012. The difference between districts will continue to be monitored in the midline and endline assessments.

Table 19: Performance by district, learner age and gender (OLS Regressions)

	(1)	(2)	(3)	(4)	(5)	(6)
	Ave. score	Ave. score	Ave. score	Letters correct	Letters correct	Letters correct
Pupil age	-0.0327		-0.00387	1.078***		1.135***
	(-0.84)		(-0.10)	(3.37)		(3.36)
Female		0.0877***	0.110***		0.992***	1.021***
		(2.77)	(3.44)		(3.11)	(3.71)
District Dummy			-0.456***			-0.784
			(-5.19)			(-1.08)
Observations	3742	4055	3447	3800	4115	3498
Adjusted R^2	0.000	0.002	0.042	0.007	0.002	0.011

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

SCHOOL AND TEACHER CHARACTERISTICS

Table 20 reports regression of school characteristics on the composite reading proficiency score (columns 1-3) and letters correct (columns 4-6). Note first that the school principal's level of education does not matter. This is not too surprising, since these children just joined grade one. Yet, we see that the official poverty quintile of the school (which reflects community level poverty) matters: pupils from schools that are classified as falling in the lowest quintile in terms of socio-economic status perform worse than those in quintile 2. The difference in outcomes

between quintile 1 and quintile 3 is also large in magnitude, although not statistically significant at conventional levels. These trend holds, after controlling for district and location. Unsurprisingly, rural schools also perform worse.

Table 20: Performance by school principal, location, and socio-economic background

	(1) Ave. score	(2) Ave. score	(3) Ave. score	(4) Letters correct	(5) Letters correct	(6) Letters correct
Principal - higher degree	0.0437		-0.0467	-0.508		-1.079
	(0.39)		(-0.38)	(-0.70)		(-1.36)
Quintile 2		0.475***	0.293*		3.442**	1.746**
		(3.06)	(1.97)		(2.52)	(2.08)
Quintile 3		0.130	0.0780		1.126	1.511
		(1.19)	(0.57)		(1.56)	(1.50)
Rural			-0.298*			-1.189
			(-1.95)			(-1.54)
District Dummy			-0.470***			-0.441
			(-4.84)			(-0.56)
Constant	-0.0681	-0.166***	0.207	4.906***	3.840***	5.306***
	(-0.81)	(-3.24)	(1.16)	(8.23)	(11.43)	(5.79)
Observations	2975	4385	2762	3018	4452	2805
Adjusted R^2	0.000	0.040	0.086	0.001	0.021	0.015

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 21 reports regression outputs of teacher characteristics on pupil performance. Surprisingly, teacher performance in the knowledge test is positively correlated with pupil performance, even after controlling for community characteristics: school quintile, location, district, as well as the randomisation strata. It is unlikely that this reflects a causal relationship between teacher quality and learner performance since learners have just joined the school. It is possible that this reflects a selection effect where both learners and teachers select themselves into better schools. Teacher age and education do not predict performance.

Table 21: Performance by teacher characteristics

	(1) Ave. score	(2) Ave. score	(3) Ave. score	(4) Ave. score	(5) Ave. score
Average Teacher Score	0.301***	0.238**	0.355***	0.355***	0.309**
	(3.45)	(2.07)	(3.82)	(3.82)	(2.37)
Age		0.00514			0.00407
		(1.11)			(0.75)
At least 3-year diploma		-0.0871			-0.0814
		(-0.88)			(-0.75)
Rural			-0.0104	-0.0104	-0.106
			(-0.12)	(-0.12)	(-0.76)
Quintile 2			0.0653	0.0653	0.0427
			(0.74)	(0.74)	(0.35)
Quintile 3			0.0778	0.0778	-0.0139
			(0.76)	(0.76)	(-0.10)
District Dummy					-0.182
					(-1.27)
Constant	-0.287***	-0.459*	-0.334***	-0.334***	-0.308
	(-7.56)	(-1.83)	(-3.72)	(-3.72)	(-0.90)
Observations	3077	1395	2787	2787	1259
Adjusted R^2	0.023	0.020	0.034	0.034	0.045

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

PARENT CHARACTERISTICS

Table 22 reports regression results of guardian characteristics on pupil test scores. As before, each column represents a different regression and standard errors are clustered at the school level. Figures 34 to 37 show the main results graphically.

These results show clearly that the home environment matters greatly. Figure 34 shows that pupils do worse if their guardian has not completed matric (there is strangely also a negative result for degree, but a very small sample gave this response). In figure 35 we can see that pupils did far worse in homes where the guardian reportedly never reads to their child; similarly for homes where the guardian reports to never check homework. Figure 36 shows that pupils do better in homes with many books. Even more interestingly, from Figure 37 we see that for

parents that believe they are responsible (and not government or teachers) their child performs far better.

There is some evidence that a child does worse if his guardian is his sister (not shown), possible because these are orphaned households. But the small sample means we shouldn't place too much value on this result. There is also suggestive evidence that for pupils whose guardians believe their child can improve in learning performed better, but this result is not strong.

Table 22: Parent characteristics and learner performance

	(1) Ave. score	(2) Ave. score	(3) Ave. score	(4) Ave. score	(5) Ave. score
Finished matric	0.248***				0.224***
	(3.66)				(3.27)
Degree	0.277				0.314
	(1.39)				(1.31)
Reads to Child		0.229***			
		(2.82)			
Homework every day		0.0836*			0.0574
		(1.67)			(1.10)
Believe - responsible for learning			0.265***		0.0805
			(3.46)		(1.32)
Books				0.0124***	0.00830*
				(2.91)	(1.85)
District Dummy					-0.317***
					(-3.40)
Quintile 1					-0.327***
					(-2.78)
Constant	-0.0554	-0.244***	-0.244***	-0.0550	0.00833
	(-0.94)	(-3.18)	(-4.90)	(-0.89)	(0.08)
Observations	2080	2043	2212	2090	1728
Adjusted R^2	0.011	0.006	0.004	0.003	0.061

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < .01$

Figure 34: Parent education and learner test scores

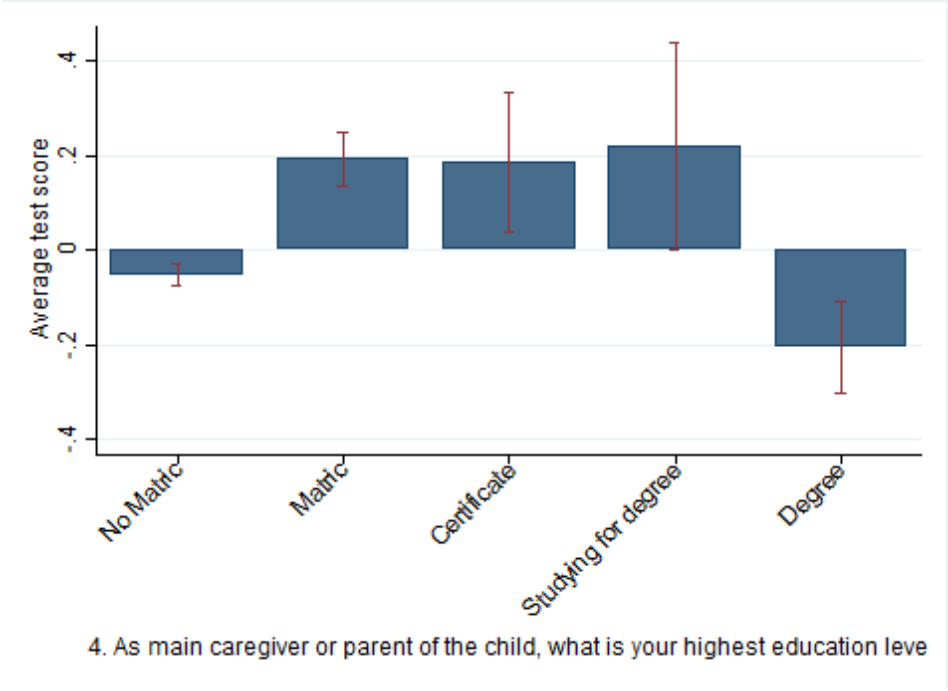


Figure 35: Parent reading to child and learner test scores

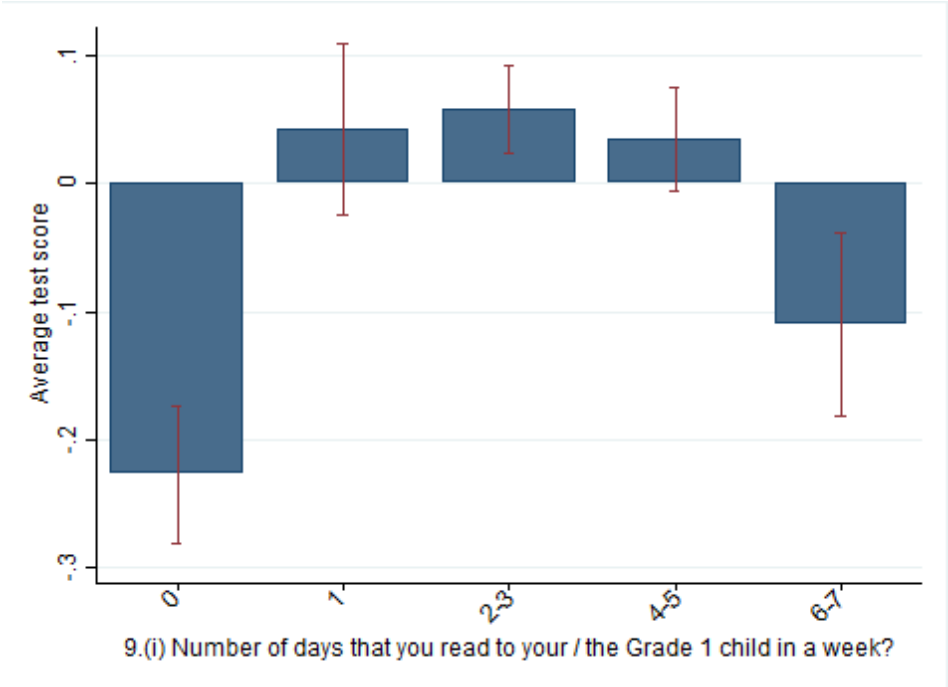


Figure 36: Parent checking homework and learner test scores

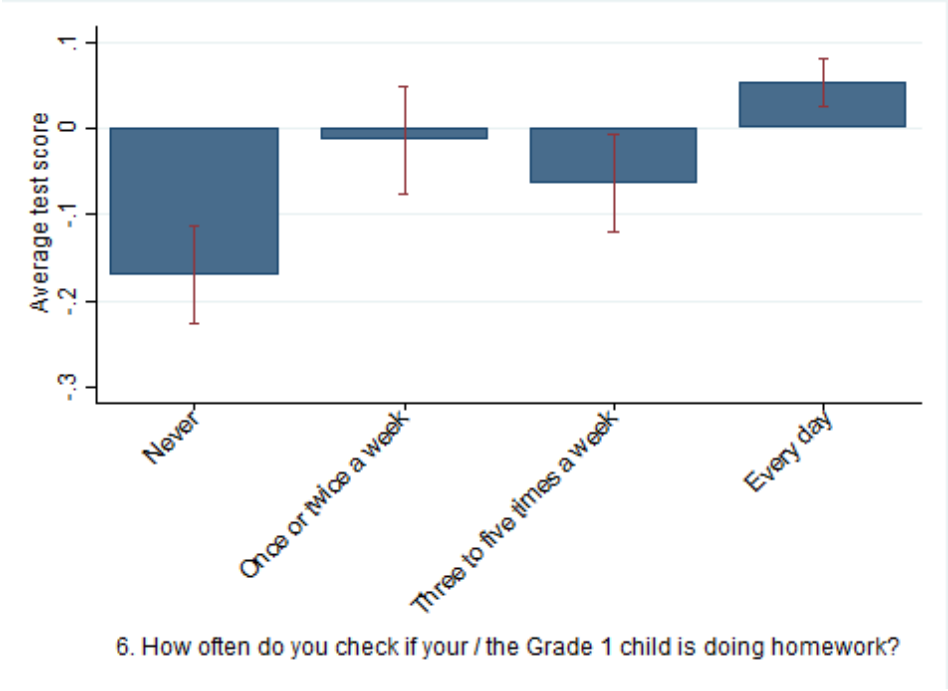
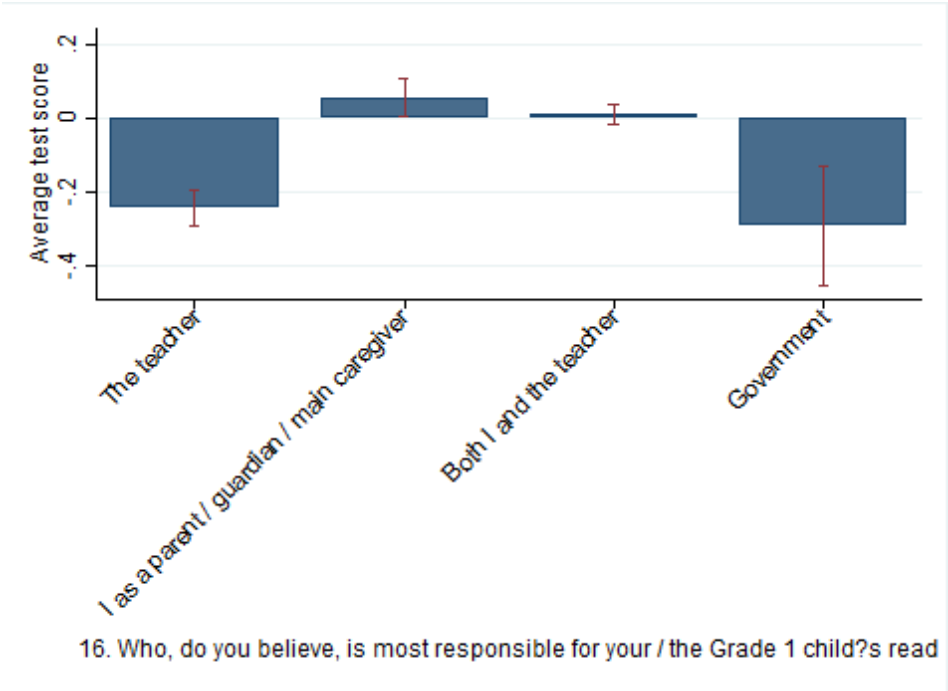


Figure 37: Parent's perceived responsibility for learning and learner test scores



Reading for own pleasure doesn't appear to matter. The following factors also do not seem to be correlated with test scores: reported frequency of meeting with teacher; attitudes regarding the importance of Setswana, beliefs about their child's learning ability, beliefs about the quality of the school.

To summarize: parents' education and involvement in their child matter. Involvement in the school or their own beliefs over learning and their child's ability doesn't matter. A sense of agency – a belief that they are important to their child's learning – does matter. All told, the self-reported home-level characteristics – guardian's education level, beliefs, and involvement in their child's reading - are strongly correlated with pupil test scores. A guardian's involvement in the school (attending SGB meetings etc.) doesn't seem that strongly correlated with reading scores.

PROGRESS REPORT ON IMPLEMENTATION OF INTERVENTIONS

EGRS Treatment 1 (training)

Treatment 1 trains the teachers on how to use the lesson plans and accompanying materials through central training sessions, each lasting 2 days, and occurring twice yearly. The first session was conducted in February 2015 and the second occurred in July 2015. Similar sessions are scheduled for 2016.

SUMMARY OF KEY EVENTS

Reference Group Meetings

Two reference group meetings were held with Provincial Officials from the Quality Assurance and Research Directorates, and with Foundation Phase Language specialists working in the Dr Kenneth Kaunda District and the Ngaka Modiri Molema District education offices. The purpose of these reference group events was primarily to ensure that the lesson plans and support materials are aligned to existing curriculum support offered in the province but also to introduce the group to the EGRS programme and approach, to garner buy-in for the programme; to critically engage with the materials used in order to strengthen them and to reflect on successes and challenges of the programme and implementation. These events are relevant to both Treatments 1 and 2 since both these treatments make use of the same lesson plans and support materials.

Training Events

- Two training events were held to train school managers and Grade 1 teachers, and to distribute materials.

- The first event was held at the Protea Hotel in Klerksdorp. Approximately half the schools attended the first session, on the 24th and 25th February 2015, and half the schools attended the second session, on the 26th and 27th February 2015.
- The second event was held at the Kedar Country Lodge in Rustenburg. Schools from Kgetleng, Maquassi Hills, Matlosana and Ramotsere Moiloa attended the first session, on the 14th and 15th July 2015. Schools from Ditsobotla, Mafeking and Rekopantswe attended the second sessions, on the 16th and 17th July 2015.
- Accurate, up-to-date data on schools, managers and teachers was gathered at the first training event, allowing for much more efficient logistics at the second event.
- At the first training, teachers were given an overview and technical understanding of the programme, as well as an introduction to classroom management, classroom environment, resources management and core methodologies. This training prepared them to implement the programme at a technical level.
- At the second training, the technical features of the programme were revised, and then teachers were given a more in-depth training on core methodologies, particularly those related to writing. Teachers were also given the opportunity to share the work done in their classrooms.
- At both trainings, school managers were introduced to the concept of supporting and monitoring teachers as they implement the programme. They were also given monitoring and support tools to assist in this process.
- Teachers and managers responded very well to the programme and materials, and it was clear that the second event really deepened the understanding of the purpose of the programme and core methodologies.
- Ongoing challenges related to Treatment 1 include:
    - Non-participation by a small number of teachers;
    - Poor time management skills of some teachers, leading to insufficient curriculum coverage;
    - The limited feedback related to implementation in this model limits the service provider's understanding of teacher challenges.

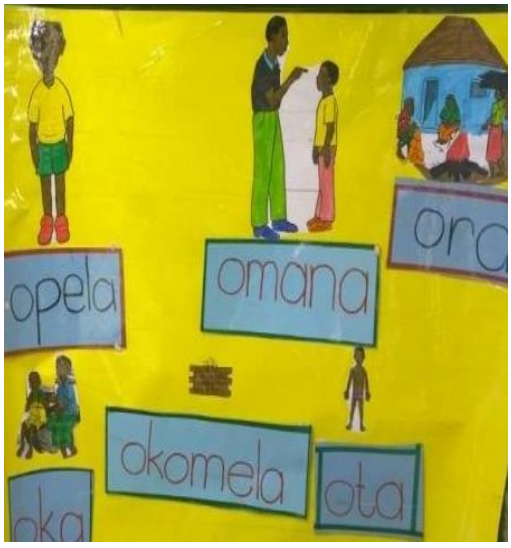
Table 23: Attendance at Treatment 1 training events

	FEB 2015	JULY 2015
School Attendance at Training	100%	92%
Grade 1 Teacher Attendance at Training	100%	85%
School Leaders Attendance at Training	79%	81%

**Note: Materials are distributed to schools that did not attend training.*

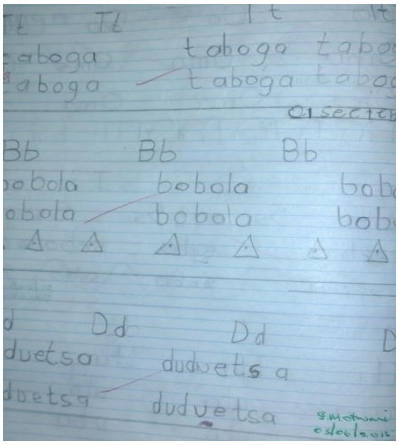
TREATMENT ONE PHOTOGRAPHS

Training: Demonstration by Teacher



Teacher's Chart: Vocabulary

Teacher's Chart: Mind Map



Learner's Work: Handwriting

EGRS Treatment 2 (coaching)

Table 24: Structure of treatment 2

Coach Name	District	Number of Schools	Number of Grade 1 Teachers
Kgomotso Phalatse	Ngaka Modiri Molema	17	26
Helen Kgobane	Ngaka Modiri Molema	18	30
Sabi Mlambo	Dr Kenneth Kaunda	15	34

SUMMARY OF KEY EVENTS

Training

Treatment 2 has 3 full time coaches, each working with a set number of schools and teachers. Coaches have held three teacher training sessions prior to the implementation of the programme for Terms 2, 3 and 4. For Term 2, this was done as 2 x half day sessions, with each coach training 3 – 5 small clusters of teachers. This was not a particularly successful model, as it took a long time to train all teachers, and the smaller groups lacked the energy and enthusiasm of the slightly larger training groups. As a result, the training model was changed after Term 2. For Terms 3 and 4, coaches held full day training sessions to prepare for the next term. Each coach held 2 – 3 larger training events at the end of the previous term, to allow teachers preparation time in the school holidays. In addition, coaches run regular Professional Working Group (PWG) training sessions during the term.

Coaching

During term times, coaches provided follow-up support to all Grade 1 teachers in participating schools. Coaches visit teachers a minimum of once per month. The support sessions include the following kinds of activities:

- Lesson demonstrations by coaches to illustrate the core methodologies
- Lesson observations by coaches
- Critical but positive feedback to teachers regarding lessons observed
- Monitoring of learner exercise and workbooks
- Monitoring of curriculum coverage
- Monitoring of learner assessment results
- Professional interaction with principal and HoDs regarding implementation

Supervision

During term times, supervision of coaches takes place on a regular basis, both on and off site. On site supervision of coaches takes place a minimum of once per term. Treatment Two supervision is characterised by:

- Accompanying coaches on teacher support visits
- Observation of the coaches in practice
- Critical but positive feedback to coaches regarding the manner in which they support teachers
- Informal discussions with teachers concerning the learning programmes and their implementation
- Informal discussions with principals and HODs concerning curriculum and assessment issues
- Monitoring of work schedules and attendance registers

Outcomes of Training and Coaching Sessions

- Successes related to Treatment Two include improvements in:
  - Teacher morale
  - Curriculum coverage
  - Pedagogical content knowledge
- Challenges related to Treatment Two include:
  - Poor-participation by a small number of teachers
  - Slow pacing by some teachers, leading to insufficient curriculum coverage
  - Difficulties related to multigrade teaching in some instances
  - High absenteeism of learners and teachers
  - Practical circumstances – large class sizes and poor infrastructure

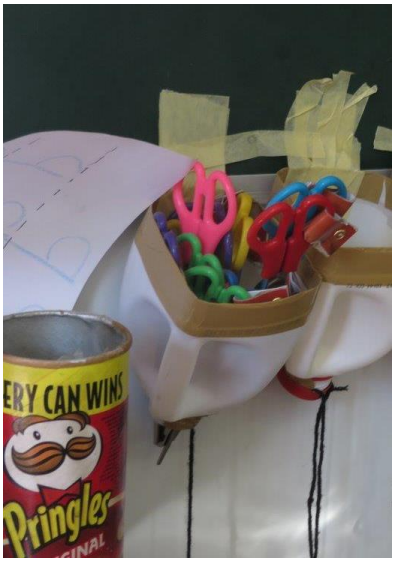
Table 24: Attendance at training events

	Term 2	Term 3	Term 4
School Attendance at Training	100%	88%	94%
Grade 1 Teacher Attendance at Training	100%	82%	93%

**Note: 'Catch-up' training sessions are held with teachers who miss the initial training sessions*

TREATMENT TWO PHOTOGRAPHS

Classroom: Word Wall



Classroom: Resource Management

Learner's Work: Writing

Group Guided Reading

EGRS Treatment 3 (parents)

SUMMARY OF KEY EVENTS

Project Launch

Principals and SGB Representatives were invited to a launch event. At the launch event, schools were introduced to the concept of a Community Reading Coach (CRC) and were asked to recruit a CRC for their school.

CRC Training Events

CRCs are regularly trained in four cluster groups: Zeerust; Lichtenburg; Mafeking and Klerksdorp. Training events focus on administration, facilitation skills and pedagogical content knowledge. The topics covered to date are as follows:

- Topic 1: Small Things Make a Big Difference: Getting the Basics Right
- Topic 2: Playing With Sounds to Support Reading
- Topic 3: Reading Pictures
- Topic 4: Letter Sounds
- Topic 5: Incidental Reading
- Topic 6: Preparing to read Books
- Topic 7: Reading Story #1

Family Training Sessions

For each topic, CRCs hold three family training sessions. The same content is covered in slightly different ways over the three sessions. For each topic, families are given a 'family card' with key information to remind them of certain behaviors and practices to implement in their homes. Key information is presented and discussed, and then activities are practiced. Families are encouraged to replicate activities at home on a regular basis.

Outcomes of Family Training Sessions

- Successes related to the family training sessions include:
    - Improvements in the knowledge of parenting and reading support skills of parents and families
    - Improved reading skills of families
    - Slowly increasing social capital of families, leading to better participation in formal school structures
- There are also anecdotal reports of improved school attendance and performance by children of participating parents.

- Challenges related to the family training sessions include:

- Recruitment of suitable CRCs. In some instances, the resident population is either not willing to work for the small volunteer stipend, or there is no suitably skilled candidate available for the position.
- Attendance of families at training sessions is an ongoing challenge. Attendance decreased over the winter months and before and after school holidays.
- New strategies to increase parent attendance include engagement with principals to try and motivate parent attendance, and a small incentive scheme in the form of a 'lucky-draw prize' per session. These strategies were implemented from August, and the results are still to be measured.

Table 25: Attendance at CRC training events

	Orientation & Topic 1	Topics 2 & 3	Topics 4 & 5	Topic 6	Topic 7
CRC Attendance at Training	98%	88%	70%	86%	88%

**Note: 'Catch-up' training sessions are held with CRCs who miss the initial training sessions*

Table 26: Attendance at parent training events

	Orientation & Topic 1	Topic 2	Topic 3	Topic 4
Parent Attendance at Training	35%	29%	21%	18%

TREATMENT THREE PHOTOGRAPHS

CRC Training: Role Play

CRC Training: Games

Family Card 1: Phonemic Awareness

Family Card 5: Incidental Reading

Family Card 7: Reading Books

Family Card 1: The Basics

NEXT STEPS IN THE PROJECT

Interventions are scheduled to continue throughout 2016. A midline data collection is taking place between the 26th of October and the 13th of November 2015. Using this data, the impacts of each intervention after one year will be measured. The endline data collection is scheduled for October/November 2016. This will allow us to measure the impacts of two years of treatment on reading outcomes at the end of grade 2. In the event of at least one of the interventions showing a significant impact on reading outcomes at the end of grade 2, we plan on using DBE administrative test data and possibly even raising funds for a further round of data collection to measure the longer-term impacts of the interventions.

REFERENCES

Andrabi, T. D. (2013). *Report Cards: The impact of providing school and child test scores on educational markets.* Unpublished mimeo.

Department of Basic Education. (2014). *Report on the Annual National Assessment of 2014.* Pretoria: Department of Basic Education.

Fleisch, B., & Schöer, V. (2014). Large-scale instructional reform in the Global South: insights from the mid-point evaluation of the Gauteng Priamry Language and Mathematics Strategy. *South African Journal of Education, 34*(3), 1-12.

Fleisch, B., Taylor, S., Schöer, V., & Mabogoane, T. (2015). The impact of a remedial reading programme on second language grade 4 students in KZN: Evidence from a randomised impact evaluation. *South African Monitoring and Evaluation Conference.* Johannesburg.

Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and Interventions to Improve Early Grade Literacy.* RTI Press.

He, F., Linden, L. L., & MacLeod, M. (2009). *A better way to teach children to read? Evidence from a randomized control trial.* Retrieved September 12, 2012, from https://webspace.utexas.edu/linden/www/Teach%20Children%20to%20Read.pdf

Howie, S., Van Staden, S., Tshele, M., Dowse, C., & Zimmerman, L. (2012). *South African Children's Reading Literacy Achievment: Summary Report.* Pretoria.

Lieberman, E. P. (2013). *Does Information Lead to More Active Citizenship? Evidence from an Education Intervention in Rural Kenya.* MIT Working Paper.

Pandey, P., Goyal, S., & Sundararaman, V. (2009). Community participation in public schools: Impact of information campaigns in three Indian states. *Education Economics, 17*(3), 355-375.

Sailors, M. H. (2010). The Effects of First- and Second-Language Instruction in Rural South African Schools. *Bilingual Research Journal, 33*, 21-41.

Spaull, N., & Taylor, S. (2015). Access to what? Creating a composite measure of educational quantity and educational quality for 11 African countries. *Comparative Education Review, 59*(1), 133-165.

Taylor, S., & Coetzee, M. (2013). *Estimating the impact of language of instruction in South African primary schools: A fixed effects approach.* Stellenbosch Economics Working Papers 21/13.