Second Early Grade Reading Study

Pre-Analysis Plan for Wave 4 round of data collection

13 December 2019

Jacobus Cilliers, Janeli Kotzé, Stephen Taylor, Brahm Fleisch and Nompumelelo Mohohlwane

Background

This project is designed as a Randomised Controlled Trial (RCT) to evaluate the impact of two alternative treatments, both aimed at improving early grade reading in English as a First Additional Language (EFAL) in the South African school system. The primary implementing partner is the South African government, in particular the Department of Basic Education (DBE). A key role is also being played by the Mpumalanga provincial education department, which is championing the project within the schools. A service provider has been appointed to run the two interventions on behalf of the DBE for the purposes of this impact evaluation. The service provider is a consortium of “Class Act” and “Molteno”, both of whom are highly involved in partnerships with government to develop Learning and Teaching Support Material (LTSM) and run literacy interventions.

The Research Team consists of academics from the University of the Witwatersrand, Georgetown University and the Department of Basic Education.

An original pre-analysis plan was written in 2015, prior to the start of all data collection. This is a revised pre-analysis plan for the final round of data collection, assessing the impact on learners after being exposed to the program for three years.

Description of Interventions

This study evaluates two different interventions, both aimed at improving early grade reading in English as First Additional Language (EFAL) to prepare learners better for the transition from home language to English as the language of instruction, which occurs in Grade 4. Both interventions focus on training and supporting teachers in the foundation phase to better teach English in the early grades.

Intervention 1: Teacher training, paper-based scripted lessons, learning and teaching support material and coaching

Both interventions aim to apply the same set of instructional practices in the teaching of EFAL in the grade 1, 2 and 3 classrooms. Both interventions therefore provide teachers with clear, scripted lesson plans, which are aligned to the South African curriculum as specified in the Curriculum and Assessment Policy Statements (CAPS) for EFAL in the Foundation Phase. The lesson plans incorporate the use of learning support materials, including the government provided workbooks as well as certain additional materials (graded reading booklets, flashcards, big books, posters, etc.), which are provided through the EGRS II. The graded reading booklets were introduced in the second year of the study (as per the recommendation of the CAPS) as 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.

Intervention 1 initially introduces the teachers on how to use the lesson plans and accompanying materials through a two-day central training session at the start of the year. During the year, cluster training sessions are held for one-day prior to the start of each academic term to additionally train the teachers on the use of certain core methodologies in the lesson plans. In addition to the training sessions, ongoing support is provided to teachers through ‘specialist’ reading coaches. The on-going support takes on the form of on-site visits by the reading coach on a monthly basis to assist teachers with the implementation of the lesson plans.

Intervention 2: Teacher training, electronic scripted lessons, learning and teaching support material and virtual coaching

Exactly the same set of instructional materials (graded reading booklets, flashcards, big books, posters, etc.) is provided to Intervention 2 schools. However, instead of providing the scripted lesson plans in a paper-based format, the scripted lesson plans are provided in an electronic format on tablets to teachers. In addition to the scripted lesson plans, the tablets also contain short video tutorials for the teachers on teaching practice, audio clips on the pronunciation of the phonics sounds, audio clips of the songs and rhymes and examples of learners' work. Furthermore, instead of a reading coach visiting the schools on a regular basis, the intervention has a virtual coach who provides on-going support to teachers through phone calls, WhatsApp and SMS messaging. The teachers in intervention 2 are initially introduced to the lesson plans and materials through a three-day central training session at the start of the year, and also receive one-day cluster training sessions at the start of each academic term.

Timeline

The program is being implemented over a period of three years, tracking learners from the start of grade 1 (February 2017) up to the end of grade 3 (November 2019). Each year a different set of teachers are exposed to the program: grade 1 teacher in year one, grade 2 teachers in year two, and grade 3 teachers in year three. There are five rounds of data collection: once prior to the start of the program (January 2017), and again at the end of every year of implementation (November 2017, 2018 and 2019). To evaluate the sustainability of the benefits of the interventions, the same learners will be assessed in Grade 4, a year after the interventions have been concluded.

Theory of Change

The Second Early Grade Reading Study aims to affect behavior change amongst teachers at a large scale, in line with the curriculum and methodologies in which teachers were trained during the teacher training at the start of the programme. The scripted lesson plans provide a mechanism to prompt the enactment of the behavior change, whereas the coaching serves as an additional mechanism to encourage fidelity to the programme.

The First Early Grade Reading Study showed that coaching was important in affecting sustained behavior change in the teaching of Home Language. In the Second Early Grade Reading Study, we are evaluating whether the same results will be obtained when applying the same programme in a different province and in a different subject (English First Additional Language). Furthermore, we are evaluating whether on-site, face-to-face coaching is essential, or technology can be utilized to reduce the costs of coaching with the same effects.

Teaching English as First Additional Language

All our interventions are grounded in the Curriculum and Assessment Policy Statements (CAPS) and aim to support teachers in the enactment of the current South African curriculum for English as a First Additional Language. The focus of the study is to inform the system on ways to strengthen teaching and learning in EFAL in the foundation phase, in order to prepare learners for the language transition to English as the language of learning and teaching (LOLT) which happens in the majority of schools in Grade 4.

A range of factors influence learning outcomes, including school and non-school variables, but instruction or instructional practice is one major influence on learning. One of the key characteristics of South African education is that the dualistic nature of learning outcomes is mirrored by dual types of instructional practice (e.g. Hoadley, 2010). It is likely that weak instructional practices have a causal impact on learning outcomes in the poorly performing part of the school system. To substantially shift achievement in the weak part of the schooling system it may be necessary to apply a comprehensive instructional change intervention, involving a set of coherent and aligned instructional inputs.

To address these complexities, the interventions provide teachers with instructional inputs such as scripted lesson plans, aligned learning materials, and support and training to teachers.

Common aspects of the two interventions

Scripted lessons specify new instructional practices, including faster-paced instruction, more appropriately sequenced content and expanded pedagogic repertoires. New expanded pedagogic repertoires specifically for primary grade reading in the FAL include systematic teaching of phonemic awareness and phonics strategies that focus on increased reading speeds or fluency; guided reading strategies; vocabulary development and strategies that improve comprehension. The scripted lessons require little additional lesson preparation from teachers, which makes the take-up of a more productive teaching practice more manageable. The faster-paced instruction (relative to typical practice) also ensures that the teachers cover the full prescribed curriculum for the year. Furthermore, scripted lesson plans free up teachers' time, because they no longer need to allocate as much time to lesson preparation. This could improve reading acquisition if teachers allocate this time to productive teaching activities, rather than leisure or unproductive teaching activities.

The role of the learning materials is to provide the appropriate resources to ensure that learners are able to develop and consolidate knowledge and skills related to English language proficiency, English reading fluency, English vocabulary development and guided reading. From Grade 2 onwards, ten titles of graded reading books will be provided for each classroom. The number of copies will be determined based on class sizes. The accompanying graded reading materials provide ample material for learners to practice decoding and reading at their level of development. It is recognized in South Africa that the opportunity to learn EFAL may be hindered by a lack of suitable materials to assist in the progression from one phase of reading acquisition to the next, and this is likely to be particularly true in African language schools, the focus of our intervention. Furthermore, teachers are required to provide regular assessment of learners' reading proficiency in order to assign learners to the appropriate graded readers and small reading groups, based on ability. The group-guided reading also provides the teacher with opportunities to provide individualized attention.

In addition to the above, coaching 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 day of training. Furthermore, the role of coaching support is to fuse capacity-building and accountability.

Given the perceived high costs of direct in-class coaching, the study tests the relative cost-effectiveness of two kinds of capacity building, the traditional face-to-face and the reduced face-to-face capacity building combined with the new alternative electronic support in the form of a tablet with video demonstration lessons and cell phone-messaging support.

Intervention 1 specifically:

The in-class support allows for modelling of the new practice on-site and the gradual development of teachers in the new practice from novice to expert. The in-class support also allows teachers to manage the emotional labour, i.e. stress, insecurity and anxiety associated with developing a new professional practice mid-career. The presence of the in-class support allows for the development of professional accountability in an environment of trust, where the coach monitors and evaluates the teachers' teaching practices in order to encourage more productive teaching practices. The on-going support from the coach also encourages the teacher to keep up with the increased pace of the scripted lesson plans throughout the course of the year.

Intervention 2 specifically:

The technology-supplemented intervention aims to provide new forms of support and guidance on teaching strategies through a range of materials, teaching guides, videos and interactive support platforms that are available at all times to the teacher. These resources are intended to encourage more productive teaching practices among teachers. Scripted lesson plans are provided in an electronic format on a tablet and are integrated with various audio and visual resources to support teachers in the teaching of EFAL. The resources are supplemented with virtual coaching through phone calls and cellphone messaging. The on-going support from the virtual coach

intends to encourage the teacher to keep up with the increased pace of the scripted lesson plans throughout the course of the year.

Evaluation design and data collection

Treatment assignment and sample selection

Through a process of elimination, we developed a sampling frame of 180 eligible schools. Beginning with 731 primary schools registered in the 2016 administrative data in the districts of Ehlanzeni and Gert Sibande, we firstly excluded relatively affluent schools (those in quintiles 4 and 5). Next, we excluded schools in which the language of instruction in the Foundation Phase was neither Siswati nor isiZulu. We also excluded schools which were missing in the 2014 Annual National Assessment (ANA) dataset.[1] We further excluded particularly small schools (fewer than 30 grade 1 learners enrolled) since many of these schools would practice multi-grade teaching rendering the grade-specific lesson plans less appropriate. We also excluded particularly large schools (more than 160 grade 1 enrolments, or more than 3 classes in grade 1, or classes with more than 60 learners in) to limit intervention costs. After all of these exclusions, 193 eligible schools remained. Using a random number generator, we then excluded a further 3 schools to remain with a sample of 190 schools. The 190 school sample included a sample of 10 replacement schools (1 in each of the strata) should the need arise to drop one of the sample schools. Thus we obtained the sampling frame of 190 schools.

To increase power and assure balance between treatment arms, we performed stratified randomization. We created 10 strata of 19 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, 8 to the control group and 1 as a replacement school. 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.

Data collection

We have conducted four rounds of data collection. The first wave of data collection took place in February 2017, where 20 grade 1 learners per school were randomly selected into the sample. Three subsequent waves of data collection took place end of Grade 1 (2017, wave 2), end of Grade 2 (2018, wave 3), and end of Grade 3 (2019, wave 4). A fifth wave of data collection is also planned for the end of Grade 4 (2020, wave 5).

Each year we visit the same 180 schools, tracking the same learners, and administering the following instruments:

- Learner assessment of English reading and language proficiency, home language reading proficiency, and mathematics.
- Basic teacher assessment in English proficiency.
- Teacher, school principal, and parent surveys
- Facility review and document inspection.

In the fourth round of data collection, a more in-depth vocabulary assessment was done with a sub-sample of learners (6 per school), in a sub-sample of 60 schools. The vocabulary assessment included an English and a home language activity that consisted of 102 vocabulary items respectively. Three tasks from the main test were also retested by the fieldworkers responsible for the vocabulary testing to allow for a measure of inter-rater reliability.

[1] The Annual National Assessments have not been administered since 2014. This is therefore the most recent standardised measure of school performance we have for the full population of schools.

In the same sub-sample of schools, a classroom observation study was conducted that entailed an observation of the English lesson and the home language lesson, an interview with the teacher and a review of learners' written work.

Learner assessment

The learner assessment at the end of Year 3 (Grade 3) of the intervention consisted of ten different tasks that assessed various home language and EFAL oral, reading, writing and comprehension skills. A number of these assessment tasks were repeated from the Year 1 assessment to be able to determine the learning gains that takes place in one year.

TABLE 1: LEARNER ASSESSMENT TASKS INCLUDED IN EACH WAVE OF DATA COLLECTION

	Construct	Baseline		Year 1		Year 2		Year 3	
		Start of Gr 1		End of Gr 1		End of Gr 2		End of Gr 3	
		L1	*English*	*L1*	*English*	*L1*	*English*	*L1*	*English*
Language Comprehension	Receptive Vocabulary		x		x		x		
	Expressive Vocabulary	x	x	x	x		x		x
	Listening Comprehension	x			x		x		x
Decoding	Phonological working memory	x							
	Phonological Awareness	x			x				
	Rapid letter naming					x		x	
	Letter-sound recognition	x			x	x		x	
	Word reading fluency	x		x	x		x		x
	Sentence reading fluency	x							
	Oral Reading Fluency (ORF)					x	x	x	x
	Reading Comprehension					x	x	x	x
	Written Comprehension							x	x
Spelling	Spelling of a CVC word				x				
	Writing two words						x		

Hypotheses under investigation

This section outlines our primary and secondary outcomes of interest, as well as heterogeneous treatment effects we plan to examine. In brackets, we indicate the source of the data as well as question number if it is based on a survey. If an outcome is based on multiple indicators, it will be aggregated using principal component analysis.

[*SPQ = School Principal Questionnaire; TRQ = Teacher Questionnaire; PRQ = Parent Questionnaire; LOBS = Lesson Observation Study*]

Primary outcomes

We have decided to construct two indices based on the two language constructs that learners of a second language have to master in the foundation phase. The first construct is language proficiency as it relates to English vocabulary development and the second relates to decoding skills. In the first grade, learners are only taught the language proficiency skills during the English lessons, whereas the decoding skills are already taught during the

Home Language lessons in the first grade. Decoding skills are only introduced in the English lesson from the second half of the second grade. The two primary outcomes are therefore:

1. **English language proficiency index**, constructed using principal component analysis with the following indicators:
    a. English expressive vocabulary test.
    b. English listening comprehension test.
2. **English decoding index**, constructed using principal component analysis with the following indicators:
    a. Word recognition
    b. Oral Reading Fluency
    c. Oral reading comprehension.
    d. Written comprehension.

Secondary outcomes

In order to test whether the treatments had any crowding-out or spillover effects on the other subject areas, we will also look at the following outcomes:

1. **Home language reading proficiency**, constructed using principal component analysis with the following indicators:
    a. Letter recognition
    b. Oral Reading Fluency
    c. Oral reading comprehension.
    d. Written comprehension.
2. Performance in a simple **mathematics test**.

In order to test for mechanisms based on our theory of change, we will also examine the impact on a series of intermediate outcomes:

3. Implementation fidelity
    a. Received training [*[TRQ 6.4]*
    b. Frequency that a coach/mentor (i) observes a teacher in EFAL *[TRQ 6.6.4]*, (ii) demonstrate teaching practice *[TRQ 6.7.4]*, or (iii) provide a compliment or praise *[TRQ 6.8.4]*
    c. Have access to graded readers in the classroom *[TRQ 4.2.1]*
    d. Total time spent teaching:
        i. English as a First Additional Language *[TRQ 3.4]*
        ii. Home Language *[TRQ 3.5]*
4. School support.
    a. Frequency that the school principal, head of department, or subject advisor (i) observes you teach EFAL *[TRQ 6.6.1 - 6.6.3]*, (ii) demonstrate teaching practice *[TRQ 6.7.1 – 6.7.3]*, or (iii) provide a compliment or praise *[TRQ 6.7.1 – 6.7.3]*
5. Teacher behavior
    a. Use of resources provided by the program, including (i) proportion of teachers who use the Graded Readers daily *[TRQ 4.2.3]*, (ii) proportion of teachers who use Lesson Plans provided by an NGO *[TRQ 4.1.3]*, and (iii) print richness of the environment *[TRQ 8]*
    b. The proportion of the teaching practices that are conducted at the correct weekly frequency *[TRQ 4.7; 4.10]*.

c. The completion of more writing exercises *[TRQ 9.3]*

6. Skill acquisition: The proportion of teaching activities that a **teacher finds difficult to implement** *[TRQ 4.8]*

Sub-group analysis.

Primary

We will examine impacts separately for **boys vs girls**. It is possible that boys benefit more, because they generally lag behind girls in reading proficiency at early grades, and benefit from the structure imposed by the program. Although we are not interested in knowing if there is a statistically significant difference between boys and girls, we want to know if there is a statistically significant impact on these two sub-groups.

We will also test if there is a statistically significant **difference between the two districts** where the program is being implemented. Since the two districts fall under different administrative areas, and also vary by socio-economic background and home language of the schools, the program could have very different impacts

Finally, we will also examine if the impacts **vary by learner ability**. For this purpose, will conduct quantile regressions, and also use a local polynomial fit to examine how the effect size varies by aggregate baseline performance in the literacy test.

Secondary

We will estimate the impacts **excluding one circuit** where teachers were on strike during the third year of implementation.

We will also test if there is a statistically significant difference between schools where the **home language is isiZulu** and schools where the home language is Swati.

Empirical strategy

The schools in our sample are randomly assigned to the three treatment groups (treatment 1, treatment 2 and control). Our main estimating equation will be:

$$y_{isb1} = \beta_0 + \beta_1 T1 + \beta_2 T2 + X'_{isb0}\,\Gamma + \rho_b + \varepsilon_{isb1}$$

Where y_{isb1} is the outcome indicator of interest (more below) for learner i in school s and strata b, T1 is the treatment dummy for treatment 1, T2 is the treatment dummy for treatment 2, ρ_b refers to strata fixed effects, X'_{isb0} is a vector of baseline controls, and ε_{isb1} is the error term clustered at the school level.

We will first estimate the model on the control data, and then select the controls that maximize the R-squared of the regression. The controls we will consider include: learner baseline performance in each sub-test (i.e. vocabulary, letter recognition, working memory, phonological awareness and word recognition), learner gender, learner age, enumerator fixed effects, the language in which the Home Language section of the test was written in, the district, and a measure of the socio-economic status of the community in which the school is located,

Additional corrections

Correction for multiple hypotheses

1. **Mean index:** To minimize the risk of over-rejection of the null, we will construct a mean index of the different metrics of reading proficiency, using principal component analysis. However, since each metric

relates to a different component in reading progression it is important to test each separately to identify blockages in reading progression.

Addressing survey attrition

We will construct a dummy indicating whether a learner attrited between baseline and midline and test if attrition was imbalanced (by regressing the attrition dummy on treatment status) and also if it was non-random (by regressing the attrition dummy on baseline test scores).

If attrition was found to be non-random and imbalanced, we can construct Lee Bounds – a conservative measure of the upper and lower bounds based on the most extreme sample selection - and also conduct Inverse Probability Weighting.

Cost effectiveness

In order to establish the cost effectiveness of each treatment we calculate the standard deviations gained per US$100 spent on treatment. This allows us to make comparisons with other studies reported on in Kremer, Brannen and Glennerster (2013). We use the estimated treatment effect sizes from the main equation. We convert costs from Rand values to US dollars using the Rand-Dollar exchange rate as at the close of the South African markets on 15 December 2019.