Moving towards cost-effective delivery models of teacher coaching: Evidence from field experiments in South Africa

Janeli Kotze, Stephen Taylor and Brahm Fleisch

RISE Conference

Abstract

Structured learning programmes have emerged as a promising way to address the low learning levels observed in many developing countries. The delivery model of these programmes matters, and on-site teacher coaching has been recommended to be highly effective, especially in early grade literacy. In this paper, we report on a series of government-led randomised experiments in South African primary schools that build on each other to test various models of teacher training and coaching. We find that sustained on-site coaching is more cost-effective (0.41 SD increase in test scores per USD 100) than either short coaching interventions (no significant impact) or centralized teacher training workshops (0.23 SD increase in test scores per USD 100). The ability to scale on-site coaching, however, is an open question. In the latest experiment, therefore, a virtual coaching programme was compared to on-site coaching. The focus of this paper is on the midline evaluation results from this experiment. After one year of intervention, virtual coaching was no less effective than on-site coaching at improving both the instructional practice of teachers and the targeted literacy outcomes of children. This points to the potential for technological innovations to enable wider rollout of coaching programmes, even in contexts where teachers are not familiar with new technologies.

1. Introduction

The value added by good teachers and good teaching to pupil learning is large, both in developed country settings (Hanushek & Rivkin, 2010; Chetty, Friedman, & Rockoff, 2014) and in developing country settings (Bau & Das, 2017; Bold, et al., 2017; Bruns & Luque, 2014), where a problem of low learning achievement is widespread. Yet it is hard to observe teacher characteristics that are consistently related to better pupil learning. It is also difficult to change teacher practice at a large scale in ways that substantially shift learning. Whilst some short-term gains may be achieved through "out-of-the-system" interventions like contract teachers (Duflo, et al., 2012) or holiday reading camps (Banerjee, et al., 2010), it seems unlikely that such interventions will be sustainable in bringing systemic change. Ultimately, changing the practices of existing state-employed teachers is the key to shifting learning outcomes in underperforming education systems.

Some have argued that the curricula in developing countries are often overly ambitious relative to the learning foundations that children begin school with (e.g. Pritchett & Beatty, 2012). In line with this idea, several of Pratham's interventions in India have achieved significant learning gains through "teaching to the right level" (Banerjee, et al., 2017). Whilst some benefits may indeed be possible through more appropriate curricula, the findings referred to above that good teachers make a large impact on learning compared to poor teachers using the same given curriculum implies that there may still be substantial room for efficiency gains through better implementation of the existing curriculum. Moreover, calls for a less ambitious curriculum are unlikely to receive much political support, especially in a context like South Africa where a "lesser" curriculum was historically associated with institutionalised racial segregation and oppression.

To a great extent, the problem lies not in the curriculum itself but in the fidelity of curriculum implementation. In South Africa, there is evidence of highly incomplete curriculum coverage (Taylor, 2011), ineffective curriculum pacing and sequencing (Hoadley, 2010), and the avoidance of more complex teaching activities (Prinsloo, 2008), which are often the very activities prescribed in the curriculum that promote individualised attention to pupils.

One common policy response is to train teachers using workshops, which typically last anywhere between half a day and a week. The implicit theory of change with this approach is that teachers will update their knowledge and automatically implement new practices in their classrooms. This may be a naïve expectation, especially for those teachers who have become comfortable with a certain set of routines over many years.

Recent systematic reviews have suggested that "structured pedagogic programmes" or "structured learning programmes" are more effective at promoting instructional change and pupil learning in mathematics and (especially) reading outcomes (Snilstveit *et al*, 2016; Popova, Evans, & Arancibia, 2016). These programmes usually combine highly specified curricula, training on instructional methods, and additional learning materials, and have a practical focus on classroom routines and activities. Recent research in South Africa, Macedonia, Bangladesh, Malawi and Kenya has further explored, and subsequently confirmed the benefits of a structured pedagogy approach to improving teaching practice (Sturtevant & Linek, 2007; Sailors, et al., 2014; Fleisch, 2016; Piper, 2017; Shaheen, et al., 2013).

Within the category of structured learning programmes, the modality of delivery of professional development also matters (Popova, et al., 2016). On-site teacher coaching has begun to receive attention in recent years as several studies have found significant positive impacts on learning outcomes (Kraft, et al., 2016). However, the high financial costs and difficulties with recruiting enough good coaches will mean that much of the research going forward should focus on developing cost-effective variations on the coaching model. A series of three experiments in South Africa, led by the Department of Basic Education in collaboration with academics, have aimed to contribute to this research agenda.

The purpose of this government-led initiative, which has become known as the Early Grade Reading Study (EGRS), is to undertake rigorous research on interventions to improve early grade reading at a systemic level, so that policy can be influenced based on evidence. The focus is on the teaching of African languages and English as a second language, particularly in public schools serving low-income communities in South Africa. The EGRS makes use of randomised control trial (RCT) designs complemented by large-sample classroom (lesson) observations and in-depth case studies. The complementary mixed-method design ensures that the study provides accurate estimates of learner outcome impacts, measures of change in instructional practice, and insights about the mechanisms of change.

In the first experiment, a "Reading Catch Up Programme" (RCUP) used on-site teacher coaches over a short period of time (11 weeks) to boost the teaching and learning of English reading amongst Grade 4 pupils whose home language was not English.[1] This programme had been reported to have had a large positive impact on English reading in the Gauteng province (Hellman, 2012). However, this evaluation was not independently administered and had no counterfactual. Fleisch, Taylor, Schöer, & Mabogoane (2017) therefore conducted an RCT in the Pinetown district of the KwaZulu-Natal province to evaluate the impact of the RCUP. The short duration of this intervention meant that, if effective, this could be implemented in a cost-effective manner through rotating coaches across schools for one term at a time.

The average impact of the RCUP was not significantly different from zero, although children with a higher baseline English proficiency did register statistically significant gains (Fleisch, et al., 2017). This experiment yielded two key lessons. Firstly, most learners did not have the language foundation to benefit from an additional support programme. This pointed to the need for interventions prior to Grade 4. Secondly, the duration of coaching was too short. Future interventions would need to last at least one or two years. These lessons informed the design of the first Early Grade Reading Study ("EGRS 1").

The EGRS 1 explicitly tested on-site coaching versus centralised teacher training, where both intervention arms received the exact same daily lesson plans and additional reading materials. Each programme was implemented in a group of 50 schools, with a further 80 schools serving as control. The programmes were administered in the North West Province and targeted reading and literacy in the home language, which in this province is Setswana.[2] The programmes were administered to the

[1] In South Africa, most children learn in their home language as the language of instruction in the first three grades and then transition to English as the language of instruction in the fourth grade.

[2] During the first three grades there are four learning areas required by the South African curriculum: Home language literacy, First Additional Language (which is usually English), Numeracy and Life Orientation.

grade 1 class of 2015, the grade 2 class of 2016 and the grade 3 class of 2017. Most teachers were therefore new to the programme each year, but the same cohort of children benefited throughout.

The centralised teacher training occurred for two days at a time, at the start of each year and midway through each year. The major cost in this intervention was the travel and accommodation for teachers. The on-site coaching intervention included a half-day cluster meeting to disseminate lesson plans and reading materials at the start of each quarter, as well as on-site coaching visits about once a month. The ratio of coaches to schools was roughly 1:17. The main cost in the coaching intervention was the salaries of coaches and this intervention was about 30% more expensive than the centralised training intervention. The same coaches administered both interventions, so the level of expertise was held constant.

After two years of interventions, the on-site coaching intervention had returned an average home language test score impact of 0.24 SD relative to control. The impact of the central training intervention was half as large at 0.12 SD. Another striking result was that the coaching intervention had a significant positive impact on English proficiency, even though this was not the direct outcome targeted by the intervention. There are two possible explanations of this "spill-over" effect: the benefits of learning to read in one's home language may have facilitated better second language acquisition, or teachers may have implemented similar methods in their English lessons as they were coached on during their home language lessons.

Lesson observations and detailed case studies indicated that the coaching intervention promoted adherence to the programme and better use of the more difficult teaching methodologies. In particular, a method called "group-guided reading", which is prescribed in the curriculum but rarely implemented in control classrooms, was much more likely to occur and created opportunities for children to handle books and read individually.

Overall, the coaching intervention remained more cost-effective (0.41 SD increase in test scores per USD 100) than central training (0.23 SD increase in test scores per USD 100). Aside from financial cost, however, there are questions about the ability to recruit and manage the larger numbers of coaches that are required for large scale implementation. At this stage, we are offering three responses to this question:

1) Just because an intervention will be difficult to implement in all primary schools throughout a country simultaneously doesn't mean it cannot be implemented in targeted schools at a level of scale that does make a difference to the system as a whole. It may well be feasible to implement coaching in about 25% of schools on a rotational basis. In the absence of alternative programmes with known evidence of impact, this seems like a prudent approach.
2) Test the impact of on-site coaching when it is implemented on a larger scale. In the next phase of EGRS 1, the government is planning on recruiting 15 coaches in the North West, instead of the three coaches used during the RCT. An experimental design will again be used to measure impact.
3) Investigate whether variations on the coaching model where costs and human resource requirements are lower can still have a similar impact on learning outcomes. For example, training lead teachers at each school to play a coaching role is one such idea. Similarly, using a virtual coach is another, and this is what motivated the second Early Grade Reading Study (EGRS 2).

The second EGRS began in 2017 in the Mpumalanga province. As was the case in EGRS 1, the grade 1 cohort is receiving three years of interventions. Two variations on coaching are being implemented, this time with the intention of improving the teaching and learning of English as a First Additional Language. The first programme uses on-site coaching, similar in design to what was found to be effective as a home language intervention in EGRS 1. The second programme uses virtual coaching where one coach is in communication with teachers in all 50 schools. Teachers in this programme receive an electronic tablet, which has the daily lesson plans on it. The same set of lesson plans are used across both intervention arms, but one group receives paper-based lesson plans and the other receives them in an electronic format.

A relevant study design to compare this to is that done in Brazil by Bruns, Costa & Cunha (2017), who evaluated the impact of a lower-cost coaching model, which used Skype to provide virtual coaching based on classroom observation feedback. This approach led to a smallish average impact on learning outcomes of between 0.04 and 0.09 SD. However, the virtual coaching was not compared to an on-site coaching intervention, meaning that the cost-effectiveness question remains unanswered.

The EGRS 2 project is still underway and the results presented in this paper are preliminary findings based on a midline survey conducted after one year of interventions. One of the challenges experienced when measuring grade 1 children in a second language is that their proficiency is still so basic that it is difficult to test. The grade 1 curriculum for English as a first additional language focuses mainly on oral language proficiency rather than on the technical aspects of reading, such as decoding, which is the focus of the home language curriculum. After one year of the interventions both intervention groups recorded positive impacts on the three subtests that focus on English oral language development (English listening, English comprehension and English vocabulary) of between 0.1 and 0.5 SD. The impacts are similar in magnitude across the on-site coaching and virtual coaching groups.

A lesson observation study in a sub-sample of 60 schools revealed that instructional practice has changed in noticeable ways in both intervention groups. There is increased use of English in lessons by teachers and learners – in control schools there is more reliance on home language even during the English lesson. Researchers also observed a more systematic approach to the development of vocabulary focusing on using words in context. Finally teachers in the intervention groups were more likely to spend less time on their lesson preparations than teachers in the control group, and were also more likely to have received curriculum support from their Head of Department (HOD) and a mentor.

The midline results of EGRS 2 presents preliminary evidence that virtual coaching together with a structured learning programme administered through tablet-based lesson plans can be an effective alternative model to deliver coaching to improve early grade reading.

2. Literature review and theory of change on teacher coaching

The structured pedagogic approach generally entails some form of teacher guideline or prescriptive lesson plans, teacher professional development as well as integrated educational material. It has been shown that the lesson plans and the quality educational materials are a necessary but not sufficient condition for large-scale instructional change in low capacity education systems. Nevertheless, without teacher learning and teacher agency, these elements of the instructional infrastructure have

little chance of transforming the everyday learning activities and tasks in the classroom. There is a growing recognition that conventional forms of in-service training, that is short-courses, workshops and accredited academic programmes, fall short of impacting on instructional practice. As such, new approaches to teacher development have emerged, many of which focus on the development of localised communities of professional practice (Chauraya & Brodie, 2017). While teacher-led communities of practice may be appropriate for high-functioning systems where the primary concern is sharing of best practice between teachers, this approach is unlikely to be effective in a low-functioning system that has few examples of 'good' practice that could be easily shared. As such, another approach to teacher development, i.e. instructional coaching, emerged as the third component of a structured pedagogic programme.

A key theory that informs the instructional coaching approach is Thomas Guskey's (2002) professional development and teacher change. Guskey's theory pivots around the sequences in the change steps. In this argument, most programmes that aim at changing attitudes and beliefs begin with steps aimed at gaining acceptance, buy-in, and generating enthusiasm from teachers prior to implementation of new instructional practices. In his view, these seldom really work. For Guskey, the most appropriate sequence for change interventions that aim to improve learning outcomes and teacher beliefs must begin by getting teachers to implement the new practice in their classroom as the first step. This is followed by appraisal of the new practice, either internal or external, followed by correction and adjustment and finally assessment of the impact on learning. Once teachers have practiced, reviewed and noted change, their attitudes and beliefs will follow.

> "The crucial point is that it is not the professional development per se, but the experience of successful implementation that changes teachers' attitudes and beliefs. They believe it works because they have seen it work, and that experience shapes their attitudes and beliefs." (p. 343)

The role of instructional coaching within this logic is that it is the best way to encourage and/or induct teachers into the new practice, the precondition for change. While focusing on educational rounds, City et al (2009) make a similar point, that is, teachers only learn to do the work by doing the work, and not by being told to do the work or told how to do the work, or told that they would be rewarded or punished for outcomes associated with the work.

Guskey's work provides a powerful theoretical rationale for coaching upon which other scholars have begun to develop a conceptual framework to unpack the elements of the coaching process. D'Amico and Stein (2002) suggested that powerful professional learning needs to be *long-term*, *experientially-based* and heavily dependent on *individualised* interaction between the novice and an expert. To this Kraft, Blazer and Hogan (2017) add the need for such capacity development to be *intensive*, with interaction taking place anywhere between once a week and once a month and *focused*, i.e. working with teachers around deliberate practice of specific skills. Through the one-on-one coaching process that facilitates 'doing the work', teachers acquire skills and knowledge and technical vocabulary associated with the new instructional practice.

While the prime purpose of coaching is to enable practice-based learning, there is a second and equally important role that coaching plays in instructional change. Since the Fullan (2007) and

Hargreaves (2005) seminal studies on the meaning of educational change, there has been a growing recognition of the significant impact, both positive and negative, that emotions have on educational change. There is an emerging body of research that suggests that instructional coaches can play a positive role in mediating the anger, insecurity and feelings of inadequacy that teachers experience as they shift their instructional practice (for an example of this in South Africa see Masterson, 2013).

It is recognised, however, that for instructional coaching to be effective, the coaches need to be carefully selected, thoroughly trained, and their role needs to be very clearly defined. There is evidence that in the absence of clearly defined roles, coaches increasingly take on administrative functions like distributing materials and collecting information (Fullan & Knight, 2011). Coaches that do not receive adequate initial training; insufficient ongoing support as well as careful monitoring and evaluation are unlikely to make an impact on instruction (Gallucci, 2010). Furthermore, Loyalka et al. (2017) provide some cautionary results by showing that professional development with follow-up visits can lead to a null result if the training is too theoretical, or too rote and passive to be useful.

3. Experimental Design and Data

To evaluate new ways in which to assist and support teachers to teach early grade reading more effectively, EGRS II was designed to evaluate the cost-effectiveness of two different ways of coaching and supporting teachers. The premise of the two interventions was to change teachers' instructional practice through the use of daily lesson plans, training and coaching, and the provision of integrated educational material. The two strategies to training and coaching which were evaluated entailed: (i) the traditional centralised training, complemented with regular on-site coaching, and (ii) the traditional centralised training, combined with an information and communication technology (ICT) component that includes tablets provided to teachers and a virtual coach which provides on-going support through instant messaging and phone calls to the teacher.

The cost effectiveness of the two interventions, as well as the mechanisms of change, is evaluated through a mixed-methods approach. The RCT design allows us to determine the relative effectiveness of the interventions through estimates of the impact on learner outputs. To supplement these results, a large-scale classroom observation study and case studies were conducted and provides further insights about the mechanisms of change.

a. Study site and school sampling

Relative to the other provinces in South Africa, Mpumalanga is a mid- to low- performing province. In the 2016 final school leaving examination, Mpumalanga ranked fifth out of the nine provinces. Unfortunately, South Africa does not currently have a standardised assessment with which to compare schools, districts and provinces in the primary phase, but various other indicators can provide an indication of the current schooling conditions. The 2016 General Household Survey reports that 94.3% of 5 to 6 year olds in Mpumalanga are currently attending an educational institution and 91.4% of Grade 1 learners reported having attended Grade R prior to Grade 1. Similarly, 98.8% of learners of a compulsory school going age (7 to 15 years old) are currently attending an educational institution. Poverty analysis shows that 28.4% of learners attending schools in Mpumalanga fall below the food

poverty line (monthly per capita income is below $30.02[3]) and 47% below the Lower Bound Poverty Line (monthly per capita income is below $44.83).

Within this context, a sampling frame of 180 eligible schools were developed. Schools were considered eligible if they were in two of the four districts (Ehlanzeni or Gert Sibande), if they were Quintile 1 – 3 schools[4] and if their language of learning and teaching was either Siswati, or isiZulu. We also excluded schools which were particularly small schools (fewer than 30 Grade 1 enrolments) since many of these schools would practice multi-grade teaching, rendering the grade-specific lesson plans less appropriate. We also excluded particularly large schools to limit intervention costs.

To increase power and assure balance between intervention arms, stratified randomisation was used in the sample selection. Ten strata of 19 similar schools were created based on school size, socio-economic status, and previous performance in the Annual National Assessments. Within each stratum, five schools were randomly assigned to each intervention group, eight to the control group and one as a replacement school. Thus 50 schools were randomly assigned to each intervention and 80 to the control. Given that we aim to 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 an intervention group with the control group and a minimum effect size of 0.23 standard deviations when comparing two intervention 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.

The classroom observations were conducted in 60 of the schools in the sample and entailed the observation of the English lesson in 20 schools in each intervention group and the control group. In each of the schools, three different types of evidence were collected: (1) information on the lessons observed; (2) the evidence of work done in learners' work or exercise books, as well as the review of various teaching documents and; (3) the information from the teacher based on an interview.[5] The classroom observations therefore allows for the analysis of richer data regarding the quality of teaching practices, curriculum coverage and teacher perceptions about the support they receive.

Furthermore, six case studies were conducted in two schools in each of the intervention groups and the control group. The case studies provide further in-depth analysis of mechanisms of change in teacher instructional practice, teacher-learner interactions and the teachers' command of the content of the lessons taught. The evidence from the three different methods of data collection will allow us to consider the main research question through three different lenses.

[3] Using 2016 average Rand – US Dollar exchange rate of R14.72 per dollar.

[4] The Quintile system in South Africa determines the funding that schools are eligible for. Quintile 1 schools are situated in the poorest areas, whereas Quintile 5 schools are situated in the wealthiest neighbourhoods. Quintile 1 – 3 schools are classified as no-fee schools, and are not allowed to charge learners any school fees. Quintile 1 – 3 schools are also widely regarded as the worst performing schools in the system.

[5] Three different instruments were employed to collect the information: (1) a lesson observation instrument; (2) a document review instrument and; (3) a teacher interview instrument.

b. Intervention Design

All teachers in the intervention schools received the same lesson plan content which was distributed either as printed lesson plans (intervention 1) or lesson plans accessible on an application on a tablet (intervention 2). The structured lesson plans are based on the South African Foundation Phase curriculum (Curriculum and Assessment Policy Statement) and provide activities (timed to the minute) which enable teachers to fulfil the requirements of the English as second language curriculum in their classrooms. The daily lesson plans are intended to provide specification of the new instructional practice including faster paced instruction, more appropriately sequenced content, and dramatically expanded pedagogic repertoires. The daily 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, lesson plans free up teachers' time, because they no longer need to allocate as much time to planning which could lead to improved reading acquisition should teachers allocate this time to productive teaching activities.

All teachers in the intervention schools also received educational material which was integrated into the lesson plans. For the most part, these resources were the same for all teachers except that intervention 2 teachers had access to method videos and sound clips via the application on the tablet. 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. It is recognized in South Africa that the opportunity to learn English 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 from both interventions also attended centralized training at the start of the school year. This training lasted two days for intervention 1 and three days for intervention 2 (because of the need to orient teachers to the tablet). In addition, all teachers attended cluster training at the start of each of the other three school terms. The focus of teacher training was to explain and revise the core teaching methodologies needed to implement the lesson plans, as well as orientation to the weekly routine. Training sessions for intervention 2 teachers also included demonstrations of how to use the tablet, and troubleshooting for tablets when they did not work optimally.

In addition to the above, coaching is provided as more intensive support to improve teacher capacity. In essence, the function of coaching is to fuse capacity building and professional accountability. 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 coach also plays the role of a ‘champion’ who keeps the teacher accountable to implementing the programme through encouragement and motivation.

Intervention 1 made use of expert reading coaches to work face-to-face with teachers in improving teacher content knowledge, instructional practice and professional confidence. Three reading coaches were employed to support the 50 schools in total, which resulted in a teacher to coach ratio of about 32 teachers in 16 schools per coach. The three reading coaches were each based in a specific area of

the province and visited each teacher in class at least three times a term. The following activities took place during the coaching visits for each teacher: lesson demonstrations, lesson observations, assessment support and discussions regarding practice. During classroom observations, coaches provided feedback on the lesson and also demonstrated the pedagogy outlined in the lesson plans in the teacher's own class. The in-class support therefore 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 therefore 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 lesson plans throughout the course of the year.

Intervention 2 entails one reading coach, based in an off-site office, coaching the teachers via instant messaging, one-on-one text messaging and telephone calls. The reading coach followed up on training focus areas with individual teachers every fortnight to improve their instructional practice in teaching English and to ensure they were covering the curriculum. The virtual coach also provided motivational support to teachers via a group message once a week. Lastly, the virtual coach sent YouTube clips and videos as additional resources for the teaching areas teachers found particularly difficult. In addition to virtual coaching, teachers in this intervention group received their lessons plans on a tablet which is accessed via an application. 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 the teachers. 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 English. Similar to intervention 1, the on-going support from the virtual coach intends to encourage the teacher to keep up with the increased pace of the lesson plans throughout the course of the year. However, given that the virtual coach will not be in the classrooms, the role of the coach focuses more on encouragement and less on direct monitoring.

Table 1: Intervention Design

	On-site coaching	**Virtual Coaching**
Lesson plans	Paper based	Electronic: Tablet
LTSM	Paper based	Paper based Includes videos, sounds clips and photos of example writing
Coaching	In classroom visit Once every three weeks.	Via cell telephone calls and instant messaging. Once every two weeks.
Training	Initial training: 2 day block training Quarterly training: 1 day at the start of each term Needs-based cluster training	Initial training: 3 day block training Quarterly training: 1 day at the start of each term Additional videos developed as required

c. Instruments and data collection

The sampling strategy resulted in 3,327 Grade 1 learners being tested at the start of their schooling career. These same learners were tracked to the end of Grade 1 when they were assessed again to evaluate the impact of the interventions after one year of implementation. Both the baseline and midline learner assessments were designed in collaboration with language test development experts and were designed on the premise that certain basic literacy skills acquired in a first language are transferred to a second language. The test therefore comprised of both home language (in this case isiZulu and siSwati) and English items.

Given that the baseline data collection was conducted at the start of Grade 1, the baseline test was designed to test learners' home language literacy skills at the start of Grade 1, as well as skills that are predictive of future literacy development. Furthermore, since the interventions are targeting English as First Additional Language as an outcome, the test included a few basic English vocabulary items to determine learners' familiarity with the language at the start of school. The baseline test included various sub-tasks including word recall, non-word recall, phonemic awareness, letter sound recognition, word recognition, listening comprehension and sentence reading in the learners' home language, as well as some English vocabulary items.

The midline test was administered at the end of Grade 1 and was designed to test the literacy skills learners are expected to have mastered in Grade 1. The South African curriculum is based on the additive bilingualism approach, which means that learners in Grade 1 were expected to have been taught decoding skills in their home language, whereas the focus in English was on mastering basic English language comprehension skills. The midline assessment assessed both these skills, although the expectation would be that the intervention will only directly impact on the English language proficiency skills. The inclusion of the Home Language items allows for the testing of any positive or negative spill over effects to the other subject areas taught.

Both the baseline and midline assessments were designed to be orally administered by the fieldworkers, and to be captured electronically on the Tangerine software. In order to test the targeted 20 learners within one school day, the test was designed to take no longer than 15 minutes to administer. Both tests were originally designed in English and then versioned to SiSwati and isiZulu. In the versioning, specific care was taken to use words which are similar in the two languages, to minimise any bias that can be introduced through the language use in the assessment. During both rounds of data collection contextual questionnaires were administered to the Grade 1 teachers, the school principals and to the parent or care giver.

d. Sample Characteristics

The sample population more or less mirrored the large school population in the province. The average age of learners in the study was 6.7 years in grade 1 and 28.6% of learners reported themselves as first language isiZulu speakers, with the other learners reporting themselves as SiSwati speakers. With regard to parent or care-giver education, 56% of learners are from household where the responding parent did not complete any secondary schooling, and a further 23% only completed secondary schooling, but have not pursued any further education. The parents were also asked some questions

on how regularly they speak English to their child. 60% of the parents in this sample reported that they sometimes speak English to their child, but 27.5% reported that they never speak English to their child.

As with the country as a whole, the teachers in the study were relatively mature with an average age of close to 50 years, with over a quarter of all teachers over 55 years old. All were female teachers. Given the ICT focus of intervention 2, all teachers were asked about the technology devices they have in their household, as well as how regularly they use certain technology features. It was interesting to note that 72% of teachers owned a smartphone and 54% owned a computer. However, only 44% of teachers made use of email, only 28% accessed educational resources on the internet and only 25% made use of Microsoft Word.

More than half the schools were classified as Quintile 1 schools (poorest), with 32.8% in Quintile 2 and only 13.5% Quintile 3. The majority of the principals in the survey described the location of the schools as remote rural, with a small number in small village, township or informal settlement in a city.

Table 2: Sample Characteristics

		Control	**On-site Coaching**		**Virtual Coaching**		**N**	**On-site vs Virtual**
		Mean	*Coeff*	*s.e.*	*Coeff*	*s.e.*		*P-value*
Learner Characteristics	Learner = Boy	0.522	0.020	(0.019)	0.026	(0.022)	3,040	0.820
	Learner Age	6.658	-0.004	(0.048)	0.004	(0.046)	3,039	0.885
	Learner Language = isiZulu	0.296	-0.038	(0.052)	-0.023	(0.054)	3,327	0.809
	Baseline Score	-0.020	-0.016	(0.067)	0.113*	(0.068)	3,327	0.075
Teacher Characteristics	Teacher Age	49.58	0.815	(1.144)	-0.252	(1.090)	306	0.372
	Older than 55	0.241	0.108*	(0.064)	0.000	(0.060)	306	0.130
	Teacher Language = Zulu	0.299	-0.02	(0.059)	-0.021	(0.052)	306	0.984
	Teacher Language = SiSwati	0.672	0.025	(0.056)	0.019	(0.052)	306	0.918
	Multi-grade Classroom	0.044	0.025	(0.034)	0.002	(0.035)	306	0.542
	Class Size	42.75	-2.368	(1.531)	-4.007***	(1.476)	306	0.317
	Teacher have at least a bachelor's degree	0.547	0.006	(0.077)	-0.019	(0.071)	306	0.779
	Teacher English Proficiency	3.821	-0.083	(0.284)	0.436	(0.289)	268	0.112
Principal and School Characteristics	Principal Age	52.55	-2.338**	(0.948)	-1.372	(0.944)	180	0.360
	Older than 55	0.375	-0.177**	(0.081)	-0.034	(0.080)	180	0.112
	Gr 1 teacher vacancies	0.063	-0.021	(0.042)	0.005	(0.042)	180	0.590
	Quintile 1 school	0.537	-0.014	(0.054)	0.001	(0.053)	180	0.801

e. Balance and attrition

Table 3 shows the results of balance tests on the various sub-tests which were conducted at the baseline. These tests evaluate whether the differences in learning outcomes among the different intervention and comparison groups are statistically significant. Each column in Table 3 is a separate regression run for each sub-task on intervention indicators, controlling for strata and district fixed

effects. It is evident that there are only two imbalances out of the ten tests run. Both these imbalances are on items which either has strong ceiling[6]- or floor[7]-effects.

Table 3: Baseline Balance Tests

	HL Vocabulary	HL Word Recall	HL Non-word Recall	HL PA	HL Compre-hension	HL Letters	HL Words	HL Sentence	HL Visual Perception	ESL Vocabulary
On-site coach	0.065	-0.008	-0.017	-0.035	-0.016	-0.033	-0.038	-0.072*	0.102	-0.034
	(0.064)	(0.051)	(0.059)	(0.067)	(0.055)	(0.097)	(0.100)	(0.042)	(0.074)	(0.064)
Virtual coach	0.166***	0.052	0.030	0.023	0.084	0.015	0.102	-0.044	0.118	0.008
	(0.063)	(0.051)	(0.084)	(0.073)	(0.054)	(0.096)	(0.121)	(0.041)	(0.076)	(0.055)
N	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327	3,327
R^2	0.095	0.009	0.011	0.015	0.012	0.019	0.044	0.006	0.045	0.037
P-value	0.158	0.245	0.573	0.448	0.0917	0.660	0.270	0.518	0.857	0.511

Note: Each column represents a separate regression on intervention dummies and stratification dummies. Standard errors are clustered at school level. * for p<.1; ** for p<.05; *** for p<.01. PA = Phonemic Awareness

Furthermore, a short English proficiency task was administered to teachers which allows a balance check on prior teacher ability. Although teacher English proficiency is not the focus of the interventions, it is likely that improved English proficiency might be the result of increased use of English during the lessons. Table 4 shows that the teachers in the virtual coaching intervention had a higher English proficiency than the other groups, although only weakly statistically significant. Taken together with the balance test on the sub-tasks, the randomisation appears to have been successful in creating balance in baseline teacher and learner performance across the intervention and comparison groups. Sensitivity analysis will be conducted to control for the imbalance on teacher proficiency.

Table 4: Balance test on Teacher Baseline English Proficiency

	Teacher Score
On-site coach	0.0254
	(0.169)
Virtual coach	0.312*
	(0.171)
N	271
R^2	0.043
P-value	0.136

Note: District and stratification dummies controlled for. Standard errors are clustered at school level. * for p<.1; ** for p<.05; *** for p<.01

Out of the 3,327 learners that were tested during the baseline assessment, 3,040 learners were retested again at the end of Grade 1, yielding an attrition rate of about 8.6%. Columns (1) and (2) in table 5 regresses intervention assignment on the probability of attriting, controlling for the

[6] The sub-test on Naming Animals
[7] The sub-test on Sentence Reading

stratification and district dummies. The results in Column (1) shows that there is no bias in attrition in any of the intervention groups. Additional learner characteristics have been included in the second regression and shows that attriters were more likely to have been the lower performing learners, but that no other learner observable were statistically significant.

Table 5: Factors related to the probability of attriting

	(1)	**(2)**
On-site coach	0.203	0.205
	(0.146)	(0.147)
Virtual coach	-0.002	0.012
	(0.089)	(0.090)
Learner = Boy		0.036
		(0.062)
Learner Age		0.028
		(0.021)
Learner test language = Zulu		0.013
		(0.153)
Baseline score		-0.076**
		(0.030)
Control Mean	0.077	0.077
N	3327	3325
Pseudo R^2	0.016	0.02
Paper=Tablet: p-value	0.18	0.21

Note: Probability of attriting regressed on intervention assignment, district and stratification dummies. Standard errors are clustered at school level. * for p<.1; ** for p<.05; *** for p<.01

f. Empirical strategy

The schools in our sample are randomly assigned to the three intervention groups (intervention 1, intervention 2 and control), and therefore 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 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 – specifically the baseline learner score, learner gender, learner age, learner Home Language, the district, and a measure of the socio-economic status of the community in which the school is located, and ε_{isb1} is the error term clustered at the school level.

Given the nature of learning a second language, the ratio of English to Home Language items increased from baseline data collection to midline data collection. We will therefore control separately for each domain of reading proficiency collected at each round of data collection: vocabulary, letter

recognition, working memory, phonological awareness, word recognition, paragraph reading and English comprehension. Controlling for each domain separately, rather than create an aggregate index of learning proficiency will increase statistical power, as well as to allow us to differentiate between the learning areas which were focused on during the Grade 1 year.

4. Findings

This section will present the findings of the main set of results by first considering the impact of the two interventions on learning outcomes. This will be followed by a discussion on the influence of the interventions on teacher instructional practice, teacher lesson preparation and instructional support to teachers by incorporating the evidence from the lesson observation study and the case studies.

a. Impact on learning

Table 6 reports the item distributions for each of the sub-tasks in both the baseline and midline assessment. The scores in the table have not been standardised and the averages should therefore be interpreted relative to the minimum and maximum values. Items 1 – 8 were assessed at baseline, whereas items 9 – 19 were assessed at midline. The table shows that in some of the sub-tasks there were floor effects (eg. baseline visual perception, midline English listening comprehension and English spelling), and some items had ceiling effects (eg. baseline non-word recall and midline expressive vocabulary). Overall, however both assessments seem to provide enough information to discriminate among learners at various parts of the performance distribution.

Table 6: Item distributions

	N	Mean	s.e.	p10	p25	p50	p75	p90	Min.	Max.
1. HL Naming the Animals	3327	7.30	0.04	4	6	7	9	10	0	12
2. HL Word Recall	3327	10.00	0.04	7	9	10	12	13	0	14
3. HL Nonword Recall	3327	4.21	0.02	3	4	5	5	5	0	5
4. HL Phoneme Isolation	3327	1.11	0.03	0	0	0	2	4	0	6
5. HL Story Comprehension	3327	2.20	0.02	1	1	2	3	4	0	4
6. HL Letter Sound Recognition	3327	6.94	0.13	0	2	5	9	18	0	30
7. HL Visual Perception	3327	1.55	0.03	0	0	1	3	4	0	10
8. ESL Vocabulary	3327	0.82	0.02	0	0	0	1	2	0	6
9. HL Expressive Vocabulary	3039	4.90	0.01	5	5	5	5	5	0	5
10. HL Words Correct	3038	5.05	0.12	0	0	1	11	17	0	18
11. ESL Phoneme Isolation: First	3038	1.72	0.02	0	0	2	3	3	0	3
12. ESL Phoneme Isolation: Last	3038	1.35	0.02	0	0	1	3	3	0	3
13. ESL Phoneme Isolation: Middle	3037	0.56	0.02	0	0	0	1	3	0	3
14. ESL Letters Correct	3040	16.66	0.31	0	3	11	26	40	0	80
15. ESL Words Correct	3035	5.02	0.13	0	0	2	6	16	0	36
16. ESL Listening	3035	2.91	0.02	1	2	3	4	4	0	5
17. ESL Listening Comprehension	3035	0.22	0.01	0	0	0	0	1	0	3
18. ESL Vocabulary	3033	0.73	0.02	0	0	0	1	2	0	2
19. ESL Writing	3029	0.24	0.01	0	0	0	0	1	0	1

The Grade 1 English curriculum mostly focusses on oral proficiency and language comprehension, with 50% of the time allocation in the curriculum being dedicated to only listening and speaking activities. A further 40% of the time is allocated to reading activities, but at Grade 1 this does not yet constitute individual reading, but rather entails the teacher reading stories to the learners. Given this, it is therefore to be expected that differences between the intervention groups should be evident in the learner results on the English listening and vocabulary tasks. As a first exploratory exercise, figures 1 and 2 shows the differences in learner performance by intervention group and suggests that learners from the two intervention groups performed consistently better than the control learners on each of these tasks.

Figure 1: Learner results on the English Listening Task

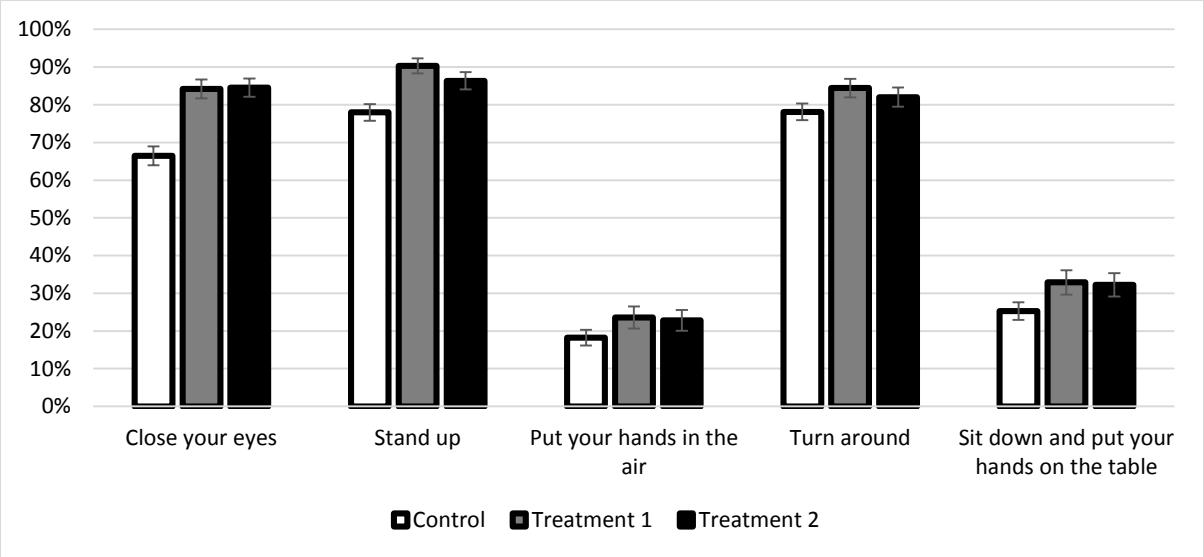
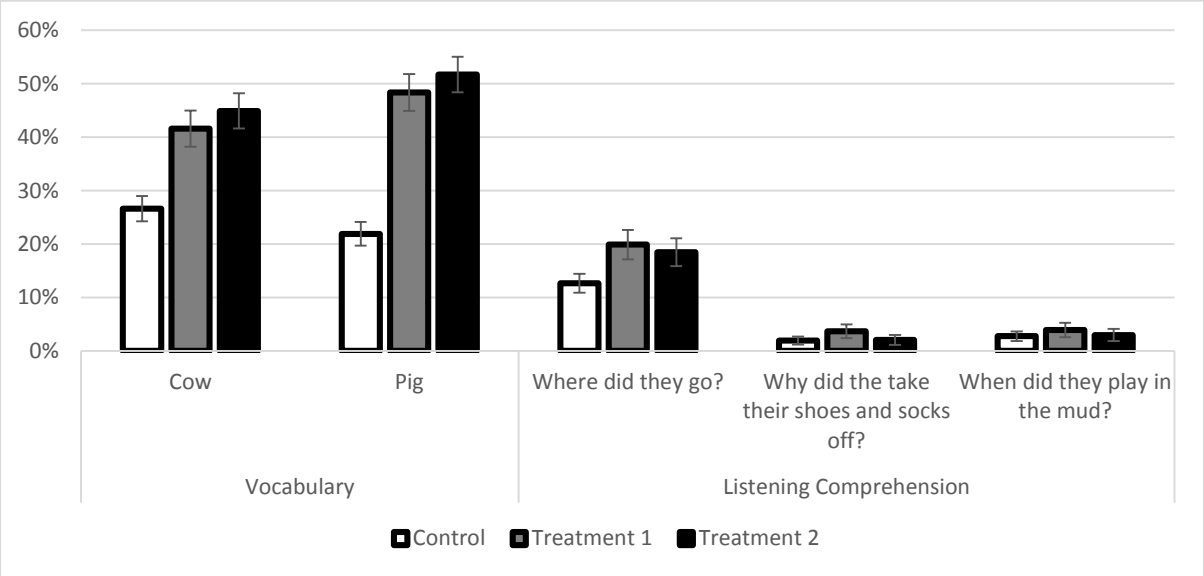


Figure 2: Learner results on the English Vocabulary and Listening Comprehension



The first major findings of the study relate to the relative performance of the three groups on the midline learner test. Separate regressions were run on each sub-test, and for the sake of comparability the scores were standardised to a mean of zero and a standard deviation of 1. The skills in columns 1 and 2 relate to home language proficiency and decoding skills, the skills in columns 3, 4 and 5 relate to English decoding skills, the skills in columns 6, 7 and 8 indicates English language proficiency skills and finally column 9 is English spelling. We see from columns 6, 7 and 8 that both interventions seem to have had an equally large impact on the various English language proficiency skills. The coefficient for the on-site coaching intervention is negative for the home language expressive vocabulary task, but given the ceiling-effects in this item, the concern of spillover effects is not too great. The coefficient for the virtual-coaching intervention is negative for home language word recognition, English letter recognition and English phonemic awareness but is weakly statistically significant. Given that the learners would have been taught decoding skills in their home language, and that the skill of decoding is considered transferable between the languages, this might suggest a negative spillover effect to the teaching of home language. In future rounds of data collection this is something which will be evaluated rigorously, if the problem persists it will warrant stronger claims on negative spill-over effects.

Table 7: Main Results

	HL Vocabulary	HL Words	ESL PA	ESL Letters	ESL Words	ESL Listening	ESL Compre-hension	ESL Vocabulary	ESL Writing
On-site Coach	-0.122**	-0.139	-0.079	-0.052	0.001	0.375***	0.190**	0.518***	-0.015
	(0.055)	(0.087)	(0.092)	(0.088)	(0.079)	(0.083)	(0.075)	(0.073)	(0.077)
Virtual Coach	0.016	-0.162*	-0.170*	-0.180**	-0.103	0.291***	0.105	0.547***	-0.082
	(0.056)	(0.083)	(0.094)	(0.077)	(0.069)	(0.083)	(0.065)	(0.073)	(0.066)
N	3,038	3,037	3,037	3,039	3,034	3,034	3,034	3,032	3,028
R^2	0.021	0.140	0.132	0.140	0.110	0.149	0.057	0.208	0.113
P-value	0.0255	0.798	0.371	0.139	0.202	0.343	0.314	0.704	0.352

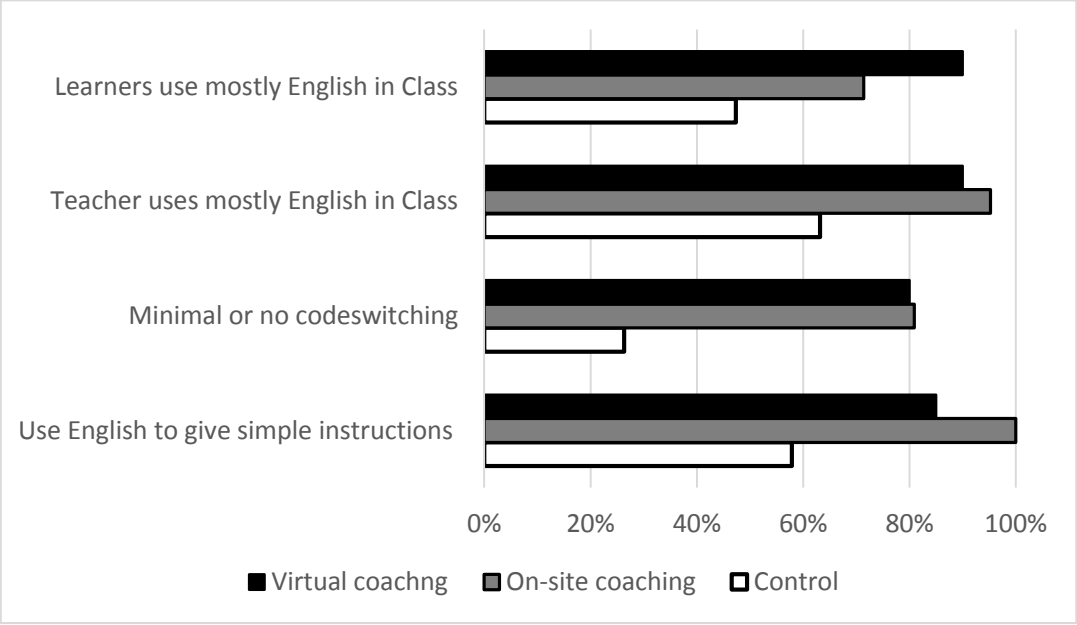
Note: Learner gender, baseline score, learner age, learner home language, district and stratification dummies controlled for. Standard errors are clustered at school level. * for p<.1; ** for p<.05; *** for p<.01

b. Teacher instructional practice change

A lesson observation study was conducted on a sub-sample of 60 schools participating in the EGRS II study to provide a more qualitative perspective on the intermediate outcomes of the two interventions. The results from these observations are suggestive rather than definitive of changes in teachers' instructional practice due to the interventions. In addition to the lesson observations, an education specialist did six case studies in a further sub-sample of schools to provide an educationalists perspective of the difference between the schools in the intervention and comparison schools. To ensure integrity in the reporting of the qualitative results, differences in instructional practices are only reported if observed in both the lesson observation study and the case studies. Three main differences were observed between the intervention and control classrooms. Firstly, the frequency of the use of English by both the learners and the teachers. Secondly, the systematic teaching of oral language proficiency in English using extended texts. Finally, the more efficient use of time which allows for greater learner exposure to the English curriculum.

Given the South African context of teachers who are not proficient or confident in speaking English being responsible for teaching learners English, one major feat was the increased use of English during the teaching of the English lesson. The observations in the control classrooms attest to English lesson being taught in the home language, with limited use of English during these lessons. In only 63% of the control classrooms observed, English was the main language spoken by the teacher, and in only 23% of the control classrooms teachers made limited use of the home language to explain the English lesson. It is evident from figure 1 that learners and teachers were more likely to use English during the English lesson, whether to answer questions, give simple instruction or in explaining the lesson.

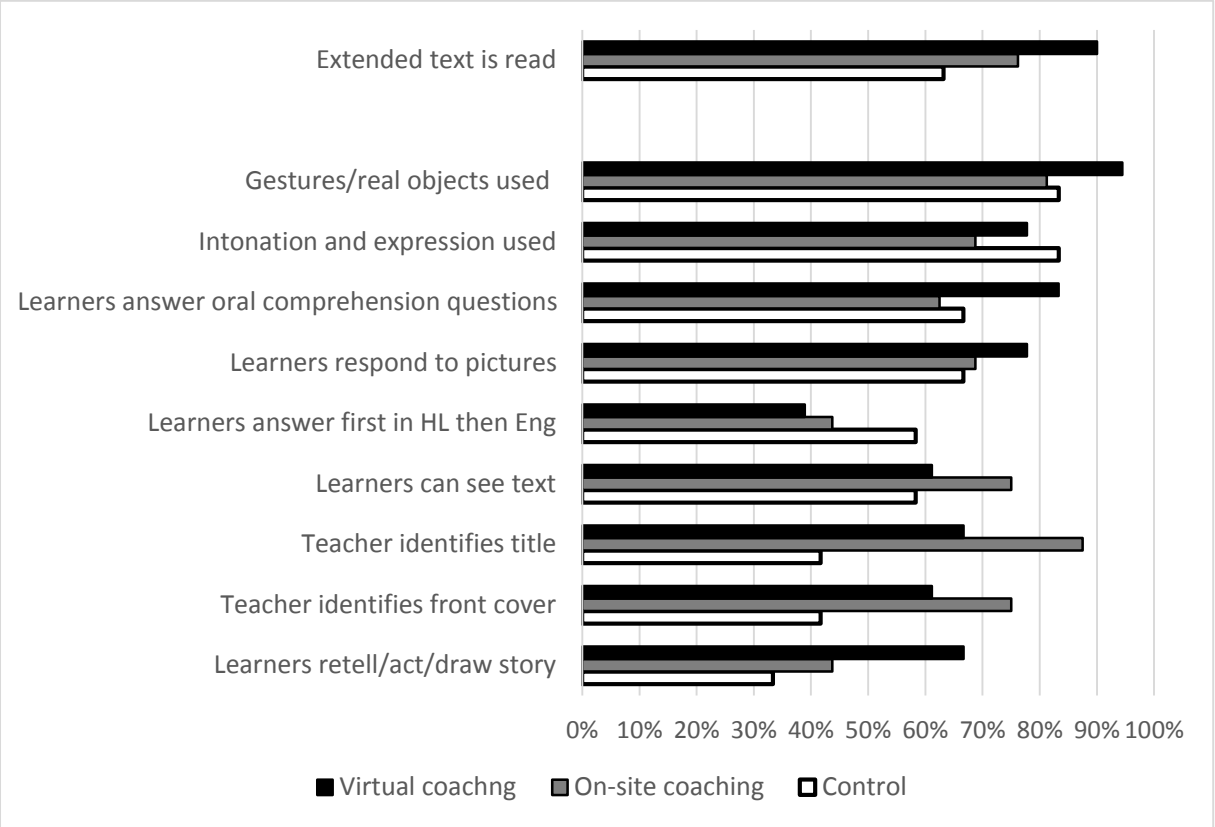
Figure 3: Frequency of English usage during the English lesson



In the first year of formal the structured lesson plans introduces English oral language proficiency and language comprehension through systematic vocabulary development and shared reading practices. Although teachers in the control schools on average introduced more vocabulary words on a given day, these words were often only repeated by the learners without much attention given to understanding the meaning. In the intervention schools teachers were more often observed using the new vocabulary words in the context of a sentence to promote understanding the meaning of new words. Differences were also seen in the shared reading practices, with 90% of the intervention 2 teachers and 76% of the intervention 1 teachers having used extended texts during their lessons, relative to only 63% of the control schools. Of the teachers that made use of extended text during their lessons, fewer intervention 1 teachers used chorusing or reading aloud together with learners and were thus more likely to be modeling fluent reading for learners as recommended in the EGRS II shared reading methodology. This finding suggests that intervention 1 teachers were working successfully with their coaches to improve their instructional practice with regards to shared reading. Shared reading should also be accompanied by teachers asking learners questions about the story to teach them English oral proficiency and to increase their English vocabulary. Intervention 2 teachers were more likely to ask their learners comprehension questions based on the story as compared to either intervention 1 or control teachers. Further differences in groups arose with reference to getting learners to retell, act out or draw in response to the story with more intervention 2 teachers (67%) getting learners involved in retelling, acting out or drawing responses to stories than control (33%) or

intervention 1 (44%) teachers. The interventions therefore seem to be expanding the teachers' pedagogic repertoires through the systematic teaching of English vocabulary and English language proficiency. The greater exposure to vocabulary development and language development through the shared reading of extended texts, should therefore lead to learners having a stronger command over the English language.

Figure 4: Instructional practices on Shared Reading



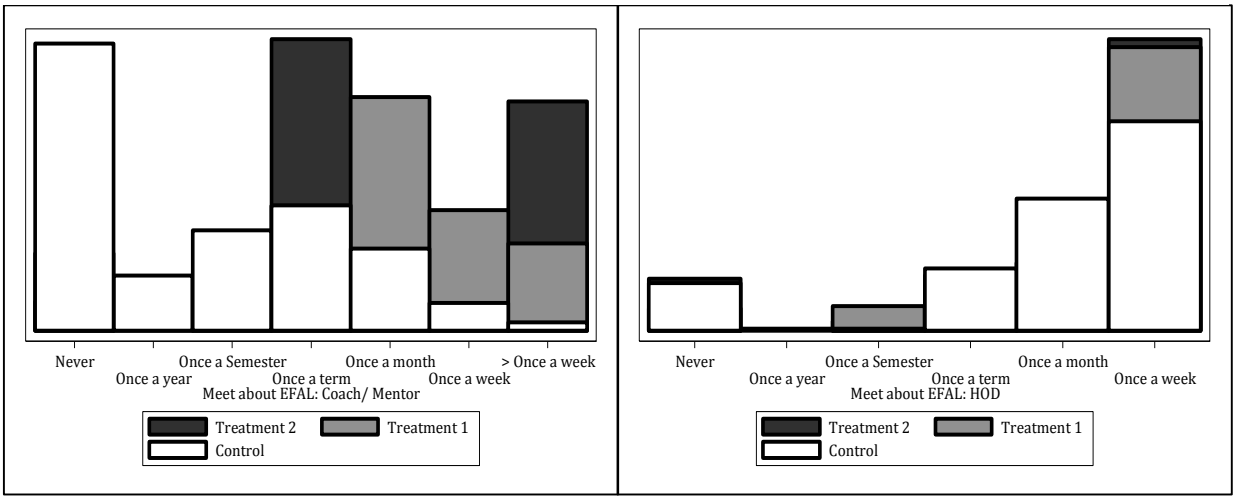
Finally, teachers in the intervention school were less likely to have a delay in starting their lesson and were also less likely to have teaching time disrupted due to handing out of books. Furthermore, teachers in the intervention schools were also more likely to have been tracking their implementation of the curriculum. Overall, 30% of the teachers observed did not stick to the time allocated for the lesson, nevertheless, intervention teachers were more likely to cover the work as planned their lesson plan than control teachers. Intervention teachers were also less likely to cut a lesson short. The majority of intervention teachers thus seemed to be spending at least the minimum amount of time on English lessons. This signifies that although the lesson plans and coaching did not appear to help teachers adhere to the time allocated for their lessons, the lesson plans/coaching do seem to have enabled teachers to cover the work as planned. Findings therefore suggest that both intervention 1 and 2 teachers are covering more of the prescribed curriculum and that their learners are thus receiving greater exposure to the English curriculum.

c. Change in teaching preparation and teacher support

It is well recognised that foundation phase teachers lack on-going instructional support in South Africa. The 2011 School Monitoring Survey indicated that foundation phase teacher received less monitoring and support from the school principal, school management team and the governance structures than their secondary counterparts (Wills, 2016). Only 45% of the 4128 teachers interviewed in the survey indicated that they had been visited by a district official responsible for curriculum advice during the year (Wills, 2016). In the districts that the EGRS II study is participating, the curriculum advisor to school ratio is around 130 school per advisor. This means that teachers on average receive a visit by the curriculum advisors once every two years. Heads of Department (HODs) in schools are further responsible for providing instructional support to teachers. However, a recent evaluation of the implementation of the curriculum found that full teaching loads of HODs, as well as inappropriate promotion practices for appointing HODs, often means that HODs do not have the capacity or ability to provide effective support to teachers (Department of Planning, Monitoring and Evaluation/ Department of Basic Education, 2017).

The presence of the in-class support provided through both the on-site coaches and the virtual coach in the EGRS II study, therefore fills a critical gap in providing instructional support. The coaches promote 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 coach to school ratio for the on-site coaches is designed to allow each teacher to be visited by a coach at least three times a term. Moveover, by removing the barrier of needing to be physically in the classroom, the virtual coach communicates with teachers on weekly basis. Figure 5 suggests that teachers in both intervention groups are much more likely to have met with a coach or mentor regarding their English teaching practices than the control group, with 42% of control group teachers indicating that they never meet with a coach or mentor regarding their English teaching. Furthermore, when asked how often teachers meet with their HOD regarding their English teaching, teachers from both intervention groups were also meeting with their HODs more regularly, with 66% and 63% of intervention 1 and intervention 2 teachers respectively mentioning that they meet with the HODs at least once a week, but only 50% of control group teachers meeting with their HOD weekly. This indicates that coaching is not only providing more instructional support to teachers directly

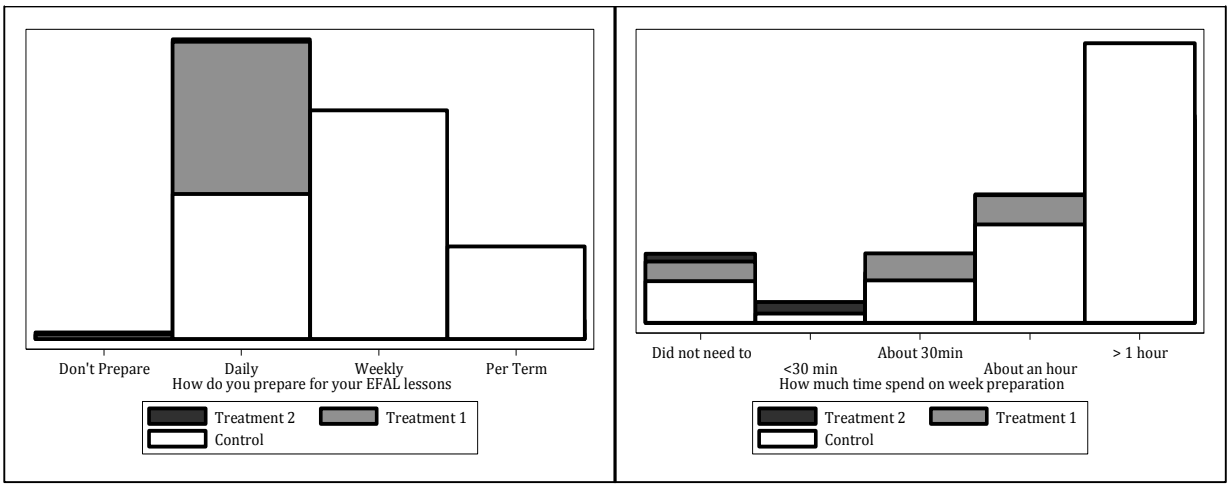
Figure 5: Instructional support



through their presence, but also indirectly through encouraging the HODs to support teachers more regularly.

With regard to lesson preparation, the theory of change of using daily lesson plans postulates that little additional lesson preparation is required from teachers, which should make the take-up of a more productive teaching practice more manageable. The use of lesson plans should therefore free up teachers' time, because they no longer need to allocate as much time to preparation. This in turn could improve reading acquisition if teachers allocate this time to productive teaching activities, rather than leisure or unproductive teaching activities. Figure 6 shows the lesson preparation practices of teachers in each of the intervention groups and indicates that teachers in the two intervention groups are much more likely to prepare on a daily basis, whereas control group teachers tend to rather prepare on a weekly basis. Furthermore, when asked how much time teacher spend their preparation, 57% of the control group teachers responded that they spend more than an hour, whereas the intervention group teachers were more likely to have responded to options of an hour or less.

Figure 6: Teacher Preparation Time



5. Discussion

The results of the EGRS I study confirmed that the coaching model of delivering teacher training is more effective than the centralized training model. Building on the experience of EGRS 1, the current study investigates the sustainability question of taking the coaching model to scale. The evidence after one year of implementation suggests that the on-site coaching and the virtual coaching interventions are equally effective in improving learner English oral proficiency. The resources involved in each of the interventions would therefore determine whether the virtual coaching model may be a more sustainable method of taking coaching to scale.

For cost estimates, the programme budget for the first year of implementation was taken, excluding any costs that were involved in the development of the programme. These estimates should therefore provide a realistic per-learner cost if these models of delivery are scaled-up. Based on these estimates the per learner cost of the on-site coaching and virtual coaching models do not differ dramatically,

and are 48 USD and 41 USD respectively per year.[8] However, a critical resource in both the on-site coaching and virtual coaching model is expert reading coaches. To support the 98 Grade 1 teachers in the 50 intervention 1 schools, three specialist reading coaches were employed, whereas one virtual coach was employed in Intervention 2 to support the a similar number Grade 1 teachers in 50 intervention 2 schools. The availability of expert reading coaches in each of the country's eleven home languages is therefore a further resource constraint that will need to be taken into account in making decisions regarding the feasibility of taking the coaching model to scale.

The basic finding from the midline evaluation, reflecting impact after one year of implementation, is that learners in the two intervention groups are significantly ahead of their control group counterparts on the core instructional outcomes associated with the Grade 1 curriculum for learning English as an additional language. The Curriculum and Assessment Policy Statement (CAPS) specifies that a maximum of three hours per week be given to the teaching of English (as the additional language) in Grade 1, with a maximum of one and half hours for listening and speaking, one and a quarter hours for reading and phonics, and 15 minutes for writing. The scope of learning in the curriculum for Grade 1 English learning is oral language development, with reading and phonics being addressed through methods such as shared reading, listening to stories and total physical response. All three of these practices essentially focus on receptive language proficiency and to a lesser extent on expressive language development practices.

For the three subtests that focus on English oral language development - English Listening (and following instruction with actions), English Listening Comprehension, and English vocabulary - the learners in the two coaching interventions did significantly better than their counterparts in the control schools. On the higher order skill sub-test that assessed children comprehension in English, learners in the intervention classrooms did only marginally better than children in the control schools, suggesting that while vocabulary development is stronger as a result of the interventions, this has not yet translated into stronger comprehension skills. That said, if oral vocabulary development in the second language is a developmental building block for reading acquisition, then the interventions may be working to improve the basic skills targeted in the grade 1 curriculum but are not yet impacting the higher order skills.

The lack of positive impact (and even weak indications of some negative effects) on home language sub-tests and on decoding outcomes (which is really only targeted in the home language curriculum) raises the possibility that the English intervention may be successful at the cost of crowding out teaching time in the other areas of the curriculum. That is, could the special attention given to English through the interventions have lead teachers to spend more time on English as a subject and less on Home Language, Mathematics and Life Skills, the three other subjects in the official curriculum? This matter will be closely evaluated in the next round of data collection to provide further information on this.

[8] The cost for implementing the on-site coaching in 50 schools totalled 182,920 USD, whereas the cost for implementing the virtual coaching programme totalled 157,152 USD. Using the average class size of 76.8 learners per school at the start of the programme results in per learner spending of 48 USD and 41 USD respectively.

The results on the other learning areas are very tentative at this stage and warrant further investigation through subsequent rounds of data collection. However, it should be noted that in EGRS 1 we found no evidence of negative effects of the home language coaching intervention on mathematics and even found a positive effect on English. Putting the two studies together, if it turns out that home language interventions have spillover benefits on second language acquisition but not the other way around, this would point to home language interventions being a more cost-effective policy response than English interventions, at least in the first three grades of school.

The large-sample classroom observation data provide important information about the mechanisms of change. The core question about the relative efficacy of the two versions of coaching can be considered through the lens of Guskey's theory of professional learning (Guskey, 2002). According to Guskey, it is practice or 'doing the work' that is the precondition to attitudinal and belief changes. The evidence from the large-scale observations suggest that both versions of coaching help teachers to implement the essential activities prescribed in the lesson plans. All teachers in the virtual coaching group were either mostly or completely following the lesson plans, fourteen of the twenty in the on-site coaching group of teachers were mostly or completely following the lesson plans. In contrast only about a third of teachers in the control were following their lesson plans. This suggests that the interventions were having a significant impact on the move from planning to execution in the lessons, meaning that learners were far more likely to experience a systematic learning experience across lessons if they were in classrooms with teachers in either intervention group.

The other major indicator of the impact of the interventions on teacher practice is the use of the English language. Far fewer control classrooms were observed where a substantial portion of the learners talked to the teachers in English. It was observed that the teachers in the intervention schools gave most of their instructions in English and hardly every used code-switching in their explanations. In the virtual coaching classrooms and the classrooms receiving on-site coaching, English was primarily used by both the teachers and the learners in the Grade 1 English lesson. While using English is not a guarantee that learners will develop deep comprehension skills, without frequent and meaningful use of the language in the lessons, substantial progress would be difficult for most learners. Both implementing the planned lessons and using the language of the subject being taught, are key ways in which the interventions appear to be shifting instructional practice, with early signs of impact of learner results.

6. Conclusion

Research focussed on input-based programmes or on interventions that circumvent the basic problem of getting public schools to function better, has either yielded disappointing results or not produced sustainable solutions for system change. But in the past half-decade, we are seeing a consensus emerge, that interventions combining different elements around the instructional core or structured pedagogic programmes can shift learning outcomes across education systems. This is achieved through better teaching by public school teachers using the official curriculum.

The new consensus has allowed for a second generation of research questions to be asked, questions that allow researchers to drill down or unpack how and why, and under what conditions structured pedagogic programmes work. This paper is part of this research movement. It was animated by the problem of instructional coaching, as part of combined structured pedagogic programmes and its

relatively high cost. Reporting on the preliminary results at the end of the first year of the study, the findings confirm that combined approaches that include coaching do indeed improve learning outcomes at scale. Furthermore, there is emerging evidence that there may be less resource intensive forms of coaching that have similar outcomes. While it is too early for definitive findings, the EGRS II learning outcomes in English vocabulary and the patterns of classroom practice signal a new possibility for policy makers to consider.

References

Araujo, M., Carneiro, P., Cruz-Aguayo, Y. & Schady, N., 2016. Teacher quality and learning outcomes in kindergarten. *The Quarterly Journal of Economics,* 131(3), pp. 1415-1453.

Banerjee, A. et al., 2017. From proof of concept to scalable policies: Challenges and solutions, with an application. *Journal of Economic Perspectives,* 31(4), pp. 73-102.

Banerjee, A. et al., 2010. Pitfalls of participatory programs: Evidence from a randomized evaluation in education in India. *American Economic Journal: Economic Policy,* 2(1), pp. 1-30.

Bold, T. et al., 2017. What do teachers know and do? does it matter? evidence from primary schools in Africa.. *Policy Research Working Paper 7956, World Bank Group.*

Bruns, B. & Luque, J., 2014. *Great teachers: How to raise student learning in Latin America and the Caribbean.* Washington, D.C.: World Bank Publications.

Chauraya, M. & Brodie, K., 2017. Learning in Professional Learning Communties: Shifts in mathematics teachers'practices. *African Journal of Research in Mathematics, Science and Technology Education ,* 21(3), pp. 223-233.

Chetty, R., Friedman, J. & Rockoff, J., 2014. Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood.. *American Economic Review,* 104(9), pp. 2633-79.

City, E. A., Elmore, R. F., Fiarman, S. E. & Teitel, L., 2009`. *Instructional round in education: A network approach to improving teaching and learning,* Cambridge: Harvard Education Press.

Department of Planning, Monitoring and Evaluation/ Department of Basic Education, 2017. *Report on the Implementation of the National curriculum Statement Grade R to 12 Focusing on the Curriculum and Assessment Policy statements (CAPS),* Pretoria: Department of Planning, Monitoring and Evaluation/ Deparment of Basic Education.

Duflo, E., Dupas, P. & Kremer, M., 2012. *School Governance, Teacher Incentives, and Pupil-Teacher Ratios: Experimental Evidence from Kenyan Primary Schools*.

Fleisch, B., 2016. System-wide improvement at the instructional core: Changing reading teaching in South Africa. *Journal of Educational Change,* 17(4), pp. 437-451.

Fleisch, B., Taylor, S., Schöer, V. & Mabogoane, T., 2017. Failing to catch up in reading in the middle years: The findings of the impact evaluation of the Reading Catch-Up Programme in South Africa. *International Journal of educational Development,* Volume 53, pp. 36-47.

Fullan, M., 2007. *The new meaning of educational change.* Routledge.

Fullan, M. & Knight, J., 2011. Coaches as system leaders. *Educational leadership,* 69(2), pp. 50-53.

Gallucci, C., Van Lare, M., Yoon, I. & Boatright, B., 2010. Instructional coaching: Building theory about the role and organisational support for professional learning. *American educational research journal,* 47(4), pp. 919-963.

Guskey, T., 2002. Professional development and teacher change. *Teachers and teaching,* 8(3), pp. 381-391.

Hanushek, E. & Rivkin, S., 2010. Generalizations about using value-added measures of teacher quality. *American Economic Review,* 100(2), pp. 267-71.

Hargreaves, A., 2005. The emotions of teaching and educational change. In: *Extending educational change.* Dordrecht: Springer, pp. 278-295.

Hellman, L., 2012. *GPLMS Intersen Catchup Programme: Analysis of Results*.

Hoadley, U., 2010. *What do we know about teaching and learning in primary schools in South Africa?,* Stellenbosch: Appendix B to Van der Berg, S; Meyer, H; Reeves, C; van Wyk, C; Hoadley, U; Bot, M; & Armstrong, P 2010. 'Grade 3 Improvement Project: Main report and Recommendations" for Western Cape Education Department.

Kraft, M., Blazar, D. & Hogan, D., 2016. The effect of teacher coaching on instruction and achievement: A meta-analysis of the causal evidence. *Review of Education Research.*

Loyalka, P. et al., 2017. *Unpacking teacher professional development*.

Masterson, L., 2013. *A case study of foundation phase teachers' experiences of literacy coaching in the GPLMS Programme,* s.l.: (Doctoral Dissertation).

Piper, B., 2017. Improving learning in Sub-Saharan Africa using rigorous research designs. *Journal of Research on Educational Effecticeness,* 10(3), pp. 535-540.

Popova, A., Evans, D. & Arancibia, V., 2016. Training teachers on the job: what works and how to measure it. *Policy Research Working Paper, no WPS 7834; Washington, D.C : World Bank Group.*

Prinsloo, C., 2008. Practice makes perfect - Limpopo study shows the value of exercising literary muscles.. *HSRC Review,* 6(4), pp. 5-6.

Sailors, M. et al., 2014. Supporting change in literacy instruction in Malawi. *Reading Research Quarterly,* 49(2), pp. 209-231.

Shaheen, R., Walsh, C., Power, T. & Burton, S., 2013. *Assessing the impact of large0scale teacher professional development (TPD) in Bangladesh: English in Action (EIA).* San Fransisco, California, AERA online paper repository.

Sturtevant, E. & Linek, W., 2007. Secondary literacy coaching: A Macedonam perspective. *Journal of Adolescent and Adult Literacy,* pp. 51(3), 240-250.

Taylor, S., 2011. Uncovering indicators of effective school management in South Africa using the National School Effectiveness Study. *Stellenbosch Economic Working Papers 10/11,* pp. 1-51.

Wills, G., 2016. *Limited support for the foundation phase: A misallocation of district resources,* Stellenbosch: ReSEP Policy Brief. Stellenbosch Univeristy.

World Bank Group, 2018. *World Development Report 2018: Learning to Realize Education's Promise,* Washington, D.C.: World Bank.