The Second Early Grade Reading Study

Year 2 Report

Learner performance after the second year of implementation

Contents

List of Figures

List of Tables

Acronyms

ANA	Annual National Assessment
CAPS	Curriculum and Assessment Policy Statement
DBE	Department of Basic Education
EFAL	English as First Additional Language
EGRA	Early Grade Reading Assessment
EGRS	Early Grade Reading Study
EGRS II	Second Early Grade Reading Study
GPLMS	Gauteng Primary Literacy and Mathematics Strategy
HOD	Head of Department
ICT	Information Communication Technology
LoLT	Language of Learning and Teaching
LTSM	Learning and Teaching Support Material
NECT	National Education Collaboration Trust
PIRLS	Progress in International Reading and Literacy Study
RAN	Rapid Automised Naming
RCT	Randomised Control Trial
RCUP	Reading Catch-Up Programme
SMT	School Management Team
USAID	United States Agency for International Development
WCPM	Words Correct Per Minute

1. Background and Context

This report builds on the information provided in the Year 1 Report of the Second Early Grade Reading Study (EGRS II). Detailed information about the situation of reading in both the Home Language, English as the First Additional Language (EFAL), as well as prior research was provided in the Year 1 report and will not be repeated here. The focus of the Year 2 report is on the 2018 data collection processes and the main results after two years of implementation. This report provides information about the implementation of the interventions in Year 2 and the analysis of the learning outcomes at the end of Year 2. Details of the study site, the school selection process and the evaluation design is contained in the Year 1 report. A comprehensive report on the Study will be made available following the multi-method data collection, analysis and interpretation at the end of Year 3 (2019).

As a brief summary, the EGRS II is a Randomised Control Trial (RCT) that evaluates two early grade reading interventions in 180 primary schools in two districts in the province of Mpumalanga, South Africa. The EGRS II was first implemented with Grade 1 teachers in 2017 and in 2018 the interventions were targeted at the Grade 2 teachers. In 2019 the Grade 3 teachers will benefit from the interventions, and across all three years of the implementation, the study focuses on measuring the causal impact on learner reading performance and unpacking the change mechanisms of a structured pedagogic programme.

Intervention design

The EGRS II focuses on the early learning of English as a second language (officially named English as First Additional Language, or EFAL, in the South African curriculum) by providing specific resources, training and on-going coaching to teachers. The interventions being trialled are based on the official government curriculum, formally referred to as the National Curriculum and Assessment Policy Statement (CAPS). As such the interventions are designed to improve and strengthen teachers' enactment of the curriculum.

Both interventions consist of three components: (1) detailed lesson plans, (2) integrated learning and teaching support material and (3) instructional coaching and training. The main difference between the two intervention models is in the delivery model of the lesson plans and the coaching support. In Intervention 1, the teachers receive a paper-based version of the lesson plans and benefit from regular on-site coaching. In Intervention 2, the teachers receive a tablet with an electronic version of the lesson plans, including various audio-visual resources and are supported through an ICT coaching model that includes telephone calls and cell phone messaging. The electronic lesson plans are provided on an application which is specifically developed for the study. This application contains additional electronic resources

such as short training videos, sound clips of the phonics sounds, songs and rhymes and examples of learners' work.

In Intervention 1, teachers receive visits from specialist reading coaches about once a month. During these visits, coaches model, support and evaluate teachers' practices and monitor implementation fidelity. Coaching in Intervention 2 involves text messaging to create virtual reading coaching practices and virtual communities of practice. The virtual reading coach uses instant messaging to communicate with teachers frequently, providing them with teaching tips on a weekly basis, answering questions on the lessons and running bi-weekly competitions to see evidence of teachers' enactment of the lesson plans.

Table 1: Comparison of intervention 1 and intervention 2

	Intervention 1	Intervention 2
Provision of lesson plans	Paper-based	Electronic On an application on a tablet
Provision of LTSM	Paper-based: - Big books - Posters - Flashcards - Writing frames	Paper-based: - Big books - Posters - Flashcards - Writing frames
Coaching	Coach visits the teacher in her classroom. Once every three weeks.	Coach contacts the teacher via telephone calls and instant messaging (Whatsapp). Once every two weeks.
Training	Initial training: 2-day block training Quarterly training: 1 day at the start of each term Needs-based training: As required	Initial training: 3-day block training Quarterly training: 1 day at the start of each term Needs-based training: Bi-weekly competitions
Core methodologies	Paper-based instructional manual	Application-based instructions, Includes videos, sounds clips and photos of example writing

2. Implementation Fidelity

As a first step to checking the success of the interventions, it is useful to consider the quality of implementation. Successful implementation would entail teachers attending the teacher training sessions, as well as teachers using the methodologies and materials in their classrooms. Administrative data, as well as information from both the teacher questionnaire and the classroom observation, are used to evaluate the fidelity of implementation.

A range of teacher development opportunities and materials were provided to all intervention teachers in Year 2. These interventions focus on supporting the teachers in teaching the EFAL curriculum by focusing on the teaching of technical reading skills and on the acquisition of critical comprehension skills. These two central parts of reading were incorporated into the structured learning programme that covers all of the CAPS required language components so that teachers cover the entire curriculum. The intention of this is to ensure that teachers would not view the EGRS as an additional programme that sits outside their normal teaching requirements.

All intervention teachers were provided with the following paper-based resources:

- Alphabet Frieze
- Display Boards
- Handwriting poster
- Big Books
- Graded readers/anthologies
- Flashcard words

In addition, Intervention 1 teachers (those supported by physical coaches) were also provided with EFAL scripted lesson plans in booklets; while Intervention 2 teachers (those supported by a virtual coach) were provided with tablets containing digital lesson plans, methodological videos and phonetic audio clips.

The implementation focused on three main teacher development strategies: firstly, quarterly teacher training to provided teachers with the materials and pedagogical skills they required to teach the EFAL curriculum (centrally located at the start of each year and then in clusters); secondly, teacher support from experienced reading coaches either in a face-to-face instructional model in Intervention 1 or through the distance engagement from a virtual coach in Intervention 2; and thirdly, needs-driven school-based workshops in Intervention 1 and teacher engagement competitions in Intervention 2. In order to provide teachers with additional school-based support, school management members (HODs in particular) are also invited to teacher training sessions; though this has met limited success.

In 2018, ten teacher training sessions took place, with the tables below providing attendance data across the interventions per year.

Terms	Categories	Intervention 1 (92 Grade 2 teachers)		Intervention 2 (85 Grade 2 teachers)	
		Numbers	Percentages	Numbers	Percentages
Term 1	Teachers at training	85	92%	82	96%
	Male teachers at training	2	2%	0	0%
	Female teachers at training	83	98%	82	100%
		(92 Grade 2 teachers)		(82 Grade 2 teachers)	
Term 2	Teachers at training	86	93%	81	99%
	Male teachers at training	2	2%	0	0%
	Female teachers at training	84	98%	81	100%
		(88 Grade 2 teachers)		(81 Grade 2 teachers)	
Term 3	Teachers at training	87	99%	72	89%
	Male teachers at training	2	2%	0	0%
	Female teachers at training	85	98%	72	100%
		(88 Grade 2 teachers)		(80 Grade 2 teachers)	
Term 4	Teachers at training	85	97%	76	95%
	Male teachers at training	2	2%	0	0%
	Female teachers at training	83	98%	76	100%

Terms	Categories	Intervention 1 (50 SMTs)		Intervention 2 (49 SMTs)	
		Numbers	Percentages	Numbers	Percentages
Term 1	SMTs at training	45	90%	40	82%
	Male SMTs at training	6	13%	9	23%
	Female SMTs at training	39	87%	31	77%
Term 2	SMTs at training	50	100%	15	31%
	Male SMTs at training	6	12%	2	13%
	Female SMTs at training	44	88%	13	87%
Term 3	SMTs at training	30	60%	17	35%
	Male SMTs at training	4	13%	1	6%
	Female SMTs at training	26	87%	16	94%
Term 4	SMTs at training	32	64%	19	39%
	Male SMTs at training	4	8%	2	11%
	Female SMTs at training	28	92%	17	89%

3. Year 2 data collection

Details of the evaluation design including an overview of the evaluation methods, quantitative estimate of impact, sampling and intervention assignment, teacher, parent and principal surveys and ethical considerations were provided in the Year 1 report and have

remained largely unchanged. As such, this section will report on specific issues related to the 2018 Grade 2 instrument design, data collection and sample attrition.

3.1. Grade 2 Instruments

The baseline, Year 1 and Year 2 learner assessment instruments were adapted from the Early Grade Reading Assessment (EGRA). The tests were designed based on the Linguistic Interdependence Hypothesis, which assumes that the languages of a bilingual reader share common proficiencies which can be transferred from the home language to the second language, given sufficient automaticity of these proficiencies in the home language (Cummins, 1979). The Year 2 tests therefore comprised of both home language (in this case isiZulu and Siswati) and English items. The baseline test was designed to test learners' home language literacy skills at the start of Grade 1 and the Year 1 test was designed to assess learners' English literacy skills at the end of Grade 1. The primary purpose of the assessments used in the evaluation is not to benchmark learner performance against curriculum requirements, but rather to determine learners' literacy abilities at the end of Grade 1 and 2. Given this focus, care was taken to minimize a floor effect. Although curriculum benchmarking has not yet been well developed for the South African languages, initial estimates of the proportion of learners that have reached certain milestones have been included in the findings.

All tests are designed to be orally administered by the fieldworkers and to be captured electronically on the Tangerine software. As was the case with the baseline and Year 1 testing, in order to test the targeted 20 learners within one school day, the tests were designed to take no longer than 15 minutes to administer.

The home language sections of the Year 2 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 used in the assessment.

Table 4: Tasks contained in the Year 2 (Wave 3) instrument

- Task 1. HL Object Naming Speed
- Task 2. HL Letter Sound Speed
- Task 3. HL Letter Sound Recognition
- Task 4.1. HL Oral Reading Fluency (ORF)
- Task 4.2. HL ORF Text Comprehension
- Task 5.1. English Decodable Word Recognition
- Task 5.2. English Sight Word Recognition
- Task 6.1. English ORF
- Task 6.2. English ORF Text Comprehension

- Task 7. "Do this" - English receptive proficiency
- Task 8. "Tell me" - English Active Vocabulary
- Task 9. "Answer me" - English Listening Comprehension
- Task 10. English Writing

The learner assessment at the end of Year 2 (Grade 2) of the intervention consisted of thirteen 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 5: Sub-tests contained in the instruments at each point in time

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

Contextual Questionnaires

A principal questionnaire, a teacher questionnaire and a home background questionnaire were further administered to gain a richer understanding of the context in which the interventions were implemented. The principal questionnaire entailed questions on school-level factors which may influence learning outcomes, whereas the teacher questionnaire (administered to all Grade 2 teachers) focussed on classroom-level factors. As was the practice in the first two rounds of data collection, a home background questionnaire was sent

home with each sampled learner and fieldworkers retrieved the filled out questionnaires from the schools at a later date. The purpose of the parent questionnaire was to collect information about general home circumstances and exposure to reading and English is the home.

3.2. Data Collection

As with the baseline and Year 1 data collection, Year 2 testing was conducted by external service providers. Testing was conducted by fourteen pairs of fieldworkers over a three week period in October/November 2018. During baseline, a random sample of 20 learners was selected to be tested in each school, and the exact same learners were re-tested again at the end of Year 1 and Year 2.

Both fieldworkers were responsible for the individual administration of the learner assessment, but one fieldworker was also responsible for administering the structured questionnaires to all the Grade 2 teachers and the school principal. The learner assessments, teacher and principal questionnaires were administered and captured using the Tangerine software. The home background questionnaires were captured and cleaned by the data collection company and the final data set was provided to the project management team at the conclusion of the data collection contract.

Of the 2,479 home background questionnaires which were returned to the schools, 2,421 of the returned questionnaires were successfully matched with learners in the sample.

Table 6: Percentage of Home Background Questionnaires (HBQs) returned

	Number learners re-tested	Number HBQs returned	Number matched	% Learners matched to HBQs
Control	1,190	1,065	1,040	87%
Intervention 1	772	706	691	90%
Intervention 2	799	708	690	86%
Total	2,761	2,479	2,421	86%

3.3. Balance at baseline

Table 7 shows that the sample was balanced on the baseline assessment at the start of Grade 1. There is a slight imbalance on only one of the sub-tasks, but since we are making 20 comparisons below, this is in line with what is expected. Moreover, the p-value of the F-test shows that we cannot reject the null for the joint significance across all the indicators. There is therefore no evidence of imbalance.

	(1) Control Mean/SE	(2) On-site Coaching Mean/SE	(3) Virtual Coaching Mean/SE	On-site vs control (1)-(2)	Virtual vs control (1)-(3)
Naming Animals in HL	7.155 [0.127]	7.310 [0.155]	7.501 [0.154]	-0.155	-0.346*
Word Recall	9.981 [0.084]	9.953 [0.093]	10.081 [0.092]	0.028	-0.099
Nonword Recall	4.208 [0.049]	4.179 [0.052]	4.237 [0.082]	0.029	-0.030
Phoneme Isolation	1.129 [0.087]	1.037 [0.092]	1.161 [0.107]	0.092	-0.032
Story Comprehension	2.179 [0.045]	2.154 [0.050]	2.263 [0.047]	0.025	-0.084
Letter Sounds Correct	6.978 [0.447]	6.784 [0.590]	7.019 [0.610]	0.194	-0.041
Words Read Correct	0.387 [0.096]	0.347 [0.103]	0.510 [0.148]	0.039	-0.123
Sentence Words Read Correct	0.051 [0.012]	0.027 [0.011]	0.034 [0.012]	0.024	0.018
Visual Perception	1.460 [0.082]	1.597 [0.111]	1.651 [0.109]	-0.137	-0.192
English Items	0.836 [0.044]	0.789 [0.063]	0.839 [0.045]	0.047	-0.003
N	1459	924	944		
Clusters	80	50	50		
F-test of joint significance (p-value)					0.782
F-test, number of observations					2383

Note. The value displayed for t-tests are the differences in the means across the groups. Standard errors are clustered at the school level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. The values displayed for F-tests are p-values.

Table 8 below shows balance on school-level characteristics. The school principals in the control schools are slightly older, and schools in the Virtual Coaching arm have a slightly larger problem of learner absence and have worse infrastructure. However, the F test for the join significant test means that we cannot reject the null that these two samples are statistically equivalent.

Table 8: Balance on school characteristics

	(1)	(2)	(3)		
		On-site	Virtual	On-site vs	Virtual vs
	Control	Coaching	Coaching	control	control
	Mean/SE	Mean/SE	Mean/SE	(1)-(2)	(1)-(3)
Principal Female	0.450	0.500	0.420	-0.050	0.030
	[0.056]	[0.071]	[0.071]		
Principal's age	52.550	50.120	51.160	2.430**	1.390
	[0.573]	[0.770]	[0.794]		
Grade 1 enrollment 2017	79.725	77.100	72.120	2.625	7.605
	[3.978]	[5.240]	[4.591]		
No. government teachers	1.950	1.960	1.860	-0.010	0.090
	[0.111]	[0.178]	[0.121]		
Vacancies of Grade 1 Educators	0.063	0.040	0.060	0.022	0.003
	[0.027]	[0.028]	[0.034]		
Problem - teacher absence	3.513	3.480	3.480	0.033	0.033
	[0.067]	[0.091]	[0.096]		
Problem - learner absence	2.975	2.900	3.180	0.075	-0.205*
	[0.080]	[0.119]	[0.089]		
Describe school maintenance	3.325	3.220	3.060	0.105	0.265*
	[0.090]	[0.125]	[0.141]		
N	80	50	50		
P-value				0.461	0.274
Number of observations				130	130

The appendix shows that evaluation arms are balanced for an additional set of indicators that were asked during the Wave 3 school principal survey.

3.4. Attrition

During the Year 2 data collection, 2,761 of the 3,327 learners who were tested during the baseline data collection were re-tested and successfully matched to their baseline results. The overall attrition rate of 17% is slightly higher than what has been found in previous studies. When breaking down the attrition rate by intervention group, the differences are not statistically significant, but from figure 1 it is clear that the attrition rate of learners in the control schools (18%) was slightly higher than the attrition rate among learners in the virtual coaching and on-site coaching schools (each at about 15%). Considering learner or school characteristic that may be significantly correlated to the likelihood of a learner having had

attrited suggests that poorer performing learners in the virtual coaching schools were more likely to not have been in the Year 2 sample.[1]

Table 8: Percentage of learners tested during baseline, Year 1 and Year 2 data collection

	Intended sample	Baseline		Year 1		Year 2	
		Tested Number	*Percentage Tested*	*Number re-tested & matched*	*Percentage re-tested & matched*	*Number re-tested & matched*	*Percentage re-tested & matched*
Control	1,600	1,459	91%	1,347	92%	1,190	82%
Intervention 1	1,000	924	92%	820	89%	772	84%
Intervention 2	1,000	944	94%	873	92%	799	85%
Total	3,600	3,327	92%	3,040	91%	2,761	83%

In figure 1, in addition to the percentage of learners in each intervention group that either remain part of the study or have attrited, we have also included the percentage of learners that were enrolled in Grade 1 and have thus repeated the grade. The percentage of repeaters in each group ranges between 8% and 10% of the sub-sample.

Figure 1: Attrition rate by intervention group

---
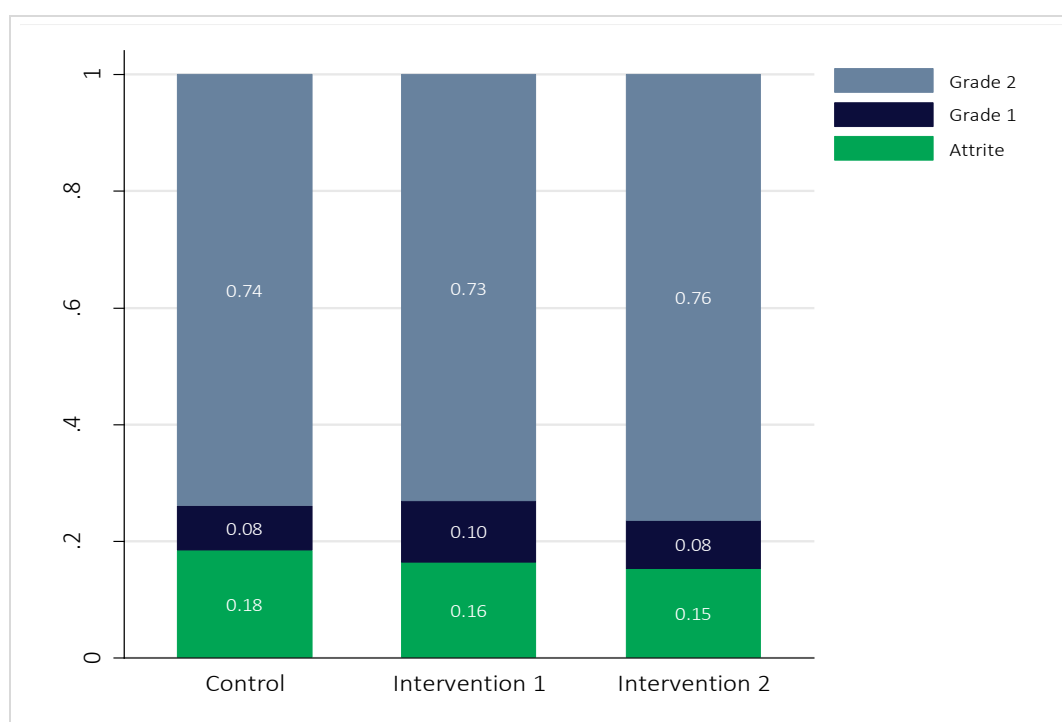
¹ See table 36 in the Appendix

Table 9 test if attrition was balanced and random across intervention arms. Column (1) shows that even though the attrition rates were slightly lower in the two intervention arms, this difference is not statistically significant at a 10% level. Columns (2) to (5) show that there is no differential attrition on age, gender, or home language of a learner. However, column (5) shows that learners in the virtual coaching schools who dropped out, performed worse at baseline compared to learners in the control schools who dropped out. There is therefore a risk that the learners that remain in the Virtual Coach sample are better-performers on average, compared to the Control. The third coefficient reported in column (5) confirms that there is indeed a slight imbalance in the remaining sample.

Table 9 Learner attrition in Wave 3 data collection

	(1) Attrite	(2) Age	(3) Gender	(4) Zulu	(5) Learning
Attrite		0.078	0.047	-0.002	-0.054
		(0.056)	(0.033)	(0.057)	(0.088)
On-site coach	-0.020	0.045	0.021	-0.038	-0.000
	(0.022)	(0.054)	(0.020)	(0.076)	(0.074)
Virtual coach	-0.029	0.043	0.021	-0.046	0.145**
	(0.021)	(0.061)	(0.023)	(0.077)	(0.071)
Attrite x On-site		-0.044	-0.056	0.117	-0.110
		(0.099)	(0.062)	(0.076)	(0.140)
Attrite x Virtual		0.058	-0.028	0.062	-0.228**
		(0.108)	(0.057)	(0.079)	(0.115)
Observations	3,327	3,327	3,327	3,327	3,327
R-squared	0.006	0.017	0.003	0.146	0.027
Mean attrition	0.182				

4. Task level learner assessment results

The learner dataset, which forms the primary evidence discussed in this report consisted of ten main tasks, three of which had two subtask sections. Unlike the learner assessment instrument designed for Wave 2, the Wave 3 instrument included more items in both Home Language and EFAL designed to assess reading skills in both languages. One of the specific reasons for the inclusion of tasks in Home Language, i.e. Siswati and isiZulu, to test for a possible crowding out effect. In this context, crowding out would involve teachers using the time for EFAL that the curriculum has earmarked for teaching literacy in the Home Language.

Table 10 provides information on the descriptive statistics of the assessment tasks administered at baseline, at the end of Year 1 and the end of Year 2. The scores in the table include the averages, standard deviation, an indication of the performance distribution, and the percentage of learners that scored zero on the task. The purpose of the table is to provide insights both to how learners on average performed relative to the maximum score

in the task but also to provide a perspective of the relative distribution of scores. The table shows that in some of the tasks there were floor effects at baseline, but that there was a good distribution of scores in the Year 2 assessment. The table therefore provides evidence that the assessment tasks provide sufficient information to differentiate learner performance across the sample distribution.

The zero scores indicate that the percentage of non-readers (i.e. the learners that could not read a single word correctly) was still remarkably high at the end of Grade 2. At the end of Grade 1, about 48% of learners could not read a single word correctly in their home language. It is disconcerting that a year later, at the end of Grade 2, 39% of the learners still did not read a single word correctly. The percentage non-readers in EFAL seem to be slightly lower, but this can because of the word length in the English language being on average shorter than in isiZulu and Siswati.

Table 10: Item descriptive statistics

		N	Mean	s.e.	p10	p25	p50	p75	p90	Min.	Max.	% zero score
Baseline	1. HL Naming the Animals	3327	7.3	0.41	4	6	7	9	10	0	12	1.30%
	2. HL Word Recall	3327	10	0.04	7	9	10	12	13	0	14	0.00%
	3. Nonword Recall	3327	4.2	0.02	3	4	5	5	5	0	5	1.50%
	4. Phoneme Isolation	3327	1.1	0.03	0	0	0	2	4	0	6	62.90%
	5. HL Story Comprehension	3327	2.2	0.02	1	1	2	3	4	0	4	8.70%
	6. HL Letter Sound Recog	3327	6.9	0.13	0	2	5	9	18	0	30	18.70%
	7. HL Words Correct	3327	0.4	0.02	0	0	0	0	1	0	6	88.30%
	8. HL Sentence Words	3327	0	0.01	0	0	0	0	0	0	6	98.50%
	9. HL Visual Perception	3327	1.6	0.03	0	0	1	3	4	0	10	35.90%
	10. EFAL Vocabulary	3327	0.8	0.02	0	0	0	1	2	0	6	50.40%
Year 1	10. HL Expressive Vocabulary	3067	4.9	0.01	5	5	5	5	5	0	5	0.00%
	10. HL Words Correct	3066	5	0.12	0	0	1	11	17	0	18	48.30%
	11. EFAL Phoneme Isolation	3066	3.6	0.06	0	0	3	6	8	0	9	28.20%
	14. Letters Correct	3068	16.7	0.31	0	3	11	26	40	0	80	12.10%
	15. EFAL Words Correct	3062	5.1	0.13	0	0	2	7	16	0	36	35.60%
	16. EFAL Listening	3062	2.9	0.02	1	2	3	4	4	0	5	8.30%
	17. EFAL Listening Comp	3062	0.2	0.01	0	0	0	0	1	0	3	82.10%
	18. EFAL Vocabulary	3060	0.7	0.02	0	0	0	1	2	0	2	51.10%
	19. EFAL Writing	3056	0.2	0.01	0	0	0	0	1	0	1	76.50%
Year 2	20. Object Naming	2765	16.2	4.8	10	13	16	19	22	0	36	0.11%
	21. HL Letter Naming	2765	18.1	8.8	5	12	19	24	29	0	36	2.43%
	22. HL Letter recognition	2765	36.6	22.6	7	19	36	53	65	0	110	3.95%
	23. HL ORF	2763	14.5	15.1	0	0	12	27	35	0	60	38.85%
	24. HL ORF Comp	2766	1.1	1.5	0	0	0	2	4	0	5	54.19%
	25. EFAL Decodable Words	2764	17.1	20.1	0	84	0	0	8	28	48	30.08%

26. EFAL Sight Words	2762	16.7	19.9	0	84	0	0	8	30	46	37.40%
27. EFAL ORF	2764	19.2	20.8	0	1	12	32	54	0	70	20.99%
28. EFAL ORF Comp	2766	1.3	1.3	0	0	1	2	3	0	9	39.18%
29. EFAL Recep Prof	2766	3.4	1.3	1.5	3	4	5	5	0	5	2.25%
30. English Rec Vocab	2763	4.3	1.3	2.5	4	5	6	6	0	6	0.58%
31. English Comp	2762	0.7	0.8	0	0	0	1	2	0	3	48.50%
32. English Writing	2766	0.6	0.7	0	0	0	1	2	0	2	46.03%
PCA: Gr2 SubTasks	2759	0	2.7	-3.1	-2.4	-0.5	2	3.9	-4.9	8.8	
PCA: Gr2 EFAL	2760	0	2.2	-2.4	-1.9	-0.5	1.5	3.2	-3.7	6.6	

Table 11 shows the mean scores of each sub-task for each of the intervention groups. The results in the table suggest that the learners in the Control group may have performed better on the Home Language items than the learners in the intervention groups, but that there are very little other observable differences on the other items. On the lower order Home Language tasks, such as object naming, letter sound speed, letter sound recognition, the On-site Coach learners were more or less on par, with the Virtual Coach learners performing worse on average compared to the Control group. For both the Home Language ORF and the five comprehension questions associated with this passage, the two intervention groups performed on average below the Control group. On the English ORF item, however, both intervention groups seem to have performed better than the Control group. These differences are address more systematically in the main results section.

Table 11: Tasks Means in Wave 3, by intervention group

		Control	On-site Coaching	Virtual Coaching
Home Language	Object Naming	16.599	16.244	15.721
		[0.235]	[0.195]	[0.251]
	Letter Naming	18.458	18.576	17.268
		[0.462]	[0.471]	[0.600]
	Letter recognition	37.083	37.839	34.706
		[1.263]	[1.373]	[1.725]
	Oral reading fluency (ORF)	15.684	13.852	13.440
		[0.782]	[0.935]	[1.110]
	ORF Comprehension	1.193	1.144	1.008
		[0.076]	[0.106]	[0.095]
EFAL	Decodable Word Recognition	17.026	17.842	16.511
		[0.933]	[1.226]	[1.469]
	Sight Word Recognition	16.469	17.611	16.271
		[0.929]	[1.142]	[1.526]
	Oral reading fluency (ORF)	18.894	19.697	19.136
		[1.025]	[1.301]	[1.650]
	ORF Comprehension	1.309	1.370	1.304
		[0.063]	[0.088]	[0.086]
	Receptive Proficiency	3.192	3.541	3.523
		[0.076]	[0.076]	[0.091]
	Expressive Vocabulary	4.177	4.422	4.379

	[0.061]	[0.083]	[0.080]
Listening Comprehension	0.617	0.730	0.658
	[0.043]	[0.061]	[0.050]
Writing	0.564	0.653	0.539
	[0.033]	[0.044]	[0.048]
N	1191	772	796
Clusters	79	50	50

Note: square brackets indicate standard errors.

4.1. Home Language Object Naming Speed

Rapid Automatised Naming (RAN) of objects and letters are measures of speed of lexical access. These tasks required learners to name out loud a set of highly familiar objects (as seen in Figure 2 below) and letters within a short time period, thereby measuring phonological processing skills. RAN tasks are predictive of reading fluency and were therefore included to determine whether the underlying skills needed for reading fluency development were similar across control and intervention groups. As there is no consistent evidence that RAN skills can be improved and that the interventions did not focus on this specific skill, the interventions were not expected to have had any effects on these sub-tests.

Figure 2: Home Language Object Naming Speed
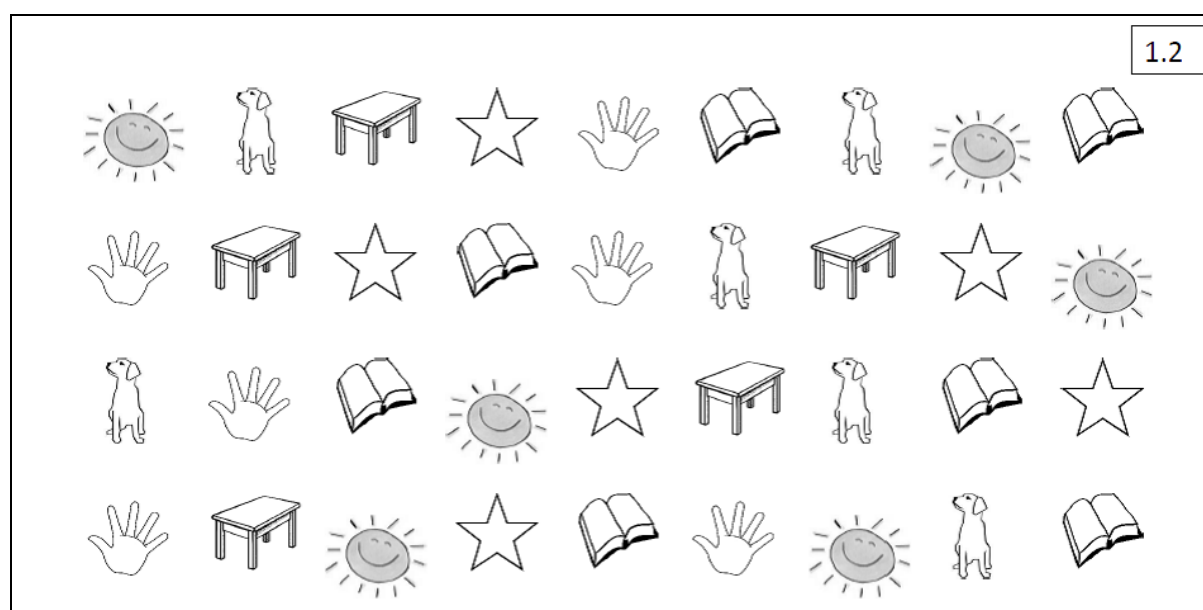
Figure 3 shows the inverse cumulative density function for Task 1. In the 20 seconds allowed for this task, at least 95% of learners in the interventions schools managed to correctly identify the first ten objects. A slightly higher percentage of the control group learners (98%) managed to do the same. Similarly, 10% of the intervention group learners managed to correctly identify 25 objects correctly in the 20 seconds, whereas 16% of the control group learners managed to reach the same threshold.

4.2. Letter Sound Recognition

In the baseline, the learners identified just fewer than seven letter sounds correct in their Home Language with about 19% of the sample scoring zero. Two years later, in the 2018 test, the average letters sounds recognised correctly increased to 36, and the percentage of learners not correctly identifying any letter sounds has gone down to 3%.

Table 12: Comparing home language letter-sound recognition

	Start of Grade 1	End of Grade 1	End of Grade 2
Control	7.0	17.7	37.1
On-site coaching	6.8	16.7	37.8
Virtual coaching	7.0	15.1	34.8

It is clear that an additional year of schooling translates into substantial gains in the number of letters sounds children can correctly identify. However, from Table 12 and Figure 4 it is evident that the three groups are not entirely similar. In Grade 2, fewer learners scored zero in control schools relative to learners in both intervention schools, and the distribution of the number of letter sounds read correctly shows a higher proportion of learners in the virtual coaching group at the lower end of the performance distribution. The relative weakness of the virtual coaching group in home language letter-sound recognition was evident at the end of Grade 1 and continues in Grade 2.

Figure 4: Home Language Letter Sound Recognition
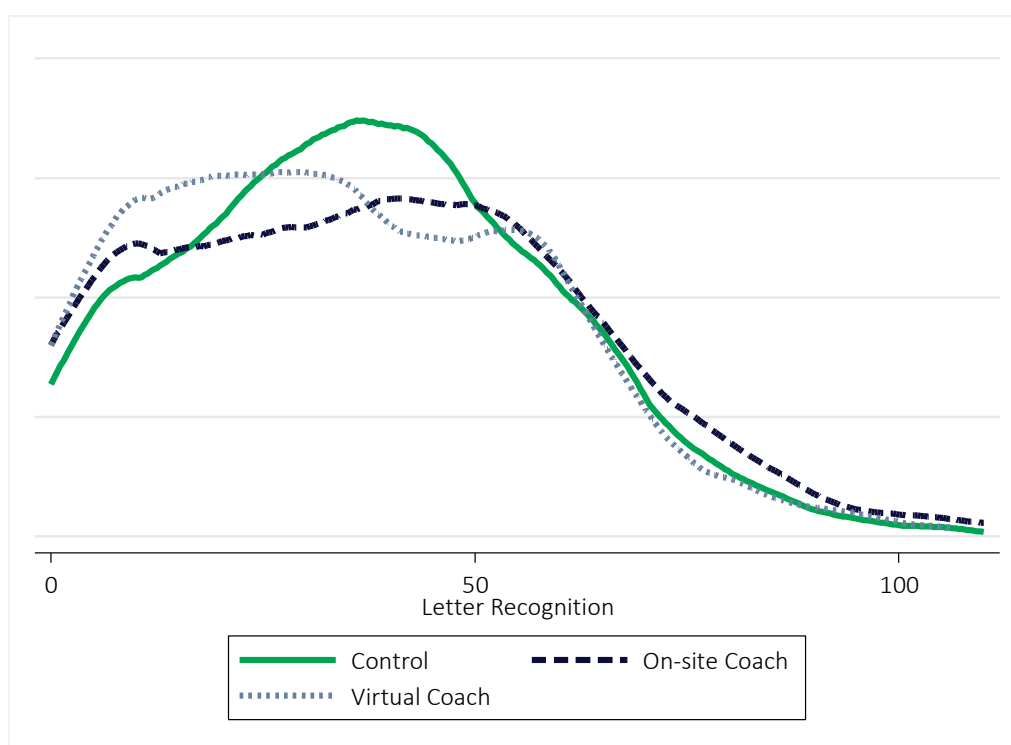
4.3.　Home Language Word Reading

One of the core assumptions implicit in the assessment is that learners' mastery of reading in English requires that they have a strong foundation of reading in their home language. The first set of assessment tasks, i.e. the object naming and home language letter recognition are the building blocks of reading in the home language. Although not part of either the on-site or virtual coaching interventions, the assessment of home language literacy skills provide important insights into how the failure to acquire home language emergent literacy skills can inhibit progress in learning to read and write in the second language.

The home language items also allow us to evaluate whether there are any positive or negative spillover effects as a result of the interventions. Since the interventions focus on teaching EFAL, it is possible that teachers may spend more time on teaching EFAL, at the cost of teaching home language, which may have a detrimental effect on home language. Alternatively, the skills teachers are equipped with are transferable to teaching home language, and may therefore enhance their teaching of home language and subsequently learners' home language outcomes.

The first and most important insights however from the test about ORF in the home language is that 39% of the entire learner sample, (34% of the control group and 41% and 44% of the intervention group) could not read a single word on the home language ORF task (Figure 4).

Figure 5 Histogram of Scores for HL Oral Reading Fluency
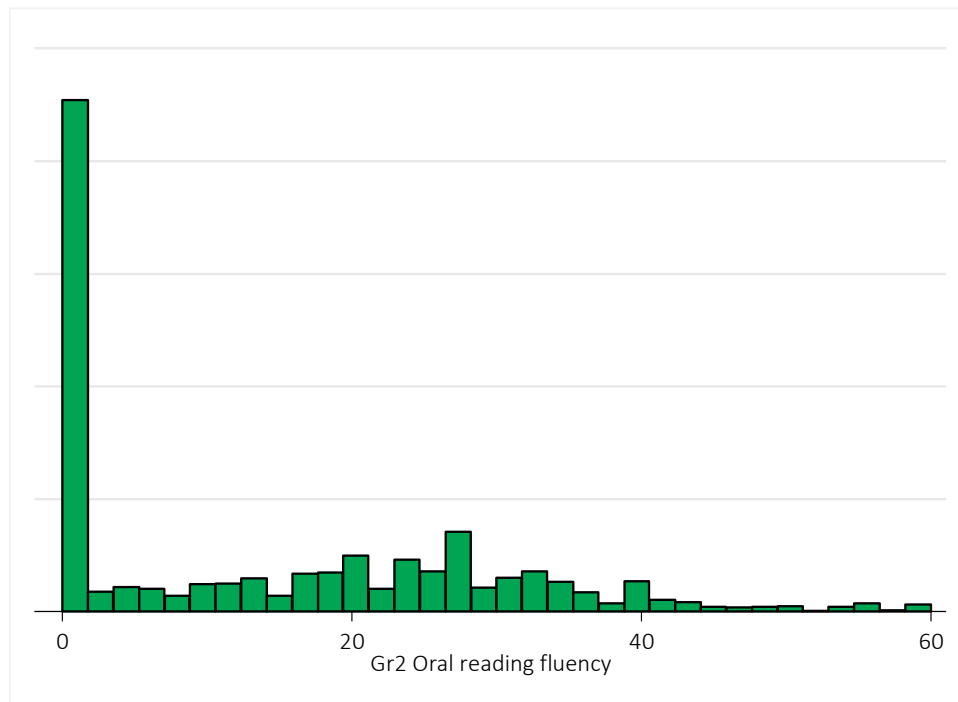
The high proportion of learners scoring zero on the home language ORF task is not unique. The NORC (2019) evaluation of the Story Powered School initiative found that 41% on the Grade 2 learners in their sample scored zero on the isiXhosa ORF task, remarkably similar to the proportion of learners in the EGRS II Grade 2 test. Notwithstanding the obvious issue of a floor effect, the consistent evidence of a very large proportion of learners at the end of Grade 2 not able to identify a single word in a home language ORF task signals the need to better understand the profile of these learners and their unique learning needs. The NORC evaluation has provided an important start with the addition of analysis of stunting in relationship to slow reading progress. The evidence from the Wave 3 assessment shows that the percentage of learners who did not read any words correctly declined substantially between the end of Grade 1 and the end of Grade 2.

Table 13: Comparing Home Language Word Reading

	Word Reading[2] End of Grade 1		ORF End of Grade 2	
	Mean	% of Zero Scores	Mean	% of Zero Scores
Control	5.5	45.7%	15.7	34.1%
On-site coaching	4.7	49.1%	13.8	41.3%
Virtual coaching	4.7	51.8%	13.5	43.6%

---

[2] Home Language ORF was not assessed at the end of Grade 1, only HL word recognition. These two measures are only loosely comparable.

Figures 6 provide insights into the comparative performance of learners across the full distribution in the three groups. The two intervention groups had a considerably higher proportion of learners that scored zero (34.1% compared to 41.3% and 43.6%) respectively. At a cut-off point of 20 words correct per minute (wcpm), the control group has a substantially higher percentage of learners achieving at least at this level. Further investigation is needed to explain these trends, specifically the extent that these results are a function of the schools' unique characteristics or an unanticipated consequence of the interventions.

Figure 6: Distribution of Home Language Oral Reading Fluency
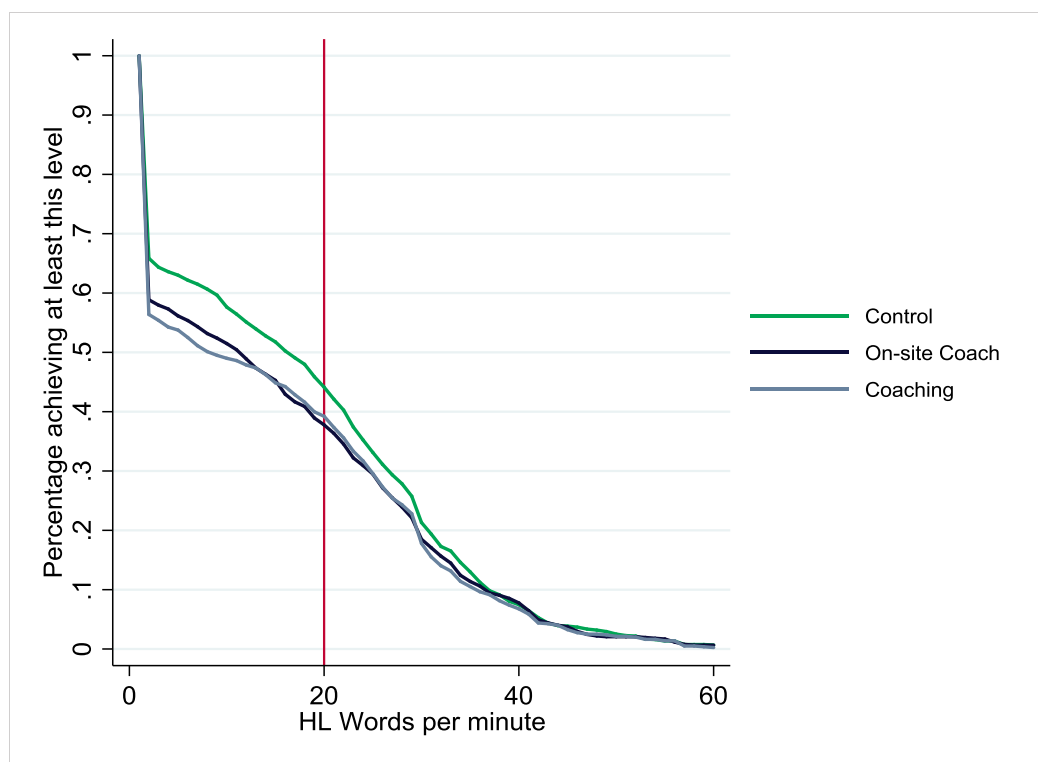
4.4. English Decodable and Sight Words

At the end of Grade 2, the first indicators of the relative impact of the interventions on English reading are learners' mastery of English decodable and sight words. Decodable words are those words that the learners can decipher using their knowledge of the letter sound relationship. Sight words are words that learners learn to recognise as whole words. These are often common words that occur frequently, and core to vocabulary in the language. The specific decodable word task in the assessment had 42 words beginning with 'dog'. The sight word task also had 42 possible words for the learners to correctly identify in 30 seconds. To

ensure that the measures are comparable to the conventional way of reporting on words read correctly, the number of words read in 30 seconds was doubled to provide a sense of the number of words read correctly in a minute.

At the end of Grade 1, the Wave 1 assessment found that on average, learners could read 4.9 of the decodable English words and 5.1 of the English sight words per minute correctly. There was no statistically significant difference observed between the control and intervention groups. (Department of Basic Education, 2018). The evidence further showed that there was no meaningful difference between learners' reading of words in home language and English. It was hypothesized that despite the priority given to home language reading in Grade 1 relative to English in the official curriculum, learners had not automated their letter-sound correspondence knowledge needed to automatically decode isiZulu and Siswati words. English word recognition did show improvement in Grade 2, but there were no meaningful differences between the control and intervention groups.

Table 14: Comparing EFAL word recognition between Wave 2 and Wave 3

	End of Grade 1		End of Grade 2	
	Decodable words	Sight words	Decodable words	Sight words
Control	5.0	5.3	17.0	16.5
On-site coaching	5.3	5.5	17.8	17.6
Virtual coaching	4.6	4.7	16.5	16.3

At the end of Grade 2, approximately six in ten learners were able to correctly recognise the word 'dog', but consistently across the majority of the initial decodable words (rat, run and can), between only 45-55% of learners could correctly recognise these words. There was a very similar pattern in the sight words, with around half of the sample not able to correctly identify common sight words such as "to" and "is", the second and third words on the word reading list. Learners in the on-site coaching schools consistently but not dramatically outperformed both learners in the control schools and learners in the virtual coaching schools.

Table 15: Percent correct in initial decodable and sight list on the respective tasks

	Decodable Words					Sight Words			
	dog	rat	run	can	box	he	to	is	with
Control	64.7%	45.2%	46.7%	41.7%	44.6%	38.0%	47.8%	54.8%	45.9%
On-site coaching	65.7%	49.7%	54.4%	43.8%	48.6%	39.6%	54.9%	54.5%	46.5%
Virtual coaching	60.1%	46.8%	44.8%	43.1%	44.1%	39.7%	47.2%	50.5%	43.7%

In the whole sample, 30% of learners could not identify a single decodable word and 37% could not identify a single sight word (e.g. 'he' or 'to'). This suggests the magnitude of learners who have not started the journey towards word reading in English.

Figure 7: Distribution of English decodable and sight words read correctly
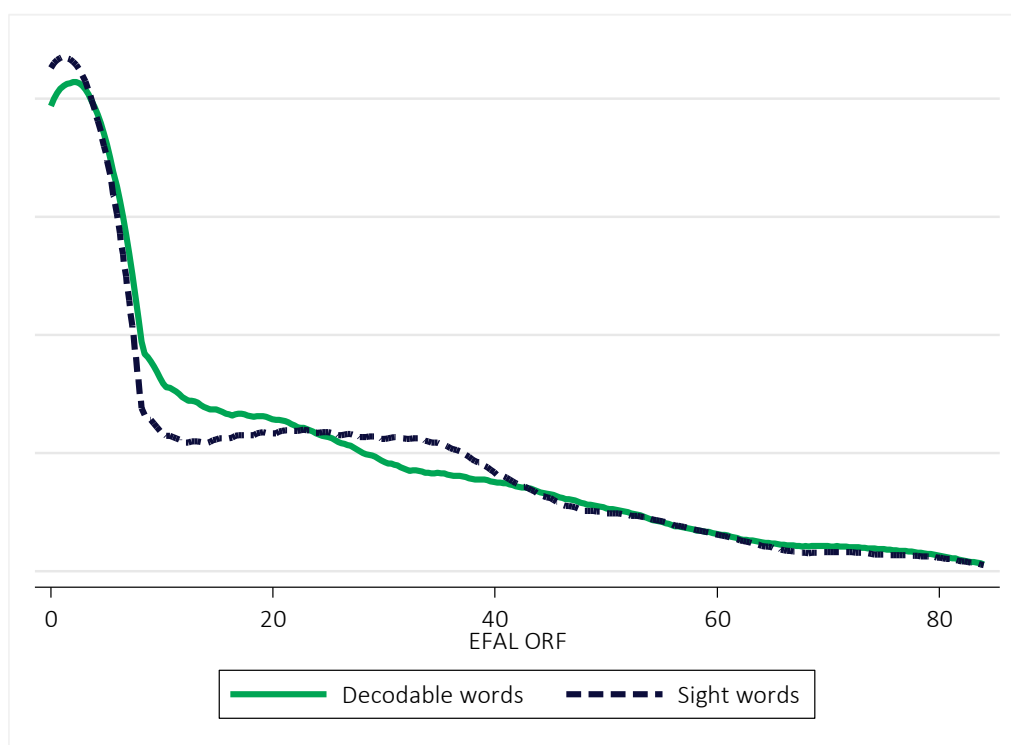
4.5. English text reading

Possibly the most important task within the Wave 3 learner assessment that could contribute to answering the core study questions is the assessment of learners' English ORF. Although there remains some question about the value of ORF and specific attained benchmarks (Kim et al 2010) as a predictor of reading difficulties, there is a growing consensus about the validity and reliability of ORF as a key measure of reading for meaning. It is now widely accepted that ability to read connected texts rapidly, accurately and with expression, is a critical component required for successful reading for understanding.

The mean scores for the learners in the three groups of schools in home language ORF and EFAL ORF are very similar. The overall mean scores suggest that learners were performing better in their EFAL ORF compared to their HL ORF, but this is most likely a function of the shorter word length in English. Further, a small and positive difference is seen between the ORF in English between learners in the on-site coaching and virtual coaching groups and those in the control schools. The important point however is that the differences are very small, and as will become evident in the main regression analyses, these differences are not statistically significant.

Table 16: Comparing HL and EFAL ORF

	HL ORF	EFAL ORF
Control	15.7	18.9
On-site coaching	13.8	19.7
Virtual coaching	13.5	19.2

In terms of the proportion of learners that scored zero on the English text reading, the percentage in the control schools is smaller than in either of the intervention groups. Approximately one in four learners could not read a single English word in a connected text, compared to less than one in five in the control schools.

Table 17 Proportion of zero scores per intervention group
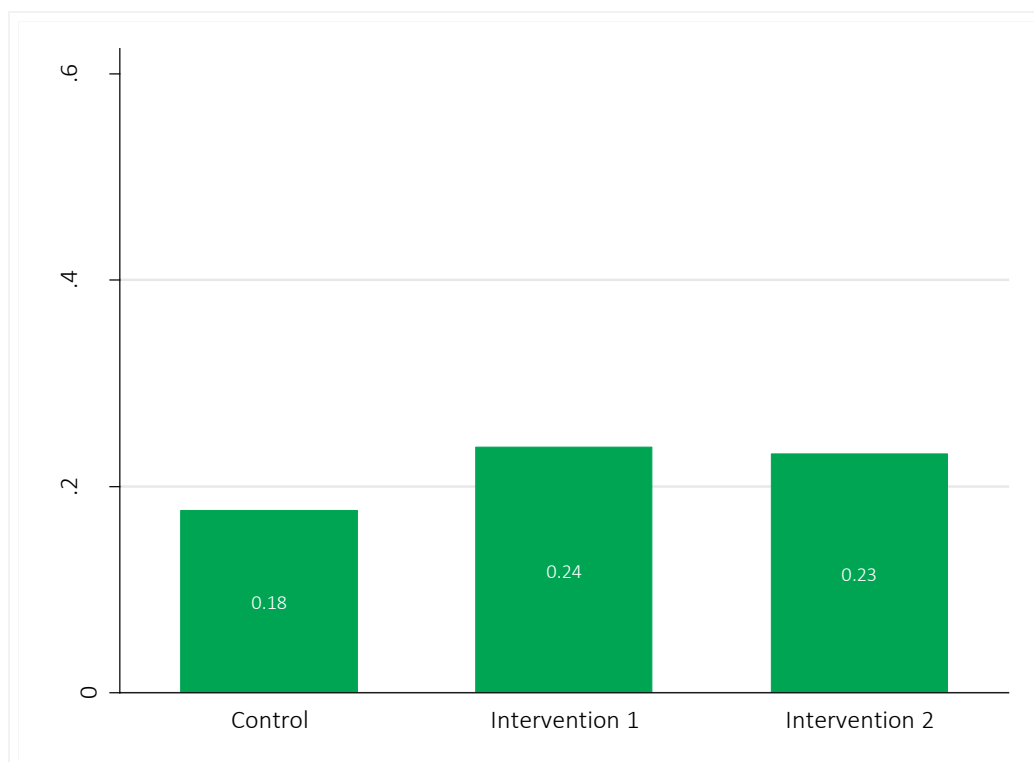
Another unusual feature emerges in an analysis of the distribution of performance for the three groups of schools. In Figure 8, the intervention group learners appear to underperform relative to learners in control schools in the performance range 1-25 wcpm. That picture changes around 35 wcpm, thereafter learners in the intervention group outperform learners in the control schools.

Figure 8: Distribution of English Oral Reading Fluency
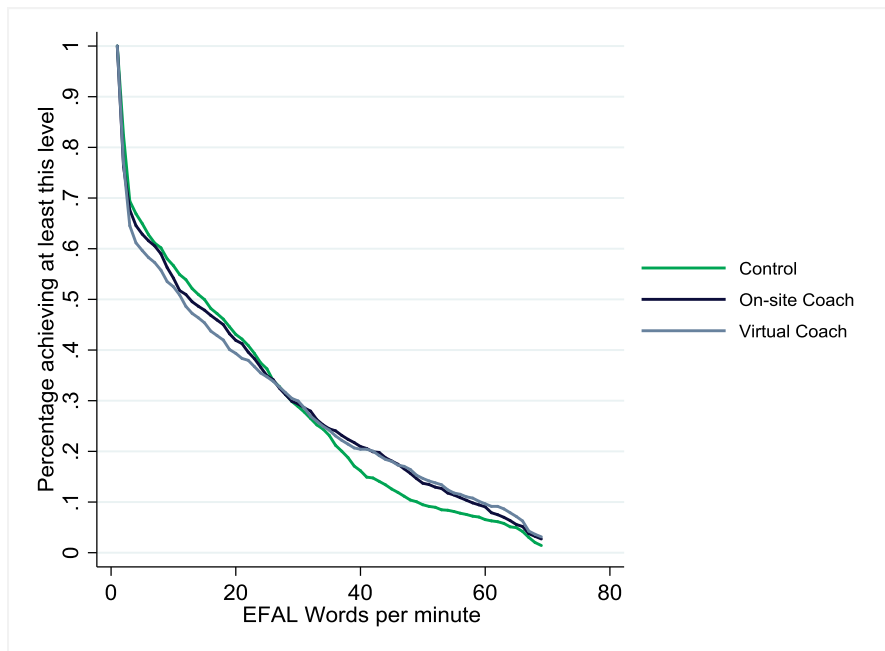
What could be expected in the absence of EFAL interventions, but with improved teaching in the home language? In the Department of Basic Education (2019) Wave 4 collection in the original Setswana study schools in the North West province, Grade 4 learners were reading at 39 WCMP in English.

4.6. English text comprehension

One of the primary concerns is the extent to which the English reading interventions improve children's reading comprehension or understanding. The comprehension task was designed to measure this. There was no time limit for the responses, but the learners were only given one minute to read the passage. Learners were subsequently only asked questions relevant to the section up to where they read. Figure 9 shows the passage that was used for the English ORF and the comprehension questions that were asked.

Figure 9: English text reading and comprehension questions

Jabu had a little dog. It was a happy dog. (10 words)
Jabu liked his dog. (14 words)
One day Jabu and his dog walked in the bush behind the house. (27 words)
The dog saw a rabbit and chased it. (35 words)
The dog got lost. (39 words)
Jabu called him but he didn't come. (46 words)
Jabu was sad. He went home. (52 words)
But before evening the little dog came back. (60 words)

Jabu was very happy. He gave his dog a bone. (70 words)

English text reading comprehension questions:
1. What is the name of the boy in the story?
2. What kind of animal did he have?
3. What did Jabu and his dog do together one day?
4. What animal did the dog chase?

There was no substantial difference in the mean scores of learners in the control, on-site coaching and virtual coaching schools (Table 18). Figure 10 below shows the comprehension answers correct in percentage terms. It suggests that the learners in the two interventions only outperformed their fellow learners towards the top end of the distribution.

Table 18: Mean scores on EFAL Reading Comprehension

	Mean	s.e.
Control	1.3	0.04
On-site coaching	1.4	0.05
Virtual coaching	1.3	0.05

Figure 10: Distribution of English comprehension scores
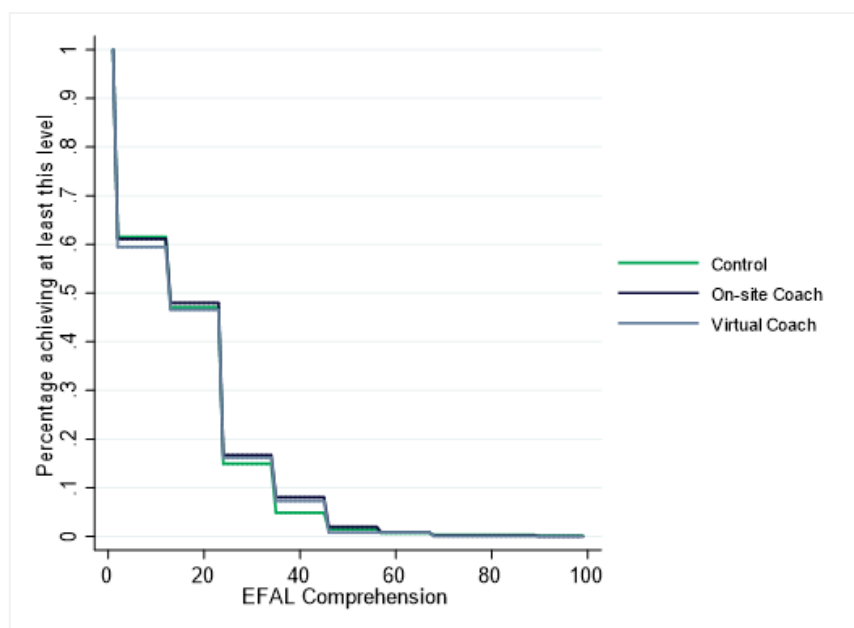
4.7.   Listening, Speaking and Comprehension in oral English

There were three tasks included in the end of Year 2 learner assessment to evaluate learners' English oral language proficiency. To measure receptive vocabulary learners were asked to respond to a set of instructions. To measure their expressive vocabulary, learners were

shown pictures of animals, and asked: "what do we call this in English?" And to measure listening comprehension learners needed to respond to a set of questions related to a short story that was read to the learner in English.

On the receptive vocabulary task, learners in the control group managed to respond correctly to 2.6 out of the five English instructions, whereas learners in the intervention groups responded correctly to three out of the five instructions. On the expressive vocabulary items, learners in both the on-site coaching and the virtual coaching were on average more likely to correctly identify the English name for a picture or a cow or pig than learners in the control group. However, very few learners in the intervention group or control group could correctly answer the comprehension questions correctly.

Table 19 Mean Scores in English Receptive, Expressive and Listening Comprehension

	Control	On-site Coaching	Virtual Coaching
English Receptive Proficiency	3.192	3.541	3.523
English Active Vocabulary	4.177	4.422	4.379
English Listening Comprehension	0.617	0.730	0.658

Given the priority given to oral language development in the CAPS, the better performance on both the receptive and expressive English oral language tasks suggested that the interventions were being impactful in improving oral language. The results show statistically significant higher scores for learners in the two intervention groups for the receptive proficiency task. This receptive task involved learner responding to a series of instructions such as put both hands in the air and draw a circle (Figure 10).
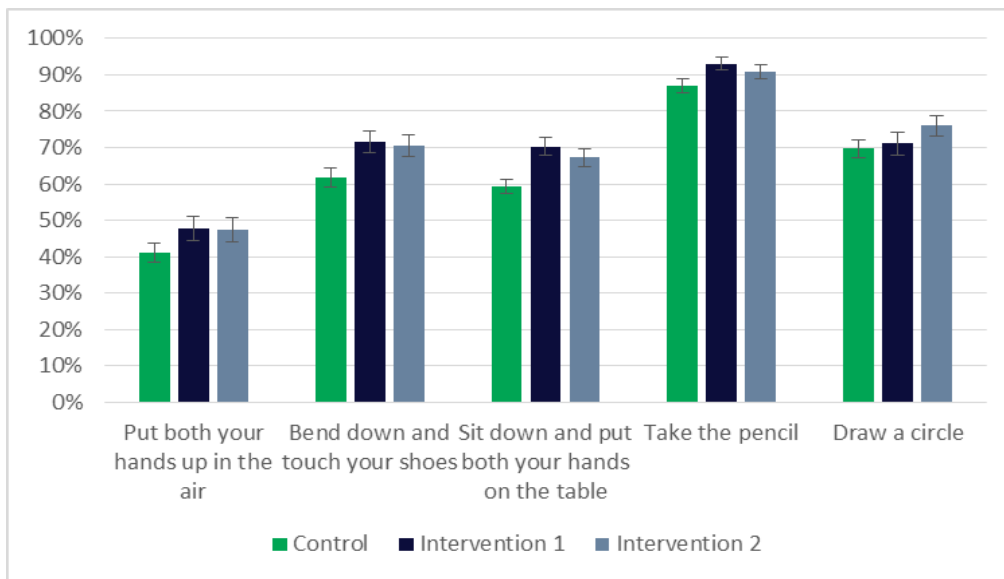
Relative to the substantial differences between interventions and control on active vocabulary (expressive vocabulary) at the end of Grade 1, the intergroup differences are not as pronounced as at the end of Grade 2 (Figure 12). This may be due to the overall higher performance of the learners on this task. On the two words highlighted in the Year 1 report, cow and pig, the proportion of learners that could correctly identify the English words for pictures of a cow and pig had increased dramatically in all three groups.

Figure 12 Means scores on expressive vocabulary
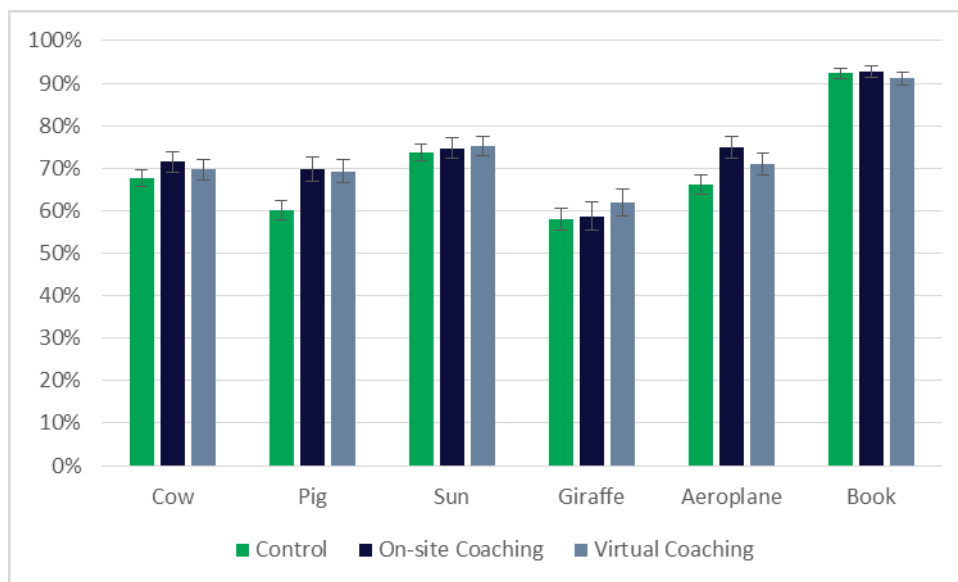
The English listening comprehension was assessed by answers to questions after the reading of a short story to learners. As with both the receptive and expressive vocabulary, learners in Grade 2 in all three groups have improved overall, with learners in the intervention groups

30

outperforming the control learners on two of the three items. However, the vast majority of learners (around 90%) in the study are not able to answer an inferential question based on an oral English story told to them. In this case, learners did not link the taking off shoes and socks to the fact that they were wet. This suggests that while a larger group of learners were able to correctly understand simple instructions and correctly identify objects in English, their mastery of even simple narrative English is limited.

Figure 13: English Listening Comprehension

In the morning Sipho and Buli ran to school in the rain. At school, they took off their wet shoes and socks. After school, they played in the mud. When they got home their Mother made them wash.

English listening comprehension questions:

Where did Sipho and Buli go in the morning?
Why did they take their shoes and socks off?
When did they play in the mud?

Figure 14 Means scores for items in the English oral comprehension task
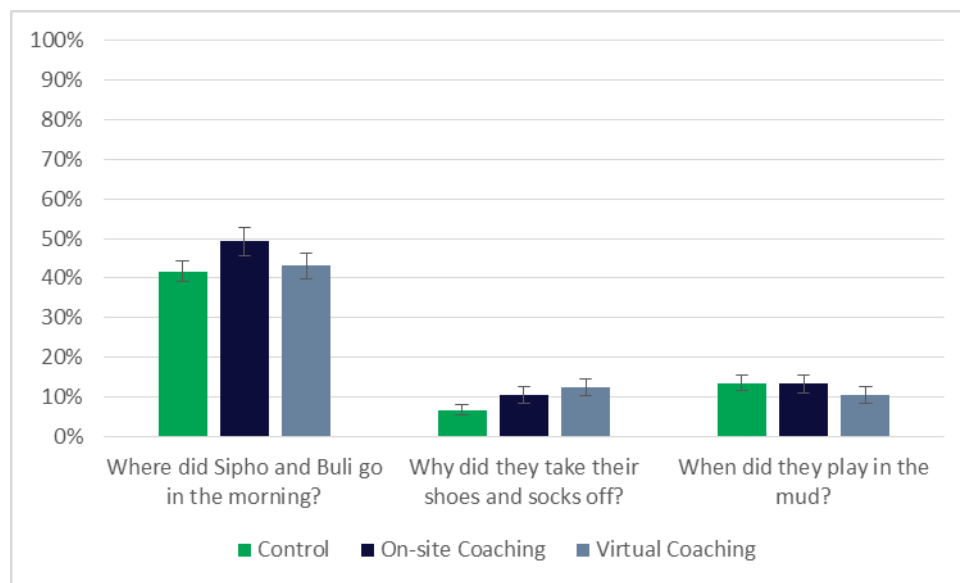
4.8.   Writing in English

Although the primary concern in the Grade 2 years is the mastery of essential skills of reading in English, particularly the decoding of phonics words, recognition of sight words, fluency of reading short simple paragraphs or stories and understanding the meaning of the passages or stories, an equally important skills that should be developed in Grade 2 is writing in English. The final task in the Grade 2 assessment consists of two items that require learners to write

single English words. The first correctly identify a verb the second a noun. On the one item, the verb completion, learners in the on-site coaching schools did better than those in control schools.

Figure 15 Means scores items on the writing in English task
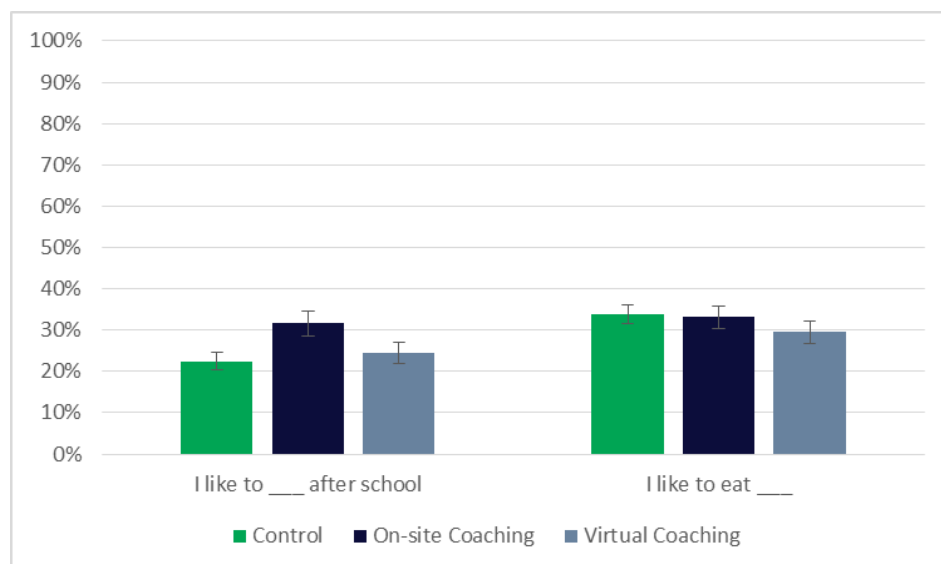
5. Main Results

5.1. Main Regression Findings

The first major finding of the study relates to the relative performance of the three groups on the Year 2 learner assessment. Two aggregate test scores were derived using Principal Component Analysis for oral vocabulary and reading proficiency respectively. For the sake of comparability, the aggregate scores were standardised to a mean of zero and a standard deviation of one. Since the attrition rate varied substantially by school, we also include learner-level weights equal to the inverse of the number of pupils surveyed, so that the sample is representative at a school level.

Table 20 shows the main regression results, including the following controls: strata, enumerator fixed effects, learner gender and the language of assessment (isiZulu vs Siswati).[3] The learner-level controls are further demeaned and interacted with treatment. The controls

---

[3] Controls were chosen so as to minimize the standard error on the treatment coefficient estimates. We iteratively added more families of controls: strata dummies, individual-level characteristics (age, gender, home language), enumerator fixed effects, and controls for performance in the different sub-tasks at baseline. The final set of controls decreased the standard errors.

help add precision to the estimates, as well as account for an incidental imbalance at baseline.

Overall, the results on oral language skills were similar to what was observed at the end of Grade 1. Column 1 in Table 20 shows that on the overall aggregate score of the vocabulary tasks both the on-site and virtual coach interventions had a positive effect. Columns 2 - 4 breaks this down and it is evident that both interventions had a positive effect on English oral language receptive and expressive skills. On the listening comprehension task, only the on-site coaching intervention had a positive and significant effect.

The more important question is if English reading proficiency also improved. Column 5-10 in Table 20 shows that neither intervention had a statistically significant positive impact on skills associated with English reading proficiency. The on-site coaching intervention had a positive effect on all the indicators – word recognition, reading fluency, comprehension, and writing – but we cannot reject the null that there is no positive impact at conventional levels of significance. In contrast, the virtual coaching intervention seems to have had a negative effect on all these indicators. On the aggregate score for English reading, the difference in magnitude between the on-site and virtual coaching interventions is statistically significant at a 10% level.

Table 20: Impact on English Oral Language and Reading Skills

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
		Vocabularly					Reading			
		Recep	Express	Listening		Decodable	Sight		ORF	
	Aggregate	Prof.	Vocab.	Compr.	Aggregate	Word	Word	ORF	Compr.	Writing
On-site	0.245***	0.409***	0.159**	0.124***	0.088	0.718	0.877	1.489	0.057	0.061
	(0.058)	(0.075)	(0.074)	(0.045)	(0.068)	(0.584)	(0.585)	(1.253)	(0.086)	(0.041)
Virtual	0.190***	0.326***	0.212***	0.034	-0.050	-0.639	-0.434	-0.585	-0.064	-0.016
	(0.057)	(0.082)	(0.071)	(0.044)	(0.075)	(0.630)	(0.682)	(1.474)	(0.085)	(0.045)
Obs	2,760	2,760	2,760	2,760	2,760	2,760	2,760	2,760	2,760	2,760
R^2	0.200	0.166	0.136	0.178	0.183	0.164	0.165	0.194	0.156	0.158
P-value	0.415	0.375	0.512	0.068	0.090	0.048	0.080	0.184	0.199	0.112
C mean	0	3.187	4.176	0.616	0	8.535	8.257	18.89	1.310	0.563

Table 21 shows the results for the Home Language tasks. Column 3 is an aggregate measure of reading proficiency – consisting of letter recognition, ORF and reading comprehension – and is again standardized to mean zero and standard deviation of one. There is now a negative impact for both interventions on Home Language literacy, although the effect size is larger and statistically significant for the virtual coaching intervention. In columns (1) and (2) the outcomes for the Rapid Automatized Naming (RAN) of objects and letters are shown. Again, there is a negative effect for both interventions. The impact on Object Naming is similar in magnitude for the two programmes and statistically significant at a 10% level. The impact of letter naming is positive (but insignificant) for the on-site coaching intervention and

negative significant for the virtual coaching intervention. It appears to suggest that there is more effective Home Language teaching in the control schools than in either of the intervention schools, although this negative effect is larger for the case of the virtual coaching intervention.

Table 21: Impact on Home Language literacy

	(1)	(2)	(3)	(4)	(5)	(6)
	Rapid Automised Naming			Home Language Reading		
	Object Naming	*Letter Naming*	*Aggregate*	*Letter Recog.*	*Reading Fluency*	*Compr.*
On-site coach	-0.487*	0.228	-0.011	1.579	-1.272	-0.012
	(0.254)	(0.586)	(0.076)	(1.694)	(1.068)	(0.108)
Virtual coach	-0.560**	-0.955*	-0.157**	-2.464	-2.534**	-0.199**
	(0.248)	(0.565)	(0.078)	(1.676)	(1.175)	(0.100)
Observations	2,760	2,760	2,759	2,760	2,759	2,760
R-squared	0.120	0.151	0.166	0.151	0.150	0.150
P-value	0.790	0.0791	0.105	0.0353	0.327	0.130
Control mean	16.60	18.45	1.01e-09	37.09	15.69	1.194

5.2. Analysis by school home language

Roughly 28% of the sample of learners are enrolled in schools where the Home Language of the school is isiZulu, compared to 72% learners who attend schools with Siswati as the Home Language. Since we saw disparate effects based on the language of assessment – English vs Home Language – it is important to investigate if the interventions have different impacts on isiZulu vs Siswati language schools.

Table 23 show that there is in fact a *negative* impact on the English reading tasks – for both intervention programmes - in the isiZulu schools, although it is larger in magnitude and only statistically significant for the on-site programme. In contrast, the impacts of both interventions on English oral proficiency are large and positive in the Siswati schools, whereas the on-site coaching intervention also had a positive effect on the English reading tasks.

Table 22: English oral language and reading proficiency, isiZulu vs Si Swati language schools

	(1)	(2)	(3)	(4)
	isiZulu		Siswati	
	Oral	*Reading*	*Oral*	*Reading*
On-site coach	-0.114	-0.340***	0.362***	0.238***
	(0.112)	(0.104)	(0.067)	(0.083)
Virtual coach	0.047	-0.240	0.252***	0.015
	(0.135)	(0.146)	(0.062)	(0.086)
Assessed wrong language+	-1.245***	-1.273***	-0.308	0.179
	(0.159)	(0.250)	(0.380)	(0.325)
Observations	758	758	2,002	2,002
R-squared	0.163	0.134	0.214	0.210
P-value	0.314	0.523	0.115	0.0168

Control mean	-0.0151	0.0576	0.00609	-0.0233

+Note that the coefficient "Assessed wrong language" refers to learners who were assessed in a different language in Wave 3, compared to Wave 2, and were assessed in a different language to the language of instruction of the school. There are 11 such learners in one Zulu school, and one learner in a Swati school. The independent variable in columns (1) and (3) is an index of receptive proficiency, vocabulary, and listening comprehension. The dependent variable in columns (2) and (4) is an aggregate indicator of word recognition, reading fluency, comprehension, and writing, constructed using principal component analysis.

In Figure 16, the tasks are represented in standard deviations for better comparability in the Siswati sub-sample. Bars that with solid colours are statistically significant at the 10% level, whereas the patterned bars are not statistically significant. The relative performance of the on-site coaching intervention relative to the control group (zero on the x axis) and the virtual coaching intervention is clear for all the English sub-tasks.

Figure 16: Standard Deviation scores for English tasks for the Siswati sub-sample
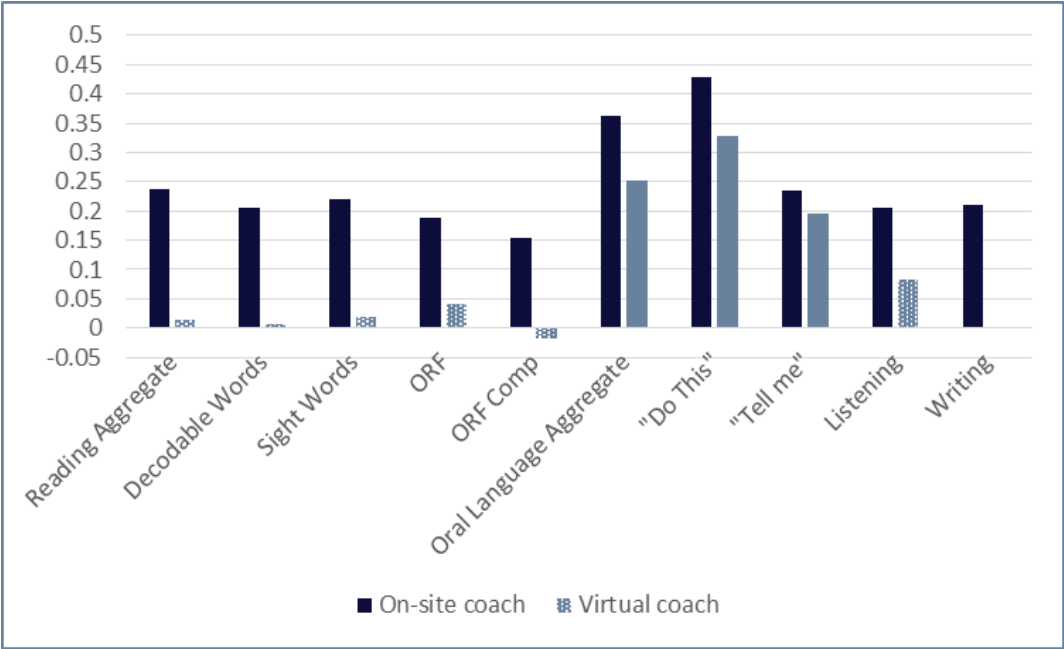
Table 24 below shows that most of the negative effect on the Home Language literacy tasks in the main sample is driven by the isiZulu sub-sample. There is, in fact, a positive (although statistically insignificant) impact of the on-site coaching intervention on Home Language reading proficiency in the Siswati-speaking schools. In columns (3) and (6) the dependent variable is an aggregate measure of reading proficiency – consisting of letter recognition, ORF, and comprehension – and is again standardized to mean zero and standard deviation of one.

Table 23: Home language oral and reading skills for isiZulu and Siswati schools

	(1)	(2)	(3)	(4)	(5)	(6)
		isiZulu			Siswati	
VARIABLES	*Object Naming*	*Letter Naming*	*Aggregate*	*Object Naming*	*Letter Naming*	*Aggregate*
On-site coach	-1.353***	-2.359**	-0.429***	-0.218	0.696	0.083
	(0.382)	(0.969)	(0.131)	(0.307)	(0.692)	(0.091)
Virtual coach	-1.396**	-2.630**	-0.356**	-0.293	-0.387	-0.084
	(0.547)	(1.076)	(0.161)	(0.247)	(0.645)	(0.087)
Assessed wrong language	-3.455***	-10.687***	-1.375***	1.481	8.501***	0.034
	(0.726)	(1.627)	(0.457)	(1.110)	(3.247)	(0.355)
Observations	758	758	757	2,002	2,002	2,002
R-squared	0.152	0.168	0.180	0.116	0.159	0.180
P-value	0.940	0.832	0.701	0.805	0.152	0.0918
Control mean	17.04	19.46	0.114	16.42	18.04	-0.0458

5.3. What explains the curious "isiZulu effect"?

What explains the surprising result of a negative performance of learners in isiZulu intervention schools, both in terms of Home Language and English literacy? There are multiple possible hypotheses, none of which we can conclusively rule out.

5.3.1. Is it imbalance?

Even though the sample as a whole is balanced, it is possible that the sub-samples of Siswati- and isiZulu-speaking learners are imbalanced, because we did not stratify by language when doing randomized assignment. Table 36 in the appendix shows that it is indeed the case that within isiZulu schools, the learners in the control performed better than learners in the two intervention arms. These differences are small and barely significant, though.

In Table 24 below we test if differences in baseline ability can explain the result, by controlling for each baseline measure of learner reading proficiency separately and interacting with treatment. The negative treatment effect of the virtual coach on aggregate reading proficiency in the Zulu sample only goes down slightly – from 0.36 to 0.3 standard deviations – and remains statistically significant. Note that floor effects at baseline means that we could not truly capture the full variation in learner ability. For example, 98.6, 100, and 99.6 percent of learners in the control, on-site coaching and virtual coaching interventions respectively could not read a single English word at baseline.

Table 24: Home language oral and reading skills for isiZulu and Siswati schools, controlling for baseline reading proficiency

| | (1) | (2) | (3) | (4) | (5) | (6) |
| | | isiZulu | | | Siswati | |
	Object Naming	*Letter Naming*	*Aggregate*	*Object Naming*	*Letter Naming*	*Aggregate*
On-site coach	-1.089**	-2.036	-0.253	-0.127	0.467	0.080
	(0.417)	(1.248)	(0.163)	(0.302)	(0.621)	(0.094)
Virtual coach	-1.780***	-1.820*	-0.300*	-0.458*	-0.769	-0.124
	(0.573)	(1.065)	(0.165)	(0.241)	(0.616)	(0.083)
Observations	758	758	757	2,002	2,002	2,002
R-squared	0.242	0.306	0.308	0.203	0.310	0.333
P-value	0.241	0.887	0.821	0.293	0.0798	0.0448
Control mean	17.04	19.46	0.113	16.42	18.05	-0.0457

Traditionally, performance on RAN assessment were interpreted to reflect a fairly stable aspect of brain functioning (Lervag & Hulme, 2009). However, there seems to be divergent evidence on whether RAN performance is enhanced by training or focussed activities – some studies have found no such effect (Lervag & Hulme, 2009; Kirby, et al., 2010), whereas other have found an effect (Fugate, 1997; De Jong & Vrielink, 2004; Vander Stappen, et al., 2018).

If we were to interpret performance in the two RAN tasks – object and letter naming – to be a measure of innate reading ability that cannot be influenced, then this would suggest imbalances in the isiZulu sample. Table 22, columns (1) and (2), shows statistically significant negative treatment effects for the virtual coaching intervention on these two tasks, and a slight negative treatment effect on object naming in the on-site coaching intervention.

Our analysis shows a different pattern of RAN in the two RAN tasks administered in Wave 3 (Figures 17 and 18). While the Siswati sub-sample show similar RAN score distributions across the intervention groups, the isiZulu sub-sample exhibits significant differences with both intervention groups performing below the control group. However, since the literature is not conclusive on whether performance on RAN can be influenced through an intervention, and since we did not test RAN at baseline, we can only interpret the scores to reflect differences in the ability of learners to achieve reading fluency, rather than systematic differences in the sample.

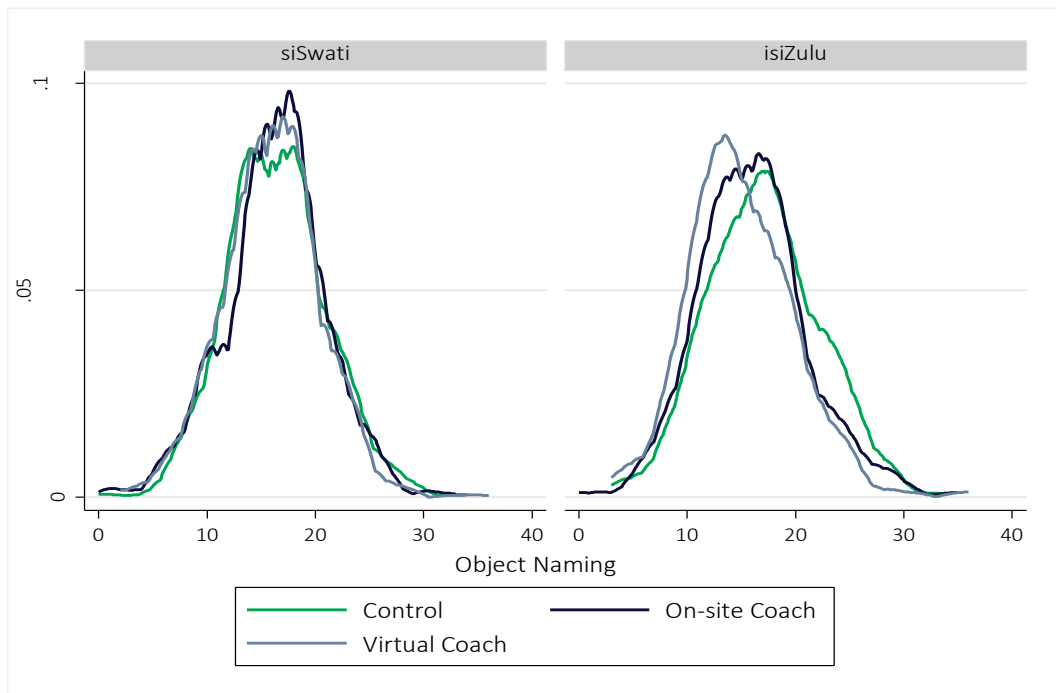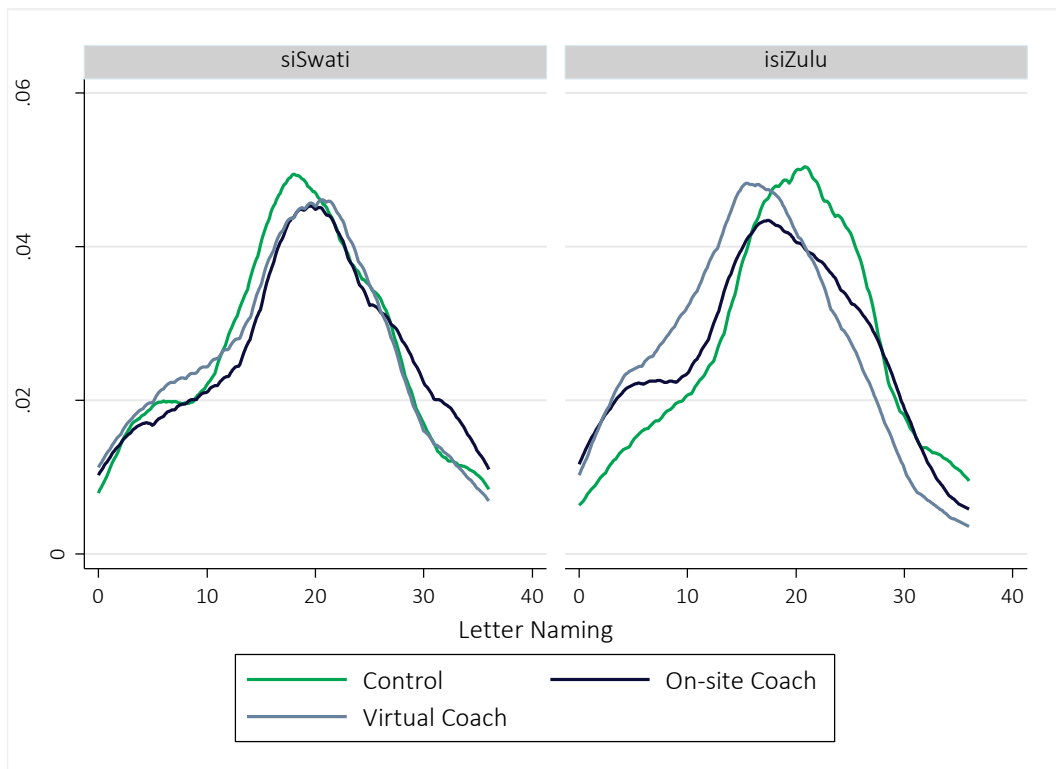*Figure 17: Grade 2 Home Language object naming (RAN) by learner language*

Figure 18: Grade 2 Home Language Naming by language

The differences in the RAN scores remain even after controlling for baseline performance (as shown in Table 24). Controlling for baseline performance assumes that the baseline

indicators have adequately capture learner actual potential. There are, however, reasons to doubt that the baseline performance will adequately account for learner ability, due the lack of inter-wave correlation, particularly for isiZulu subsample.

Table 25 shows that the alpha scores for the baseline isiZulu sample are much lower in the intervention groups than in the Siswati sample. The alpha in the isiZulu Intervention 2 group in Wave 2 is also lower. Table 26 further shows that the inter-wave correlation with the baseline performance has little correlation for the isiZulu sub-sample with subsequent performance in Waves 2 and 3. It is possible that this may be driven by the smaller sample size of the isiZulu sample.

Table 25: Alpha scores for the test sub-tasks

		Siswati	isiZulu
	Control	0.478	0.432
Wave 1	*Intervention 1*	0.487	0.381
	Intervention 2	0.415	0.213
	Control	0.686	0.604
Wave 2	*Intervention 1*	0.693	0.557
	Intervention 2	0.682	0.668
	Control	0.841	0.850
Wave 3	*Intervention 1*	0.853	0.851
	Intervention 2	0.846	0.857

Table 26 Correlation matrix for task items by language

		Siswati			isiZulu		
	Wave 1	1			1		
Control	*Wave 2*	0.446	1		0.265	1	
	Wave 3	0.444	0.799	1	0.287	0.637	1
	Wave 1	1			1		
Intervention 1	*Wave 2*	0.493	1		0.168	1	
	Wave 3	0.458	0.788	1	0.266	0.564	1
	Wave 1	1			1		
Intervention 2	*Wave 2*	0.428	1		0.337	1	
	Wave 3	0.377	0.774	1	0.379	0.711	1

The data suggest that there may have been some differences between the two language groups at baseline that were not factored into the sampling process. We have further shown in this analysis that the RAN by language and intervention groups shows some unusual patterns. While there is substantial similarities between the three groups in the Siswati sub-sample, there seems to have been slight differences in the isiZulu subsample. Finally, we have

reason to believe that the baseline scores of the isiZulu learners may not be accurate enough to control for learner ability sufficiently. This has been demonstrated by the low alpha score in the baseline assessment and the low inter-wave correlation of the baseline score with the Wave 2 and 3 scores.

5.3.2. Contextual differences: Are the isiZulu and Siswati schools very different?

A further question to consider is whether there are systematic differences between the isiZulu and Siswati language schools? And whether these characteristics may moderate the effect size? To explore these questions further, we will investigate whether there is something else about the characteristics of the isiZulu language schools driving the result, rather than merely the language.

Table 27 shows the differences between the isiZulu and Siswati schools. The majority of the teachers in both groups of intervention schools are in their forties and fifties, with the average age for the Siswati schools being almost 50 years old, and the average age in the isiZulu schools being 47.5 years old. There is, however, a significant difference in the average number of years of teaching, with teachers in Siswati schools having about 3.2 more years of experience. Teachers in these schools are also more likely to be trained to teach foundation phase classes and teach all subjects in their Grade 2 classes. Possibly one of the most telling indicators of the differences between these two groups of schools is that teachers in Siswati schools report staying much closer to their school.

While there is no significant difference in the average class size of between 42 to 43 learners per class, the average number of learners absent in any particular day is dramatically higher in the isiZulu schools (5.3 per day) compared to the Siswati schools (2.3 per day). A lack of parent support also appears to be a bigger problem in the Siswati schools.

Table 27: Contextual differences between home language isiZulu and Siswati schools.

	(1) Siswati		(2) isiZulu		T-test Difference
	N/[Clusters]	Mean/SE	N/[Clusters]	Mean/SE	(1)-(2)
Age	341	49.821	105	47.581	2.240**
	[130]	[0.438]	[49]	[0.818]	
Female	341	0.988	105	0.990	-0.002
	[130]	[0.006]	[49]	[0.010]	
Years teaching	341	19.716	105	16.457	3.258***
	[130]	[0.553]	[49]	[0.841]	
Years teaching F. Phase	341	16.147	105	14.019	2.128*
	[130]	[0.580]	[49]	[0.940]	
Bachelors degree or more	325	0.532	93	0.505	0.027
	[128]	[0.028]	[49]	[0.053]	
Multi grade	341	0.091	105	0.067	0.024
	[130]	[0.018]	[49]	[0.025]	
Formally trained in F. phase	335	0.833	103	0.728	0.105**

	[129]	[0.021]	[49]	[0.048]	
Teach all FP subjects	341	0.921	105	0.752	0.168***
	[130]	[0.019]	[49]	[0.055]	
Class size	341	42.537	104	43.279	-0.742
	[130]	[0.768]	[49]	[1.363]	
No. learners absent on ave day	341	2.238	104	5.346	-3.109**
	[130]	[0.142]	[49]	[1.317]	
Stay close to school	341	0.601	105	0.362	0.239***
	[130]	[0.033]	[49]	[0.051]	
Time to school < 15 min.	341	0.548	105	0.400	0.148**
	[130]	[0.029]	[49]	[0.053]	
Teacher test - % correct	341	0.587	105	0.601	-0.014
	[130]	[0.012]	[49]	[0.021]	
How much of a problem?					
Shortage Language Workbooks	339	3.540	98	3.398	0.142
	[130]	[0.054]	[49]	[0.135]	
Shortage Readers	338	2.663	100	2.590	0.073
	[130]	[0.077]	[49]	[0.145]	
Learner absence	341	2.997	100	2.620	0.377***
	[130]	[0.058]	[49]	[0.100]	
Overcrowding	341	2.768	100	2.580	0.188
	[130]	[0.088]	[49]	[0.132]	
Lack parental support	340	2.141	100	1.860	0.281**
	[130]	[0.074]	[49]	[0.105]	
Lack of learner language skills	336	2.420	98	2.582	-0.162
	[130]	[0.065]	[49]	[0.126]	

These differences are mirrored in the school principal survey. Notably, the isiZulu schools are less likely to be considered to be "located in an area that attracted teachers to the school". This would explain the relatively higher teacher turnover: teachers in the isiZulu schools are younger and less experienced.

Table 28: Teacher Implementation, full sample

	HL Lesson Plans	EFAL Lesson Plans	Uses NGO Lesson Plans	EFAL Graded Readers	HL Graded Readers
On-site Coach	1.677	0.005	0.399***	0.180***	-0.038
	(1.918)	(0.005)	(0.053)	(0.053)	(0.064)
Virtual Coach	-1.199	-0.012	0.239***	0.204***	-0.065
	(1.239)	(0.013)	(0.069)	(0.055)	(0.066)
Observations	446	438	446	446	446
R-squared	0.015	0.027	0.163	0.080	0.061
P-value	0.109	0.155	0.0100	0.629	0.720
Control mean	2.995	0.995	0.495	0.607	0.617

These differences should only matter if it means that the quality of implementation in these schools was different. We turn to this below. The results on the intermediate outcomes are, however, what to be expected. In both intervention groups, the schools are more likely to

use NGO-provided lesson plans and EFAL graded readers. However, there is no difference in the likelihood of using home language graded readers.

When we disaggregate by language, we see that in the isiZulu schools, the on-site coaching teachers say they are much more likely to use NGO lesson plans than teachers in the virtual coaching group. Teachers who receive on-site coaching are also more likely to use EFAL graded readers than the control group. However, the teachers in the isiZulu on-site coaching group indicated that they are marginally less likely to use Home Language graded readers than the isiZulu control schools.

Table 29: Teacher Implementation, isiZulu schools sample only

	HL Lesson Plans	EFAL Lesson Plans	Uses NGO Lesson Plans	EFAL Graded Readers	HL Graded Readers
On-site Coach	2.262	0.002	0.552***	0.269**	-0.177*
	(6.020)	(0.008)	(0.106)	(0.116)	(0.105)
Virtual Coach	1.544	-0.059	0.171	0.232	-0.006
	(6.554)	(0.053)	(0.157)	(0.139)	(0.111)
Observations	105	100	105	105	105
R-squared	0.036	0.093	0.446	0.129	0.112
P-value	0.911	0.272	0.00860	0.787	0.213
Control mean	6.765	1	0.314	0.451	0.431

Teachers in the Siswati intervention schools are more likely to use NGO lesson plans and EFAL graded readers than teachers in the Siswati control schools, suggesting a difference in use of core resources between isiZulu and Siswati teachers in the virtual coaching group. The isiZulu teachers in the virtual coaching group are as likely to use NGO lesson plans and EFAL graded readers as the control teachers who received neither. In suggests that isiZulu teachers are not really engaging with these two key intervention resources.

Table 30: Teacher implementation in Siswati schools

	HL Lesson Plans	EFAL Lesson Plans	Uses NGO Lesson Plans	EFAL Graded Readers	HL Graded Readers
On-site Coach	2.258	0.005	0.309***	0.146**	-0.017
	(1.651)	(0.006)	(0.060)	(0.059)	(0.074)
Virtual Coach	-0.287	-0.005	0.193***	0.174***	-0.092
	(0.611)	(0.010)	(0.072)	(0.063)	(0.073)
Observations	341	338	341	341	341
R-squared	0.027	0.030	0.120	0.068	0.073
P-value	0.0768	0.245	0.0772	0.577	0.368
Control mean	1.669	0.993	0.559	0.662	0.683

The second section of the teacher interview explores what teachers say about their enactment of key teaching methodologies or techniques. Group-guided reading is one of the

most important approaches that teachers have to provide guidance to individual learners in both decoding and comprehension. Shared reading provides important opportunities for modelling reading practice and vocabulary development, and creative writing is a key link to consolidating the reading process.

All three groups of teachers reported the same use of individual reading and phonics as a teaching technique, but teachers in the on-site coaching were more likely to use ability grouping and shared reading. Teachers in both groups reported regular use of group guided reading and creative writing.

Table 31: Frequency of use of various pedagogic techniques

	Reading HL	Reading EFAL	Group by ability	GGR	Spelling Test	Phonics	Shared Reading	Creative Writing
On-site Coach	0.026	0.161	0.072**	0.255*	0.115	-0.152	0.303***	0.236**
	(0.100)	(0.117)	(0.036)	(0.146)	(0.133)	(0.096)	(0.096)	(0.110)
Virtual Coach	0.094	0.152	0.056	0.332**	0.147	-0.111	0.143	0.225**
	(0.087)	(0.103)	(0.034)	(0.131)	(0.129)	(0.103)	(0.110)	(0.108)
Observations	427	434	446	441	434	441	439	438
R-squared	0.025	0.040	0.042	0.023	0.024	0.025	0.031	0.035
P-value	0.514	0.941	0.553	0.598	0.822	0.665	0.106	0.921
Control mean	4.387	4.171	0.883	3.875	3.255	4.307	3.916	3.693

Teachers in the intervention groups were less likely to find various key techniques to be difficult including group guided reading, phonics, shared reading and creative writing. This is in contrast to the findings in the Early Grade Reading Study 1, were teachers in the intervention schools found these techniques just as difficult as those in the control schools. For both intervention groups, teachers reported the correct frequency of phonetic repetition and the print-richness of the environment is higher. An analysis of the workbooks also attest to the more frequent writing of full paragraphs in both intervention groups.

Table 32: Teachers views on pedagogic techniques

	GGR is difficult	Phonics is difficult	Shared Reading is difficult	Creative Writing is difficult	Correct phonetic repet	Print richness	# of written acts	# full sentences	# full paragraph
On-site	-0.229**	-0.311***	-0.353***	-0.456***	0.283***	0.517***	-4.254**	-0.930	0.475**
	(0.105)	(0.087)	(0.096)	(0.108)	(0.052)	(0.086)	(1.738)	(0.640)	(0.239)
Virtual	-0.161	-0.240**	-0.272***	-0.321***	0.130**	0.397***	-3.851**	-0.597	0.480*
	(0.100)	(0.093)	(0.091)	(0.114)	(0.060)	(0.094)	(1.752)	(0.614)	(0.277)
Obs	419	439	439	433	446	446	443	442	441
R^2	0.031	0.051	0.052	0.056	0.090	0.162	0.056	0.044	0.038
P-value	0.530	0.429	0.439	0.249	0.00949	0.233	0.846	0.601	0.988
C mean	2.479	1.870	2.046	2.550	0.520	-6.27e-10	26.75	7.154	0.902

The overall pattern however masks large differences between teachers in the isiZulu and Siswati schools. In the Siswati schools teacher spent more time practicing individual reading of English, which was not the case in the isiZulu schools.

Table 33: Frequency of use of techniques in isiZulu schools

	Freq of GGR	Freq of Phonics	Freq of Shared Reading	Freq of Creative Writing	GGR is difficult	Phonics is difficult	Shared Reading is difficult	Creative Writing is difficult
On-site Coach	-0.201	0.161	0.571***	0.035	-0.010	-0.192	-0.149	-0.392*
	(0.376)	(0.219)	(0.192)	(0.268)	(0.191)	(0.157)	(0.181)	(0.212)
Virtual Coach	0.013	-0.151	-0.189	-0.096	-0.129	-0.072	0.041	-0.475*
	(0.325)	(0.195)	(0.239)	(0.286)	(0.179)	(0.249)	(0.272)	(0.241)
Observations	102	102	102	101	95	101	99	99
R-squared	0.048	0.093	0.157	0.082	0.121	0.108	0.127	0.130
P-value	0.597	0.179	0.00161	0.646	0.559	0.617	0.502	0.745
Control mean	4	4.063	3.833	3.833	2.340	1.857	2.020	2.571

In contrast, teachers in both interventions in the Siswati intervention schools reported more frequent use of group guided reading, phonics teaching, shared reading and creative writing, and found these techniques less difficult in than teachers in the control schools.

Table 34: Teachers view on pedagogy in Siswati schools

	Freq of GGR	Freq of Phonics	Freq of Shared Reading	Freq of Creative Writing	GGR is difficult	Phonics is difficult	Shared Reading is difficult	Creative Writing is difficult
On-site Coach	0.389**	-0.263**	0.247**	0.289**	-0.300**	-0.346***	-0.407***	-0.481***
	(0.158)	(0.107)	(0.111)	(0.122)	(0.120)	(0.098)	(0.109)	(0.125)
Virtual Coach	0.483***	-0.216*	0.192	0.315***	-0.202*	-0.294***	-0.330***	-0.297**
	(0.143)	(0.112)	(0.119)	(0.115)	(0.117)	(0.108)	(0.102)	(0.130)
Observations	339	339	337	337	324	338	340	334
R-squared	0.046	0.057	0.038	0.043	0.055	0.067	0.082	0.070
P-value	0.527	0.647	0.594	0.828	0.413	0.597	0.492	0.160
Control mean	3.833	4.389	3.944	3.645	2.528	1.875	2.056	2.542

The implementation of the lesson plans and core resources can be seen through the responses on the more frequent use of group guided reading and share reading, as well as the teachers' attitudes towards these techniques. But the real significant differences was between teachers in the isiZulu and Siswati samples. Specifically, teachers in the isiZulu interventions schools appear far more like teachers in the control group when it comes to their reported use of new techniques and attitudes towards those using these techniques. It

might be that the take-up of the new techniques in the isiZulu schools may be linked to the unique contextual factors common to the isiZulu schools in Mpumalanga.

One of the original hypotheses that were tested was relative number of hours schools used for Home Language and EFAL. It was suggested that since the intervention focused on English, the majority of intervention schools would opt for greatest number of hours in this language, and the lowest possible hours in the Home Language. The evidence, however, shows no difference in number of hours that the school principals say the school dedicates to Home Language vs English.

5.4. Learner-level heterogeneity[4]

The descriptive statistics on the specific sub-tasks suggested that the interventions may have benefitted learners at the top end of the distributions more. Figure 19 below explores this notion further by showing quantile regressions on the aggregate score for English reading proficiency, controlling for strata fixed effects. It confirms results from the inverse cumulative distribution functions: the impact of the on-site coach intervention is really driven by the top-end of the distribution. There is no impact for the bottom 10% of learners (because everyone scored zero). The effect size becomes positive and increases in magnitude starting at the sixth decile, and there is only a positive statistically significant effect for the top 20% of learners. The narrow confidence intervals for the bottom deciles is due to the fact that everyone at the bottom scored close to zero.

Figure 20 shows that there is a similar trend for the virtual coaching intervention: learners in the top quintile improved their learning by 0.08 SD, and there is a far smaller effect size for learners in the bottom decile. The pattern is slightly non-monotonic, though, with the smallest effect size for learners in the middle of the distribution.

---

[4] Heterogeneity was tested on learner gender and learner age as well, but no significant interaction effects were found.

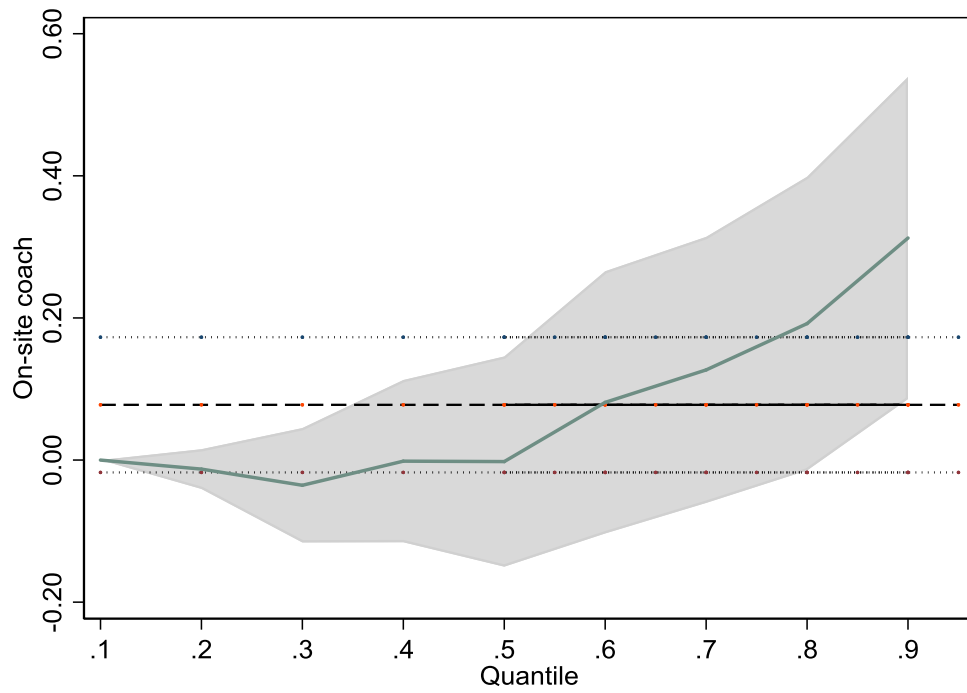*Figure 19: Quantile regression of the On-site Coach intervention on aggregate English reading proficiency*

Figure 20: Quantile regression of the Virtual Coach intervention on aggregate English reading proficiency
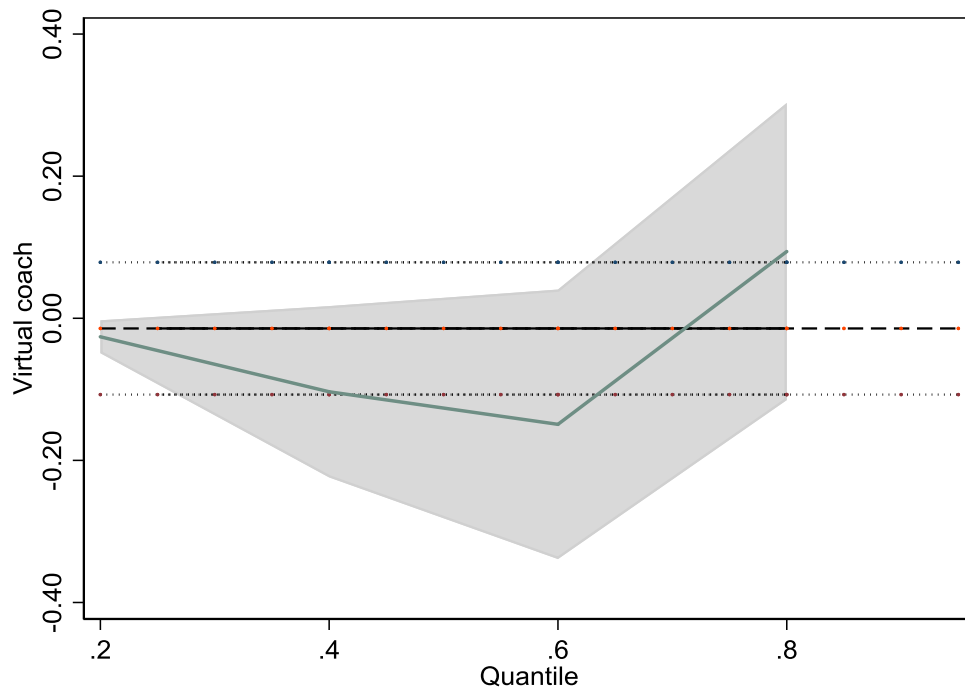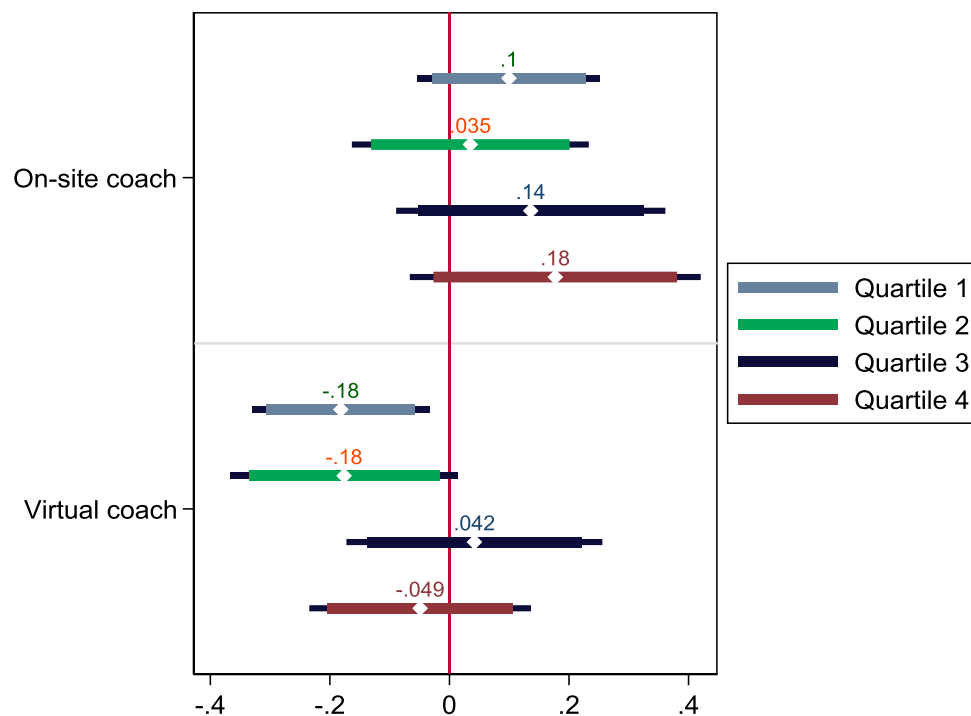
Figure 21 shows impacts on English literacy, by learner performance in the baseline reading assessments. It shows that in the virtual coaching intervention, learners who scored in the bottom half at baseline are *negatively* affected by the intervention. Beyond that, the effect sizes do not vary much by baseline reading proficiency.

Figure 21: Impacts on English reading proficiency, by quartile of baseline reading proficiency

It is not surprising, then, that the impacts on English literacy are largest at the top end of the learner distribution. Figure 21 below shows the different proportion of learners that reach key benchmarks in English ORF, restricted to the Siswati Grade 2 learner sub-sample. At each of the key benchmarks (40 and 45 for emergent literacy, and 65 for reading for meaning), the On-site and Virtual Coaching groups outperforming the control learners.

Figure 22 Percentage of Grade 2 learners English oral reading fluency, Siswati only
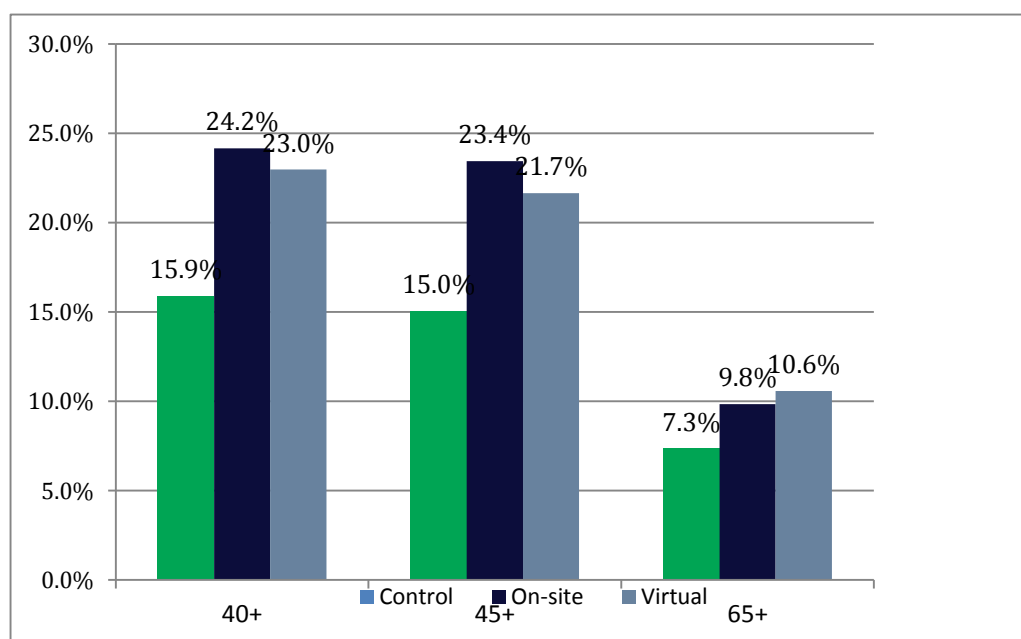
Turning to impacts on Home Language literacy, the large ceiling effects for ORF, makes it impossible to run quantile regressions. However, showing the intervention effects for different cut-offs of number of words read: at least one word, at least 10 words, at least 20 words, and at least 30 words. For all cut-offs and for both interventions there is a negative effect. However, for both interventions, the magnitude is largest at the cut-off of one word. In other words, a larger proportion of learners in the intervention arms cannot read a single word. This pattern is most pronounced in the virtual coaching intervention group.
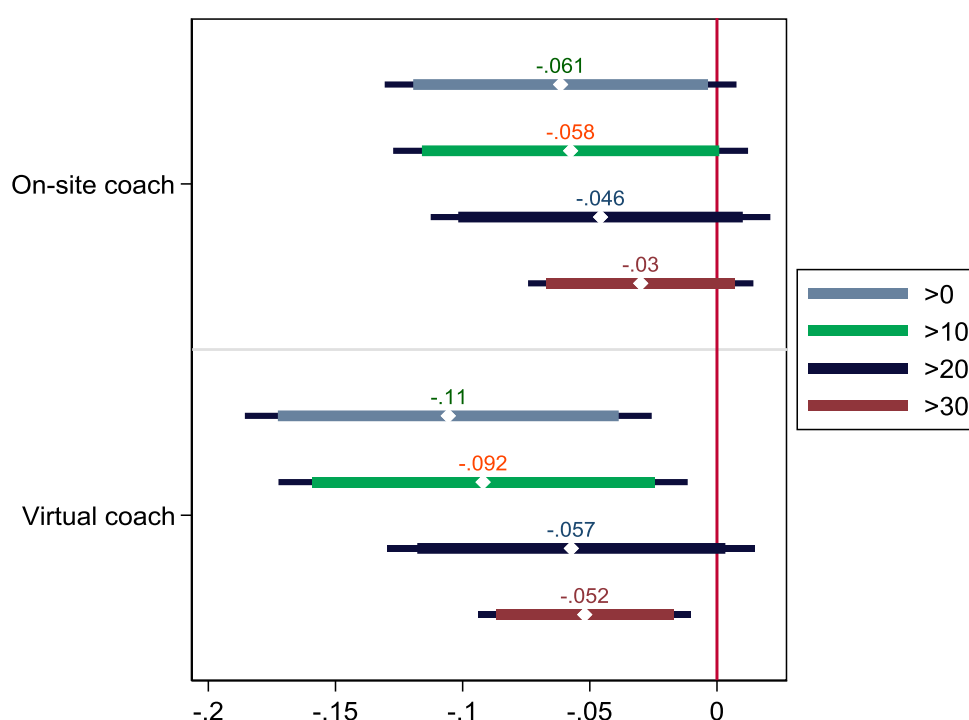
5.5.   Case studies findings

The questions that animated the 2018 EGRS II case studies in the Mpumalanga focused on better understanding curriculum coverage, the *how* and the *why* of the coaching process and the enactment of one particular methodology, i.e. group guided reading.

Botha and Schollar (2019) found that the structured pedagogic programme was more often than not about formal coverage and less about substantive and deep changes in instructional practices. They provide evidence that there is a disjuncture between the need to be fully curriculum compliant, particularly around the number of weeks of teaching required and the reality in schools, which gives teachers considerably less real time for teaching. This left teachers in a situation where they felt that they always need to catch up. Botha and Schollar also observed the challenges associated with how teachers had to address the needs of the slower learners within a policy context designed to limit grade repetition. Arguably, the most important insight from their case studies related to differential uptake of specific methodologies, as expressed as activities within the lesson plans. The lesson plan activities in English oral language development and writing were simple and easy to follow and were rapidly incorporated into teachers' daily instructional repertoire. In contrast, and consistent with the experience with the Home Language programme, teachers struggle to master the methodology of group guided reading, with the majority of learners reading at frustration

level, rather than 'just right' level and poor classroom management associated with those not in the guiding reading group.

Alsofrom's (2018) qualitative case study of coaching explored the psychological and professional aspects of coaching. Her main finding is exemplified by a quote from one of the teacher interviews: "We are afraid of the change, not knowing that it will help us." Her research shows that fear is the prevalent emotion associated with large-scale interventions and policy in general. And fear is the major barrier to real teacher learning. She shows that, in the coaching literature, 'trust' between the teacher and coach is often discussed as transactional – as a simple foundation necessary only for the transfer of skills and knowledge. The relationship is not discussed as an intrinsic and central part of coaching, valuable in its own right. The feeling of support, which emerges from trust, ultimately opens up a space for making mistakes and taking risks. Thus, this research shows that it is the 'feeling of support' that has the power to influence teachers to change their practice.

6. Discussion

From the descriptive statistics of the English language tasks we see very little difference between the control school means and those of the mean task scores for the two interventions. This is confirmed in the main regression result tables, in the table without controls, the main results with all the main variables and the main results controlling specifically key variables. These analyses would suggest that beyond the oral language skills that showed promising effects in Year 1, the interventions are not impacting positively on phonic and sight word reading, ORF, and reading comprehension and (b) that because neither intervention is showing positive impact, there is no conclusive difference in effectiveness between the two intervention models.

The basic results, however, obscure substantially different patterns when we look at the two languages, isiZulu and Siswati separately. The descriptive statistics show a very substantial difference in the mean English task scores between the learners tested in the two languages. This is confirmed in the regression tables with the Siswati group showing a consistent positive statistically significant impact across most of the English tasks for the on-site coaching group only.

The third main set of findings related to how the interventions performed for different parts of the learner distribution. This is first evident in an analysis of the percentage of learners that reach certain ORF benchmarks, particularly, 40, 45 and 65 wcpm. Using the descriptive statistics, we observe that at all three benchmarks, both interventions have a substantially higher percentage of learners achieving at the level than in the control group. This however is not confirmed in the quartile regression analysis, based on learner baseline performance.

How are we to interpret these results in light of the two main research questions after two years of the intervention? On the question of if, and to what extent the structured programme works to improve English reading and writing, we cannot provide a definitive answer at this point. Some evidence (whole sample) suggests that there is no difference between the performance of the control and in the intervention schools, but when disaggregated and examining the Siswati sub-sample only (which represents about 72% of the total sample), we see that the on-site coaching group of schools is performing consistently better than the control schools on all the English tasks. This is not true for the virtual coaching schools. However, we have no clear explanation at this point for why there should be such a different pattern for the isiZulu intervention schools relative to the Siswati intervention schools.

The other major insight relates to the fact that the interventions appear to be working most effectively for the stronger learners/stronger schools and specifically for the on-site coaching model. There are at least two ways that we could explain this. First, that the intervention is only at the right level for the 20-30% of the top learners and certainly not appropriate for schools were the vast majority of learners are scoring zero on the Home Language and English text reading tasks. In Grade 2, the learning programmes which are aligned to CAPS, are moving forward and concentrating on higher levels of learning activities and are thus not really providing these learners with what they need to start the journey towards reading. The second is that the schools at the middle or bottom of the distribution would take more than one year to master the approach and that impact would only be visible after a number of years.

References

De Jong, P. & Vrielink, L., 2004. Rapid automatic naming: Easy to measure, hard to improve (quickly). *Annals of dyslexia,* 54(1), pp. 65-88.

Fleisch, B. & Christie, P., 2004. Structural Change, Leadership and School Effectiveness/Improvement: Perspectives from South Africa. *Discourse: Studies in the Cultural Politics of Education,* 25(1), pp. 95-112.

Fleisch, B. & Schoër, V., 2014. Large-scale instructional reform in the Global South: Insights from the mid-point evaluation of the Gauteng Primary Language and Mathematics Strategy. *South African Journal of Education,* 34(3), pp. 1-11.

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.

Fugate, M., 1997. Letter training and its effect on the development of beginning reading skills. *School Psychology Quarterly,* Volume 12, pp. 170-192.

Hoadley, U., 2016. *A review of the research literature on teaching and learning in the foundation phase in South Africa,* Stellenbosch: Research on Socioeconomic Policy, Department of Economics, Stellenbosch University.

Kirby, J., Georgioiu, G., Martinussen, R. & Parrila, R., 2010. Nming speed and reading: From prediction to instruction. *Reading Research Quarterly,* 45(3), pp. 341-362.

Lervag, A. & Hulme, C., 2009. Rapid automised naming (RAN) taps a mechanism that places constraints on the development of early reading fluency. *Psychological Science,* 20(8), pp. 1040-1048.

Vander Stappen, C., Reybroeck & M.V, 2018. Phonological awareness and rapid automatized naming are independent phonological competencies with specific impacts on word reading and spelling: an intervention study. *Frontier in Psychology,* Volume 9, p. 320.

Appendix

Imbalance on Sample, by Language

Table 35 tests for balance separately for the sample of learners who are in isiZulu and Siswati schools. Columns (1) and (3) show the mean difference between the on-site coaching and control schools, whereas columns (2) and (4) show the difference between the virtual coaching and the control schools. A positive number indicates that the average is higher in the control than the intervention group. There seems to be slight imbalance in the sample of isiZulu learners: learners in the control perform better than learners in the intervention groups at baseline for the total number of sentence words read correct. The p-value of the F-statistic which is lower than 0.1 means that we can now reject the null of joint significance of all the sub-tasks, when comparing either intervention group with control. This is not the case for the Siswati sub-sample.

Table 35: Balance tests per task: isiZulu vs Siswati group sub-sample

	(1)	(2)	(3)	(4)
	isiZulu sample		Siswati sample	
	On-site vs control	Virtual vs control	On-site vs control	Virtual vs control
	(1)-(2)	(1)-(3)	(1)-(2)	(1)-(3)
Naming Animals in HL	0.200	-0.306	-0.302	-0.400*
Word Recall	0.047	0.163	0.020	-0.208
Nonword Recall	0.031	0.062	0.028	-0.061
Phoneme Isolation	0.409*	0.018	-0.039	-0.044
Story Comprehension	-0.014	0.021	0.041	-0.127*
p/m Number of Letter Sounds Correct	1.585	1.458	-0.380	-0.566
Number of Words Read Correct	0.182	-0.119	-0.020	-0.148
Sentence Words Read Correct	0.059**	0.055**	0.010	0.004
Visual Perception	-0.024	-0.037	-0.184	-0.248*
English Items	-0.008	0.087	0.070	-0.037
F-test of joint significance (p-value)	0.053*	0.069*	0.476	0.223
F-test, number of observations	697	680	1686	1723

Note. The value displayed for t-tests are the differences in the means across the groups. Standard errors are clustered at the school level. ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. The value displayed for F-tests are p-values.

Table 36: Balance on school level characteristics

	(1) Control	(2) I 1	(3) I 2	(4) Total	Difference (1)-(2)	Difference (1)-(3)
Principal age	52.861 [0.626]	51.380 [0.860]	51.540 [0.756]	52.078 [0.423]	1.481	1.321
Principal – years of experience	8.428 [0.858]	7.900 [0.914]	9.547 [1.047]	8.593 [0.542]	0.528	-1.118
Enrollment - Foundation phase	231.215 [11.467]	223.510 [15.568]	213.490 [13.964]	224.175 [7.703]	7.705	17.725
Proportion teachers vacant	0.052 [0.022]	0.000 [0.000]	0.023 [0.016]	0.029 [0.011]	0.052*	0.029
Teachers absent	3.423 [0.083]	3.520 [0.100]	3.340 [0.116]	3.427 [0.056]	-0.097	0.083
Learners absent	2.974 [0.087]	3.160 [0.083]	3.080 [0.117]	3.056 [0.056]	-0.186	-0.106
Shortage materials	2.558 [0.119]	2.673 [0.125]	2.540 [0.143]	2.585 [0.074]	-0.115	0.018
Low parental involvement	2.487 [0.104]	2.300 [0.138]	2.360 [0.156]	2.399 [0.074]	0.187	0.127
School disruption	3.269 [0.096]	3.420 [0.103]	3.380 [0.106]	3.343 [0.059]	-0.151	-0.111
Scholar transport	3.658 [0.074]	3.646 [0.096]	3.840 [0.066]	3.706 [0.046]	0.012	-0.182*
Lack of training	3.679 [0.065]	3.540 [0.111]	3.583 [0.111]	3.614 [0.052]	0.139	0.096
Days not at school	0.772 [0.147]	0.540 [0.149]	0.420 [0.159]	0.609 [0.089]	0.232	0.352
Days not full school day	1.380 [0.254]	0.640 [0.148]	1.040 [0.204]	1.078 [0.134]	0.740**	0.340
Stay close to school	0.481 [0.057]	0.460 [0.071]	0.500 [0.071]	0.480 [0.038]	0.021	-0.019
Stay elsewhere in weekend	1.620 [0.089]	1.580 [0.107]	1.720 [0.114]	1.637 [0.059]	0.040	-0.100
stay < 15 minutes from school	0.456 [0.056]	0.440 [0.071]	0.520 [0.071]	0.469 [0.037]	0.016	-0.064
School located in good area	0.595 [0.056]	0.680 [0.067]	0.480 [0.071]	0.587 [0.037]	-0.085	0.115
Access to internet	0.747 [0.049]	0.860 [0.050]	0.840 [0.052]	0.804 [0.030]	-0.113	-0.093
No reading books for library	0.709 [0.051]	0.660 [0.068]	0.680 [0.067]	0.687 [0.035]	0.049	0.029
Also teach	0.662 [0.054]	0.620 [0.069]	0.740 [0.063]	0.672 [0.035]	0.042	-0.078
F. Phase teachers retired since 2017	0.544 [0.168]	0.440 [0.091]	0.260 [0.085]	0.436 [0.082]	0.104	0.284
State of buildings	2.911 [0.116]	2.720 [0.140]	2.960 [0.140]	2.872 [0.075]	0.191	-0.049
N	80	50	50	180		