# User guide

EGRS II – Wave 1-5 Panel

Data was collected in South Africa in two districts in the Mpumalanga province as a part of the Second Early Grade Reading Study (EGRS II) evaluation. The randomised controlled trial (RCT) testing an LTSM and coaching intervention was run by the Department of Basic Education in Partnership with the Mpumalanga Department of Education and the University of the Witwatersrand, funded by USAID.

<u>Experimental design</u>

Two interventions were tested as a part of this RCT. Intervention 1 was LTSM+ paper-based lesson plans and in-person coaching, whilst Intervention 2 was LTSM+ digital lesson plans on tablets and virtual coaching. In total 180 schools were a part of the evaluation, with 80 schools in the control group and 50 in each of the treatment arms. At each school, 20 learners were selected to be tested. In some schools, fewer than 20 learners were included in the baseline sample if there were fewer than 20 children in the grade or due to other issues relating to fieldwork or data capturing and matching [1].
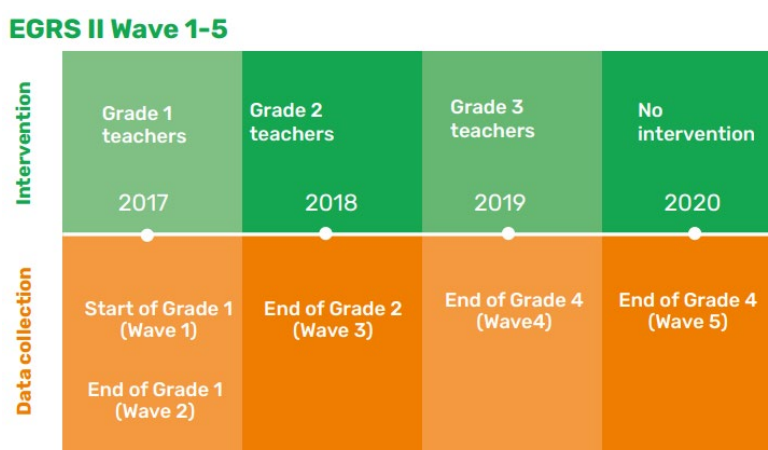
More information about the evaluation can be obtained from the EGRS website and the following reports and papers:

Website:

- [Early Grade Reading Study (education.gov.za)](#)

Paper:

Cilliers, J., Fleisch, B., Kotze, J., Mohohlwane, N., Taylor, S., & Thulare, T. (2022). Can virtual replace in-

person coaching? Experimental evidence on teacher professional development and student

---

[1] Some examples of possible issues include, learners that were not able to be matched back to the learner identification data (linking file) or running out of time at the school to complete all of the assessments.

learning. *Journal of Development Economics*, *155*, 102815.

https://doi.org/10.1016/j.jdeveco.2021.102815

Wave 1 – Baseline Report:

- EGRS II Baseline Report_Final_WEB.pdf (education.gov.za)

Wave 2 – Year 1 Report:

- EGRS II Year 1 Report - Final Midline Report (education.gov.za)
- Case study (EFAL): EGRS II EFAL Case Study 2017_Final_WEB.pdf (education.gov.za)

Wave 3 – Year 2 Report:

- The Second Early Grade Reading Study: Year 2 Report (Wave 3 – Draft report)
- Qualitative Report: How and why does coaching work to improve teaching practices in the EGRS II? An examination of causal mechanisms
- Case study (Instructional Practices): EGRS II Case Study 2018_Final_WEB.pdf (education.gov.za)

Wave 4 – Year 3 Report:

- Technical Report: The Second Early Grade Reading Study: Year 2 Report (Wave 3 – Draft report)
- Fieldwork Report: Data collection for the Early Grade Reading Study II in Mpumalanga (EGRS II) (Wave 4), Dec 2019
- Case study (Virtual Coaching): 2019_Virtual Coaching Model Summary REVISED.pdf (education.gov.za)
- EFAL Classrooms in 6 Schools: EGRS II-Case Study 2019.pdf (education.gov.za)

Wave 5 – Year 4 Report:

- EGRS II Technical Report (education.gov.za)

Data collection

A total of 3 327 learners were assessed, surveyed and matched in the first wave. These learners were then tracked over the following 4 years, with the number of learners that were reassessed and matched in the following years recorded below.

Table 1: Learners assessed and dates of assessment for the five waves of data collection

Wave	Learner observations	Data collection dates	Grade (main)
1	3 327[2]	23 Jan 2017 – 31 Mar 2017	Grade 1
2	3 068	17 Oct 2017 – 14 Nov 2017	Grade 1
3	2 764	22 Oct 2018 – 9 Nov 2018	Grade 2
4	2 684	27 Oct 2019 – 15 Nov 2019	Grade 3
5	2 410	4 Nov 2020 - 11 Dec 2020	Grade 4

[2] A total of 3482 learners were tested, but not all of them were matched.

In each wave learners' assessments were completed, as well as a combination of contextual surveys covering principals, teachers and parents (care-givers). Table 2 gives an indication of which assessments and questionnaires were conducted in each wave.

Table 2: Overview of the types of assessments, surveys completed and datasets merged for each wave:

Wave	Learner Assessment Oral (LA) [HL and L1]	Learner Assessment Written (LAW)	Learner details (LF)	Linking form (LF)	Teacher Questionnaire (TQ)	Teacher Assessment (XX)	Principal Questionnaire (PQ)	Home Background (Parent) Questionnaire (HBQ)	Linked EMIS Masterlist information (S)	
1	x		x	x	x	x	x	x	x	
2	x				x	x	x	x		
3	x	x			x	x	x	x		
4	x	x			x	x	x			
5	x				x					

Each variable in the dataset is named using w[X]_[YY]_[Variable name]. Here [X} indicates the number of the wave. Then [YY] indicated the survey instrument from which this variable was obtained. Lastly, the [Variable name] identifies the relevant question or variable in the survey.

Identifiers

Learner: The unique_id is an anonymised identifier that uniquely identifies each of the 3,327 learners in the panel. This unique_id allows for the tracking of each individual learners over the 5 waves.

School: School names and EMIS numbers are not included in the dataset, instead, an anonymised school_id can be used to recognize schools. This school id is set to the id of the school of the learners in Wave 1, as in general children were not tracked in general when they moved schools. There are a total of 180 unique school_ids in the dataset.

Principal: Principal_id These came from the Principal questionnaires and were matched across waves. Principals were surveyed in Waves 1-4, but not in Wave 5. It is therefore possible to identify schools where the principal (or the person who answered the principal questionnaire) changed. A note that for Wave 2, there seemed to be numerous schools where another staff member stepped in for the principal to complete the questionnaire, and it is therefore recommended that analysis of changes in principals be conducted using Wave 1, 3 and 4 only.

Teacher: Teacher_id's come from the teacher survey. This data tends to be incomplete and is not necessarily a perfect indicator of which class a learner belongs to. These were assigned separately for each wave. The same teachers cannot therefore be linked across waves.

The table shows the number of unique teacher id's across each wave:

	Wave 1	Wave 2	Wave 3	Wave 4
Number of teachers	278	306	439	283

The table below shows the number of enumerators on each of the assessments and surveys:

Survey/Assessment	Number of enumerators				
	Wave 1	**Wave 2**	**Wave 3**	**Wave 4**	**Wave 5**
Learner Assessment (Oral)	39	31	34	39	34
Learner Assessment (Wr.)				39	33
Teacher Questionnaire	28	27	31	36	
Principal Questionnaire	25	25	34	32	

Learner Assessments

In each wave, a set of age-appropriate assessments were selected for inclusion in the learner assessment. The EGRS II evaluation focussed mainly on supporting teachers in English as a First Additional Language (EFAL), and thus a large share of the assessment, particularly in higher grades was on EFAL outcomes. The home language (HL), which was also used as the language of instruction in the first three grades was either siSwati or isiZulu.

In Table 3, the different constructs that were assessed are indicated. For each wave of data collection there is an indicator in the table that shows whether learners completed a task that assessed this construct in either home language or EFAL, by indicating the task number of the task in the assessment. A note that for the numbering in Wave 5, the written assessment started on Task 7, even though there were only 5 tasks in the oral assessment (so there is no task 6).

Table 3: Learner Assessment tasks in the various waves of data collection

Skill	Construct	Baseline (W1) Start of Gr 1		Year 1 (W2) End of Gr 1		Year 2 (W3) End of Gr 2		Year 3 (W4) End of Gr 3		Year 4 (W5) End of Gr 4	
		HL	EFAL	HL	EFAL	HL	EFAL	HL	EFAL	HL	EFAL
Language comp.	Receptive Vocabulary		9		8						
	Expressive Vocabulary	1	9	1	8		8		6		4
	Listening comprehension	4			7		9		7		5
	Receptive proficiency				6		7				
Decoding	Phonological working mem.	2									
	Phonological awareness	3			3						
	Rapid Letter naming					2		1			
	Letter sound recognition	5			4	3		2			
	Word reading fluency	6		2	5		5		4		2
	Sentence reading fluency	7									
	Oral Reading Fluency (ORF)					4.1	6.1	3.1	5.1	1.1	3.1
	Reading comprehension					4.2	6.2	3.2	5.2	1.2	3.2
	Written comprehension							8	9	7	9
Writing & spelling	Writing a CVC word				9						
	Writing two words						10				
	Receptive vocabulary									8	10
Visual & Cognitive	Visual perception	8									
	Rapid Object Naming					1					
Maths	Arithmetic							10 (Wr 3)		11 (Wr 5)	

Note: The numbers shown in the table represent the number of the task in that assessment.

A summary table of the key statistics is provided in Table 4 to Table 8. For tasks with multiple measures (eg. ORF at 60s, at 180s and accuracy in %) only the key measure is reported in the table.

The naming of the outcome variables shown in the table and used in the dataset follows a consistent convention of the following form w[Wave]_LA_SS_[Lan]_[Type]_[Measure]_[Item], where:

- Wave – Indicates that data collection Wave
- Lan – Indicates whether the task was testing home language (H) or EFAL (E) or Mathematics (M).
- Type – This is shortened for to indicate the construct or general skill being tested
- Measure – (Optional) is used where additional information is needed to differentiate within the task, usually because there are numerous outcome measures included for a single task (eg. letter sound knowledge in 30 seconds or per minute)
- Item – This gives the number of the item in the assessment

<u>Notes: Data quality and data changes</u>

Standardisation and cleaning of all outcome variables: A check was completed on all outcome variables and these were standardised across the 5 waves (where feasible). In some cases, the type of assessment was the same across waves, but the standard or difficulty of the assessment was increased as the learners progressed to higher grades.

Timed tasks – Waves 1-4

One particular type of data entry error was found and corrected, which means that a small number of observations (generally <15) will have different outcomes to those in previously released datasets and reports. The main changes were within timed tasks, including rapid letter naming, letter sound recognition, word recognition (or word reading fluency) and oral reading fluency.

The most common change in Wave 1 – 4 was where an enumerator appeared to have stopped the assessment due to the learner providing an incorrect answer repeatedly, however the task was not auto-stopped, and as a result, the remaining items in the assessment were counted as correct. Particularly in the earlier waves such errors tended to be concentrated among a small number of enumerators.

The following tasks were affected:

- Wave 1: Task 5
- Wave 2: Task 4
- Wave 3: Task 3; 4.1 ; 5.1 ; 5.2; 6.1;
- Wave 4: Task 2, 4

For a small number of observations, there are clear data entry issues, where the outcome is impossible within the time taken for the assessment. However, it is not possible to tell what the correct outcome should have been. These observations were set to missing in the per-minute measures.

The following tasks were affected:

- Wave 1: Task 5
- Wave 2: Task 4
- Wave 4: Task 2, 4

Timed tasks - Wave 5

In addition, there was missing data for some of the timed tasks, where a total was calculated, however, the item level information is missing. The original total is included in this dataset, though it was not in previous versions (set to 0 in the previous version). However, it must be noted that it is not possible to get more detailed information at the letter or word level for these affected observations. In some cases, the 60s totals exist, but the 180s total is also missing, as this information appears to have not been stored correctly.

The following tasks were affected:

- Task 1 (7 obs.)
- Task 2 (9 obs.)

- Task 3.1 (13 obs.) -Items level information and # correct at 180s are missing

In Wave 5 Task 1 (HL ORF), there were 4 observations that had Number of items attempted and correct at time = 60s, but the time remaining was missing. Assumed that the learners took the full time, so set time remaining to 0.

Word reading (Task 6) – Wave 1

There are some selected enumerators, that may have accidentally, inverted the data entry. Specifically, the following 3 enumerators: 1010 ; 1016; 1013

It is highly likely that another two also used an inverted entry system for a portion, but not all, of their observations. 1031; 1029

The created value for this task (w1_LS_ss_H_wre_6) was set to missing for the first three enumerators.

Given the similarity in skill for Task 6 and 7, the differences in the outcomes are striking. It is recommended that this variable be excluded from analyses.

Phonetical awareness (Task 3) – Wave 2

The total phonemic awareness score across Task 3.1, 3.2 and 3.3 was a duplication of Task 3.3. This has been updated.

Mathematics (Task 10 – Written 3) – Wave 4

In previous version the calculated variable here was a duplication of the Task 9 outcome.

Wave 3 Parent's Questionnaire: There were some instances of missing being coded as 0's in the Parent questionnaire in Wave 3, so it is impossible to tell in hindsight, which values are 0 and which missing. Where possible avoid using this information and rather use the Wave 1 data for household-level variables.

Table 4: Table for Home Language Measures from the Learners Assessment

Renamed Var	Extra label	Wave	N	Mean	SD	Min	Max	p10	p25	p50	p75	p90
Expressive Vocabulary												
w1_LA_ss_H_voc_1		W1	3327	7.30	2.37	0.0	12.0	4.0	6.0	7.0	9.0	10.0
w2_LA_ss_H_voc_1		W2	3067	4.90	0.37	0.0	5.0	5.0	5.0	5.0	5.0	5.0
Phonemic Awareness - First Sound												
w1_LA_ss_H_paf_3		W1	3327	1.11	1.79	0.0	6.0	0.0	0.0	0.0	2.0	4.0
Rapid letter naming (per min)												
w3_LA_ss_H_rln_pm_22		W3	2764	54.47	26.36	0.0	135.0	15.0	36.0	57.0	72.0	87.0
w4_LA_ss_H_rln_pm_12		W4	2684	62.48	28.05	0.0	180.0	24.0	45.0	63.0	81.0	99.0
Letter sound recognition (per min)												
w1_LA_ss_H_lsr_pm_52		W1	3325	8.12	10.91	0.0	145.7	0.0	1.5	4.5	10.5	19.5
w3_LA_ss_H_lsr_pm_32		W3	2764	36.32	22.20	0.0	112.6	7.0	18.0	35.0	52.0	65.0
w4_LA_ss_H_lsr_pm_22		W4	2683	43.96	22.34	0.0	129.4	13.0	28.0	44.0	60.0	72.0
Word recognition												
w1_LA_ss_H_wre_6	Six words	W1	3201	0.20	0.83	0.0	6.0	0.0	0.0	0.0	0.0	0.0
w1_LA_ss_H_sre_7	Two 3-word sentences	W1	3327	0.04	0.38	0.0	6.0	0.0	0.0	0.0	0.0	0.0
w2_LA_ss_H_wre_pm_22	Per minute	W2	3066	6.43	10.09	0.0	170.0	0.0	0.0	1.2	11.0	19.2
Oral reading fluency (per min)												
w3_LA_ss_H_orf_pm_4_12		W3	2762	14.20	14.82	0.0	82.3	0.0	0.0	11.0	27.0	34.0
w4_LA_ss_H_orf_pm_3_13		W4	2492	21.94	16.24	0.0	80.9	0.0	5.8	23.2	34.4	43.3
w5_LA_ss_H_orf_pm_1_13		W5	2390	24.08	18.50	0.0	86.5	0.0	4.0	25.0	38.0	49.0
ORF Comprehension												
w3_LA_ss_H_orc_4_2		W3	2766	1.13	1.52	0.0	5.0	0.0	0.0	0.0	2.0	4.0
w4_LA_ss_H_orc_3_3		W4	2684	2.30	1.91	0.0	5.0	0.0	0.0	3.0	4.0	5.0
w5_LA_ss_H_orc_1_2		W5	2390	3.45	2.55	0.0	8.0	0.0	0.0	4.0	6.0	7.0

Table 5: Table for Home Language Measures from the Learners Assessment

Renamed Var	Extra label	Wave	N	Mean	SD	Min	Max	p10	p25	p50	p75	p90
Listening Comprehension												
w1_LA_ss_H_lco_4		W1	3327	2.20	1.16	0.0	4.0	1.0	1.0	2.0	3.0	4.0
Written Comprehension												
w4_LA_ss_H_wco_8		W4	2669	2.30	1.91	0.0	6.0	0.0	0.0	2.0	4.0	5.0
w5_LA_ss_H_wco_7		W5	2352	2.57	2.08	0.0	6.0	0.0	0.0	3.0	4.0	5.0
Word match (Receptive vocabulary)												
w5_LA_ss_H_vor_8		W5	2352	2.55	2.26	0.0	10.0	0.0	1.0	2.0	4.0	6.0
Word Recall												
w1_LA_ss_H_wrc_2_1		W1	3327	10.00	2.33	0.0	14.0	7.0	9.0	10.0	12.0	13.0
Nonword Recall												
w1_LA_ss_H_nwr_2_2		W1	3327	4.21	1.13	0.0	5.0	3.0	4.0	5.0	5.0	5.0
Rapid object naming (in 20s)												
w3_LA_ss_H_ron_1		W3	2764	16.24	4.77	0.0	36.0	10.0	13.0	16.0	19.0	22.0
Visual Perception												
w1_LA_ss_H_vsp_8		W1	3327	1.55	1.64	0.0	10.0	0.0	0.0	1.0	3.0	4.0

Table 6: Table for EFAL Measures from the Learners Assessment (Vocabulary and other skills)

Renamed Var	Variable info	Wave	N	Mean	SD	Min	Max	p10	p25	p50	p75	p90
Receptive and Expressive vocabulary												
w1_LA_ss_E_voc_9		W1	3327	0.82	1.03	0.0	6.0	0.0	0.0	0.0	1.0	2.0
w2_LA_ss_E_voc_81		W2	3060	0.74	0.83	0.0	2.0	0.0	0.0	0.0	1.0	2.0
Expressive Vocabulary												
w3_LA_ss_E_voc_8		W3	2762	4.31	1.27	0.0	6.0	2.5	3.5	4.5	5.5	6.0
w4_LA_ss_E_voc_6		W4	2684	3.29	1.78	0.0	6.0	1.0	2.0	3.0	5.0	6.0
w5_LA_ss_E_voc_4		W5	2405	17.12	7.44	0.0	30.0	5.0	12.0	19.0	23.0	26.0
Phonemic Awareness												
w2_LA_ss_E_paf_3_1	First Sound	W2	3066	1.72	1.34	0.0	3.0	0.0	0.0	2.0	3.0	3.0
w2_LA_ss_E_pal_3_2	Last Sound	W2	3066	1.35	1.36	0.0	3.0	0.0	0.0	1.0	3.0	3.0
w2_LA_ss_E_pam_3_3	Middle Sound	W2	3065	0.56	1.02	0.0	3.0	0.0	0.0	0.0	1.0	3.0
w2_LA_ss_E_pat_31	Total	W2	3065	3.63	3.10	0.0	9.0	0.0	0.0	3.0	6.0	8.0
Listening Comprehension												
w2_LA_ss_E_lco_71		W2	3062	0.22	0.52	0.0	3.0	0.0	0.0	0.0	0.0	1.0
w3_LA_ss_E_lco_9		W3	2761	0.66	0.83	0.0	3.0	0.0	0.0	0.3	1.0	2.0
w4_LA_ss_E_lco_7		W4	2684	0.99	1.07	0.0	4.0	0.0	0.0	1.0	1.0	3.0
w5_LA_ss_E_lco_5		W5	2405	1.30	1.20	0.0	4.0	0.0	0.0	1.0	2.0	3.0
Letter sound recognition (per min)												
w2_LA_ss_E_lsr_pm_42		W2	3060	17.44	19.25	0.0	198.8	0.0	3.1	12.0	27.0	40.0
Receptive Proficiency ('Do this' tasks)												
w2_LA_ss_E_rpr_6		W2	3062	2.91	1.32	0.0	5.0	1.0	2.0	3.0	4.0	4.0
w3_LA_ss_E_rpr_7		W3	2766	3.38	1.31	0.0	5.0	1.5	2.5	3.5	4.5	5.0
Word match (Receptive vocabulary, vocabulary and grammar)												
w5_LA_ss_E_vor_10_2		W5	2352	1.77	1.84	0.0	8.0	0.0	0.0	1.0	3.0	4.0
w5_LA_ss_E_vow_10_1		W5	2352	3.82	2.48	0.0	8.0	0.0	2.0	4.0	6.0	7.0
English Writing												
w2_LA_ss_E_wri_9	CVC word	W2	3056	0.23	0.42	0.0	1.0	0.0	0.0	0.0	0.0	1.0
w3_LA_ss_E_wri_10		W3	2766	0.58	0.69	0.0	2.0	0.0	0.0	0.3	1.0	2.0

Table 7: Table for EFAL Measures from the Learners Assessment (Reading and decoding)

Renamed Var	Variable info	Wave	N	Mean	SD	Min	Max	p10	p25	p50	p75	p90
Word Recognition (per min)												
w2_LA_ss_E_wre_pm_52		W2	3062	6.14	10.39	0.0	140.0	0.0	0.0	2.0	8.0	18.0
w3_LA_ss_E_wre_pm_52		W3	2761	16.44	19.05	0.0	92.4	0.0	0.0	8.0	29.0	44.0
w4_LA_ss_E_wre_pm_42		W4	2683	23.66	21.23	0.0	132.9	0.0	2.4	22.0	40.0	54.0
w5_LA_ss_E_wre_pm_22		W5	2383	28.12	23.38	0.0	105.0	0.0	4.0	27.0	47.0	60.0
Decodable Word Recognition (per min)												
w2_LA_ss_E_wrd_pm_5_12		W2	3062	5.85	11.88	0.0	240.0	0.0	0.0	2.0	8.0	16.0
w3_LA_ss_E_wrd_pm_5_12		W3	2763	16.60	19.48	0.0	84.8	0.0	0.0	8.0	28.0	46.0
Sight Word Recognition (per min)												
w2_LA_ss_E_wrs_pm_5_22		W2	3062	6.44	12.45	0.0	210.0	0.0	0.0	2.0	7.5	20.0
w3_LA_ss_E_wrs_pm_5_22		W3	2761	16.27	19.36	0.0	100.8	0.0	0.0	8.0	30.0	46.0
Oral reading fluency (per min)												
w3_LA_ss_E_orf_pm_6_12		W3	2763	18.78	20.99	0.0	116.7	0.0	1.7	11.0	31.0	49.0
w4_LA_ss_E_orf_pm_5_13		W4	2444	28.65	27.85	0.0	153.1	0.0	1.6	22.0	49.4	68.8
w5_LA_ss_E_orf_pm_3_13		W5	2385	36.79	34.19	0.0	164.3	0.0	1.4	31.0	64.0	83.0
ORF Comprehension												
w3_LA_ss_E_orc_6_2		W3	2766	1.33	1.33	0.0	9.0	0.0	0.0	1.0	2.0	3.0
w4_LA_ss_E_orc_5_2		W4	2684	1.07	1.46	0.0	5.0	0.0	0.0	0.0	2.0	4.0
w5_LA_ss_E_orc_3_2		W5	1801	2.60	2.31	0.0	8.0	0.0	1.0	2.0	4.0	6.0
Written Comprehension												
w4_LA_ss_E_wco_9		W4	2669	1.43	1.25	0.0	4.0	0.0	0.0	1.0	2.0	3.0
w5_LA_ss_E_wco_9		W5	2352	1.31	1.64	0.0	7.0	0.0	0.0	1.0	2.0	4.0

Table 8: Table for Mathematics Measures from the Learners Assessment

Renamed Var	Variable label	Wave	N	Mean	SD	Min	Max	p10	p25	p50	p75	p90
Mathematics sums												
w4_LA_ss_M_mat_10		W4	2668	2.39	1.87	0.0	10.0	0.0	1.0	2.0	3.0	5.0
w5_LA_ss_M_mat_11		W5	2352	2.78	2.01	0.0	10.0	1.0	1.0	2.0	4.0	6.0