



Longitudinal Study: Baseline 2024

Submitted to Funda Wande

November 2024

Prepared by JET Education Services



RESEARCH

IMPLEMENTATION

MONITORING & EVALUATION

CONTENTS

TABLE OF FIGURES	4
LIST OF TABLES	4
EXECUTIVE SUMMARY	6
Introduction	6
Research design and sampling.....	6
Fieldwork.....	6
Instruments used.....	6
Data analysis	7
Findings.....	8
Participation.....	8
Test Functioning Rasch Analysis	8
MARKO-D Grade R.....	8
MARKO-D Grade 1	8
Learner achievement results	9
MARKO-D Grade R.....	9
MARKO-D Grade 1	9
EGMA	9
Conclusions and recommendations	12
Test functioning.....	12
General conclusions and recommendations.....	12
Conclusions and recommendations for Mathematics.....	12
Conclusions and recommendations for Reading.....	13
ACRONYMS AND ABBREVIATIONS	14
1 BACKGROUND.....	15
1.1 Background.....	15
1.2 Terms of reference and purpose of the research	15
2 METHODOLOGY	17
2.1 Research design	17
2.2 Sampling.....	18



2.3	Fieldwork	19
2.4	Data analysis	19
2.4.1	Rasch analysis.....	19
2.4.2	Learner achievement analysis.....	21
2.5	Methodological norms: Validity and Reliability	21
2.5.1.	Validity.....	21
2.5.2.	Reliability.....	23
2.6	Ethical guidelines	23
2.6.1	Beneficence.....	23
2.6.2	Informed consent and assent.....	23
2.6.1	Confidentiality.....	23
2.6.2	Ethical test use	23
3	TEST INSTRUMENTS	25
3.1	Test specifications	25
3.1.1	EGRA	25
3.1.2	South African EGMA	26
3.1.3	The South African MARKO -D	28
4	FINDINGS	31
4.1	The samples	31
4.1.1.	EGRA	31
4.1.2.	EGMA	33
4.1.3.	MARKO-D (Sepedi).....	34
3.2	Test Functioning: Rasch analysis	36
3.2.3	MARKO-D (Sepedi) Grade R.....	36
3.2.3	MARKO-D (Sepedi) Grade 1.....	37
4.3.	Learner achievement results	39
4.3.1	MARKO-D (Sepedi).....	39
4.3.2	EGMA	41
	Number identification subtest.....	41
4.3.3	EGRA	49
5	CONCLUSIONS AND RECOMMENDATIONS	61
5.1	Test functioning	61
5.2	General conclusions and recommendations	61
5.3	Conclusions and recommendations for Mathematics	62
5.3.1.	Mastering the counting rhyme, ordinality principal and number relationships	62
5.3.2.	Part-part-whole and class inclusion principles.....	63
5.4	Conclusions and recommendations for Reading	65
	REFERENCES	66



Table of figures

<i>Figure 1: Report structure</i>	15
<i>Figure 2: Test development cycle</i>	18
Figure 3: Aspects of validity	22
<i>Figure 4: Ethical guidelines in research</i>	24
<i>Figure 5: Item-person map for Grade R MARKO-D</i>	37
<i>Figure 6: Item-person map for Grade 1 MARKO-D</i>	39
<i>Figure 7: Grade R achievement by level</i>	40
<i>Figure 8: Grade 1 achievement per level</i>	41

List of tables

Table 1: Longitudinal design	17
Table 2: Expected number of assessments per the contract	18
Table 3: Actual number of assessments	18
Table 4: Distribution of learner across two grades: EGRA sample	31
Table 5: Distribution of learner across two grades by gender and grade: EGRA sample.....	31
Table 6: Distribution of learners across two grades by age and grade: EGRA sample	33
Table 7: Distribution of learners across two grades: EGMA sample	33
Table 8: Distribution of learners across two grades by gender and grade: EGMA sample	33
Table 9: Distribution of learners across two grades by age and grade: EGMA sample	34
Table 10: Distribution of learners across two grades: MARKO- D sample	34
Table 11: Distribution of learners across two grades by gender and grade: MARKO-D sample..	35
Table 12: Distribution of learners across two grades by age and grade: MARKO-D sample	35
Table 13: Rasch Summary Results Grade R	36
Table 14: Rasch Summary Results Grade 1.....	38
Table 15: Distribution of correct numbers read per minute (NRPM) by grade.....	42
Table 16: Proportions of correct answers of number discrimination by Grade	43
Table 17: Proportions of correct number pattern recognised by Grade.....	44
Table 18: Proportion of correct addition (level 1) sums by grade.....	45
Table 19: Proportion of correct addition (level 2) sums by grade.....	45
Table 20: Proportion of correct subtraction (level 1) sums by grade.....	46
Table 21: Proportion of correct subtraction (level 2) sums by grade.....	47
Table 22: Is number recognition speed associated with accuracy?	47
Table 23: Is number recognition speed associated with addition and subtraction (level 1)?	48
Table 24: Is the number recognition speed associated with addition and subtraction (level 2)?	48
Table 25: Distribution of correct letter sounds read per minute (CLPM) by grade.....	49
Table 26: Distribution of letter sounds benchmark scores by grade	50
Table 27: Total average score per each letter indicator	51



Table 28: Distribution of CWPM by grade	52
Table 29: Distribution of word reading benchmark scores by grade	52
Table 30: Total average score per each word reading indicator	53
Table 31: Distribution of correct word reading in a passage per minute (CWPM) by grade	54
Table 32: Distribution of word reading in a passage benchmark scores by grade.....	55
Table 33: Distribution of correct comprehension questions by grade.....	56
Table 34: Is letter reading accuracy associated with speed?	56
Table 35: Is word reading accuracy associated with speed?	57
Table 36: Is passage word reading accuracy associated with speed?	57
Table 37: Is word reading accuracy and speed associated with comprehension?.....	58
Table 38: Is word reading accuracy associated with letter sound accuracy?.....	58
Table 39: Is total letter sound attempts associated with letter sound error rate?.....	59
Table 40: Is a total word reading attempts associated with word reading error rate?	59
Table 41: Is error rate associated with comprehension?	60



Executive summary

Introduction

This study aims to assess the achievement of learners taught through Funda and Bala Wandé programmes in 30 schools in Limpopo. JET Education Services was appointed by Funda Wandé to conduct the first round of this longitudinal study of learner achievement.

For the first round, this was done empirically:

- through a baseline assessment to determine the initial achievement levels of the 30 project schools;
- through an error analysis and by identifying performance levels using MARKO-D data; and,
- by undertaking an analysis of Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) data.

Research design and sampling

The study uses a longitudinal, quasi-experimental design to track changes in learner achievement over three time points: baseline (Grade R and Grade 1 learners in 2024, midline (Grade 1 and 2 learners in 2025), and end line (Grade 2 and 3 learners in 2026).

The longitudinal design enables tracking changes over time but could be vulnerable to cohort effects due to inconsistent tracking from one grade to the next. While attrition management mitigates this risk, potential bias remains.

The study relies on DBE-developed EGRA and EGMA, along with the independently developed and validated MARKO-D (Sepedi), which adheres to international assessment standards. A minimum of nine learners per grade, per school were assessed using the MARKO-D., EGMA and EGRA. Absent learners were replaced using random selection where possible.

Fieldwork

Learner assessments were conducted between 9 and 24 April 2024. Fieldwork was preceded by rigorous training and competency assessments for all team members. Teams, led by experienced team leaders, adhered to strict test administration protocols and quality assurance measures.

The study adhered to ethical principles, including beneficence (HPCSA, 2008a), informed consent/assent (Jamieson and Lake, 2013), and confidentiality (HPCSA, 2008a). Ethical test use was ensured by using valid and reliable instruments (Foxcroft and Roodt, 2013).

Instruments used

Early grade reading assessments, like EGRA, are crucial for identifying and addressing literacy challenges in young learners. While EGRA has been adapted for the South African context, its implementation and the PIRLS 2021 results highlight the urgent need for rigorous assessment and evidence-based interventions to improve literacy outcomes, especially in under-resourced schools.



The South African EGRA assesses four key foundational literacy skills:

- 1. Letter Sound Recognition:** Evaluates phonemic awareness, which is important for decoding words.
- 2. Word Reading:** Assesses vocabulary knowledge and word recognition.
- 3. Oral Reading Fluency:** Evaluates reading speed and accuracy.
- 4. Reading Comprehension:** Assesses understanding of the text.

The South African EGMA assesses foundational (Grades 1-3) numeracy skills, including number recognition, number discrimination, number operations (addition and subtraction), and pattern recognition. These skills are crucial for building a strong foundation in mathematics, as they assess skills that are essential for more advanced mathematical reasoning. This assessment provides valuable insights into learners' mathematical understanding, informing educational practices and policy decisions. However, its limitations include the potential impact of language barriers, cultural diversity, and the narrow focus on early numeracy skills. Additionally, questions remain about the reliability and validity of the assessment, particularly for disadvantaged learners.

The South African MARKO-D is a diagnostic tool that assesses young children's early mathematical understanding, specifically in arithmetic and number concepts. Unlike traditional assessments, it focuses on conceptual understanding rather than procedural skills. The MARKO-D assesses children's mathematical development across five levels:

- Level 1. Counting sequence/rhyme:** Concrete counting skills
- Level 2. Ordinality principal:** Understanding and identifying "more than" and "less than."
- Level 3. Relational understanding of numbers:** Basic addition and subtraction.
- Level 4. Part-part-whole relationships:** Decomposing and recomposing numbers.
- Level 5. Class inclusion and hierarchical relationships:** Abstract mathematical thinking.

By identifying a child's level, educators can tailor instruction to support their mathematical progress.

The Sepedi MARKO-D (the Sepedi-language adaptation) is a reliable and valid adaptation of the South African MARKO-D. It includes items specifically designed to assess higher-level mathematical understanding, particularly part-part-whole relationships and the concept of zero.

Data analysis

A Rasch analysis was conducted on the MARKO-D (Sepedi) assessment. As standard Department Basic Education versions of EGRA and EGMA were used, further reliability and validity testing was not performed on these data.

In this study, Rasch analysis is used to empirically validate the tests. This method improves measurement precision by ensuring items contribute meaningfully to the construct. It assesses item difficulty and learner ability, determining if items align with the intended item. The more the learner's ability exceeds the item difficulty, the more likely it is that a person will answer the item correctly (Bond & Fox, 2015).

A validation process was undertaken to ensure the soundness of test interpretations. Learner test results together with other data contribute to a broader evaluation of Funda and Bala Wande's impact on



education. Test scores are part of a larger data interpretation, and their support, or lack thereof, should be considered within the project's context. The evidence suggests validity, particularly item and person reliability and construct validity for MARKO-D.

A descriptive analysis of overall and subtest scores was conducted across grades. Inferential statistics were used to explore deeper patterns, but limited sample sizes prevented reliable school comparisons. Results are presented by grade, with comparisons only valid where sample sizes are comparable.

Findings

Participation

A total of 585 learners from the sample 30 schools, participated in the EGRA and EGMA. The MARKO-D tests was also conducted in the same sample of 30 schools but from 594 learners. In all the assessments the number of learners was almost equally distributed between Grade R and Grade 1, Sepedi was the language of learning and teaching in all schools. Female learners constituted a slightly higher proportion than male learners in both grades. The majority of Grade R learners were 5 and 6 years old, aligning with DBE admission policies. Grade 1 primarily consisted of 6-year-olds, but also included a significant number of 7-year-olds, and a smaller number of 8 and 9-year-olds. This suggests potential grade repetition or delayed school entry for older learners in Grade 1.

Overaged learners require close monitoring. Their presence highlights the significance of school readiness, a cornerstone for academic success and long-term development. Socioeconomic factors, particularly poverty, hinder school readiness for many South African children. Addressing this requires targeted interventions and comprehensive assessments that consider cognitive, social, and emotional dimensions.

Test Functioning Rasch Analysis

MARKO-D Grade R

The Rasch analysis of the MARKO-D for Grade R showed high reliability for both person and item measures. INFIT and OUTFIT values were close to ideal, indicating good item fit, but with some gaps in the person ability distribution that were not represented by the available items. The item-person map revealed a slightly challenging assessment for Grade R learners especially for concepts like part-part-whole relationships and class inclusion, which are typically developed later, and aligns with developmental theory

MARKO-D Grade 1

The Rasch analysis of the MARKO-D for Grade 1 showed high reliability for both person and item measures. INFIT and OUTFIT values were close to ideal, indicating good item fit, but with some gaps in the person ability distribution that were not represented by the available items (especially at level 2). The item-person map revealed that the assessment was well-targeted for Grade 1 learners. The easiest items were Level 1 counting tasks, which align with developmental theory.



Learner achievement results

MARKO-D Grade R

Grade R learners generally struggle with mathematics in April especially as complexity increased. There were no significant gender or age-related differences in performance, except for two seven-year-olds who performed differently. To improve, these learners could benefit from targeted support in foundational concepts like counting and ordinality, as well as gradually introducing relational number strategies like "counting on" and "counting all" throughout the year.

Grade R learners struggled with math in April, especially as tasks got harder. Gender and age didn't significantly impact performance, except for two seven-year-olds. They need support in basic concepts like counting and ordinality, and should gradually learn strategies like "counting on" and "counting all" to improve addition and subtraction skills.

MARKO-D Grade 1

Grade R learners struggled with math in April, especially with more complex tasks. Gender and age didn't significantly affect performance, except for two older learners. Surprisingly, Grade 1 learners performed worse on zero-related questions. Both grades need support in basic math concepts and strategies. Similar to Grade R learners, Grade 1 learners would benefit from focused support to strengthen their understanding of essential mathematical concepts

EGMA

Number Identification subtest: Most Grade R learners struggled, with over half unable to answer correctly and only a small fraction achieving two or more correct answers. Grade 1 learners showed improvement, with a higher percentage of correct answers. However, a significant portion of both grades still struggled, especially with larger numbers.

Missing Number subtest: Most Grade R learners struggled to recognize number patterns. While Grade 1 learners showed improvement over Grade R, just under a quarter were still struggling, 40.4% could identify one pattern, and 24.5% could identify two or more, indicating overall development in number pattern recognition.

Addition and Subtraction Level 1 subtests: More than half of Grade R learners struggled with Level 1 addition, solving five or fewer problems. Grade 1 learners showed improvement, with a quarter unable to answer any problem and just under two-third solving up to five problems and 13.6% solving six to twenty, indicating developmental progress in basic addition.

Addition and Subtraction Level 2: Most Grade R learners (97.5%) couldn't solve Level 2 subtraction problems. While Grade 1 showed slight improvement, a majority (92.8%) still struggled. Both grades faced significant challenges with higher-level subtraction tasks, which is common at this stage of learning.

Inferential statistics:

Association of number recognition speed and accuracy: Grade R learners show a weak association between number discrimination speed and recognition accuracy. In contrast, Grade 1 learners exhibit a moderate positive relationship between these two skills, suggesting some connection but not a perfect one.



Association of number recognition speed with addition and subtraction (level 1): In both Grade R and Grade 1, there's a weak or negligible relationship between number reading speed and both addition and subtraction accuracy at Level 1. This suggests that reading speed is not a strong predictor of arithmetic performance at this stage.

Association of number recognition speed with addition and subtraction (level 2): Grade R shows a weak positive relationship between reading speed and Level 2 addition, but a moderate negative correlation with subtraction. In Grade 1, a weak positive correlation appears between reading speed and subtraction. Overall, the relationship between reading speed and Level 2 addition and subtraction is weak across both grades, suggesting limited impact.

Letter sounds subtest

Most Grade R learners (67.1%) couldn't identify any letter sounds in a minute. Grade 1 showed improvement, with fewer unable to identify any sounds. A smaller proportion (28%) could identify 10 or fewer letter sounds per minute. Grade 1 had a wider range of performance, with some identifying 11-20 sounds per minute, indicating more advanced letter sound recognition.

Most Grade R learners (67.1%) were "non-fluent" in letter sound recognition. Grade 1 showed improvement, with fewer "non-fluent" learners (25.3%) and a notable portion (20.1%) classified as "developing accuracy" readers.

Grade 1 learners performed better than Grade R learners in all reading measures, showing faster recognition, greater accuracy, and fewer errors. This indicates clear developmental progress from Grade R to Grade 1. To improve Grade R performance, targeted interventions are needed. Many Grade R and Grade 1 learners stopped the assessment early, suggesting a need for further investigation into the reasons for this.

Word reading subtest

Most Grade R learners (86.3%) couldn't read any words in a minute. Grade 1 showed improvement, with fewer unable to read any words (63.5%) and some (6.4%) reading 11-20 words per minute. Most Grade R learners (86.3%) were "non-fluent" readers. Grade 1 showed improvement, with fewer "non-fluent" learners (63.5%) and some classified as "low-accuracy readers" (35.2%). However, no learners in either grade were "accurate readers."

Grade R learners read faster than Grade 1 learners, but Grade 1 learners were more accurate. Both grades had low accuracy, indicating a need for improvement. Many learners in both grades stopped the assessment early due to the stop rule, suggesting a need for further investigation.

Passage reading subtest

Grade 1 learners were more accurate and faster at reading passages than Grade R learners. However, many learners in both grades struggled with passage reading, especially Grade R learners, where the largest proportion (23.1%) couldn't read any words correctly.

Grade 1 learners were more proficient in word reading than Grade R learners. While Grade 1 showed improvement in passage word reading, many (51.2%) were still "low-accuracy readers." However, there was a significant increase in "accurate readers" from Grade R to Grade 1. Grade R learners read faster and more accurately than Grade 1 learners. However, Grade R made more errors per minute. This might be due to the smaller sample size. Fewer Grade R learners (10 out of 13) completed the comprehension questions compared to Grade 1 (74 out of 80).



Comprehension subtest

Grade 1 learners generally performed better than Grade R learners on comprehension questions. However, both grades had a significant proportion of low scores, indicating a need for improvement in comprehension skills. While Grade 1 learners performed slightly better, the difference wasn't statistically significant.

Inferential statistics

Letter sound accuracy associated with speed

Both Grade R and Grade 1 showed a positive relationship between letter sound accuracy and speed. Grade 1 had a stronger correlation coefficient than Grade R. The slight difference between Pearson's and Spearman's coefficients suggests a non-linear relationship, especially in Grade R.

Word reading speed associated with accuracy

Both Grade R and Grade 1 showed a positive correlation between word reading accuracy and speed. Grade R had a stronger correlation than Grade 1. The difference between Pearson's and Spearman's coefficients suggests a non-linear relationship, especially in Grade 1.

Passage reading accuracy associated with speed

Both Grade R and Grade 1 showed a strong positive correlation between passage word reading accuracy and speed. Grade R had a stronger relationship than Grade 1. The strong, consistent correlation suggests a clear linear and monotonic pattern between accuracy and speed for both grades.

Word reading accuracy associated and speed associated with comprehension?

Grade R showed a weak relationship between reading speed/accuracy and comprehension. Grade 1 showed a slightly stronger relationship for passage reading, but overall, the association remains limited.

Passage reading accuracy and letter sound accuracy

Grade R showed a moderate positive relationship between passage reading accuracy and letter sound accuracy. In contrast, Grade 1 showed a weak, inconsistent relationship between these variables. The relationship between word reading accuracy and letter sound accuracy was stronger in Grade R than in Grade 1. The overall pattern of association was complex and inconsistent across the two grades.

Total letter sound attempts associated with letter sound error rate

Grade R showed a moderate negative relationship between letter sound attempts and error rates. Grade 1 showed a strong negative relationship, indicating that more attempts were associated with fewer errors.

Total word reading attempts associated with word reading error rate

Grade R showed a strong negative correlation between word reading attempts and error rates. Grade 1 showed a moderate negative correlation, indicating that more attempts were associated with fewer errors, but the relationship was weaker than in Grade R.

Both Grade R and Grade 1 showed a negative relationship between attempts and error rates in letter sound and word reading tasks. This suggests that more practice leads to improved accuracy. The stronger correlation in Grade 1 for letter sounds indicates that this relationship strengthens with development.



Error rate associated with comprehension

Grade R showed a weak to moderate positive correlation between letter sound error rates and comprehension. Grade 1 showed a weak negative relationship. This indicates a complex and inconsistent relationship between letter sound accuracy and comprehension across the two grades.

Grade R showed a negligible correlation between word reading error rates and comprehension, while Grade 1 showed a moderate negative relationship. Both grades showed a moderate negative correlation between passage error rate and comprehension. Overall, reading error rates were generally associated with lower comprehension, especially in word reading and passage accuracy. The relationship between letter sound error rate and comprehension was more complex, with a weak positive relationship in Grade R and a weak negative relationship in Grade 1.

Conclusions and recommendations

Test functioning

The MARKO-D instrument is reliable and valid for Sepedi-speaking Grade R and 1 learners. While minor limitations exist, its robustness and cross-linguistic availability make it suitable for assessing learner achievement and enabling comparisons within South Africa.

General conclusions and recommendations

To address the challenges of age disparities and school readiness, it is recommended that Funda Wande monitor under- and over-aged learners closely. Early learning interventions and resource allocation are crucial to mitigate the impact of poverty on educational outcomes.

To improve the evaluation, it is recommended that the 2024 Grade R cohort and the 2024 Grade 1 cohort be tracked as part of the Grade 1 and Grade 2 sample respectively in 2025 and the current assessment schedule (April) is maintained. Additional biographical data, including practitioner and teacher qualifications and socio-economic context should also be collected. For the midline evaluation we recommend the inclusion of case studies to supplement quantitative analysis.

The EGRA and EGMA assessments were conducted early in the academic year, so the poor performance aligns with curriculum expectations. However, the results provide valuable insights for targeted interventions. Conducting an error analysis and mapping the Bala Wande materials against the MARKO-D levels can also provide deeper insights. Additionally, exploring unexpected outcomes through case studies can inform future strategies.

Conclusions and recommendations for Mathematics

The MARKO-D analysis shows that Grade R learners struggle with part-part-whole and class inclusion concepts and there is a weak correlation between reading speed and early arithmetic skills. These concepts are foundational for addition and subtraction.

Focusing on foundational skills like counting rhyme and ordinality is crucial for developing higher-level mathematical abilities and conceptual progression is essential for early numerical development (Fritz et al., 2023). To help teach part-part-whole concepts, hands-on activities such as Nature Basket Exploration, Story Box Activities, Stacking and Nesting, Base Ten Blocks, Digital Resources, are recommended.



Conclusions and recommendations for Reading

Learners struggled with the EGRA, most likely due to the timing of the year (April) in which testing took place. Grade R learners struggled particularly in reading. Grade 1 learners outperformed Grade R learners in letter-sound reading but both grades showed weaknesses in word-reading accuracy.

It is recommended that: pre-reading activities for Grade R should be implemented to improve reading readiness; there should be a focus on developing phonological and phonemic awareness in both grades through play-based activities; targeted interventions should be provided to address individual needs, especially in Grade R; and, curriculum expectations and instruction should align with specific goals for each grade level.

By addressing these areas, schools can improve early literacy skills and set learners.



Acronyms and abbreviations

AERA	American Educational Research Association
APA	American Psychological Association
CAPS	Curriculum and Assessment Policy Statement
CLPM	Correct Letters Per Minute
CWPM	Correct Words Per Minute
DBE	Department of Basic Education
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
FP	Foundation Phase
HPCSA	Health Professions Council of South Africa
JET	JET Education Services
LoLT	Language of Learning and Teaching
LPM	Letters per minute
NCME	National Council on Measurement in Education
NRPM	Numbers correctly Read Per Minute
WPM	Word Per Minute



1 Background

1.1 Background

In April 2024, Funda Wande commissioned JET Education Services (JET) to conduct an independent assessment of educational outcomes in schools supported by the Funda Wande and Bala Wande Programmes in the Polokwane districts.



Figure 1: Report structure

The report is organized as follows:

1. **Background:** This section outlines the terms of reference, research questions, and overall purpose of the study.
2. **Research Design:** This section covers the longitudinal design of the study, highlighting its characteristics and limitations. It details the development of instruments, sampling procedures, equivalence testing between intervention and matched schools, and the approaches for fieldwork and data analysis. Ethical considerations are also discussed.
3. **Test Framework:** This section provides specifications for each test instrument, with a focus on the multilevel nature of the tests and alignment with the curriculum. Test validity is examined, and reliability estimates for each test are reported.
4. **Findings and Results:** The results of the study are presented here, with an analysis for each grade and subject, segmented by school groupings and individual schools.
5. **Conclusions and Recommendations:** The final section offers conclusions drawn from the study and presents recommendations based on the findings.

1.2 Terms of reference and purpose of the research

The aim of the study was to establish, through empirical, evidence-based research, the achievement of learners taught through Funda Wande and Bala Wande programmes at the 30 intervention schools. The objectives of the study were as follows:

- Ascertain the baseline achievement of the achievement of the 30 project schools.
- Perform analysis that would allow for error analysis using the MARKO-D data, including the levels of performance identified in the MARKO-D

- Perform analysis that would allow for meaningful analysis using the Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA)

In this report, the results of the **baseline assessment** done in April 2024 are discussed.

In order to improve the readability of the report, some stylistic adaptations were used and should be noted:

- Key findings are bolded and/or presented in graphics.
- Statistically relevant data is reported in tables and summarised in graphs.
- Numbers reported in text and figures are reported to one decimal point.
- References are included as a separate section at the end of the report.
- Specific explanatory notes are included as footnotes.



2 Methodology

2.1 Research design

The study employs a classic longitudinal design to track changes in learner achievement over time. This quasi-experimental design assesses cohorts across two grades (see Table 2) at three time points: baseline, midline, and endline.

Table 1: Longitudinal design

Year	2024	2025	2026
Grade R	EGRA, EGMA, MARKO-D		
Grade 1	EGRA, EGMA, MARKO-D	EGRA, EGMA, MARKO-D	
Grade 2		EGRA, EGMA, MARKO-D	EGRA, EGMA, MARKO-D
Grade 3			EGRA, EGMA, MARKO-D

While this design enables the examination of longitudinal trends, the analysis is susceptible to cohort effects, as not all learners are tracked consistently from Grade R to Grade 2 or from Grade 1 to Grade 3. This introduces a potential limitation, as observed effects may be influenced by cohort-specific factors. However, this risk is mitigated to some extent by managing attrition rates; lower attrition reduces the likelihood of bias due to sample loss.

Learner assessments form the foundation of this study, utilising versions of the EGRA and EGMA developed by the Department of Basic Education (DBE). Additionally, the MARKO-D (Sepedi) is an independently developed psychometric instrument currently in the final stages of development. The MARKO-D (Sepedi) has undergone a rigorous validation process to ensure adherence to international standards for educational assessment, specifically those outlined in the “Standards for Educational and Psychological Testing” (AERA, APA, & NCME, 2014).

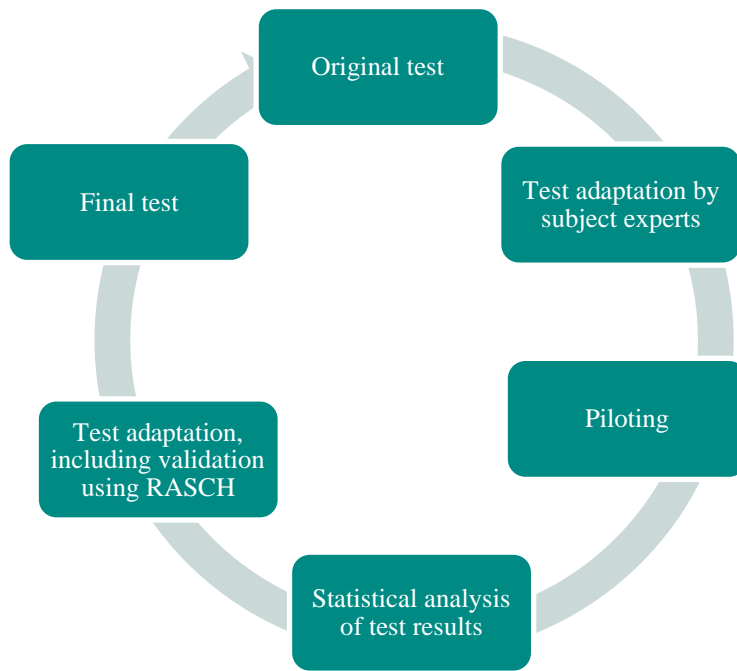


Figure 2: Test development cycle

2.2 Sampling

In 2024, a minimum of nine learners per grade, per school, were selected for the assessment. These learners participated in the EGRA, EGMA, and the MARKO-D assessments. In cases where a learner was absent on the second day of testing for the EGRA, efforts were made to replace the individual with a reserve participant, where feasible. The tables below details the expected versus actual number of assessments conducted in 2024.

Table 2: Expected number of assessments per the contract

	EGRA	EGMA	MARKO D	Assessments
Grade R	270	270	270	810
Grade 1	270	270	270	810

Table 3: Actual number of assessments

	EGRA	EGMA	MARKO D	Assessments
Grade R	289	294	296	879
Grade 1	293	291	298	882



2.3 Fieldwork

Learner assessments were conducted between 9 and 24 April 2024. The commencement of fieldwork was delayed due to the timing of contract finalisation.

Prior to field deployment, both fieldwork teams underwent rigorous training on test administration protocols and procedures. To qualify as assessors, all team members were required to pass a live testing competency assessment. Each team was composed of a team leader and several fieldworkers. Team leaders were responsible for addressing daily operational challenges, ensuring strict adherence to test administration protocols, and performing quality assurance checks on test materials.

2.4 Data analysis

A Rasch analysis was conducted on the MARKO-D (Sepedi) assessment, as it remains in the final stages of development. Given that the standard DBE versions of the EGRA and EGMA assessments were employed, additional reliability or validity testing was not performed on these data.

2.4.1 Rasch analysis

Georg Rasch formulated a probabilistic model of measurement in 1960 in which he used the principle:

a person having a greater ability than another person should have a greater probability of solving any item of the type in question, and similarly, one item being more difficult than another means that for any person the probability of solving the second item is the greater one (Rasch, 1960, p.117).

In this study, Rasch's probabilistic model of measurement is used to empirically substantiate the validity argument for each for the tests. In other words, Rasch will be used to establish how well the data fits the expected statistical model of behaviour. While it is acknowledged that factor analysis and internal consistency reliability have a place in instrument development, the advantage that Rasch analysis offers is that the precision of measuring constructs is substantially improved (Boone, 2016). At the heart of Rasch analysis is the principle of unidimensionality, or whether if items are contributing in a meaningful way to the construct we would like to measure (Bond & Fox 2015). Thus, poor or misfitting items are possibly measuring a construct contrary to the construct under exploration and thus need further exploration to determine the source of misfit.

Furthermore, item difficulty and learner ability are evaluated. Item difficulty is defined as the position on a trait variable in which the learner has a 50% probability of a correct response. The more the learner's ability exceeds the item difficulty the more likely it is that a person will answer the item correctly (Bond & Fox, 2015).

Fits statistics

Construct validity expresses the idea that the recorded values for items are representations of a single underlying trait, thus speaking to the idea of unidimensionality. Within the Rasch framework, fit statistics provides a quality control mechanism to establish how well the data conform to the Rasch model, and if



the data do not fit the model then questions around whether the same construct is measured are raised (Boone, 2016). While there are a number of fit indices that may be used in the context of this study, mean squares (MNSQ) are used. The fit statistics are the mean squares or average of the squared residuals of each item.

There are two infit statistics: infit and outfit. Infit statistics gives more weight to persons whose ability levels are close to the item difficulties. On the other hand, outfit statistics gives equal weight to all persons including outliers. Practically, a choice can be made to pay more attention to infit values since those are not affected by outliers (Linacre, 2002). Ideally fit statistics should fall between 0.5 and 1.5 to be considered sufficient for measurement (Linacre, 2005).

Reliability statistics

In terms of reliability statistics, the real person and real item separation reliabilities are evaluated (Smith, 2003). The separation reliabilities are similar to measures of internal consistency in that a value between 0 and 1 is obtained. The interpretation of the separation reliabilities is the same as when evaluation internal consistency reliability, in that a higher value is advantageous (Andrich, 1982). The person reliability index in Rasch analysis indicates the replicability of the order of persons on the person-item map if this sample of persons were given a parallel set of items measuring the same construct. This assumes that the sample is big enough for persons to be spread along the ability continuum but also suitable and sufficient items that would demonstrate such a hierarchy of ability. While the item reliability indicates the replicability of the order of item if the same items were given to a different sample of respondents with the same number of respondents and they behaved in the same way. A high item reliability indicates that the test has sufficient items along the pathway of easy to difficult and that we can expect consistent inferences about easy and difficult items.

It is also important to explore the person and item separation indices which reveal how well items separate persons (or learners in the context of this report) and how well items are able to separate persons. Ideally, item separation indices should be greater than 3 while person separation indices should be greater than 1.5 (Boone, 2017 & Noltemeyer, 2017).

Item-maps

The item-person map provides a visual representation of the spread of items on one hand and person ability on the other. On the right-hand side of the continuum are the items, with the persons displayed on the left. Ideally, the persons should form a standard normal curve, as one would expect persons of high and low ability to be at the ends, with the majority of the persons in the middle. The person-item map is intended to assist researchers in (Boone, 2017 & Noltemeyer, 2017):

- Assessing the strengths and weaknesses of the instrument;
- Explore the hierarchy of items;
- Compare theory with the observed data.



2.4.2 Learner achievement analysis

A comprehensive descriptive analysis was conducted on both overall scores and subtest scores across grades. Additionally, inferential statistical analyses were performed to explore deeper patterns in the data. However, the sample sizes per school were insufficient to support reliable comparative analyses between schools. Although results are presented by grade, comparisons are only appropriate where sample sizes across grades are comparable.

2.5 Methodological norms: Validity and Reliability

2.5.1. Validity

In the *Standards for Educational and Psychological Testing*¹ (AERA, APA & NCME, 2014), validity is referred to in the publication's opening sentence, indicating the importance of the concept in testing. The definition of validity put forth by Kelley (1927), namely that a test can be considered to have validity "if it really measures what it purports to measure" (Kelley, 1927, p. 14), is expanded by psychologists such as Sattler (2008) who points out that validity further contributes to the appropriateness and relevance of the conclusions based on the test results. This is more aligned to the classic definition of validity put forth by Messick (1989) and supported in the *Standards*, i.e. that validity is an "integrated evaluative judgment" (p.16) of the degree theory and evidence supports "the adequacy and appropriateness of interpretations and actions based on test scores" (Messick, 1989, p.5).

Borsboom et al., 2004) and Lissitz & Calizo (2012) argue that the conceptualisation of validity is based on two fundamental assumptions: 1) There is something to measure (a construct or trait); and 2) The changes in the construct cause changes in the measurement result. This proposition fits with Bond & Fox's (2015) assertion that in order to achieve measurement status, there should be an empirically tested causal relationship between the construct and the measurement result.

¹ Henceforth, referred to as Standards.



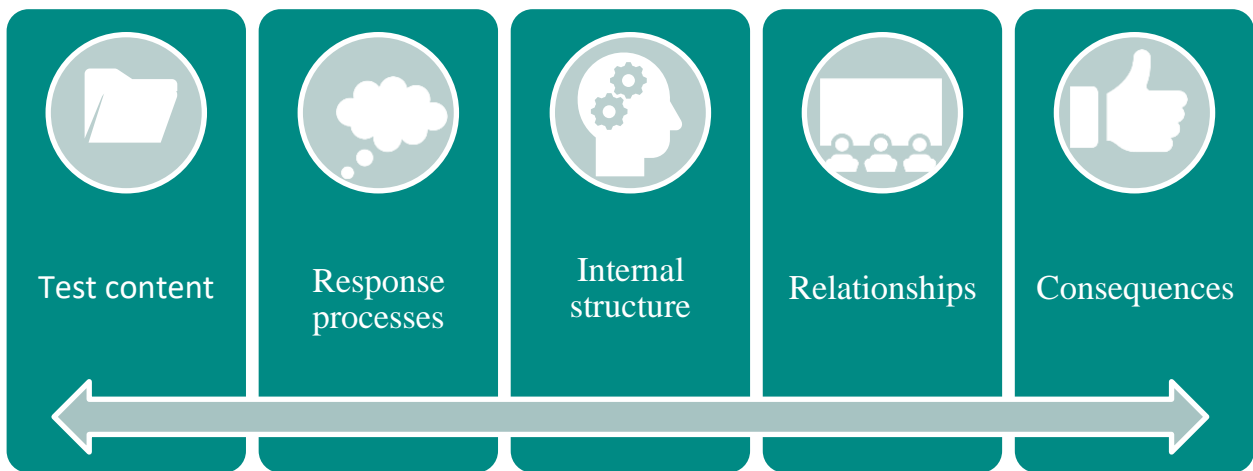


Figure 3: Aspects of validity

Validity in the form of response processes include the fit between the construct and the detailed response produced by a test taker (AERA, APA & NCME, 2014). It also includes processes to ensure the extent to which markers are consistent in applying specific criteria in the marking of responses. This is sometimes referred to as interrater reliability. In the assessment process we allow for a 10% difference between scores given by master coders and assessors for scoring of the EGRA and EGMA. Although the percentage 90% gives you the percentage agreement between moderator and marker, it is not a reliability coefficient since it tells you nothing about the testing procedure itself (Sattler, 2008).

Construct validity is the idea that recorded achievements are reflections of a single underlying trait or construct (Bond & Fox, 2007). This is referred to as unidimensionality. The dimensionality analysis is done through analysis of the fit values for each item. These are discussed in the results sections under each test. However, a short note on differential item functioning is needed. Differential item functioning analysis also forms part of validity evidence that falls under internal structure. A DIF analysis will need to be undertaken to ensure that differences in scores between learners tested on the MARKO-D (Afrikaans, English, Sesotho and isiZulu) and MARKO-D (Sepedi) are not due to differences in translations. This, however, falls outside the scope of this study.

The validation process involves gathering evidence regarding the soundness of the proposed interpretation for each purpose of the test. In this study, the learner test results contributes, together with other data, to evaluating the contribution made or not made by Funda and Bala Wandu to improve education. This goes beyond the direct interpretation of the test results. In this regard, test scores form only part of a bigger interpretation of available data, and the support or lack of support that test results render, should be considered in the bigger context of the project.

Based on the above evidence presented, the conclusion is drawn that the instruments show validity, especially interrater reliability in training of assessors on the EGRA and EGMA and construct validity of the MARKO-D.



2.5.2. Reliability

Reliability addresses the extent to which the results are free from measurement error (Gronlund, 1998) and speaks to the consistency of scores, which are obtained by the same individuals when they are requested to complete the assessment on different occasions (Anastasi & Urbina, 1997). Furthermore, reliability is important, because, unless results are stable, one cannot expect the results to be valid. Internal consistency is a pre-requisite for construct validity, where one would expect a high item-total correlation since items measuring the same construct contributes to the total score of a test (Kline, 1993). In the context of this project internal consistency reliability is considered to be a vital element for interrogating the psychometric properties of the assessment.

2.6 Ethical guidelines

2.6.1 Beneficence

In line with the ethical principle of beneficence (Health Professions Council of South Africa (HPCSA), 2008a), assessment techniques and instruments have been used that have been shown in research and practice to be beneficial in similar circumstances.

2.6.2 Informed consent and assent

According to the *Children's Act Guide for Health Professionals* (Jamieson and Lake, 2013), Grade 3 and 6 learners are still too young to consent to assessment. Informed consent were, therefore, obtained from parents. Furthermore, Grades R and 1 learners were asked to provide informed assent to be tested,

2.6.1 Confidentiality

Parents as well as learners were informed that their individual information would be kept confidential. In line with the HPCSA guidelines for confidentiality (HPCSA, 2008a), learners were also informed of circumstances in which the best interest principle overrides their rights to confidentiality, for example, in the prevention of maleficence or harm to themselves or others.

2.6.2 Ethical test use

Ethical test use implies that the test is valid and reliable in the context in which it will be used. Ethical test use includes ensuring that tests developed are valid before reporting results. (Foxcroft and Roodt, 2013). Each test is discussed in the test instrument section that follows.



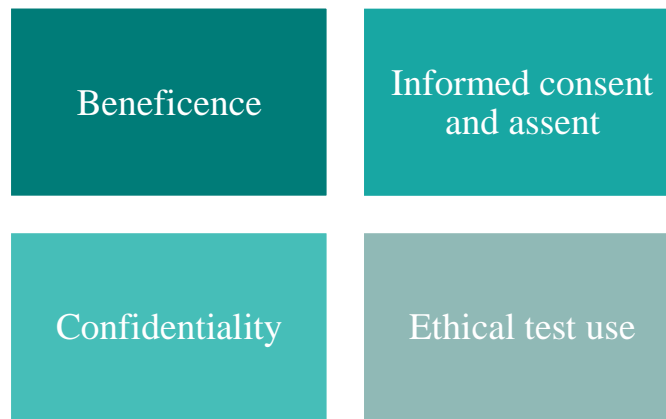


Figure 4: Ethical guidelines in research



3 Test instruments

3.1 Test specifications

3.1.1 EGRA

Early grade reading assessments are critical tools designed to evaluate and enhance the literacy skills of young learners, reflecting a growing recognition of the importance of foundational literacy in a child's educational journey. The most notable assessment in this domain, the EGRA, was introduced in 2006 to measure basic literacy skills in low- and middle-income countries. Since then, EGRA has gained prominence as a standard method for identifying learners at risk of reading difficulties and guiding instructional strategies (Pritchett & Viarengo, 2009; USAID, 2006).

Although EGRA was developed as an instrument to evaluate foundational literacy skills in young learners, specifically targeting low- and middle-income countries (Gove & Wetterberg, 2011), in South Africa, literacy challenges are acute, exacerbated by persistent socio-economic disparities and the historical legacy of apartheid (Spaull, 2013). To address this issue, EGRA was introduced to monitor reading proficiency in Grades 1 to 3 across all 11 official languages of instruction, providing critical data to guide targeted literacy interventions (RTI International, 2009).

Since its implementation, EGRA has prompted diverse discussions regarding its suitability for various educational settings, the practicalities of large-scale data collection, and how these elements may influence the quality and reliability of the results (Pritchett & Viarengo, 2009; RTI International, 2009). In response to these contextual challenges, adaptations have been made to EGRA to enhance its relevance and efficacy for South African learners, ultimately aiming to support more effective literacy interventions (Piper, Zuilkowski, & Mugenda, 2014).

The urgency of addressing early literacy gaps in South Africa is underscored by findings from the Progress in International Reading Literacy Study (PIRLS) 2021, which reported that 81% of Grade 4 learners in South Africa are unable to read for meaning in any language (Mullis et al., 2021). This stark statistic highlights the pressing need for rigorous literacy assessments like EGRA, which enable educators and policymakers to track progress and implement evidence-based interventions. Furthermore, initiatives such as the Early Grade Reading Studies (EGRS) have demonstrated promising results in improving literacy outcomes, particularly in under-resourced schools, underscoring the importance of well-trained teachers and accessible resources in fostering early reading skills (Cilliers et al., 2019).

The South African EGRA subtests

The South African EGRA includes four subtests designed to evaluate foundational literacy skills crucial for early language development. Here is a summary of each subtest, with its purpose in relation to language acquisition:

1. Letter Sound Recognition

In this subtest, learners are asked to identify the sounds associated with different letters of the alphabet. This subtest evaluates learners' phonemic awareness, an essential skill for decoding words in reading. Phonemic awareness enables learners to connect sounds with written symbols, forming the basis for fluent reading and spelling (RTI International, 2009).

2. Word Reading

In this subtest, learners are presented with a list of common words and are asked to read them aloud. The purpose of this subtest is to assess learners' ability to recognise and pronounce familiar words accurately and fluently. Word reading skills are fundamental for reading comprehension, as they indicate the learner's familiarity with vocabulary and ability to recognise words without needing to decode each one phonetically (Gove & Wetterberg, 2011).

3. Oral Reading Fluency

In this subtest, learners are required to read a passage aloud within a specified time (i.e. one minute). The rate and accuracy of their reading are recorded, which provides an indication of their reading fluency. Reading fluency is essential for comprehension, as fluent readers can focus on understanding the text rather than decoding each word individually. Higher fluency levels are strongly linked to better comprehension outcomes (Piper, Zuilkowski, & Mugenda, 2014).

4. Reading Comprehension

Following the oral reading fluency passage, learners answer questions about the content they just read. This subtest assesses their reading comprehension, a complex skill that relies on a range of abilities including vocabulary knowledge, syntax, and fluency. Comprehension is the ultimate goal of reading, as it reflects the learner's ability to understand and interpret written text, which is vital for academic success and overall language development (Mullis et al., 2021).

3.1.2 South African EGMA

The South African EGMA is an important diagnostic tool designed to evaluate the foundational numeracy skills of children in the early grades, particularly in Grades 1 to 3. Developed through a collaboration between USAID, RTI International, and the (DBE) in South Africa, the EGMA was adapted from an international framework to meet the specific needs of South African learners. Its development aimed to create a reliable and contextually relevant measure of early mathematical competencies, recognising the importance of early numeracy as a predictor of future academic success. As such, the EGMA focuses on key early mathematical skills, including counting, number recognition, simple arithmetic operations, and the ability to understand number patterns (RTI International, 2016).

The primary use of the EGMA is to assess the numeracy skills of young children, providing insights into their mathematical understanding at a crucial stage of development. The assessment tool is widely utilised by educators, policymakers, and researchers to evaluate the effectiveness of educational practices and



track learning outcomes. It serves as a valuable instrument for identifying gaps in learners' mathematical knowledge, enabling targeted interventions that aim to address these deficiencies. Furthermore, the EGMA is a crucial resource for evaluating the performance of South African learners in comparison to other countries that use similar assessment tools, facilitating international benchmarking (Dixon, 2017). It also plays an integral role in informing curriculum and instructional reforms, as well as teacher training initiatives, by providing data on areas where learners most commonly experience difficulties (Fleisch, 2016).

Despite its significant contributions to educational assessment, the EGMA faces certain limitations that may affect its accuracy and scope. One such limitation is the multilingual and multicultural context of South Africa. Given that many learners may be assessed in a language that is not their first language, the results may not fully reflect their true mathematical abilities, as language proficiency can influence performance on the test (Spaull, 2013). Furthermore, while the EGMA has been adapted to be culturally relevant to South Africa, the assessment may still struggle to fully account for the diverse cultural and regional differences that exist within the country's educational system (Van der Berg, 2017). These contextual factors could compromise the validity of the results and limit the generalisability of the findings.

Another limitation of the EGMA is its narrow focus on early mathematical competencies. While this focus is critical in the early stages of education, it may not fully capture the range of cognitive, social, and emotional factors that influence a child's learning process (Fleisch, 2016). Additionally, the EGMA is primarily concerned with assessing basic numeracy skills and does not extend to more advanced mathematical abilities that are needed in later grades. This restricted focus may not provide a complete picture of a learner's mathematical potential, especially as they progress through the education system.

Finally, while the EGMA provides valuable data on early numeracy, questions regarding the reliability and validity of the assessment remain, particularly for learners from disadvantaged backgrounds or those with special educational needs. Variations in educational resources, teaching quality, and access to support services may contribute to disparities in test performance, making it difficult to interpret the results in a fully equitable manner (Spaull, 2013).

The South African EGMA subtests

The South African EGMA comprises several subtests designed to assess various foundational mathematical competencies in young learners. The main subtests of the EGMA include number recognition, number operations (addition and subtraction), and pattern recognition. Each of these subtests is integral to the development of foundational mathematics knowledge in early learners, as they assess skills that are essential for more advanced mathematical reasoning.

1. The number recognition subtest

The **number recognition subtest** assesses a child's ability to recognise and identify numbers. This skill is foundational as it forms the basis for understanding numerical concepts such as quantity, magnitude, and order. The ability to recognise numbers is a prerequisite for performing more complex mathematical operations and understanding the place value system, which is central to arithmetic (Fleisch, 2016). The



development of number recognition in early education is crucial, as it influences a child's ability to interpret mathematical symbols, engage with numerical problems, and participate effectively in more advanced mathematical learning.

2. Number discrimination subtest

The **Number Discrimination** subtest in the EGMA is designed to evaluate a child's ability to recognise and differentiate between numerical values. The primary aim of this subtest is to assess foundational number sense by determining whether learners can identify which of two numbers is larger, which is an essential skill for understanding numerical relationships. This skill supports further mathematical learning, as it underpins more complex concepts such as ordering, comparing quantities, and ultimately, performing operations like addition and subtraction.

3. Number operations subtest

The number operations subtest, which includes tasks on **addition (level 1 and 2)** and **subtraction (level 1 and 2)**, tests learners' fluency with basic arithmetic operations. This is directly related to the development of arithmetic skills, which are fundamental to almost all areas of mathematics. Proficiency in addition and subtraction in the early grades is critical, as it serves as a building block for more complex operations such as multiplication, division, and algebra in later years (Dixon, 2017). Early mastery of these operations ensures that children can handle increasingly complex mathematical tasks with confidence and precision.

4. Pattern recognition subtest (also known as the Missing numbers subtest)

The **pattern recognition** subtest evaluates a child's ability to identify and extend number patterns. Recognising patterns is vital for the development of mathematical thinking, as it underpins many mathematical concepts such as sequences, functions, and algebraic reasoning. Pattern recognition enhances a learner's ability to identify relationships between numbers, predict outcomes, and make connections between different mathematical ideas (Van der Berg, 2017). This subtest is important for fostering the cognitive skills necessary for more abstract reasoning in mathematics, as pattern recognition forms the basis for understanding mathematical structures in both number theory and geometry.

3.1.3 The South African MARKO -D

The South African MARKO-D test (MindMuzik, in press), adapted from the German "MARKO" (Mathematik und Arithmetik-Konzept) developed by A. Fritz and colleagues, is a diagnostic tool that assesses young children's early mathematical understanding, specifically in arithmetic and number concepts. Intended for children aged 4 to 8, the MARKO-D focuses on foundational concepts, such as counting, number relationships, and basic operations, to identify areas of conceptual understanding that may need support. Unlike assessments that prioritise rote or procedural skills, the MARKO-D provides insight into children's grasp of essential mathematical principles, which are critical for future learning (Fritz, Ehlert, & Balzer, 2013).

Purpose

The MARKO-D test aims to evaluate children's conceptual knowledge of early arithmetic, highlighting both strengths and potential gaps in foundational understanding. Fritz et al. (2013) argue that assessing these



early concepts is essential, as a solid understanding of arithmetic in the formative years can significantly impact later mathematical achievement. By identifying areas where children may struggle, the MARKO-D allows educators to implement interventions that address specific cognitive needs and promote a strong foundation for future learning.

Use

Teachers, educational psychologists, and researchers use the MARKO-D as both a diagnostic tool and a guide for instructional planning. The test's insights into a child's understanding of mathematical concepts inform educators' decisions on instructional approaches and interventions, especially for learners at risk of struggling with mathematics. As a targeted assessment, the MARKO-D supports individualised learning pathways, which are critical for addressing early difficulties and fostering mathematical confidence (Fritz, 2018).

Specifications

The MARKO-D consists of structured tasks that assess core arithmetic concepts, such as counting, number recognition, addition, subtraction, and part-whole relationships. Unlike procedural assessments that emphasise correct answers, the MARKO-D focuses on revealing a child's understanding of number concepts and relationships. This allows educators to assess mathematical reasoning more holistically, capturing not only a child's performance but also their conceptual progression (Fritz et al., 2013).

Levels

The developmental levels of the MARKO-D correspond to stages in a child's understanding of arithmetic concepts, capturing progression from concrete counting skills to more abstract mathematical reasoning:

- **Level 1 - Counting sequence/rhyme:** At this initial stage, children demonstrate the ability to count objects and recognise numbers in sequence. Their understanding is largely concrete, focusing on physical counting rather than abstract reasoning.
- **Level 2 – Ordinality principal:** In this phase, children begin to understand that numbers can represent quantities and start to compare different sets of objects. They can identify "more than" and "less than" relationships and make judgments about quantity.
- **Level 3 -Relational understanding of numbers:** Children at this level can perform basic addition and subtraction by combining or separating groups of objects. Their understanding of these operations remains grounded in physical manipulation rather than symbolic representation.
- **Level 4 – Part-part-whole relationships:** At this level, children grasp the concept of decomposing and recomposing numbers, recognising that numbers can be divided into parts or combined into wholes. This part-whole awareness is fundamental for arithmetic fluency and lays the groundwork for algebraic reasoning (Fritz et al., 2013).
- **Level 5 - Class inclusion and hierarchical relationships:** At the highest level, children begin to understand that numbers can belong to multiple sets simultaneously and start to grasp hierarchical relationships within numbers, such as subsets. Fritz (2018) highlights this level as a transition to abstract mathematical



thinking, where children recognise numbers as part of different numerical groups. This level of understanding is essential for complex arithmetic and algebraic reasoning in later grades.

The MARKO-D's developmental levels allow educators to identify each child's current stage of understanding, providing a roadmap for guiding learners through increasingly complex arithmetic concepts. This targeted approach supports early interventions, which are crucial for building a strong foundation for future mathematics learning (Fritz, Ehlert, & Balzer, 2013).

The MARKO-D (SEPEDI)

The MARKO-D (Sepedi) is the Sepedi-language adaptation of the South African MARKO-D test. Over a three-year period, it has undergone pilot testing in various Bala Wandé sample groups, consistently demonstrating both reliability and construct validity through Rasch analysis.

This Sepedi version serves as an enhanced adaptation of the South African MARKO-D, incorporating additional level 4 (part-part-whole) items to facilitate assessment beyond the foundational quintile 1, 2, and 3 school levels. Furthermore, it includes items specifically designed to assess conceptual understanding of the number zero, expanding the test's scope to capture more advanced mathematical comprehension.



4 Findings

4.1 The samples

4.1.1. EGRA

Data were collected from a sample of 30 schools, encompassing a total of 582 learners distributed across Grade R (289 learners) and Grade 1 (293 learners). Sepedi was reported as the language of learning and teaching (LOLT) in all participating schools. Table 3 presents the distribution of learners across these two grades.

Table 4: Distribution of learner across two grades: EGRA sample

Grade	Frequency	Percent
Grade R	289	49.7%
Grade 1	293	50.3%
Total	582	100.0%

The data indicate a nearly equal distribution of learners across Grade R and Grade 1, with 49.7% of learners in Grade R and 50.3% in Grade 1.

In terms of gender distribution, female learners slightly outnumbered male learners in both grades, comprising 50.2% in Grade R and 52.9% in Grade 1, compared to male learners who represent 49.8% and 47.1% in the respective grades. This gender distribution difference was minimal, indicating a relatively balanced gender representation in both grade levels.

Table 5: Distribution of learner across two grades by gender and grade: EGRA sample

Gender	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Female	145	50.2%	155	52.9%
Male	144	49.8%	138	47.1%
Total	289	100.0%	293	100.0%

In Grade R, the majority of learners were aged 5, comprising 48.1% of the total population. The second largest age group were 6-year-olds, who constituted 44.6% of the learners. There was a minimal representation of 7-year-olds and no presence of learners aged 8 or 9. This distribution reflects adherence to the's DBE admission policies for Grade R, which aim to standardise the age range for early education placement.



The majority of Grade 1 learners were 6 years old, comprising 60.4% of the grade. There was a significant number of 7-year-olds, making up 31.7% of the grade. A smaller number of learners were 8 or 9 years old.

In this sample, there was a notable overlap in the age range of learners between Grade R and Grade 1, with both grades including 5- and 6-year-olds. Grade R primarily consisted of learners within this age group, whereas Grade 1 accommodated a wider age range, extending to 7-, 8-, and even 9-year-olds. The presence of older learners in Grade 1 suggests instances of grade repetition or delayed school entry. These overaged learners were likely to be late starters or repeaters, whose progress should be monitored closely. Monitoring the number of overaged learners is crucial for several reasons:

1. **Significance of School Readiness**

School readiness serves as the cornerstone of a child's educational pathway, influencing both immediate academic success and long-term developmental outcomes. Research indicates that children who commence formal schooling with adequate preparation display enhanced cognitive abilities and experience more favourable academic trajectories throughout adolescence and adulthood (Duncan et al., 2007; Heckman & Karapakula, 2019). Beyond individual achievement, school readiness confers benefits that extend across generations, fostering improved prospects for future offspring (Atmore, 2019). Ensuring readiness at school entry is thus essential for holistic educational and social development.

2. **Socioeconomic Influences on School Preparedness**

In the South African context, poverty remains a major barrier to school readiness. Approximately 60% of children in the country live in poverty (Maluleke, 2020), which has profound implications for their educational preparedness. Socioeconomic challenges often result in a lack of access to early learning opportunities, leaving up to 50% of children unprepared to meet the necessary benchmarks upon entering Grade 1 (Janse Van Rensburg, 2015). This reality underscores the urgent need for targeted interventions that provide enriched early learning experiences to counteract the adverse effects of poverty and support school readiness for all children (Moses, Van der Berg, & Rich, 2017).

3. **Conceptual Frameworks for School Readiness**

The assessment of school readiness in South Africa requires a nuanced approach, recognising its complex, multifaceted nature. School readiness encompasses a range of cognitive, social, and emotional dimensions, each of which contributes to a child's ability to engage effectively in the learning environment (Walker & Avant, 2011). Adapting conceptual frameworks on school readiness to produce accessible, non-technical resources for educators—particularly those in under-resourced settings—is essential for improving educational outcomes and supporting the professional needs of teachers (Dormer et al., 2022).



Table 6: Distribution of learners across two grades by age and grade: EGRA sample

Age	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
4	20	6.9%	0	0.0%
5	139	48.1%	20	6.8%
6	129	44.6%	177	60.4%
7	1	0.003%	93	31.7%
8	0	0.0%	2	0.7%
9	0	0.0%	1	0.3%
Total	289	100.0%	293	100.0%

4.1.2. EGMA

Data was collected from a sample of 30 schools, encompassing a total of 585 learners distributed across Grade R (294 learners) and Grade 1 (291 learners). In all participating schools, Sepedi was identified as the LOLT. The distribution of learners across these two grades is presented in the figure below.

Table 7: Distribution of learners across two grades: EGMA sample

Grade	Frequency	Percent
Grade R	294	50.3%
Grade 1	291	49.7%
Total	585	100.0%

In both Grade R and Grade 1, female learners constitute a slightly higher proportion than male learners, with 52.4% and 52.6% female representation in each grade, respectively, compared to 47.6% and 47.4% for male learners. This gender distribution reveals only a marginal difference between female and male learners across both grades, indicating no substantial gender imbalance.

Table 8: Distribution of learners across two grades by gender and grade: EGMA sample

Gender	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Female	154	52.4%	153	52.6%
Male	140	47.6%	138	47.4%
Total	294	100.0%	291	100.0%



The age distribution across Grade R and Grade 1 reveals notable patterns with implications for early educational planning. In Grade R, the majority of learners are 5 years old (49.3%), followed by 6-year-olds (42.5%), with very few learners aged 7, 8, or 9. In contrast, Grade 1 predominantly comprises 6-year-olds (59.1%), though a significant proportion are 7 years old (33.7%), with smaller numbers aged 8 or 9. There is an overlap in age groups between the grades, as both include 5- and 6-year-old learners. However, the broader age range in Grade 1, alongside the presence of older learners in both grades, suggests instances of grade repetition or delayed entry. Given the potential impact of overaged learners on early educational acquisition, as evidenced in EGRA data, it is crucial to monitor these cases carefully to support timely school entry and progression through the early grades.

Table 9: Distribution of learners across two grades by age and grade: EGMA sample

Age	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
4	21	7.1%	0	0.0%
5	145	49.3%	18	6.2%
6	125	42.5%	172	59.1%
7	1	0.3%	98	33.7%
8	1	0.3%	2	0.7%
9	1	0.3%	1	0.3%
Total	294	100.0%	291	100.0%

4.1.3. MARKO-D (Sepedi)

Data was collected from a sample of 30 schools, comprising a total of 594 learners across Grade R (298 learners) and Grade 1 (296 learners). Sepedi was reported as LOLT in all participating schools. The distribution of learners between these two grades is illustrated in the figure below.

Table 10: Distribution of learners across two grades: MARKO- D sample

Grade	Frequency	Percent
Grade R	298	50.2%
Grade 1	296	49.8%
Total	594	100.0%

In both Grade R and Grade 1, female learners represent a slightly higher proportion than male learners, comprising 51.7% and 52.0% of each grade, respectively, compared to 48.3% and 48.0% for male learners. This distribution indicates only a minor gender difference across both grades, suggesting an overall balanced gender representation.



Table 11: Distribution of learners across two grades by gender and grade: MARKO-D sample

Gender	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Female	153	51.7%	155	52.0%
Male	143	48.3%	143	48.0%
Total	296	100.0%	298	100.0%

The age distribution observed in Grade R and Grade 1 reveals significant patterns that have implications for early educational planning and policy. In Grade R, the majority of learners are 5 years old (47.6%), followed by 6-year-olds (44.3%), with only a small proportion of learners aged 7, 8, or 9. In contrast, Grade 1 is predominantly composed of 6-year-olds (59.1%), with a notable proportion of 7-year-olds (33.6%), and smaller numbers of 8- and 9-year-olds. This age overlap between the two grades—particularly the presence of 5- and 6-year-old learners in both—indicates a potential for grade repetition or delayed school entry. The broader age range in Grade 1 further underscores this point. Given the demonstrated impact of overaged learners on early educational outcomes, as shown in EGRA data, it is imperative to closely monitor these cases to ensure timely school entry and optimal progression through the early grades.

Table 12: Distribution of learners across two grades by age and grade: MARKO-D sample

Age	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
4	22	7.4%	0	0.0%
5	141	47.6%	19	6.4%
6	131	44.3%	176	59.1%
7	2	0.7%	100	33.6%
8	0	0.0%	2	0.7%
9	0	0.0%	1	0.3%
Total	296	100.0%	298	100.0%



3.2 Test Functioning: Rasch analysis

3.2.3 MARKO-D (Sepedi) Grade R

Summary of results for Grade R

The summary statistics for the Rasch analysis on The MARKO-D for Grade R are presented for both the learners (persons) and the test items. Both the person reliability (0.86) and the item reliability (0.99) were above the acceptable threshold of 0.80, indicating that the test was consistently measuring what it is supposed to measure. The INFIT and OUTFIT Mean-squares values were both close to 1, which was ideal. INFIT and OUTFIT values higher than 1 suggest that the scale may contain an item/s that may be harder to predict and does not fit well with the rest of the test, meaning it might not be measuring the same thing as the other items. On the other hand, values below 1 suggest that the scale may contain an item/s which is too predictable, which could mean there is some redundancy in the test data.

Table 13: Rasch Summary Results Grade R

Focus area	Result	Comment
INFIT MNSQ Persons	1.00	Sufficient for measurement
OUTFIT MNSQ Persons	1.05	Sufficient for measurement
INFIT MNSQ Items	1.00	Sufficient for measurement
OUTFIT MNSQ Items	1.05	Sufficient for measurement
Person Reliability	0.86	High reliability
Item Reliability	0.99	High reliability
Misfitting items	0	Items should be investigated and possibly reformulated

The item-person maps indicated that both the item difficulty and person ability distributions closely resembled normal distributions, with some gaps in the person ability distribution that were not represented by the available items. Notably, the mean of the item difficulties was slightly higher than the mean of the person abilities, which suggests that the MARKO-D assessment may be slightly too challenging for Grade R learners, particularly given the timing of the assessment in April of the Grade R year.

Further analysis revealed that the most difficult items for Grade R learners corresponded to Level 4 (part-part-whole) and Level 5 (class inclusion and hierarchical relationships) cognitive tasks. According to developmental theory, the understanding of part-part-whole relationships and class inclusion typically begins to emerge in children around ages 4 to 6, with more complex hierarchical relationships generally developing around ages 6 to 7 (Piaget, 1952). These cognitive abilities align with the expected developmental trajectory, as children in Grade R (typically aged 5 to 6) may not yet have fully mastered these concepts, which would explain the difficulty they experience with these items.



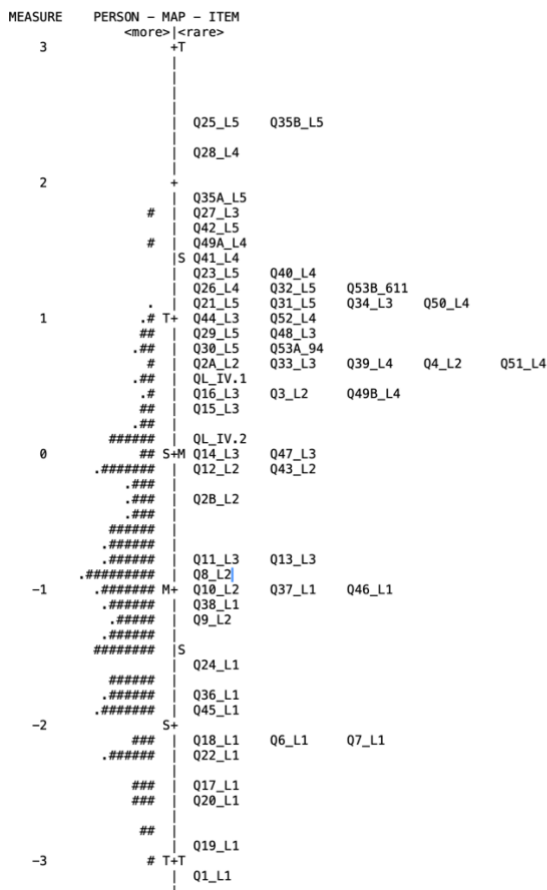


Figure 5: Item-person map for Grade R MARKO-D

3.2.3 MARKO-D (Sepedi) Grade 1

Summary of results for Grade 1

The Rasch analysis summary for the MARKO-D in Grade 1 presents statistics for both the learners (persons) and the test items. With person reliability at 0.89 and item reliability at 0.99, both values exceeded the acceptable threshold of 0.80, demonstrating that the test reliably measures its intended constructs. The INFIT and OUTFIT Mean-square values were both near 1, which is ideal. INFIT and OUTFIT values exceeding 1 suggest the presence of items that are less predictable and may not align well with the overall test, indicating a potential mismatch with the intended measurement. Conversely, values below 1 suggest that some items might be overly predictable, which could lead to redundancy in the test data.



Table 14: Rasch Summary Results Grade 1

Focus area	Result	Comment
INFIT MNSQ Persons	1.00	Sufficient for measurement
OUTFIT MNSQ Persons	1.07	Sufficient for measurement
INFIT MNSQ Items	99	Sufficient for measurement
OUTFIT MNSQ Items	1.09	Sufficient for measurement
Person Reliability	0.89	High reliability
Item Reliability	0.99	High reliability
Misfitting items	0	Items should be investigated and possibly reformulated

The item-person maps indicated that both the item difficulty and person ability distributions closely resembled normal distributions, with some gaps in the person ability distribution that were not represented by the available items (especially at level 2). Notably, the mean of the item difficulties and the mean of the person abilities were virtually the same, which suggests that the MARKO-D assessment was well targeted for Grade 1 learners.

Further analysis revealed that the most easiest items for Grade 1 learners corresponded to Level 1 (counting rhyme) cognitive tasks. According to developmental theory, understanding of counting sequences are foundational to early childhood mathematics education, as they help children develop essential number sense and numeracy skills. Activities that involve counting, recognising patterns, and quantifying objects are crucial for establishing a strong mathematical foundation. Educators often employ manipulatives in teaching, as these hands-on tools actively engage learners, providing them with opportunities to explore counting sequences through play and practical experiences (Clements & Sarama, 2014).

A central concept related to counting sequences is one-to-one correspondence, where each number in a counting sequence is paired with a specific quantity. This understanding is vital for reinforcing the concept of cardinality, the principle that the last number in a counting sequence represents the total quantity of items being counted (Gelman & Gallistel, 1978). Developing a strong grasp of one-to-one correspondence lays the groundwork for more advanced mathematical concepts, such as addition and subtraction. The Rasch analysis is indicative of well-developed counting sequences and one-to-one correspondence.



Grade R learners would benefit from targeted support to reinforce their mastery of foundational mathematical concepts, including the counting rhyme and the ordinality principle. As the academic year progresses, instruction could gradually incorporate relational number strategies, such as "counting on" and "counting all," to facilitate their ability to solve basic addition and subtraction problems.

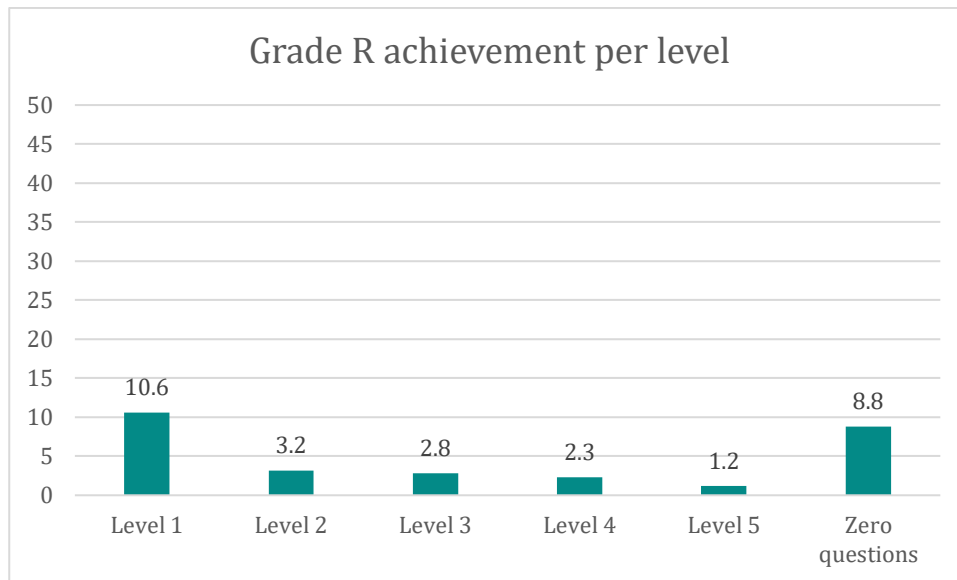


Figure 7: Grade R achievement by level

MARKO-D Grade 1

As anticipated, Grade R learners exhibited low achievement percentages in April of their Grade R year, with scores ranging from 6.8% at Level 1 to 0.9% at Level 5. Additionally, performance declined consistently as task complexity increased across levels. Differences in achievement between genders were neither statistically significant nor meaningful. With the exception of two seven-year-olds, whose performance diverged from expected patterns, no significant differences in achievement were observed across age groups.



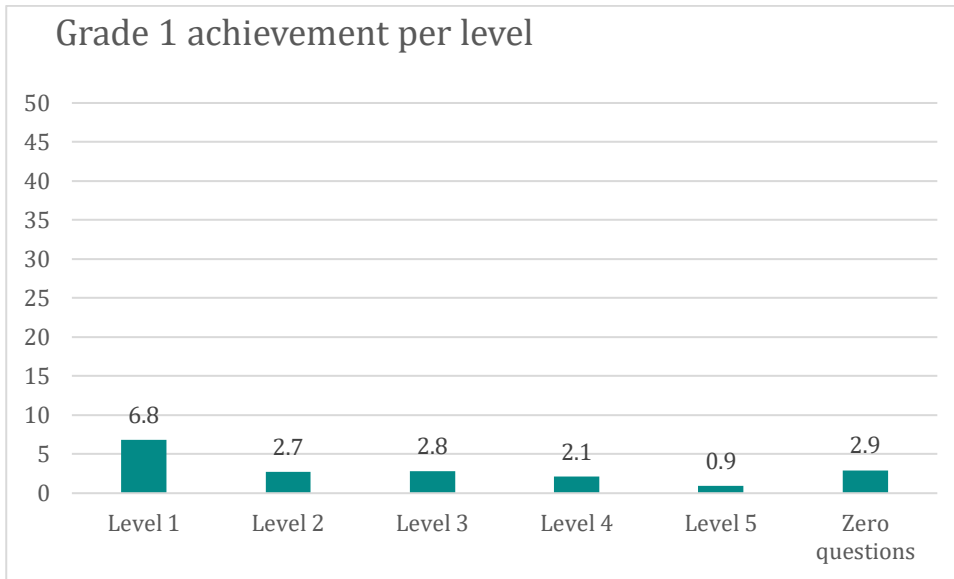


Figure 8: Grade 1 achievement per level

Interestingly, the Grade 1 learners score well below the Grade R learners on the questions relating to zero. This is a phenomena for further research.

Similar to Grade R learners, Grade 1 learners would benefit from focused support to strengthen their understanding of essential mathematical concepts, such as the counting rhyme and the ordinality principle. As the year advances, introducing relational number strategies like "counting on" and "counting all" could help them in solving basic addition and subtraction problems. However, instruction should also extend their range of understanding to encompass part-part-whole relationships.

4.3.2 EGMA

Number identification subtest

The Number Identification subtest is timed (60 seconds) with no stop rules, and it consists of 20 items that increase in difficulty. The first three items of the subtest include the numerals 0, 9, or one other single-digit number. The next 12 items consist of two-digit numbers from 10 to 99, and the last five items are three-digit numbers from 100 to 999. Learners are asked to say each number aloud.

A substantial proportion of Grade R learners (51.0%) were unable to read any numbers, while another significant segment (43.2%) could read only up to five numbers per minute. A small percentage of Grade R learners demonstrated the ability to read more than five numbers per minute.

In Grade 1, there was a marked improvement in number reading per minute (NRPM) compared to Grade R. Only 8.2% of Grade 1 learners were unable to read any numbers, a notable reduction from Grade R levels. The majority (58.8%) could read up to five numbers per minute, a slight increase relative to Grade R. Additionally, 33.0% of Grade 1 learners could read more than five numbers per minute, indicating enhanced fluency in number reading.

This data underscores the developmental progression in number reading fluency between Grade R and Grade 1. While the majority of Grade R learners faces challenges with number reading, Grade 1 learners exhibited substantial gains, with a greater proportion reading numbers at an increased rate. This was an



impressive feat for Grade 1s in April of the Grade 1 year, since the curriculum required number recognition up to 99 (numbers 1-15) by the end of the Grade 1 year.

Table 15: Distribution of correct numbers read per minute (NRPM) by grade

Numbers Read Per Minute (NRPM)	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	150	51.0%	24	8.2%
5 or fewer numbers	127	43.2%	171	58.8%
6 - 10 numbers	11	3.7%	61	21.0%
11-15 numbers	4	1.4%	35	12.0%
16 - 20 numbers	2	0.0%	0	0.0%
Total	294	100.0%	291	100.0%

Number discrimination subtest

The Number Discrimination subtest is an untimed test of 10 items with a stop rule after four successive errors. Each item consists of a set of two numbers, one of which is greater than the other. The first item is a set of one-digit numbers, the next five items are sets of two-digit numbers, and the last four items are three-digit numbers. Learners state the higher of each set of two numbers (pointing at the correct number is insufficient evidence for scoring).

The majority of Grade R learners (51.8%) gave no correct answers. A considerable proportion (32.9%) managed to get one answer right. Only a small percentage achieved two or more correct answers.

While significantly higher correct answer proportions compared to Grade R learners across all categories were evident, a substantial portion of both Grade R (51.8%) and Grade 1 (21.5%) learners provided no correct answers. A significant portion of Grade 1s (21.5%) gave no correct answers, however Grade 1 shows improvement compared to Grade R. The highest proportion of correct answers for Grade 1 is one (38.9%). A noticeable percentage achieved two or more correct answers, indicating better number discrimination skills. It should be noted that only the first 6 questions fell within the range 0-100 and the first 2 questions within the range 0-25.



Table 16: Proportions of correct answers of number discrimination by Grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	74	51.8%	57	21.5%
1 answer	47	32.9%	103	38.9%
2 answers	7	4.9%	26	9.8%
3 answers	5	3.5%	13	4.9%
4 answers	3	2.1%	20	7.6%
5 answers	3	2.1%	21	7.9%
6 answers	2	1.4%	17	6.4%
7 answers	1	0.7%	2	0.8%
8 answers	0	0.0%	6	2.3%
9 answers	1	0.7%	0	0.0%
10 answers	0	0.0%	0	0.0%
Total	143	100.0%	265	100.0%

Number Patterns (Missing Numbers) subtest

The Missing Number subtest is an untimed test of 10 items with a stop rule after four successive errors. The items are presented as four horizontally aligned boxes, three of which contain numbers and one of which is empty (the target missing number). Eight of the items increase in number from left to right; two of the items decrease in number from left to right. Items 1, 2, and 6 increase by one (in a set of one-, two-, and three-digit numbers, respectively). Items 3, 4, 5, and 8 increase by tens, hundreds, twos, and fives, respectively. Items 7 and 9 decrease by twos and tens, respectively. The last item with numerals within the range of 1–20 increases by fives, but does not begin with a multiple of five. Learners are asked to state the number that belongs in the empty box.

A majority of Grade R learners (67.1%) were unable to recognise any correct number patterns. A smaller percentage (23.8%) successfully identified one correct pattern, while fewer than 10% of learners could recognise two or more correct patterns.

In contrast, while a considerable portion of Grade 1 learners (23.8%) also struggled to identify any correct patterns, their performance showed clear improvement over Grade R. The largest proportion of Grade 1 learners (40.4%) identified one correct pattern, and a notable segment (24.5%) demonstrated more advanced pattern recognition skills by identifying two or more correct patterns. This progression reflects an overall development in number pattern recognition from Grade R to Grade 1.



Table 17: Proportions of correct number pattern recognised by Grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	96	67.1%	63	23.8%
1 answer	34	23.8%	107	40.4%
2 answers	8	5.6%	65	24.5%
3 answers	2	1.4%	20	7.6%
4 answers	3	2.1%	5	1.9%
5 answers	0	0.0%	4	1.5%
6 answers	0	0.0%	1	0.4%
7 answers	0	0.0%	0	0.0%
8 answers	0	0.0%	0	0.0%
9 answers	0	0.0%	0	0.0%
10 answers	0	0.0%	0	0.0%
Total	143	100.0%	265	100.0%

Addition (level 1) subtest

The **Addition and Subtraction Level 1 subtests** are timed tests (60 seconds) consisting of 20 items each that increase in difficulty. No addends are greater than 10, and no sums are greater than 19. The subtraction problems are the inverse of the addition problems.

The majority of Grade R learners (56.6%) were unable to solve any Level 1 addition problems correctly. A smaller portion (42.7%) managed to solve five or fewer problems accurately, while only a minimal percentage (0.7%) successfully solved between six and twenty problems.

In comparison, Grade 1 learners demonstrated notable improvement, though 24.9% still could not solve any problems correctly. A majority of Grade 1 learners (61.5%) could solve up to five problems correctly, and a meaningful subset (13.6%) achieved accuracy on six to twenty problems. This shift indicates a developmental progression in basic addition skills from Grade R to Grade 1.



Table 18: Proportion of correct addition (level 1) sums by grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	81	56.6%	66	24.9%
5 or fewer answers	61	42.7%	163	61.5%
6-20 answers	1	0.7%	36	13.6%
Total	143	100.0%	265	100.0%

Addition (level 2) subtest

The Addition and Subtraction Levels 2 subtests are untimed tests consisting of five items each that increase in difficulty, with a stop rule of four successive errors. Addition Level 2 is not given to learners who receive a score of zero for Addition Level 1, and Subtraction Level 2 is not given to learners who receive a score of zero for Subtraction Level 1. No sums are greater than 70. The subtraction problems are the inverse of the addition problems.

An overwhelming majority of Grade R learners (96.7%) were unable to solve any Level 2 addition problems correctly, with only a minimal number (2 out of 61) managing to solve one or more problems.

While Grade 1 learners demonstrated a modest improvement, a large majority (88.9%) still could not correctly solve any Level 2 problems. Fewer than 15% of Grade 1 learners succeeded in solving at least one problem, indicating that, although some advancement is evident, proficiency in Level 2 addition remains limited at this stage.

Table 19: Proportion of correct addition (level 2) sums by grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	59	96.7%	169	88.9%
1 answer	1	1.6%	7	3.7%
2 answers	0	0.0%	11	5.8%
3 answers	0	0.0%	1	0.5%
4 answers	1	1.6%	2	1.1%
5 answers	0	0.0%	0	0.0%
Total	61	100.0%	190	100.0%

Subtraction (level 1) subtest

As indicated for Addition (level 1) subtest, the **Addition and Subtraction Level 1 subtests** are timed tests (60 seconds) consisting of 20 items each that increase in difficulty. No addends are greater than 10, and no sums are greater than 19. The subtraction problems are the inverse of the addition problems.



A substantial majority of Grade R learners (72%) were unable to correctly solve any Level 1 subtraction problems, while only 27.3% were able to solve up to five problems accurately. A minimal fraction (0.7%) demonstrated the ability to solve between six and twenty problems.

In comparison, Grade 1 learners exhibited improved performance, although a considerable proportion (58.1%) still could not solve any problems correctly. A notable 38.5% of Grade 1 learners succeeded in solving up to five problems, while a smaller subset (13.6%) managed to solve between six and twenty problems. This progression suggests a developmental improvement in basic subtraction skills from Grade R to Grade 1, though challenges remain prevalent.

Table 20: Proportion of correct subtraction (level 1) sums by grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	103	72.0%	154	58.1%
5 or fewer answers	39	27.3%	102	38.5%
6-20 answers	1	0.7%	9	3.4%
Total	143	100.0%	265	100.0%

Subtraction (level 2) subtest

As indicated in the Addition (level 2) subtest, the **Addition and Subtraction Levels 2** subtests are untimed tests consisting of five items each that increase in difficulty, with a stop rule of four successive errors. Addition Level 2 is not given to learners who receive a score of zero for Addition Level 1, and Subtraction Level 2 is not given to learners who receive a score of zero for Subtraction Level 1. No sums are greater than 70. The subtraction problems are the inverse of the addition problems.

An overwhelming majority of Grade R learners (97.5%) were unable to solve any Level 2 subtraction problems correctly, with only a single learner managing to solve one problem.

Although there is slight improvement in Grade 1, the majority of learners (92.8%) still could not solve any Level 2 subtraction problems accurately, and fewer than 10% succeeded in solving at least one problem.

The performance disparity between Grade R and Grade 1 learners is notable; however, the vast majority in both grades encountered substantial difficulty with Level 2 subtraction, underscoring a shared challenge in mastering higher-level subtraction tasks at this stage. This is not surprising in April of the Grade 1 year.



Table 21: Proportion of correct subtraction (level 2) sums by grade

Correct Answers	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	39	97.5%	103	92.8%
1 answer	1	2.5%	6	5.4%
2 answers	0	0.0%	1	0.9%
3 answers	0	0.0%	0	0%
4 answers	0	0.0%	1	0.9%
5 answers	0	0.0%	0	0.0%
Total	40	100.0%	111	100.0%

Inferential statistics

The **Pearson's correlation coefficient** measures the linear relationship between two variables. Values range from -1 (perfect negative correlation) to 1 (perfect positive correlation). A value close to 0 indicates a weak or no linear relationship.

The **Spearman's rank correlation coefficient** measures the monotonic relationship between two variables. It assesses how well the relationship between the variables can be described by a monotonic function (one that always increases or always decreases). Values also range from -1 to 1, with the same interpretation as Pearson's coefficient.

Association of number recognition speed and accuracy

For Grade R learners, both Pearson and Spearman correlation coefficients are low, indicating a weak association between number discrimination speed and recognition accuracy. This suggests limited alignment between the rate of number discrimination and the accuracy of number recognition at this developmental stage.

In contrast, Grade 1 learners exhibit moderate Pearson and Spearman correlation coefficients, indicating a moderately positive relationship between number discrimination speed and recognition accuracy. This suggests that, generally, as number discrimination speed improves, recognition accuracy also tends to increase; however, the correlation remains far from perfect, highlighting that other factors may influence these skills.

Table 22: Is number recognition speed associated with accuracy?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
0.1797	0.1448	0.5388	0.5625



Association of number recognition speed with addition and subtraction (level 1)

In Grade R, the relationship between number reading speed and Level 1 addition accuracy appears to be negligible, as evidenced by very low correlation coefficients (both Pearson and Spearman are close to zero). A weak negative correlation is observed between number reading speed and Level 1 subtraction accuracy, indicating that as reading speed increases, subtraction accuracy tends to decrease slightly. However, this relationship remains weak and lacks substantial significance.

Similarly, in Grade 1, there is no significant relationship between number reading speed and addition accuracy at Level 1, as indicated by the very low correlation coefficients. A weak positive relationship is found between number reading speed and Level 1 subtraction accuracy, suggesting that as reading speed increases, subtraction accuracy also tends to increase slightly. Again, this relationship is weak and does not demonstrate a strong or meaningful association.

Overall, the relationship between number reading speed and performance in both addition and subtraction at Level 1 is weak for both Grade R and Grade 1 learners. These findings imply that number reading speed is not a robust predictor of success in these arithmetic operations at this developmental stage.

Table 23: Is number recognition speed associated with addition and subtraction (level 1)?

Subtest	Grade R		Grade 1	
	Pearson	Spearman	Pearson	Spearman
Addition L1 Correct	-0.0465	-0.0175	0.0512	-0.002
Subtraction L1 Correct	-0.2962	-0.2655	0.1356	0.2006

Association of number recognition speed with addition and subtraction (level 2)

In Grade R, there appears to be a weak positive relationship between number reading speed and Level 2 addition accuracy, as indicated by both Pearson and Spearman correlation coefficients, which are slightly positive but close to zero. In contrast, a moderate negative correlation exists between number reading speed and Level 2 subtraction accuracy, suggesting that as reading speed increases, subtraction accuracy may decrease for Grade R learners.

The relationship between number reading speed and Level 2 addition and subtraction performance becomes more nuanced when comparing Grade R to Grade 1 learners. For Grade R, a slight positive association with addition is observed, while in Grade 1, a weak positive association appears between reading speed and subtraction accuracy. Nevertheless, the overall relationship remains weak across grades, indicating that reading speed has minimal impact on Level 2 addition and subtraction accuracy in early learners.

Table 24: Is the number recognition speed associated with addition and subtraction (level 2)?

Subtest	Grade R		Grade 1	
	Pearson	Spearman	Pearson	Spearman
Addition L2 Correct	0.1636	0.2275	0.1141	0.1016
Subtraction L2 Correct	-0.3411	-0.3337	0.1034	0.1451



4.3.3 EGRA

Letter sounds subtest

The letter sounds subtask assesses learners' ability to recognise letters and accurately produce their corresponding sounds. In this task, learners are presented with a sheet of letters and instructed to read aloud as many as they can within a one-minute time frame, emphasising both speed and accuracy. The EGRA administrator monitors the time and records the number of correct letters identified per minute (CLPM).

A substantial majority (67.1%) of Grade R learners were unable to correctly identify any letter sounds within a one-minute timeframe. A smaller proportion (28%) could identify 10 or fewer letter sounds per minute, while a very limited number (less than 4%) could identify more than 10 letter sounds per minute.

In comparison, Grade 1 learners showed an improvement, with a smaller proportion (25.3%) unable to identify any letter sounds per minute. The distribution of CLPM was more evenly spread among Grade 1 learners than in Grade R, indicating a wider range of performance. A notable proportion of Grade 1 learners (21.8%) could identify between 11 and 20 letter sounds per minute, reflecting a more advanced level of letter sound recognition compared to their Grade R counterparts.

Table 25: Distribution of correct letter sounds read per minute (CLPM) by grade

CLPM (correct letter sounds per minute)	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	194	67.1%	74	25.3%
10 or fewer letters	81	28.0%	82	28.0%
11 - 20 letters	10	3.5%	64	21.8%
21-30 letters	2	0.7%	40	13.7%
31-40 letters	2	0.7%	18	6.1%
41-50 letters	0	0.0%	11	3.8%
51-60 letters	0	0.0%	3	1.0%
61-70 letters	0	0.0%	1	0.3%
71-80 letters	0	0.0%	0	0.0%
81-90 letters	0	0.0%	0	0.0%
91-100 letters	0	0.0%	0	0.0%
Total	289	100.0%	293	100.0%

A substantial majority (67.1%) of Grade R learners were classified as "non-fluent" in letter sound recognition. A smaller proportion (31.1%) fell into the "low-accuracy reader" category. Only a very limited



number of Grade R learners (1.7%) were categorised as "developing accuracy" readers, and none were classified as "accurate readers."

In contrast, Grade 1 learners demonstrated some improvement in letter sound recognition. While still a considerable proportion (25.3%) were classified as "non-fluent," the majority (48.5%) were considered "low-accuracy readers." A notable portion (20.1%) were classified as "developing accuracy" readers, and a small but significant percentage (6.1%) were identified as "accurate readers."

Table 26: Distribution of letter sounds benchmark scores by grade

Letter Sounds: CLPM benchmarks	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Non-fluent	194	67.1%	74	25.3%
Low-accuracy reader	90	31.1%	142	48.5%
Developing accuracy	5	1.7%	59	20.1%
Accurate reader	0	0.0%	18	6.1%
Total	289	100.0%	293	100.0%

Grade 1 learners exhibited higher mean scores across all reading performance measures compared to Grade R learners. This suggests that Grade 1 learners generally demonstrated faster letter sound recognition, greater accuracy, and fewer errors in identification. Notably, there was a significant difference in both accuracy and error rates between the two groups, with Grade 1 learners showing considerably higher accuracy and lower error rates. The differences in Letters Per Minute (LPM) and Incorrect LPM were also substantial, indicating that Grade 1 learners were able to attempt more letter sounds and make fewer incorrect attempts than their Grade R counterparts.

This data revealed a clear developmental progression in letter sound recognition from Grade R to Grade 1, with a higher percentage of learners achieving higher proficiency levels in Grade 1. The findings underscored the need for targeted interventions to enhance letter sound recognition, especially for Grade R learners.

At the point a child scored zero for the first ten letter sounds, a stopping rule was applied. As a result, 194 Grade R learners discontinued the assessment, while 95 learners continued. In Grade 1, 74 learners stopped at this stage, leaving 219 to complete the assessment.



Table 27: Total average score per each letter indicator

Summary variables	Grade R		Grade 1	
	N	Mean	N	Mean
LPM (total attempts - speed)	95	14.11	219	24.12
CLPM (total attempts correct)	95	5.95	219	16.98
Incorrect LPM (total attempts incorrect)	95	8.16	219	7.14
Accuracy (percent of attempts correct)	95	38.29	219	64.12
Error rate (percent of attempts incorrect)	95	61.72	219	35.89

Word reading subtest

The word reading subtest presents learners with a grid of words that are appropriate for their grade level, and that they are likely to have encountered previously. Learners are instructed to read aloud as many words as they can within one minute. The EGRA administrator times the session, documenting any errors made by the learner during the reading. The final score is reported as the number of correct words read per minute (CWPM).

A substantial majority (86.3%) of Grade R learners were unable to read any words correctly within one minute. A smaller proportion (13.7%) could read 10 or fewer words per minute, and no Grade R learners were able to read more than 10 words per minute.

In contrast, a significant percentage (63.5%) of Grade 1 learners also could not read any words correctly within one minute. However, a larger proportion of Grade 1 learners (30.1%) could read 10 or fewer words per minute compared to Grade R. Additionally, a small but noteworthy percentage (6.4%) of Grade 1 learners could read between 11 and 20 words per minute.



Table 28: Distribution of CWPM by grade

CWPM (correct words per minute)	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	82	86.3%	139	63.5%
10 or fewer words	6	6.3%	66	30.1%
11-20 words	4	4.2%	11	5.0%
21-30 words	2	2.1%	3	1.4%
31-40 words	1	1.1%	0	0.0%
41-50 words	0	0.0%	0	0.0%
51-60 words	0	0.0%	0	0.0%
61-70 words	0	0.0%	0	0.0%
71-80 words	0	0.0%	0	0.0%
81-90 words	0	0.0%	0	0.0%
91-100 words	0	0.0%	0	0.0%
Total	95	100.0%	219	100.0%

A significant majority (86.3%) of Grade R learners were classified as "non-fluent" readers, with a smaller proportion (10.5%) categorised as "low-accuracy readers." Only a very small percentage (2.1%) of Grade R learners were classified as "developing accuracy" readers, and none were considered "accurate readers."

In Grade 1, a smaller proportion (63.5%) of learners were categorised as "non-fluent" readers, while the majority (35.2%) fell into the "low-accuracy reader" category. A very small percentage (1.4%) of Grade 1 learners were classified as "developing accuracy" readers, and, similarly to Grade R, no learners were classified as "accurate readers."

Table 29: Distribution of word reading benchmark scores by grade

Word reading: CWPM benchmarks	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Non-fluent	82	86.3%	139	63.5%
Low-accuracy reader	10	10.5%	77	35.2%
Developing accuracy	2	2.1%	3	1.4%
Accurate reader	1	1.1%	0	0.0%
Total	95	100.0%	219	100.0%



Grade R learners exhibited a slightly higher words per minute (WPM) rate compared to Grade 1 learners. However, Grade 1 learners demonstrated a higher correct words per minute (CWPM) rate, indicating that although Grade R learners attempted to read more words, Grade 1 learners were more accurate in their word reading. The error rate was significantly higher among Grade R learners than their Grade 1 counterparts. Both grades showed relatively low accuracy in word reading, indicating a need for further development of these skills.

Similarly, the stopping rule resulted in 82 out of 95 Grade R learners discontinuing the assessment, with only 13 continuing. In Grade 1, 139 learners stopped at this point, while 80 learners continued with the assessment.

Table 30: Total average score per each word reading indicator

Summary variables	Grade R		Grade 1	
	N	Mean	N	Mean
WPM (total attempts - speed)	95	9.44	219	9.20
CWPM (total attempts correct)	95	1.82	219	2.44
Incorrect WPM (total attempts incorrect)	95	7.62	219	6.75
Accuracy (percent of attempts correct)	95	2.23	219	2.92
Error rate (percent of attempts incorrect)	95	89.79	219	79.08

Passage reading subtest

The passage reading subtask in EGRA assesses both the speed and accuracy with which a learner reads. It is a crucial component of the assessment, as it integrates foundational reading skills with the learner's ability to fluently and effortlessly read a given passage. Learners are provided with a brief written passage on a familiar topic and are instructed to read aloud "quickly but carefully." They are allotted one minute to complete the task, beginning from the moment they start reading. The EGRA administrator times the reading session, recording any errors made during the reading process. The learner's performance is quantified as the number of correct words per minute (CWPM).

Overall, Grade 1 learners demonstrated higher correct words per minute (CWPM) scores compared to Grade R learners, indicating that Grade 1 learners were generally more proficient in both the speed and accuracy of word reading within a passage. Despite this, a significant proportion of learners from both Grade R and Grade 1 exhibited low CWPM scores, suggesting that passage reading skills were still in the process of development at these grade levels. In Grade R, the majority of learners (23.1%) were unable to read any words correctly within a minute, highlighting that a substantial group of Grade R learners were still acquiring fundamental passage reading skills. In contrast, only 7.5% of Grade 1 learners failed to read



any words correctly, reflecting a notable advancement in their passage reading abilities compared to their Grade R counterparts.

Table 31: Distribution of correct word reading in a passage per minute (CWPM) by grade

CWPM (correct words per minute)	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	3	23.1%	6	7.5%
10 or fewer words	1	7.7%	28	35.0%
11-20 words	3	23.1%	15	18.8%
21-30 words	3	23.1%	12	15.0%
31-40 words	1	7.7%	7	8.8%
41-50 words	2	15.4%	6	7.5%
51-60 words	0	0.0%	2	2.5%
61-70 words	0	0.0%	3	3.8%
71-80 words	0	0.0%	1	1.3%
81-90 words	0	0.0%	0	0.0%
91-100 words	0	0.0%	0	0.0%
Total	13	100.0%	80	100.0%

Grade 1 learners generally exhibited higher word reading proficiency compared to Grade R learners, as evidenced by the lower percentage of Grade 1 learners classified as "non-fluent." A significant proportion of Grade R learners (23.1%) fell into the "non-fluent" category, indicating a greater challenge in typing proficiency at this stage. In contrast, the majority of Grade 1 learners (51.2%) are categorised as "low-accuracy readers," suggesting improvement over Grade R but also indicating that further development was needed in passage word reading skills. Additionally, there was a notable increase in the proportion of learners classified as "accurate readers" from Grade R to Grade 1, signalling a significant progression in passage word reading proficiency between these two grades.



Table 32: Distribution of word reading in a passage benchmark scores by grade

CWPM benchmarks	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
Non-fluent	3	23.1%	6	7.5%
Low-accuracy reader	1	7.7%	41	51.2%
Developing accuracy	3	23.1%	16	20.0%
Accurate reader	6	46.2%	17	21.3%
Total	13	100.0%	80	100.0%

Grade R learners demonstrated higher average words per minute (WPM) and correct words per minute (CWPM) in passage word reading than Grade 1 learners. However, Grade R learners also exhibited a higher average incorrect words per minute (incorrect WPM) compared to their Grade 1 counterparts, indicating that they made more errors in reading passage words. Despite this, Grade R learners showed a higher accuracy rate and a lower error rate overall than Grade 1 learners. This apparent contradiction may be attributed to the smaller sample size in Grade R (N=13) compared to Grade 1 (N=80), as smaller sample sizes often result in greater variability in the data.

Of the 13 Grade R learners, only 10 proceeded to the comprehension questions, whereas 74 Grade 1 learners continued with this portion of the assessment.

Comprehension subtest

Reading comprehension is the ultimate objective of reading—understanding and interpreting the content that is read. It is a multifaceted skill that necessitates proficiency in various foundational reading abilities. The comprehension subtask is closely linked with the passage reading subtask, as it assesses the learner’s ability to understand the material they have just read. Based on the portion of the passage the learner successfully reads, the EGRA administrator poses up to five questions related to the content of the story. The administrator records the number of questions answered correctly, providing an indication of the learner's comprehension level.

In general, Grade 1 learners outperformed Grade R learners in terms of the number of correct answers on comprehension questions, as evidenced by the distribution across the different performance categories. However, a substantial proportion of both Grade R and Grade 1 learners scored relatively low on these questions, indicating that there was considerable room for improvement in their comprehension skills. For Grade R learners, the most frequent score was one correct answer, while for Grade 1 learners, the most common scores were one and two correct answers. This pattern suggests that Grade 1 learners were performing marginally better than their Grade R counterparts, though the difference was not statistically significant.



Table 33: Distribution of correct comprehension questions by grade

Number of questions correct	Grade R		Grade 1	
	Frequency	Percent	Frequency	Percent
None	3	30.0%	17	23.0%
1	3	30.0%	23	31.1%
2	2	20.0%	22	29.7%
3	1	10.0%	7	9.5%
4	0	0.0%	2	2.7%
5	1	10.0%	3	4.1%
Total	10	100.0%	74	100.0%

Inferential statistics

The Pearson's correlation coefficient measures the linear relationship between two variables. Values range from -1 (perfect negative correlation) to 1 (perfect positive correlation). A value close to 0 indicates a weak or no linear relationship.

The Spearman's rank correlation coefficient measures the monotonic relationship between two variables. It assesses how well the relationship between the variables can be described by a monotonic function (one that always increases or always decreases). Values also range from -1 to 1, with the same interpretation as Pearson's coefficient.

Letter sound accuracy associated with speed

Both Pearson's and Spearman's correlation coefficients were positive for both Grade R and Grade 1 learners, indicating a generally positive relationship between letter sound accuracy and speed. Specifically, as learners increased their speed in attempting letter sounds, their accuracy also tended to improve. The correlation coefficients for Grade 1 learners were higher than those for Grade R learners, suggesting a stronger association between accuracy and speed among older learners.

The difference between Pearson's and Spearman's coefficients for each grade might indicate that the relationship between accuracy and speed was not perfectly linear. Spearman's coefficient, which measures monotonic relationships, was slightly lower than Pearson's in Grade R, suggesting some non-linearity.

Table 34: Is letter reading accuracy associated with speed?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
0.8235	0.602	0.9108	0.9318



Word reading speed associated with accuracy

For both Grade R and Grade 1, there was a positive correlation between word reading accuracy and speed, suggesting that as learners read faster, they tended to be more accurate, and vice versa.

The correlation coefficients were higher for Grade R compared to Grade 1, indicating a stronger association between accuracy and speed in younger learners.

The difference between Pearson's and Spearman's coefficients suggests that the relationship between accuracy and speed might not be perfectly linear, especially in Grade 1.

Table 35: Is word reading accuracy associated with speed?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
0.9506	0.7282	0.8179	0.581

Passage reading accuracy associated with speed

For both Grade R and Grade 1 learners, there was a robust positive correlation between passage word reading accuracy and speed, suggesting that as learners increased their reading speed, their accuracy also improved, and conversely. The correlation coefficients for Grade R were exceptionally high, indicating a stronger relationship between accuracy and speed in younger learners compared to Grade 1. Both Pearson's and Spearman's coefficients demonstrate strong and consistent associations for both grades, revealing a clear linear and monotonic pattern between accuracy and speed.

Table 36: Is passage word reading accuracy associated with speed?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
0.9903	0.9805	0.918	0.8556

Word reading accuracy associated and speed associated with comprehension?

The correlation coefficients between WPM (word), CWPM (word), WPM (passage), CWPM (passage), and comprehension were relatively low for Grade R learners, suggesting a weak or negligible relationship between reading speed/accuracy and comprehension. This implies that, for these young learners, increased reading speed or accuracy did not necessarily correspond to improved comprehension.

In Grade 1, while Spearman's correlation showed a slightly stronger indication of a relationship between passage reading speed and accuracy, the overall evidence for a significant association remained limited.



Table 37: Is word reading accuracy and speed associated with comprehension?

	Grade R		Grade 1	
	Pearson	Spearman	Pearson	Spearman
WPM (word)	0.0671	0.1646	0.1714	0.231
CWPM (word)	0.0023	0.3763	0.3471	0.2149
WPM (passage)	0.3836	0.3021	0.2613	0.5007
CWPM (passage)	0.4288	0.477	0.3858	0.599

Passage reading accuracy and letter sound accuracy

There was a moderate positive correlation between passage reading accuracy and letter sound accuracy in Grade R, as indicated by both Pearson's (0.4023) and Spearman's (0.3759) coefficients. This suggests a moderate relationship between these two variables in this grade level.

In contrast, the correlation between passage reading accuracy and letter sound accuracy in Grade 1 was weak and inconsistent, with a slightly negative Pearson correlation (-0.1342) and a weak positive Spearman correlation (0.0846). This indicates a minimal relationship between these variables in Grade 1.

The data revealed a complex relationship between word reading accuracy and letter sound accuracy across the two grades. While there was a moderate to strong positive correlation for word reading accuracy in Grade R, this relationship weakened considerably in Grade 1. Additionally, the correlation between passage reading accuracy and letter sound accuracy was inconsistent across the two grades, suggesting differing patterns of association.

Table 38: Is word reading accuracy associated with letter sound accuracy?

	Grade R		Grade 1	
	Pearson	Spearman	Pearson	Spearman
Accuracy (passage reading)	0.4023	0.3759	-0.1342	0.0846
Accuracy (word reading)	0.5945	0.7044	0.1717	0.241

Total letter sound attempts associated with letter sound error rate

The correlation analysis revealed a moderate negative relationship between total letter sound attempts and letter sound error rates for Grade R learners, as indicated by Pearson's (-0.5349) and Spearman's (-0.545) coefficients. This suggests that Grade R learners who attempt more letter sounds tended to exhibit lower error rates, albeit with a moderate effect size.

In Grade 1, there was a strong negative correlation between total letter sound attempts and letter sound error rates, demonstrated by Pearson's (-0.6651) and Spearman's (-0.8192) coefficients. This stronger association indicated that Grade 1 learners who attempted a higher number of letter sounds generally showed significantly reduced error rates, underscoring a more robust relationship between these variables in this grade level.



Table 39: Is total letter sound attempts associated with letter sound error rate?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
-0.5349	-0.545	-0.6651	-0.8192

Total word reading attempts associated with word reading error rate

The data revealed a moderate to strong negative correlation between total word reading attempts and word reading error rates for Grade R learners, with Pearson’s coefficient at -0.6245 and Spearman’s at -0.6528. This indicates that Grade R learners who attempted more words generally exhibited lower error rates, suggesting a substantial relationship between reading attempts and accuracy at this developmental stage.

In contrast, Grade 1 learners demonstrated a moderate negative correlation between total word reading attempts and word reading error rates, reflected by Pearson’s coefficient of -0.4333 and Spearman’s of -0.4842. While the correlation remained negative, indicating that increased word reading attempts were associated with reduced error rates, the relationship was less pronounced than in Grade R.

Analysis of letter sound and word reading error rates for both grades underscored a consistent negative relationship between total attempts and error rates in letter sound and word reading tasks. This trend suggested that increased practice in letter sound and word reading tasks correlated with improved accuracy. The stronger correlation in Grade 1 for letter sound tasks implied that the link between practice and accuracy strengthened as learners develop, particularly in foundational reading skills.

Table 40: Is a total word reading attempts associated with word reading error rate?

Grade R		Grade 1	
Pearson	Spearman	Pearson	Spearman
-0.6245	-0.6528	-0.4333	-0.4842

Error rate associated with comprehension

In Grade R, there was a weak to moderate positive correlation between letter sound error rates and comprehension, as reflected in Pearson (0.2244) and Spearman (0.3282) coefficients. This indicated a possible positive relationship wherein higher error rates in letter sound recognition may be associated with slightly lower comprehension.

For Grade 1, a weak negative correlation was observed between letter sound error rates and comprehension, with Pearson (-0.1518) and Spearman (-0.1962) coefficients, suggesting a minimal inverse relationship where increased letter sound errors were very slightly associated with higher comprehension.

In examining the relationship between word reading error rates and comprehension, Grade R showed a negligible correlation, with Pearson (-0.0045) and Spearman (0.1532) coefficients, indicating no meaningful association. Conversely, Grade 1 demonstrated a moderate negative correlation, with Pearson



(-0.4021) and Spearman (-0.3774) coefficients, highlighting a more evident relationship where higher word reading errors correlated with lower comprehension.

Both grades displayed a moderate negative correlation between passage error rate and comprehension, as shown by Pearson and Spearman coefficients. This consistent relationship implied that higher passage reading errors were strongly associated with reduced comprehension.

In summary, the data suggests that reading error rates generally corresponded with lower comprehension, particularly in word reading and passage accuracy. However, the association between letter sound error rate and comprehension appeared more nuanced, showing a weak positive relationship in Grade R and a weak negative relationship in Grade 1.

Table 41: Is error rate associated with comprehension?

	Grade R		Grade 1	
	Pearson	Spearman	Pearson	Spearman
Error rate (letter sound)	0.2244	0.3282	-0.1518	-0.1962
Error rate (word reading)	-0.0045	0.1532	-0.4021	-0.3774
Error rate (passage)	-0.4529	-0.3586	-0.4368	-0.4448



5 Conclusions and recommendations

5.1 Test functioning

The Rasch analysis of the MARKO-D instrument revealed no significant concerns regarding its reliability or validity when used with Sepedi-speaking Grade R and Grade 1 learners. Based on the evidence, including reliability metrics, fit statistics, and construct-related validity, the instrument is considered robust for measuring learner achievement.

A slight floor effect in Grade R is anticipated, as is a potential ceiling effect in Grade 3. However, given the extensive body of research supporting the MARKO-D across various international contexts, the instrument is not only well-suited for analysing learner achievement—such as conducting error analysis to inform targeted interventions rather than generalised approaches—but also facilitates comparisons across diverse learner populations. Furthermore, the availability of published versions of the MARKO-D in other African languages enhances its utility by enabling cross-linguistic comparisons within South Africa, providing valuable insights into regional variations in learner performance.

5.2 General conclusions and recommendations

School readiness and overaged learners in Grades R and 1

To address the educational challenges associated with age disparities in Grade R and Grade 1, it is recommended that the Funda Wande programme closely monitor under-and-overaged learners, particularly those who enter school late or repeat grades. Ensuring school readiness at entry is critical, as it strongly influences both immediate academic success and long-term outcomes. Socioeconomic barriers, notably poverty affecting around 60% of South African children, hinders readiness for formal education and contributes to the high number of unprepared learners. Targeted interventions that enhance early learning experiences and provide resources for educators, especially in under-resourced areas, are essential to support cognitive, social, and emotional preparedness. Prioritising these initiatives can help mitigate the effects of poverty on educational outcomes, foster equity in school readiness, and ultimately improve long-term educational trajectories across socio-economic groups.

We recommend the inclusion of detailed case studies in the midline evaluation to investigate the experiences of both under-aged and over-aged learners. Particular attention should be given to exceptional cases, such as the Grade R learner who demonstrated fluent reading abilities. Analysing the factors contributing to such high performance in unexpected circumstances—commonly referred to as "positive deviance"—is critical. Understanding these drivers can provide valuable insights that could inform strategies to enhance reading achievement across all learners.

Further Recommendations for the 2025 Midline

Given the limitations posed by the small sample size, it is recommended that the 2023 Grade R cohort be tracked as part of the Grade 1 sample in 2025. Similarly, the 2024 Grade 1 sample should guide the selection of the Grade 2 cohort for 2025. To ensure consistency and control for curriculum exposure, it is further recommended that the timing of assessments remain aligned with the current schedule (i.e.,



April). This approach will help standardize data collection and enhance the comparability of results across years.

To enhance the depth of analysis, it is recommended that additional biographical data be collected during the 2025 midline. This could include information such as practitioner and teacher qualifications, as well as the socio-economic context of the schools. However, it is important to acknowledge that the sample size of these groups will significantly influence the types of analyses that can be conducted. If the sample size is insufficient for reliable quantitative analysis, qualitative methods such as case studies could be incorporated to provide meaningful insights.

Overall Interpretation of results on the EGMA, EGRA, and MARKO-D Assessments

It is important to consider that the EGRA and EGMA assessments were conducted in April, early in the academic year for learners in Grades R and 1. Consequently, the poor performance observed aligns with curriculum expectations at this stage. Nonetheless, the results provide valuable insights into the baseline abilities of learners and highlight critical starting points for targeted interventions.

To further enrich the interpretation of these findings, conducting an error analysis and mapping the Bala Fande materials against the MARKO-D levels are recommended. Additionally, exploring unexpected outcomes—such as cases where Grade R learners outperformed Grade 1 learners on "zero questions"—through detailed case studies would offer valuable context and inform future strategies for improvement.

5.3 Conclusions and recommendations for Mathematics

5.3.1. Mastering the counting rhyme, ordinality principal and number relationships

Providing Grade R learners with focused support can strengthen their understanding of essential mathematical concepts, such as the counting rhyme and the ordinality principle. As the year advances, gradually introducing relational number strategies like "counting on" and "counting all", can further aid their ability to tackle simple addition and subtraction tasks.

Grade R learners demonstrate a weak association between number discrimination speed and recognition accuracy. In contrast, Grade 1 learners exhibit a moderate positive relationship between these two skills, indicating a developing but not fully established connection. These findings, coupled with the observed weaknesses in number recognition achievement, underscore the importance of implementing targeted interventions in Grade R. Specifically, activities that reinforce counting principles, such as counting rhymes, number songs, and games that emphasise the ordinal principle, should be prioritized. These approaches have the potential to strengthen foundational numerical competencies and support the progression of early mathematical skills.

Similar to Grade R learners, Grade 1 learners would benefit significantly from targeted interventions aimed at strengthening their understanding of foundational mathematical concepts, including counting



rhymes and the ordinality principle. As the academic year advances, introducing strategies such as "counting all" and "counting on" can further support learners in developing proficiency in basic addition and subtraction tasks. Instructional approaches should also extend to fostering an understanding of part-part-whole relationships, which are critical for building a solid foundation in arithmetic.

However, findings indicate a weak or negligible correlation between number reading speed and the accuracy of both addition and subtraction tasks at Level 1 for learners in Grades R and 1. This suggests that number reading speed is not a reliable predictor of arithmetic performance. Consequently, instructional focus should prioritize conceptual understanding of mathematical principles over the speed of task completion. Emphasising comprehension of underlying numerical concepts rather than timed performance will likely yield more substantial and enduring improvements in mathematical achievement.

5.3.2. Part-part-whole and class inclusion principles

As noted in the RASCH analysis and confirmed in the analysis of the MARKO-D levels for grade R, learners struggle most with Level 4 (part-part whole) and 5 (class inclusion) activities. The theoretical framework underpinning the development of part-part-whole and class inclusion concepts in early mathematics education is grounded in various developmental theories and models. One prominent model is the conceptual numerical development framework proposed by Fritz et al. (2013) for the MARKO-D, which outlines six hierarchical levels of numerical sophistication. This model emphasises a conceptual progression, with children advancing from simpler to more complex numerical understandings. Notably, it aligns with the idea that children's learning is characterised by overlapping waves, whereby abilities may coexist and develop concurrently, rather than through discrete, sequential stages (Fritz et al., 2013). This perspective underscores the fluid and dynamic nature of cognitive development in early childhood mathematics.

The part-part-whole model is a foundational framework in early mathematics education, particularly for understanding the concepts of addition and subtraction in grades R and 1. This model emphasises the relationship between a whole and its constituent parts, facilitating the development of fundamental arithmetic skills. To effectively teach this model, a variety of hands-on activities can be employed to deepen children's understanding of how parts combine to form a whole and vice versa.

The findings indicate that in Grade R, there is a weak positive correlation between reading speed and Level 2 addition, but a moderate negative correlation with subtraction. In contrast, Grade 1 learners exhibit a weak positive relationship between reading speed and subtraction. Overall, the relationship between reading speed and Level 2 addition and subtraction remains weak across both grades, suggesting a limited impact of reading speed on arithmetic performance.

These results align with the Fritz et al. (2003) developmental model, which posits that foundational skills, such as the counting rhyme and the ordinality principle, must at least be well established before more advanced numerical competencies—such as part-part-whole reasoning and class inclusion—can fully develop. This underscores the importance of prioritising early interventions that focus on strengthening lower-level skills to support the progression of higher-level mathematical abilities.



Hands-On Activities for younger learners

Nature Basket Exploration

Creating a nature basket provides an engaging way to develop children's understanding of sizes and quantities, while also incorporating the part-part-whole concept. By using natural items of varying sizes, children can engage in sorting activities that help them recognise and compare concepts such as "big" and "little." This activity supports the development of one-to-one correspondence, matching, and sorting skills—core components of early mathematical learning (Bodrova & Leong, 2007).

Story Box Activities

Incorporating storytelling with physical objects reinforces the part-part-whole model by connecting abstract ideas with tangible experiences. By gathering objects mentioned in a story and using them during the reading, children can visually and physically link the parts of the story to the whole narrative. This active learning strategy not only enhances comprehension but also deepens children's engagement with the material (Cohen, 2016).

Mathematical Stacking and Nesting Using household items that can stack or nest, such as measuring cups or bowls, provides children with a tactile experience of part-part-whole relationships. Manipulating these objects allows children to see how different parts combine to form a whole, as well as how a whole can be divided into smaller parts. This kinesthetic approach fosters cognitive development by encouraging children to explore mathematical concepts through hands-on discovery (Clements & Sarama, 2014).

While some Grade 1 learners have demonstrated proficiency in Level 4 and Level 5 items, there remains room for further development. It is important to recognise that it is typical for children aged 6 to 7 years to be in the process of developing Level 4 and Level 5 concepts, while simultaneously consolidating their understanding of Level 1 to Level 3 concepts. This developmental trajectory aligns with the broader theoretical framework of cognitive development, where children progress through various stages of mathematical understanding at differing rates (Piaget, 1952; Van Oers, 2017). Here follows some activities for supporting the development of the part-part-whole principle as it applies to addition and subtraction.

Application of part-part-whole in addition and subtraction for older learners

Using base ten blocks

Base ten blocks and other manipulatives are effective tools for teaching addition and subtraction within the part-part-whole framework. By presenting a total number of blocks and asking children to separate them into smaller groups, educators can help children visualise the inverse relationship between addition and subtraction. This activity not only clarifies basic mathematical concepts but also enhances problem-solving and critical thinking skills (Baroody, 2004).

Incorporating digital resources

Digital resources, such as interactive games and activities that focus on addition and subtraction, can complement traditional teaching methods. These tools often use visual and auditory stimuli to reinforce the part-part-whole concept, catering to diverse learning styles. For example, digital ten frames can help



children manipulate numbers in a visual context, making the abstract concept of part-part-whole more concrete and accessible (Nicol & Newton, 2013).

5.4 Conclusions and recommendations for Reading

As expected learners struggled with the EGRA. This is possibly due to the timing of the year at which testing took place, especially for the Grade R learners. A focus on pre-reading activities are recommended to improve readiness to start formal reading instruction in Grade 1.

In the letter-sound reading test, Grade 1 learners consistently outperformed Grade R learners across all reading measures, demonstrating faster recognition, higher accuracy, and fewer errors. This progression reflects the expected developmental trajectory from Grade R to Grade 1. To address the gaps observed in Grade R learners, targeted interventions focusing on phonological and phonemic awareness are strongly recommended, with an emphasis on recognizing Grade R-level sounds. Once learners have mastered these foundational skills, Grade 1 instruction should progress to decoding words and combining sounds to form words.

Phonological awareness involves the broader ability to recognize and manipulate sound structures in spoken language, such as syllables, onsets, rimes, and rhymes. It is a critical precursor to reading readiness. Developmentally appropriate tasks for Grade R and Grade 1 learners include clapping syllables in words, identifying rhyming words in stories or songs, and sorting objects by their initial sounds.

Phonemic awareness, a more specific subset of phonological awareness, focuses on isolating and manipulating individual phonemes—the smallest units of sound in language. Tasks that develop phonemic awareness in young learners include segmenting words into individual sounds (e.g., /c/-/a/-/t/ for "cat"), blending sounds to form words, and substituting phonemes to create new words (e.g., changing the /m/ in "mat" to /h/ to form "hat").

Both phonological and phonemic awareness are critical for developing decoding and spelling skills. These foundational abilities can be fostered through play-based, engaging activities that align with the developmental stages of the learners. Implementing such interventions early can establish a strong foundation for long-term reading success.

While Grade R learners demonstrated faster word-reading speed compared to Grade 1 learners, the latter group exhibited greater accuracy in their reading. Both grades, however, showed low overall accuracy, highlighting areas for improvement. It is important to note that these findings align with curriculum expectations: the Grade R curriculum focuses on recognizing their names and high-frequency words, while the Grade 1 curriculum introduces decoding of words with familiar sounds. Therefore, the performance of both Grade R and Grade 1 learners is consistent with expectations for the first term of the school year.

As mentioned earlier, incorporating phonological and phonemic awareness activities would benefit both grades, helping to reinforce foundational literacy skills and enhance reading proficiency over time.



References

- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. Washington DC: American Educational Research Association, American Psychological Association & National Council on measurement in Education.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). New Jersey: Prentice Hall.
- Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR. 20 index, and the Guttman scale response pattern. *Education Research and Perspectives*, 9, 95-104.
- Atmore, E. (2019). Early Childhood Development in South Africa: Progress since the end of apartheid. *Journal of Social Sciences*, 55(2), 121-134.
- Boone, W. J. (2016). Rasch analysis for instrument development: Why, when, and how? *CBE-Life Science Education*, 15 (4), 1-7.
- Boone, W. J., & Noltemeyer, A. (2017). Rasch analysis: A primer for school psychology researchers and practitioners. *Cogent Education*, 4 1416898.
- Baroody, A. J. (2004). *The development of young children's number and operation sense*. ERIC Digest.
- Bodrova, E., & Leong, D. J. (2007). *Developing self-regulation in early childhood education*. National Association for the Education of Young Children.
- Bond, T.G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental measurement in human sciences*. New Jersey: Lawrence Erlbaum Associates Incorporated.
- Bond, T.G., & Fox, C. M. (2015). *Applying the Rasch Model: Fundamental measurement in human sciences* (3rd ed.). New York: Routledge.
- Borsboom, D., Mellenbergh, G.J. & van Heerden, J. (2004). The concept validity. *Psychological Review*, 111, 1061-1071.
- Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach*. Routledge.

- Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2019). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources, 54*(3), 785-813.
- Clements, D. H., & Sarama, J. (2014). *Learning and teaching early math: The learning trajectories approach*. Routledge.
- Cohen, J. (2016). The power of storytelling: Using stories to enhance learning in early childhood education. *Journal of Early Childhood Education, 42*(1), 54-62.
- Dormer, E., Walker, S., & Avant, K. (2022). Frameworks for school readiness: Application in low-resource educational environments. *Education Policy Journal, 18*(3), 221-239.
- Duncan, G. J., Dowsett, C. J., Claessens, A., et al. (2007). School readiness and later achievement. *Developmental Psychology, 43*(6), 1428-1446.
- Department of Basic Education. (2011). *National Curriculum Statement (NCS), Curriculum and Assessment Policy Statement Grade 1-3 Mathematics* (DBE, 2011). Pretoria: Government Printing Works.
- Dixon, P. (2017). *Education in South Africa: A critical perspective*. Routledge.
- Fleisch, B. (2016). The challenges of early grade mathematics in South Africa. *South African Journal of Education, 36*(1), 1-11.
- Foxcroft, C. & Roodt, G. (2013). *Introduction to psychological assessment in the South African context (4th ed)*. Cape Town: Oxford University Press Southern Africa.
- Fritz, A. (2018). Children's mathematical thinking: From early numeracy to the development of mathematical understanding. *Educational Studies in Mathematics, 97*(1), 1-20
- Fritz, A., Ehlert, A., & Balzer, L. (2013). Development of mathematical concepts as basis for an elaborated mathematical understanding. *South African Journal of Childhood Education, 3*(1), 38-67.
- Fritz, A., Ginsburg, H. P., & Fisher, M. H. (2013). A conceptual framework for the development of numerical thinking in early childhood. *Early Childhood Research Quarterly, 28*(4), 648-658. <https://doi.org/10.1016/j.ecresq.2013.01.003>
- Health Professions Council of South Africa (2008a). *Guidelines for good practice in the health care professions. Confidentiality: Protecting and providing information. Booklet 10*. Pretoria: Health Professions Council of South Africa.



- Heckman, J. J., & Karapakula, G. (2019). The Perry Preschoolers at late midlife: A study in design-specific inference. *NBER Working Paper No. 25888*.
- Gelman, R., & Gallistel, C. R. (1978). **The child's understanding of number**. Harvard University Press.
- Gove, A., & Wetterberg, A. (2011). *The Early Grade Reading Assessment: Applications and interventions to improve basic literacy*. Research Triangle Park, NC: RTI Press.
- Gronlund, N. E. (1998). *Assessment of learner achievement* (6th ed.). Boston: Allyn and Bacon.
- Jamieson, L. & Lake, L. (2013). *Children's Act guide for health professionals (5th ed)*. Cape Town: Children's Institute, University of Cape Town.
- Janse Van Rensburg, C. (2015). Poverty and education in South Africa: The effect of socio-economic status on educational outcomes. *Journal of Education Studies, 14*(1), 34-48.
- Kelley, T.L. (1927). *Interpretation of educational measurements*. New York: MacMillan.
- Kline, P. (1993). *The handbook of psychological testing*. Routledge: London.
- Maluleke, M. (2020). Understanding poverty and inequality in South Africa. *South African Journal of Economic and Management Sciences, 23*(1), a3306.
- Messick, (1989). Validity. In R.L. Linn (Ed.). *Educational Measurement* (3rd ed). New York: MacMillan pp. 13-103.
- Moses, E., Van der Berg, S., & Rich, K. (2017). A society divided: How unequal schooling reproduces social inequality in South Africa. *Stellenbosch Economic Working Papers, 7*, 1-20.
- Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2021). *PIRLS 2021 International Results in Reading*. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.
- Nicol, C., & Newton, J. (2013). The impact of digital resources on early mathematics learning. *Computers & Education, 60*(1), 20-28.
- Linacre, J.M. (2002). What do Infit and Outfit and Standardised mean? *Rasch Measurement Transactions, 16*(2), 878.
- Linacre, J. M. (2005). *A user's guide to Winsteps*.
- Piaget, J. (1952). *The origins of intelligence in children*. International Universities Press.



- Piper, B., Zuilkowski, S. S., & Mugenda, A. (2014). Improving reading outcomes in Kenya: First-year effects of the Primary Math and Reading (PRIMR) initiative. *International Journal of Educational Development*, 37, 11-21.
- Pritchett, L., & Viarengo, M. (2009). *The effect of early grade reading assessments on educational quality*. In R. L. McGlinchey & R. A. Harn (Eds.), *Identifying and intervening with reading difficulties* (pp. 21-42).
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks Paedagogiske Institut.
- RTI International. (2009). *Early Grade Reading Assessment Toolkit*. Washington, DC: United States Agency for International Development.
- RTI International. (2016). *EGMA: Early grade mathematics assessment in South Africa*. USAID.
- Sattler, J.M. (2008). *Assessment of children. Cognitive foundations* (5th ed). San Diego: Jerome M. Sattler Publisher Inc.
- Smith, R. M. (2003). *Rasch measurement models: Interpreting WINSTEPS and FACETS outputs*. Maple Grove: JAM Press.
- Spaull, N. (2013). Poverty & privilege: Primary school inequality in South Africa. *International Journal of Educational Development*, 33(5), 436-447.
- Spaull, N. (2013). *South Africa's education crisis: The quality of education in South Africa 1994-2011*. Center for Development and Enterprise.
- Van der Berg, S. (2017). Education and inequality in South Africa. *South African Journal of Economics*, 85(4), 455-485.
- Van Oers, B. (2017). *Learning mathematics in early childhood education: The role of number sense and mathematical reasoning*. Springer.
- Walker, L. O., & Avant, K. C. (2011). *Concept analysis as a strategy for developing theory*. Philadelphia, PA: Mosby Elsevier.
- USAID. (2006). *Early grade reading assessment: A tool for measuring skills*.

