

Pre-analysis plan: Long term impacts of a successful foundational literacy program in South Africa

October 2nd, 2025

Jacobus Cilliers, Nompumelelo Mohohlwane, Kolosa Nonkenge, Stephen Taylor, Sharon Wolf, Zamangwe Zwane

1 INTRODUCTION

This pre-analysis plan outlines the methodology for evaluating the long-term impacts of a foundational literacy program, the Early Grade Reading Study I (EGRS I), implemented in South Africa between 2015 and 2017. The project seeks to understand whether initial improvements in literacy skills from early-grade interventions translate into sustained academic progress and completion of high school. We will track students approximately 10 years after the program began, measuring outcomes in 2025 when non-repeating students would be in Grade 11, augmenting this data with administrative data on grade attainment and high school completion in 2026 and 2027.

Research questions.

1. Does the program have sustained impacts on home language (Setswana) literacy skills 10 years after the start of the intervention? Did these gains translate into improved English literacy and overall academic performance?
2. Does the early literacy program increase educational attainment, specifically the probability of reaching grade 11 and completing high school with university eligibility?
3. What is the program's cost-effectiveness when accounting for intervention costs and estimated lifetime earnings gains from improved educational attainment?

Motivation. Foundational literacy programs represent a major investment priority for international donors and developing country governments. However, the cost-effectiveness of these programs remains uncertain. While many interventions demonstrate short-term learning gains, rigorous evidence on whether these gains translate into economically meaningful outcomes—such as increased educational attainment and future earnings—is scarce. This evidence gap is increasingly consequential as development budgets contract and funders demand stronger evidence of cost-effectiveness relative to competing interventions in health, infrastructure, and other sectors.

The primary challenge in assessing cost-effectiveness is that the returns to foundational literacy programs are inherently delayed: benefits accrue over decades as improved early skills translate into educational attainment, labor market entry, and lifetime earnings. Most evaluations track students for

only 2-3 years, forcing cost-benefit analyses to rely on strong assumptions about fade-out or persistence that often prove incorrect. Consequently, there is limited causal evidence on the long-run impacts of early gains in foundational literacy and whether investments in this sector generate returns that justify their costs.

This study addresses this evidence gap by tracking students from a foundational literacy intervention for over 10 years, through completion (or non-completion) of secondary schooling. High school completion is a critical outcome for cost-benefit analysis in the South African context, where labor market returns to secondary and tertiary education are substantial (Patrinos & Psacharopoulos, 2018; Branson & Leibbrandt, 2013). Given the high returns to education in South Africa's labor market, a key determinant of future lifetime earnings is whether a child passed high school, especially if their exam performance meets the minimum requirements for applying to universities. By observing actual educational attainment rather than projecting from short-term test scores, we can provide credible estimates of the intervention's long-term impact on economically meaningful outcomes and, consequently, its cost-effectiveness.

In addition to the expected financial benefits, it is possible that improved performance in early grades could improve a child's self-esteem and overall life satisfaction. These are important outcomes in themselves, independent of future earnings, and contribute to a comprehensive assessment of the intervention's welfare impacts.

2 BACKGROUND AND PROGRAM DESCRIPTION

2.1 PROGRAM DESCRIPTION

The Early Grade Reading Study (EGRS) I, implemented **between 2015 and 2017**, was a randomized evaluation of a **structured pedagogy program** aimed at improving the teaching of home-language literacy. The intervention provided teachers with **curriculum-aligned lesson plans and learning materials**, such as graded booklets, flash cards, and posters. Teachers also received support through an external reading **coach** who would visit the school at roughly **a monthly basis** to observe teaching, provide constructive feedback, and sometimes demonstrate effective teaching practices.

The program targeted **Grade 1 teachers in 2015, Grade 2 teachers in 2016, and Grade 3 teachers in 2017**. This design ensured that the same student cohort could potentially benefit from the intervention for up to three consecutive years if they did not repeat a grade. The quality of implementation was high, with a large percentage of treated teachers reporting access to lesson plans and graded reading booklets. Notably, classroom observations revealed an **improvement in teaching practices**, particularly the implementation of **group-guided reading**, which allowed students to practice reading and receive individual attention.

2.2 SAMPLE AND EVALUATION DESIGN

The intervention was implemented in 130 Quintile 1–3 schools, serving primarily poor communities in North-West province. Schools were assigned to treatment using stratified randomization across 10 strata of 13 schools each, based on school size, socio-economic status, and prior performance in the Annual National Assessments. Within each stratum, five schools were randomly assigned to treatment and eight to the control group. Overall, **50 schools were assigned to treatment and 80 were assigned as control**. Randomization achieved balance across treatment and control groups on baseline characteristics, validating the experimental design (Cilliers et al. 2020).

Data was collected from 20 randomly sampled students in each school. These students were surveyed and assessed in home-language literacy across four waves of data collection beginning at the end of Grade 1, again at the ends of Grades 2, 4, and 7. Each round of data collection also included teacher and head teacher surveys.

2.3 PREVIOUS RESULTS

After two years of exposure, the intervention improved student home language literacy by 0.24 standard deviations (SDs) (Cilliers et al. 2020). A long-term evaluation, tracking these students for seven years (four years after the program ended), revealed sustained improvements in home language Oral Reading Fluency (0.19 SDs) and written comprehension (0.16 SDs). These gains translated into improved English written comprehension (0.16 SDs), which were not observed immediately after the program ended (Stern et al, 2024). Additionally, treated students showed improved grade progression, being approximately 9 percentage points more likely to reach Grade 7.

3 DATA

Analysis relies on both administrative data and primary data collection, scheduled for September 2025.

3.1 ADMINISTRATIVE DATA

Administrative data provides information on the school and grade each child was enrolled in for every year since 2016, as well as their performance in the end-of-year exam. (These exams are not standardized across schools.) This allows us to measure dropouts and grade repetition over time. We will also get access to performance in the national, standardized, end-of-high-school exam. The EGRS sample students will write these exams in 2026, if they did not repeat a grade. The National Senior Certificate (NSC) has four categories of passes, which provides access to different types of tertiary institutions (see appendix). All this data is available for *all* children who were in the sampled schools in 2016, not only the randomly sampled 20 students.

3.2 PRIMARY DATA COLLECTION

In addition to the administrative data, we are performing follow-up data collection September 1st to mid-November 2025—when non-repeating students will be in Grade 11—tracking the panel of roughly

2,600 students assessed at the beginning of grade one in 2015. During these school visits we will assess and survey students, and survey their language teachers and the school principal. Table 1 provides a breakdown of all the constructs that will be covered in this data collection.

Table 1. Constructs Measured in Grade 11 Data Collection

Construct	Indicators	Instrument	Source
English Literacy	<ul style="list-style-type: none"> • Oral Reading Fluency (ORF) • Oral Comprehension • Written Comprehension • Vocabulary 	Student oral and written assessment	Progress in International Reading Literacy Study/Early Grade Reading Assessment (EGRA)
Home Language Literacy (Setswana)	<ul style="list-style-type: none"> • Oral Reading Fluency (ORF) • Oral Comprehension • Written Comprehension • Vocabulary 	Student oral and written assessment	EGRA
Science Performance	<ul style="list-style-type: none"> • 8 questions 	Student written assessment	Trends in International Mathematics and Science Study (TIMSS)
Math Literacy	<ul style="list-style-type: none"> • 11 questions 	Student written assessment	Adapted from TIMSS and the National Senior Certificate (NSC)
Student Engagement	<ul style="list-style-type: none"> • Teacher: 7 Likert scale questions • Student: 24 Likert scale questions 	Teacher survey and student written questionnaire	Adapted from the Lessons in Character Program
Self-Esteem	<ul style="list-style-type: none"> • 24 Likert scale questions 	Student written questionnaire	Self-Esteem Questionnaire from DuBois et al (1996), MacArthur Ladder
Reading Enjoyment	<ul style="list-style-type: none"> • 24 Likert scale questions 	Student written questionnaire	Adapted from the Programme for International Student Assessment
Aspirations	<ul style="list-style-type: none"> • 8 questions 	Student written questionnaire	Adapted from Young Lives, and TIMSS

Construct	Indicators	Instrument	Source
Socio-Economic Status	• 4 questions	Student oral interview	Early Grade Reading Study Wave IV
Household Composition	• 15 questions	Student oral interview	Adapted from TIMSS and the Cape Area Panel Study
Home Learning Environment	• 9 questions	Student oral interview	The Southern and Eastern Africa Consortium for Monitoring Educational Quality, Early Grade Reading Study prior waves, TIMSS
School Quality	• Principal: 6 questions • Student: 1 Likert scale question	Principal survey, student written assessment, and school context survey	Early Grade Reading Study Wave IV
Life Satisfaction	• 1 question	Student written questionnaire	Satisfaction with Life Scale

3.3 TRACKING STRATEGY

To track the students assessed in earlier rounds of data collection, we merged our original sample with Department of Basic Education administrative records. This enables us to identify the school and grade that they are currently enrolled in, or were enrolled in before they dropped out. We were able to locate **87.6% of the original sample (2,556 learners in total)** in the administrative data. Out of this sample, 2,110 are located in 307 secondary schools and 13 primary schools in North-West or Gauteng provinces. We plan to visit all 307 secondary schools and a randomly selected subset of four primary schools.

We are not surveying all students still in primary school because: (i) they cannot complete high-school level assessments in math and science; (ii) we have ethical concerns about potentially traumatizing students by highlighting their academic difficulties; and (iii) excluding 9 students will not materially affect results, particularly since we can weight the 4 surveyed students to represent all 13 in relevant analyses.

4 OUTCOMES

4.1 PRIMARY OUTCOMES

- 1) **Obtained NSC Bachelor's pass or higher by 2027.** This outcome captures whether a student achieved the minimum qualification required to apply to any university degree program in South Africa. This is the most policy-relevant outcome as it determines access to tertiary education, which has substantial labor market returns. This outcome implicitly captures both reaching Grade 12 and passing the exam at the requisite level.¹
- 2) **Setswana literacy index.** Index constructed using the aggregation method proposed by Anderson (2008). The constituent indicators are: Oral Reading Fluency (ORF), Oral Comprehension, and Written Comprehension. Each component will be standardized to mean 0, SD 1 using the control group distribution. The index is the weighted average of these standardized components, weighted by the inverse of the variance-covariance matrix. For missing values, we will impute the control group mean, unless the observation is missing for all indicators (in which case the index will be coded as missing).
- 3) **Overall academic performance.** Index constructed using the aggregation method proposed by Anderson (2008). The three constituent indicators are: English written comprehension, mathematical literacy, and science literacy

No multiple testing correction will be applied to these two primary outcomes. We have clear a priori hypotheses for both, they measure fundamentally different constructs (terminal qualification vs. current literacy skills), and limiting analysis to three primary outcomes is already conservative.

4.2 SECONDARY OUTCOMES

Family A: Educational Attainment and Performance

1. Reached Grade 12 by 2027
2. Passed NSC (any level) conditional on reaching Grade 12
3. Type of pass obtained (ordered categorical: no pass, Higher Certificate pass, Diploma pass, Bachelor's pass)
4. Years of schooling completed

Family B: Learning by Subject

1. English written comprehension (proportion of questions answered correctly)
2. Mathematical literacy (proportion of questions answered correctly)

¹ For an earlier draft of the paper, this outcome will be replaced by grade 11 attainment by 2025.

3. Science literacy (proportion of questions answered correctly)

Family C: Potential Mechanisms

These outcomes are measured contemporaneously in Year 11 and may help explain the persistence (or lack thereof) of treatment effects:

1. Student engagement
2. Aspirations
3. Self-esteem
4. Love of reading

Indices will be constructed by taking the mean across the constituent indicators, using the same standardization and missing data procedures as for the primary literacy index.

Family D: Other Outcomes

1. General life satisfaction
2. Parental investment in a child's education

Visualization of Treatment Effects Across Educational Pipeline

We will also present a graphical representation of treatment effects across the educational pipeline. The figure will show the proportion of students in treatment and control groups reaching successive educational milestones: Grade 7, Grade 10, Grade 11, Grade 12, NSC pass (any level), NSC Diploma pass or higher, and NSC Bachelor's pass or higher. This visualization will illustrate at which points in the educational trajectory treatment effects emerge, persist, or fade. The figure will include confidence bands and will show both the absolute proportions and the treatment-control gaps at each stage.

4.3 MULTIPLE TESTING CORRECTION:

For secondary outcomes, we will apply the False Discovery Rate (FDR) correction within each family using the Benjamini-Hochberg procedure with $\alpha = 0.10$. We will report both uncorrected p-values and FDR-adjusted q-values.

5 EMPIRICAL ANALYSIS

5.1 MAIN ESTIMATING EQUATIONS

Equation (1): Analysis Using Primary Data Collection Sample

With the original EGRS cohort of students who we directly survey, we estimate the following equation:

$$y_{isb} = \beta_0 + \beta_1 T_{i,s} + X'_{isb} \Gamma + \rho_b + \varepsilon_{isb} \quad (1)$$

Where y_{isb1} is the outcome indicator of interest for learner i in strata b who was enrolled in primary school s at baseline, T is the treatment dummy; ρ_b refers to strata fixed effects, X'_{isb0} is a vector of controls measured at baseline, and ε_{isb1} is the error term clustered at the primary school level, which is the level of randomization.

Our main specification will include a parsimonious set of controls: strata fixed effects, student gender, and a baseline composite literacy score. In addition, we will use the post-double selection method proposed by Belloni et al. (2014) to select additional control variables that might be predictive of the dependent variable or treatment indicator. These potential controls include: pupils' parents' education, district dummy, performance in the most recent Annual National Assessments (ANA), a community-level wealth index, and average secondary school attendance rate in the community surrounding the school.

This specification will be used for all outcomes measured in the primary data collection (Primary Outcome 2 and all Secondary Outcomes in Families B, C, and D).

Equation (2): Analysis Using Full Administrative Data Sample

When analyzing Primary Outcome 1 (Bachelor's pass or higher) and Secondary Outcomes in Family A (educational attainment), we will use administrative data covering all students originally enrolled in treatment and control schools, not just the 20 students per school who were individually tracked and assessed. For this broader sample, we do not have individual-level baseline covariates, so we estimate:

$$y_{isb} = \beta_0 + \beta_1 T_{i,s} + \beta_2 y_{s,0} + \rho_b + \sigma_{Sec} + \varepsilon_{isb} \quad (2)$$

Where $y_{s,0}$ is the school-level average home language literacy at baseline, calculated from the 20 sampled students in each school who were assessed in 2015.

5.2 HETEROGENEOUS TREATMENT EFFECTS

We will test whether treatment effects vary by student and school characteristics using the following interaction model:

$$y_{isb} = \beta_0 + \beta_1 T_{i,s} + \beta_2 W_{i,s} + \beta_3 (W \times T)_{i,s} + X'_{isb} \Gamma + \rho_b + \varepsilon_{isb} \quad (3)$$

Where $W_{i,s}$ is the mediating factor of interest β_3 captures the interaction effect.

Moderating variables to be examined:

1. **School quality**, proxied by the average end-of-high school exam performance of the secondary school the student attends, averaged across the years 2021 to 2025. While this variable is measured post-treatment, it is unlikely to have been affected by the intervention given that treated students only reach Grade 12 in 2026. The interpretation is: do students who attend higher-quality secondary schools maintain treatment effects better? This tests whether the quality of later schooling environment moderates the persistence of early literacy gains.

2. Student gender.

Interpretation and limitations: The coefficient β_3 captures heterogeneity in long-term treatment effects but cannot distinguish between: (i) heterogeneous short-term treatment effects in grades 1-3 that persisted over time, or (ii) homogeneous short-term effects but heterogeneous persistence over time. Without intermediate measurements of the moderators or mechanisms, we cannot cleanly separate these interpretations. For school quality in particular, the interpretation is clearest: since secondary school quality only becomes relevant after primary school, this interaction more plausibly captures differential persistence rather than heterogeneous initial effects.

5.3 ATTRITION ANALYSIS

We will conduct the following analyses to assess whether attrition is balanced and random:

1. **Overall balance in attrition rates:** We will test whether the probability of attriting from the sample is balanced across treatment and control groups by estimating:

$$Attrite_i = \alpha_0 + \alpha_1 T_{i,s} + \rho_b + v_{is}$$

where $Attrite_i$ is an indicator for whether student i was not successfully tracked, and v_i is the error term clustered at the primary school level.

2. **Differential selection into attrition:** We will test whether students who attrite differ systematically between treatment and control groups by estimating:

$$y_{isb} = \gamma_0 + \gamma_1 T_{i,s} + \gamma_2 Attrite_i + \gamma_3 (T \times Attrite)_{i,s} + \rho_b + \varepsilon_{isb}$$

where y_i represents baseline characteristics (baseline academic performance index, gender, and age). The coefficient of interest is γ_3 , which captures differential selection into attrition between treatment and control groups.

We will interpret the threats of attrition using standards developed by the What Works Clearinghouse (WWC, 2017), established by the U.S. Department of Education's Institute of Education Sciences, classify evaluation studies as having either "high" or "low" attrition based on a combination of overall and differential attrition. The WWC standards also account for an important trade-off between overall and differential attrition—namely, that a study can have a higher overall rate of attrition if it has a low rate of differential attrition.

3. **Balance among non-attriters:** Using the specification above, γ_1 tests whether treatment and control groups remain balanced among the non-attrited sample.
4. **Lee bounds:** We will calculate Lee (2009) bounds in both directions (upper and lower) to assess the sensitivity of our main treatment effect estimates to non-random attrition. These bounds provide a range of plausible treatment effects under worst-case assumptions about the direction of selection.

6 COST-BENEFIT ANALYSIS

A central objective of this study is to assess whether investments in foundational literacy interventions generate returns that justify their costs. We will conduct a cost-benefit analysis that combines the observed treatment effects on educational attainment with estimates of labor market returns to education, similar to approaches taken by Chetty et al (2011) and Heckman (2010).

Cost and benefit estimates will incorporate the following: (a) direct program costs per student, (b) costs of schooling, including both government and parental investments; (c) costs of attending university; and (d) impact of treatment on number of years of schooling.

For expected future earnings, we will apply earnings returns from existing literature on South African labor markets (Patrinos & Psacharopoulos, 2018). Specifically, we will use Mincerian returns to: (a) grade attainment; (b) completing secondary school; (c) completing tertiary education. And we will use data on university drop-out rates to estimate the probability that a student who achieves a bachelor's pass actually receives a degree to estimate the impact on the expected probability of completing a university degree.

Using these estimates, we will calculate the per student Net Present Value of both costs and benefits, using a social discount rate of between 3 and 7 percent, and assuming that individuals will work until the age of 65. We will also present both private returns (to the individual) and social returns (including tax revenue)

We will clearly acknowledge several limitations of this approach:

- **Partial equilibrium:** We cannot account for general equilibrium effects if the intervention were scaled.
- **Unobserved benefits:** Our estimates exclude potential non-labor market benefits (health, civic participation, intergenerational effects)
- **Attribution:** Current earnings differences by education level may partly reflect selection rather than pure causal effects of education

7 REFERENCES

- Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, 103(484), 1481-1495.
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *The Review of Economic Studies*, 81(2), 608-650.
- Chetty, Raj, John N. Friedman, Nathaniel Hilger, Emmanuel Saez, Diane Whitmore Schanzenbach, and Danny Yagan. "How does your kindergarten classroom affect your earnings? Evidence from Project STAR." *The Quarterly journal of economics* 126, no. 4 (2011): 1593-1660
- Cilliers, J., Fleisch, B., Prinsloo, C., & Taylor, S. (2020). How to improve teaching practice? An experimental comparison of centralized training and in-classroom coaching. *Journal of Human Resources*, 55(3), 926-962.
- Heckman, James J., Seong Hyeok Moon, Rodrigo Pinto, Peter A. Savelyev, and Adam Yavitz. "The rate of return to the HighScope Perry Preschool Program." *Journal of public Economics* 94, no. 1-2 (2010): 114-128
- Lee, D. S. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *The Review of Economic Studies*, 76(3), 1071-1102.
- Stern, Jonathan MB, Matthew CH Jukes, Jacobus Cilliers, Brahm Fleisch, Stephen Taylor, and Nompumelelo Mohohlwane. "Persistence and emergence of literacy skills: Long-term impacts of an effective early grade reading intervention in south africa." *Journal of Research on Educational Effectiveness* (2024): 1-22

8 APPENDIX

Table 2. Description of National Senior Certificate (NSC) Pass Categories

Type of Pass	Type of Institution	Requirements
NSC Pass	-	• 40% minimum for home language • 40% for another 2 subjects
Higher Certificate Pass	Specific college and technical institution-based higher certificate courses and diploma programs	• 40% in Home Language • At least 40% in two other subjects • A minimum of 30% for three other subjects

Type of Pass	Type of Institution	Requirements
Diploma Pass	Universities of technology, private colleges, and TVET institutions	<ul style="list-style-type: none"> • 40% in Home Language
 • 40% for three other High Credit subjects (other than home language)
 • 30% for two other subjects
Bachelor's Pass	Minimum requirement to apply for any university degree program	<ul style="list-style-type: none"> • 40% for Home Language
 • A minimum of 50% for four other High Credit subjects
 • At least 30% for two other subjects