JOLLY PHONICS AND JOLLY GRAMMAR
Impact Evaluation in Namibia

2025 Report

Elzeth Zastrau[1], Jessica Mary-Ann Jansen[2], Renate Olga Shikwaya[3]

Genesis Analytics, South Africa[1], University of Namibia (UNAM)[2&3], School of Education, Early Childhood Education and Care, Khomasdal Campus

Analysis Report

July 2025

G:ENESIS UNLOCKING VALUE

UNAM UNIVERSITY OF NAMIBIA

ACKNOWLEDGEMENTS

The successful execution of this project would not have been possible without the generous support, guidance, and contributions of many individuals and organizations. We are immensely grateful to the Gates Foundation for their generous funding and continued support of this work. We would like to thank Izzy Boggild-Jones from the Gates Foundation for her valuable guidance and engagement throughout the study and guiding us to gain permission to access the data and have it published on the UNAM website. We thank the Roger Federer Foundation for their support in availing the tablets for the fieldwork. We acknowledge the Genesis Analytics team for their rigorous planning, data analysis and thoughtful inputs under the leadership of Elzeth Zastrau. Our sincere appreciation goes to Professor Brahm Fleisch for his expert guidance and insights.

We extend our sincere gratitude to the Minister, Hon. Dr. Sanet Steenkamp, and the Ministry of Education, Arts and Culture (MoEAC), now the Ministry of Education, Innovation, Youth, Sports, Arts & Culture, for their leadership and support throughout the study. We are deeply grateful to Ms. Ronel Bosch, Ms. Wenthworth, Mr. Eiman, and the EMIS team for their invaluable assistance and collaboration. We would like to acknowledge the support of Senior Education Officers, Dr. Paulina Hamukonda, Mr. Heinrich Mathews, Mr. Naftal Gabriel and Ms. Laimi Haikali - for their coordination and contributions throughout the study. We are also immensely grateful to the fieldwork teams, principals, HOD's, teachers and learners in Oshana and Oshikoto region for their participation, engagement, and warm welcome, without which this study would not have been possible.

We acknowledge the University of Namibia, the logistical and administrative coordinators, Dr. Pamela February and Dr. Beausetha Bruwer and the fieldwork team leaders, Ms Jacolynn Van Wyk, Ms Serena Hoeses, Ms Helga Kaurora, Ms Ivy Heyman Van Wyk, Mr Andre Kandundu and Mr Gustav Gustav.

Elzeth Zastrau

elzethz@genesis-analytics.com

Dr Jessica Mary-Ann Jansen

e-mail: jansenj@unam.na

Dr Renate Olga Shikwaya

e-mail: rshikwaya@unam.na

Table of Contents

Tables / Figures / Boxes

Abbreviations

Abbrev.	Meaning
C	Comparison
cwpm	Correct words per minute
EFL	English as a Foreign Language
EGRA	Early Grade Reading Assessments
ESL	English as a Second Language
HoDs	Heads of Departments
ITT	Intent-to-Treat
NN	Nearest Neighbour
NSAT	National Standardised Achievement Tests
OLS	Ordinary Least Squares
ORF	Oral Reading Fluency
PCA	Principal Component Analysis
PSM	Propensity Score Matching
SACMEQ	Southern and Eastern Africa Consortium for Monitoring Educational Quality
SEO	Senior Education Official
T	Treatment
TOT	Treatment-on-the-Treated
UAE	United Arab Emirates
UNAM	University of Namibia

1 | Executive Summary

This report presents the findings of an independent impact evaluation of the Jolly Phonics and Grammar programmes in Namibia, commissioned to assess the programme's effectiveness in improving foundational English reading skills among early-grade learners. The evaluation focused on Grade 3 learners from the 2024 academic year, sampled from 100 schools equally divided between the Oshana region (which had been trained on Jolly Grammar 2 for Grade 3s) and the Oshikoto region (which had not received Jolly Grammar 2 training at the time of the survey). The study employed a Propensity Score Matching (PSM) approach to estimate average treatment effects on Early Grade Reading Assessment (EGRA) performance. Four primary models were employed, varying in their matching specifications to test the robustness of results.

Across all models, the most consistent and robust result was a positive and statistically significant effect of the Jolly programmes on learners' Word Reading. In Model 1, the estimated effect was 3.58 points ($p < 0.01$), and remained significant at 3.19 points in Model 2 ($p < 0.01$). Even under stricter matching in Model 4, the effect persisted at 2.20 points ($p < 0.05$), confirming the programme's targeted benefit on this outcome.

However, the programme did not demonstrate a significant impact on other key literacy domains (Letter Sounds, Phonemic Awareness, Oral Reading Fluency (ORF), or Reading Comprehension). In some instances, results for fluency were unexpectedly negative, although this result was inconsistent across matching models and disappeared in stricter models. Further sensitivity checks presented in the Annexe reinforce these findings. Only models demonstrating both improved covariate balance and strong propensity score overlap yielded consistent and significant impacts on Word Reading. Where positive effects were observed for other outcomes, they were often accompanied by poor balance, weakening their credibility.

These findings may suggest that while the Jolly Grammar programme effectively strengthens decoding, it does not currently support more complex reading processes. Teacher reports echoed this trend, highlighting high levels of satisfaction with the training, but also pointing to gaps in support from seniors, limited instructional time, and challenges with deeper pedagogical concepts like grammar and comprehension strategies.

Several factors may explain the modest and selective impact. First, the study faced key design constraints: it was not randomised, and there were no baseline EGRA scores, limiting the ability to definitively attribute differences between regions to the Jolly programmes. Although statistical methods such as multivariate regression and PSM were applied to account for observable differences, unmeasured variables - such as teaching quality, learner motivation, or classroom implementation fidelity - could still bias estimates. Second, teacher training was often too brief to consolidate complex content, and ongoing classroom support appeared uneven across regions. Moreover, the availability of materials and mentoring differed substantially between schools, with some teachers reporting no access to key resources or follow-up guidance.

These limitations notwithstanding, the consistent positive effect on Word Reading provides encouraging evidence that the Jolly programme is capable of building foundational decoding skills, particularly in multilingual contexts where English is not the first language. However, the lack of impact on broader reading outcomes suggests that decoding instruction alone is insufficient. The findings indicate a need for a more integrated approach to literacy instruction - one that includes comprehension, fluency, and vocabulary-building strategies.

Based on these results, the report recommends a series of refinements alongside the full-scale expansion. These include extending the training duration to allow deeper engagement with the

curriculum; rebalancing training content to better support comprehension and fluency instruction; increasing support and coaching for teachers in the classroom; and integrating differentiated teaching strategies to support learners at varying ability levels. The evaluation also calls for improved monitoring and the incorporation of longitudinal study designs, including baseline assessments, to better track learner progress over time and attribute improvements to the Jolly programmes.

In conclusion, the Jolly Phonics and Grammar programmes show promise in enhancing early literacy in Namibia, particularly in strengthening word reading skills. However, realising its full potential will require a more comprehensive instructional approach, stronger implementation support, and more rigorous future evaluations. With these enhancements, Jolly Phonics and Grammar could play a meaningful role in addressing the country's foundational literacy challenges.

2 | Introduction

This data analysis report focuses on an independent impact evaluation conducted to assess the effectiveness of the Jolly Phonics and Grammar programme in enhancing early-grade English reading skills in Namibia. The study aimed to evaluate the impact of the programme by comparing reading outcomes between two groups of schools: those in the Oshana region, where Grade 2 and 3 teachers had been trained in Jolly Grammar 1 and 2, and comparison schools in the Oshikoto region where teachers had had no prior exposure to Jolly Grammar 1 or 2.

The analysis is based on data collected from 100 schools - 50 schools in the comparison group and 50 schools in the treatment group. In each school, one Grade 3 teacher was surveyed, and individual Early Grade Reading Assessments (EGRA) were administered to a sample of 20 learners. The evaluation was conducted in partnership with the University of Namibia (UNAM).

The analysis will offer evidence-based recommendations into the effectiveness, scalability, and potential areas for improvement in the implementation of the Jolly Phonics and Grammar programmes in Namibia.

2.1 Background

2.1.1 The Namibian Literacy Landscape: A Persistent Challenge

The state of early grade literacy in Namibia is not a recent concern but a deep-seated, systemic challenge that has been documented for over a decade. Evidence from independent, large-scale assessments points to a crisis in foundational reading skills, which has profound implications for the nation's educational trajectory and its long-term development goals as articulated in Vision 2030[1]. This persistent challenge frames the context for all current and future educational interventions.

Learner Performance: Insights from SACMEQ, NSAT and EGRA

Namibia's participation in the Southern and Eastern Africa Consortium for Monitoring Educational Quality (SACMEQ) has provided a crucial regional benchmark for its educational performance. The 2010 SACMEQ assessment of Grade 6 learners revealed that Namibia's reading proficiency score was 496.9, which was below the 15-country mean of 511.8[2]. More alarmingly, the same 2010 assessment found that 38.7% of Namibian Grade 6 students were unable to read for meaning in English, the official language and eventual medium of instruction[3].

These findings are echoed in Namibia's National Standardised Achievement Tests (NSAT), conducted in Grades 5 and 7. The 2015 NSAT showed that a staggering 87% of Grade 7 learners performed in the "below basic" and "basic" achievement categories for English[4]. More recent data from the 2018 NSAT report revealed that learners performed poorly on reading comprehension tasks, with achievement rates of only

[1] Office of the President. (2004). *Namibia Vision 2030: Policy framework for long-term national development (Main Document)*. Government of the Republic of Namibia. Available: https://www.npc.gov.na/wp-content/uploads/2021/11/vision_2030.pdf

[2] Southern and Eastern Africa Consortium for Monitoring Educational Quality. (2010). SACMEQ III Project results: Pupil achievement levels in reading and mathematics. Working document No. 1. Harare: SACMEQ. Available: https://nicspaull.com/wp-content/uploads/2011/04/wd01_sacmeq_iii_results_pupil_achievement.pdf

[3] Ibid.

[4] Shaakumeni, S. N., & Mupupa, S. P. (2019). National standardised achievement tests: A source of reliable formative feedback for teachers. *The Namibia CPD Journal for Educators*, 5, 131-161.

23% in the Oshikoto region and a national average of just 14%[5]. Together, SACMEQ and NSAT confirm a widespread and persistent failure to equip learners with essential reading skills by the end of primary school.

In response, and aligned with the goals of Vision 2030, Namibia piloted the EGRA in 2012. The pilot aimed to validate a national EGRA tool for wider use and was conducted in three languages - Khoekhoegowab, Oshindonga, and English as a Second Language (ESL). The EGRA pilot moved the conversation from a general concern about "poor reading" to a specific diagnosis of which foundational skills were weakest. It revealed widespread difficulties among Grade 1–3 learners in recognising letter names and sounds, reading fluently, and comprehending text [6]. The pilot further found that learners who had attended pre-primary school performed better than their peers who had not, while girls, on average, performed better than boys. Across all three regions, the pilot pointed to a clear and urgent need to improve literacy instruction with a specific focus on developing comprehension skills[7].

While large-scale assessments like SACMEQ and NSAT highlight Namibia's persistent early literacy challenges, there is limited publicly available EGRA data and results for Namibia. This constrains the ability to diagnose specific gaps in foundational reading skills. Using the EGRA for this impact evaluation of the Jolly programmes presents a timely opportunity to address these gaps in knowledge.

2.1.2 The Jolly Learning Methodology

The Jolly Phonics and Jolly Grammar programmes offer a comprehensive, structured approach to early literacy acquisition, grounded in the principles of systematic synthetic phonics[8] and aligned with the "Science of Reading"[9] [10]. Developed by Sue Lloyd and Sara Wernham and published by Jolly Learning Ltd., a leading UK educational publisher, these programmes are designed to build strong foundational English literacy skills from Pre-Primary through to the end of Year/Grade 6[11].

The Jolly Phonics programme serves as the initial programme, focusing on teaching children the 42 main letter sounds of English (phonemes) in isolation, along with their corresponding letters or letter combinations (graphemes). Children are then taught to blend these phonemes to read words (e.g., blending d-o-g to read "dog") and to segment spoken words into phonemes to support spelling[12]. This direct and explicit instruction contrasts with analytic phonics, which typically begins with a whole word and encourages children to analyse its component parts[13].

---

[5] Alumbungu, M. N., & Mpofu, N. (2025). Examining the reading comprehension pedagogical practices developed by ESL teachers in Namibian primary schools. Language and Education, 39(1), 1-20.
[6] Ibid.
[7] Haifidi, E. (2019, May 19–22). An evaluation of Early Grade Reading Assessment in Namibia [Conference presentation]. SAAEA Conference, Gaborone, Botswana. Directorate of National Examinations and Assessment, Ministry of Education, Arts and Culture, Windhoek, Namibia. Available: https://slideplayer.com/slide/17338120/
[8] Jolly Learning (2025). *What is Jolly Phonics?* Available: https://jollylearning.com/#:~:text=Jolly%20Phonics%20is%20a%20programme,using%20the%205%20key%20skills.
[9] This term refers to a consensus from decades of multidisciplinary research in cognitive psychology, linguistics, and neuroscience on how humans learn to read.
[10] Jolly Learning. (2025). *A Comprehensive Programme.* Available: https://jollylearning.com/jolly-phonics
[11] Summers Issa, S. (2022, February). *Namibia: National Jolly Phonics rollout – Executive summary report.* Available: https://www.jollyfutures.org/cache/docs/attWBwcmzsDc0yqM3.pdf
[12] Jolly Learning. (2025). *What is Jolly Phonics?* Available: https://jollylearning.com/jolly-phonics
[13] Johnston, R. S., McGeown, S., & Watson, J. E. (2012). Long-term effects of synthetic versus analytic phonics teaching on the reading and spelling ability of 10 year old boys and girls. *Reading and writing*, 25(6), 1365-1384.

Jolly Grammar builds on the Jolly Phonics foundation by continuing systematic literacy instruction through primary school. It introduces key grammar, punctuation, and spelling concepts in a structured, multi-sensory way - often incorporating colour-coding and actions to represent parts of speech[14]. The programme is structured with two lessons per week covering grammar, punctuation, and spelling. The "grammar" component broadly encompasses parts of speech, language structure, punctuation, and word meaning, with definitions evolving throughout the programme. Spelling lessons cover a wide range of concepts and patterns, providing children with rules to apply in their spelling, alongside weekly word lists and dictation exercises.

Taken together, the Jolly Phonics and Jolly Grammar programmes reflect the Science of Reading's emphasis on the importance of explicit, systematic teaching of phonemic awareness and phonics. The programmes' tight integration of method, sequence, and reinforcement distinguishes them from less structured approaches and positions them as effective tools for improving foundational literacy outcomes.

2.1.3 Evidence of Efficacy: A Review of Key Impact Evaluations

The Jolly Phonics programme - which has been adopted in over 150 countries[15] - is supported by a body of research ranging from small-scale classroom interventions to large, multi-year government initiatives.

The Clackmannanshire Study: A Longitudinal Analysis of Synthetic Phonics

Among the most frequently cited pieces of evidence is the longitudinal research conducted by Dr. Rhona Johnston and Joyce Watson in Clackmannanshire, Scotland[16]. This seminal study, initiated in the late 1990s, was designed to compare the effectiveness of a systematic synthetic phonics programme, which heavily utilised Jolly Phonics principles and materials, against two other methods: a traditional analytic phonics programme and an analytic phonics programme supplemented with phonemic awareness training.

The initial phase of the research involved approximately 300 Primary 1 children who received 20 minutes of daily phonics instruction over 16 weeks. The results were striking. The group taught with synthetic phonics demonstrated significantly superior performance in both reading and spelling. On average, they were reading and spelling at a level seven months ahead of their chronological age and approximately seven months ahead of the children in the two analytic phonics groups[17].

However, the most powerful findings emerged from the seven-year follow-up study, which tracked the children's progress to the end of their primary schooling[18]. This longitudinal data revealed that the initial advantage conferred by synthetic phonics was not only sustained but dramatically amplified over time. By the end of Primary 7, the children from the original synthetic phonics cohort were, on average, 3 years and 6 months ahead of their chronological age in word reading. Their spelling skills were 1 year and 9 months ahead of their chronological age, and their reading comprehension was 3.5 months ahead.

---

[14] Jolly Futures. (2022, February). *Namibia: National Jolly Phonics rollout – Executive summary report.* Available: https://www.jollyfutures.org/cache/docs/attWBwcmzsDc0yqM3.pdf

[15] Jolly Learning. (2025). *Jolly Phonics: The world's leading phonics programme*. Available: https://jollylearning.com/jolly-phonics

[16] Johnston, R. S., & Watson, J. E. (2004). Accelerating the development of reading, spelling and phonemic awareness skills in initial readers. Reading and Writing, 17(4), 327-357. Available: https://link.springer.com/article/10.1023/B:READ.0000032666.66359.62

[17] Johnston, R. S., & Watson, J. E. (n.d.). Putting phonics teaching to the test. Available: https://jolly2.s3.amazonaws.com/Case%20Studies/Clackmannanshire%20Case%20Study%20D0.pdf

[18] Johnston, R. S., & Watson, J. E. (2005). The effects of synthetic phonics teaching on reading and spelling attainment: A seven year longitudinal study (Vol. 11). Edinburgh: Scottish Executive.

The West Dunbartonshire Literacy Initiative: A Large-Scale Application

Following the promising results from Clackmannanshire, the West Dunbartonshire Literacy Initiative provided a real-world test of synthetic phonics at an unprecedented scale[19] [20]. Led by psychologist Dr. Tommy MacKay, this ambitious 10-year project (1997-2007) was implemented across an entire local education authority. The multiple-component intervention involved 58 of the authority's nurseries and primary schools, ultimately assessing nearly 70,000 children.

To validate the phonics component, a quasi-experimental supporting study was conducted, directly comparing a synthetic phonics programme with a traditional analytic phonics programme in 18 primary schools. This sub-study found significant gains for the synthetic phonics group on measures of both word and non-word reading, confirming the findings from Clackmannanshire in a different context.

By the conclusion of the research phase, the initiative claimed to have effectively eliminated functional illiteracy. The success of this large-scale, systemic application of principles championed by Jolly Phonics provided powerful evidence that the results observed in controlled research settings could be replicated across an entire, and notably disadvantaged, educational system with sufficient institutional commitment.

International Case Studies: Performance in Diverse Educational Contexts

The evidence base for Jolly Phonics extends well beyond the UK, with numerous impact evaluations conducted in a variety of international and linguistic settings[21]. Many of these studies focus on its effectiveness for children learning English as a Second Language (ESL) or English as a Foreign Language (EFL), a context of critical importance for a programme with global reach.

- **Pakistan:** A significant pilot study was conducted in Lahore in partnership with Idara-e-Taleem-o-Aagahi and involved 209 students in 9 low-cost schools[22]. Using an experimental design with a control group, the study found that students taught with the Jolly Phonics course achieved a significantly higher reading age and were able to read more complex texts compared to the control group, which was taught using a traditional whole-word approach.
- **Nigeria:** Multiple studies in Nigeria have reported improved literacy outcomes as a result of the Jolly Phonics programme. A study in Anambra State involving 100 primary two pupils found that the experimental group using Jolly Phonics "had a better performance on the reading and spelling tests" compared to the control group[23]. Similar studies in Akwa Ibom State, with 168

[19] MacKay, T. (2006). The West Dunbartonshire Literacy Initiative: The Design, Implementation and Evaluation of an Intervention Strategy to Raise Achievement and Eradicate Illiteracy. Phase I Research Report. Dumbarton: West Dunbartonshire Council. ISBN: 0-906938-12-0

[20] MacKay, T. (2007). Achieving the vision: The final research report of the West Dunbartonshire Literacy Initiative. West Dunbartonshire Council Educational Services. ISBN: 978-0-9537736-6-4. Available: https://www.researchgate.net/publication/268386849_Achieving_the_Vision_The_Final_Research_Report_of_the_West_Dunbartonshire_Literacy_Initiative

[21] Please find academic studies on the Jolly Phonics programme here: https://jollylearning.com/research/research-1

[22] Zaidi, S., Naeem, H., & Naheed, S. (n.d.). *Early teaching of English through synthetic phonics approach using Jolly Phonics program.* Research Society of Synthetic Phonics, in collaboration with Idara-e-Taleem-o-Aagahi, Jolly Learning UK, & Phonics Club Pvt Ltd. Available: https://jolly2.s3.amazonaws.com/Research/Ita%20Lahore.pdf

[23] Umezinwa, R. N., & Udogu, C. O. (2018). Effect of Jolly Phonics on primary school pupils' achievement on reading and spelling skills in Anambra state. *European Journal of Research and Reflection in Educational Sciences.* Vol, 6(6). Available: https://www.idpublications.org/wp-content/uploads/2018/11/Paper-Effect-of-Jolly-Phonics-on-Primary-School-Pupils.pdf

primary one pupils, and in Benue State, with 62 pupils, concluded that the programme led to an improvement in pupils' reading abilities [24] [25].

- **United Arab Emirates (UAE):** A comparative study investigated the impact of integrating the Jolly Phonics Lessons mobile app into literacy lessons for kindergarten students (ages 4-6)[26]. The results showed that the group using the Jolly Phonics app outperformed pupils receiving literacy instruction through traditional methods, reporting "dramatic growth in letter-sound recognition and letter formation" and a "substantial increase in phoneme blending and segmenting abilities" for the Jolly Phonics group.

The breadth of these international studies demonstrates that the programme's effectiveness is not confined to a single cultural or linguistic context. The consistent positive results across countries as diverse as Pakistan, Nigeria, and the UAE, and specifically with ESL learners, provide a strong argument for the programme's applicability in Namibia.

2.1.4 Jolly Phonics and Grammar in Namibia: Implementation Timeline

Since 2017, Namibia has demonstrated a strong commitment to integrating Jolly Phonics and Grammar into its national education system.

Initial Pilot and Adoption of Jolly Phonics in Oshana (2017-2019)

In June 2017, Jolly Phonics was piloted in six primary schools in the Oshana Region to enhance learners' literacy skills in the Junior Primary phase[27]. An evaluation of the Oshana pilot was conducted in November 2017 by Junior Primary Senior Education Officers from the Oshana Regional Council. On average, 24 learners were tested at each piloting school, comprising 12 learners from the experiment group (classrooms in which Jolly Phonics was piloted) and 12 learners from the control group (classrooms where Jolly Phonics was not piloted). The evaluation report indicated that learners in the "experiment group" demonstrated improved reading abilities across various components compared to the "control group". Following the pilot, **Jolly Phonics was adopted and rolled out in 2019 to all Grade 1 classes in the Oshana Region[28].**

Progression to Jolly Grammar in Oshana (2021)

In November 2021, the Directorate of Education, Arts and Culture introduced Jolly Grammar 1 and 2 in Oshana to Grade 2 and 3 teachers, respectively[29]. This aimed to ensure progression and continuity

---

[24] Ekpo, C. M., Udosen, A. E., Afangideh, M. E., Ekukinam, T. U., & Ikorok, M. M. (2007). Jolly phonics strategy and the ESL pupils' reading development: A preliminary study. In *1st Mid Term Conference University of Ibadan*. Available: https://jolly2.s3.amazonaws.com/Research/Nigeria%20-%20University%20Uyo%20Research.pdf

[25] Udu, T. T. (2025). EFFECTS OF JOLLY PHONICS ON BASIC II PUPILS' INTEREST AND ACHIEVEMENT IN READING COMPREHENSION IN PUBLIC SCHOOLS IN MAKURDI, BENUE STATE. Nigerian Journal of Literacy and English Education, 2(3), 43-54. Available: https://www.nijolee.net.ng/index.php/home/article/view/80

[26] Almansoori, N., Ogdol, R., & Alteneiji, A. (2024). The impact of integrating Jolly Phonics Lessons application into English literacy lessons on UAE preschoolers' phonics skills. *Journal of Childhood, Education & Society*, 5(1), 41-60. Available: https://core.ac.uk/reader/604053830

[27] Oshana Regional Council, Directorate of Education, Arts and Culture. (2017). *Jolly Phonics evaluation report: 28 November–01 December 2017, Oshana Region*. Republic of Namibia. Available: https://www.jollyfutures.org/cache/docs/uncategorized/Osana_Jolly_Evaluation_report-November_2017_revised_23_January_2018.pdf

[28] Summers Issa, S., & Moyo, S. (2024, March). *The national rollout of Jolly Grammar 1 in Namibia: February 2024*. Jolly Futures CIC. Edited by E. Haifidi, Ministry of Education, Arts and Culture (MoEAC). Available: https://www.jollyfutures.org/cache/docs/attWBwcmzsDc0yqM3.pdf

[29] Summers Issa, S. (2021, November). Jolly Grammar Training: Oshana Region, Namibia November 2021.

from Jolly Phonics. 117 Grade 2 teachers, 130 Grade 3 teachers, and 32 officials received training. The training consisted of a 2-day course covering progression from Jolly Phonics to Jolly Grammar, the key principles of Jolly Phonics, and various aspects of spelling and grammar lessons from Jolly Grammar 1 and 2. Teachers were also instructed on how to use the Pupil and Teacher Books and big books.

While the 2021 training was well-received, with high participation and positive feedback, several challenges were noted in the training report. A significant challenge was that many teachers lacked a foundational understanding of Jolly Phonics, necessitating more time on basic phonics principles during the Jolly Grammar training. This also meant some activities were explained rather than practiced. Technical issues with speakers, the English syllabus not being provided in advance, and the Jolly Grammar materials being purchased in precursive script were also noted as issues.

National Rollout of Jolly Phonics (2022)

At the start of 2022, Jolly Phonics was officially adopted and rolled out for Grade 1 in all government primary schools in Namibia[30]. To launch the rollout, the Ministry of Education planned and provided a 5-day training on Jolly Phonics, Jolly Grammar (1 and 2), and the Jolly Monitor to 57 Senior Education Officials (SEOs). In addition, 2,665 Grade 1 teachers received 2.5 days of Jolly Phonics training across 20 locations in the 14 regions. Training was delivered in cohorts of 50 teachers. Over 13 million Namibian Dollars worth of Jolly Phonics materials were donated by Jolly Learning Ltd.

While the training had high attendance rates, enthusiastic participation and strong support from regional officials, there were delays in material arrival, errors in the quantity of materials ordered and low attendance of Heads of Departments (HoDs) at the training. The low attendance of HoDs likely impacted their ability to support teachers in implementing Jolly Phonics.

National Rollout of Jolly Grammar 1 (2024)

In early 2024, Jolly Grammar 1 was introduced nationally, with Grade 2 teachers receiving training in Jolly Phonics and Jolly Grammar 1 [31]. The aim was to focus on the progression from Jolly Phonics to Jolly Grammar 1 in Grade 2, and strengthen implementation through capacity building of HoDs. 2,742 Grade 2 teachers were trained in 2.5-day sessions, while 651 HoDs were trained in 2-day sessions. HoDs received 5 years' access to the Jolly Monitor App for monitoring, and 2,742 copies of the Jolly Grammar 1 Handbook (print edition, adapted for Namibia's handwriting policy) were procured for Grade 2 teachers.

National Rollout of Jolly Grammar 2 (2025)

In early 2025, Jolly Grammar 2 was rolled out nationally to Grade 3 teachers[32]. 2,678 Grade 3 teachers received 2.5 days of training in their respective regions, and 225 HoDs were trained in 2-day sessions at four centralised locations. While physical handbooks for Jolly Grammar 2 did not arrive in time for training, the Jolly Monitor App licence was procured for HoDs. During the training, concerns were raised about Grade 1 and 2 teachers not implementing Jolly Phonics and Grammar 1 effectively, impacting Grade 3 learners' readiness.

[30] Summers Issa, S. (2022, February). *Namibia: National Jolly Phonics rollout – Executive summary report*. Available: https://www.jollyfutures.org/cache/docs/attWBwcmzsDc0yqM3.pdf
[31] Summers Issa, S., & Moyo, S. (2024, March). *The national rollout of Jolly Grammar 1 in Namibia: February 2024*. Jolly Futures CIC. Edited by E. Haifidi, Ministry of Education, Arts and Culture (MoEAC). Available: https://www.jollyfutures.org/cache/docs/attWBwcmzsDc0yqM3.pdf
[32] Summers Issa, S., & Moyo, S. (2025, April). *The national rollout of Jolly Grammar 2 in Namibia: January-February 2025*. Jolly Futures.

Box 1: Implementation Modalities, Monitoring and Evaluation of the Jolly Programmes

Implementation Modalities:

- **Instructional Time:** While specific daily/weekly hours for Jolly Phonics and Grammar implementation are not explicitly defined, the structure of "two lessons a week" for Jolly Grammar indicates that the programmes' curriculum is integrated into the regular school timetable rather than being an extracurricular activity.
- **Teacher Practices:** Teachers are expected to actively use the provided Jolly Phonics and Grammar materials, despite not always having physical copies (personal communication with D. Mupopiwa, 17 July 2025). Following the 2021 training of Grade 3 teachers in Oshana, physical handbooks were not always available. Instead, the regional office distributed soft copies of the teacher's book and learner's book to principals via email, with the expectation that principals would share these with their teachers. There was no follow-up to verify if all teachers printed or effectively utilised these materials. In addition, the regional office advised teachers to have learners write activities in their exercise books rather than making copies from the handbook, due to the impracticality of making copies for all students annually.

Monitoring and Evaluation of Implementation:

The regional offices, particularly SEOs and HoDs, are primarily responsible for monitoring and evaluating the programme's implementation. However, according to D. Mupopiwa (personal communication, 17 July 2025) the regional office's monitoring and evaluation efforts are primarily reactive and based on self-reported data from schools. Schools submit "need analyses" and bimonthly reports on syllabus coverage, detailing any challenges they face. Based on self-reported challenges and analysis of academic results, the regional office develops programmes for school visits and training. To date, there has been limited monitoring or evaluation of the implementation of the Jolly programmes at the school level. In other words, the Jolly Phonics and Grammar programmes have not been consistently tracked to assess implementation fidelity or impact over time.

2.1.5 Leveraging Oshana's Early Implementation of Jolly Phonics and Grammar

The national rollout of Jolly Phonics and Jolly Grammar in Namibia has followed a staggered implementation schedule across regions, creating a valuable opportunity for evaluating the impact of the programmes on early grade English reading skills. In particular, the earlier and more sustained implementation in the Oshana region allows this region to function as a treatment group, enabling a comparison of learning outcomes with other regions where exposure to Jolly Grammar at Grade 2 and 3 level began only with the national rollout in 2024 and 2025.

Oshana was the first region to pilot Jolly Phonics in 2017 and rolled it out to all Grade 1 classrooms by 2019. In 2021, the region took an additional step by introducing Jolly Grammar training to Grade 2 and 3 teachers. Although this training was limited in scale - typically involving only one teacher per school[33] - and not as comprehensive as the later national implementation, it nonetheless provided early exposure to key Jolly Grammar concepts and materials. Regional officials supported dissemination by circulating soft copies of the training resources to school principals, enabling informal adoption beyond the initially trained cohort.

By contrast, schools in other regions did not receive any Jolly Grammar training or resources for Grade 2 or 3 prior to 2025. These schools can thus serve as a natural counterfactual, having had no formal exposure to Jolly Grammar at the Grade 2 and 3 level during the period under evaluation. This variation in timing provides a rare and timely opportunity to assess whether earlier exposure to the Jolly Phonics

[33] Personal communication with D. Mupopiwa, 17 July 2025

and Grammar programmes contributed to improved learner outcomes in Oshana - particularly as Namibia seeks to strengthen foundational literacy as part of its education reform agenda.

2.1.6 Evaluation Design

The evaluation capitalises on the earlier implementation of Jolly Phonics and Jolly Grammar in Oshana to estimate the programme's impact on English reading outcomes. Given the absence of baseline data, a matched comparison group design will be employed, focusing on the 2025 cohort of Grade 4 learners, who were in Grade 2 and 3 in 2023 and 2024, respectively.

In Oshana, this 2025 Grade 4 cohort would have been taught in Grade 2 and 3 by teachers who had likely received Jolly Grammar training in 2021 and materials since then. In contrast, the teachers of the 2025 Grade 4 cohort in other regions would not have received training in or materials for Jolly Grammar 1 or 2. In addition, the Oshana Grade 4 2025 cohort would have been taught by Jolly Phonics trained teachers in Grade 1, while the Grade 1 teachers teaching the same cohort in other regions would have only been trained in Jolly Phonics the same year that they taught the Grade 4 2025 cohort (2022).

By selecting matched regions, the evaluation design allows for meaningful comparison and measurement of the programme's impact. Oshana's neighbouring region, Oshikoto, was selected for comparison. The following shared characteristics informed this choice:

- **Geographic and Administrative Similarity:** Both regions are in north-central Namibia, share borders, and operate under similar educational structures, supporting consistent programme delivery.
- **Linguistic and Cultural Context:** Both are predominantly Oshiwambo-speaking, with learners typically acquiring English as a second or third language.
- **Educational profiles**, which include comparable indicators such as enrollment, infrastructure, teacher qualifications, and learner outcomes, provide a strong baseline for evaluation.
- **Implementation Timing:** Oshikoto only received Jolly Grammar 1 and 2 training in 2024 and 2025, with no opportunity for classroom application amongst the Grade 4 2025 cohort, making it a natural comparison group.

This design allows the evaluation to attempt to estimate the causal impact of early exposure to the Jolly Grammar programmes by comparing reading outcomes between learners in Oshana and those in matched schools in Oshikoto - a region which shares a similar context but experienced delayed implementation.

2.2 Purpose of the Evaluation

The purpose of the evaluation is to determine if the implementation of the Jolly programmes improves early-grade English reading in Namibia.

Specifically, the evaluation aims to:

- To **evaluate the effectiveness** of the Jolly Phonics and Grammar programmes in **enhancing early-grade English reading outcomes**.
- To **address concerns** about the **absence of independent evidence** validating the programme's impact.
- To **offer evidence-based recommendations** for potential expansion, modification, or redesign of the programme.

2.3 Report Outline

The report is structured into the following sections:

- **Sampling Methodology and Survey Design:** Outlines the overall research approach, including the sampling strategy, target population, and the design of the data collection instruments. It also details how the survey was implemented and monitored to ensure data quality and reliability.
- **Data and Analysis Methodology:** Specifies the final sample size of the dataset used for analysis and details the main outcome variables measured. It further outlines the various analytical approaches employed to assess and determine the impact of the programme, including any specific statistical models or comparative methods utilised.
- **Descriptive Statistics:** Presents an overview of key findings from the teacher and student surveys. This includes student demographics, background characteristics, and educational context, as well as insights into how classes are structured, planned, and delivered by teachers.
- **Impact Estimation:** Provides a comparative analysis of reading outcomes across treatment and comparison groups. This includes statistical tests to assess the effectiveness of the Jolly programmes.
- **Discussion:** Sums up the findings of the report, contextualising the impact estimation within the programme setting.
- **Recommendation on Effectiveness and Scalability:** Offers practical, evidence-based recommendations derived from the findings. These are intended to inform programme improvement, policy decisions, and potential scale-up strategies.
- **Annexes:** Contains supplementary materials, including the full survey instruments, additional tables, technical notes on the methodology, and any other relevant documentation to support transparency and replicability.

3 | Sampling Methodology and Survey Design

3.1 Sample Design and Implementation

3.1.1 Sample Design

The sample design aims to achieve a balance between the treatment and comparison groups, allowing for later matching while retaining randomisation to avoid introducing sampling bias. A stratified random sampling approach was therefore implemented, with crucial filtering steps to ensure balance. The sampling frame comprises schools in both the Oshana and Oshikoto regions, denoted as treatment and comparison for ease of analysis. Before sampling, any schools in the sampling frame with missing geographic coordinates, geotype or school grade information were dropped.

Based on prior literature, we decided to focus on balancing the sample on school size and distance to the regional office[34]. To achieve this, we constructed strata from these variables by categorising them into five categories through a means clustering algorithm. We randomly allocated the sample across strata so that the sample size per stratum in both treatment and comparison groups was as close as possible (within a one-unit difference, where the sample size in a stratum was insufficient). Our final sample balance checks on school size and distance to the regional office revealed that the sample was balanced on school size (average 45.32 for comparison versus 45.58 for treatment) and less balanced on distance (average of 25.7km for comparison versus 18.9km for treatment).

An important part of achieving balance involves filtering steps:

- Private schools were excluded from the sample;
- Schools more than 100km away from the regional office were excluded, as treatment schools generally were closer to the respective regional office than comparison schools;
- Schools with a total enrolment larger than 150, as comparison schools were generally smaller than treatment schools;
- One stratum where there were no treatment schools was excluded.

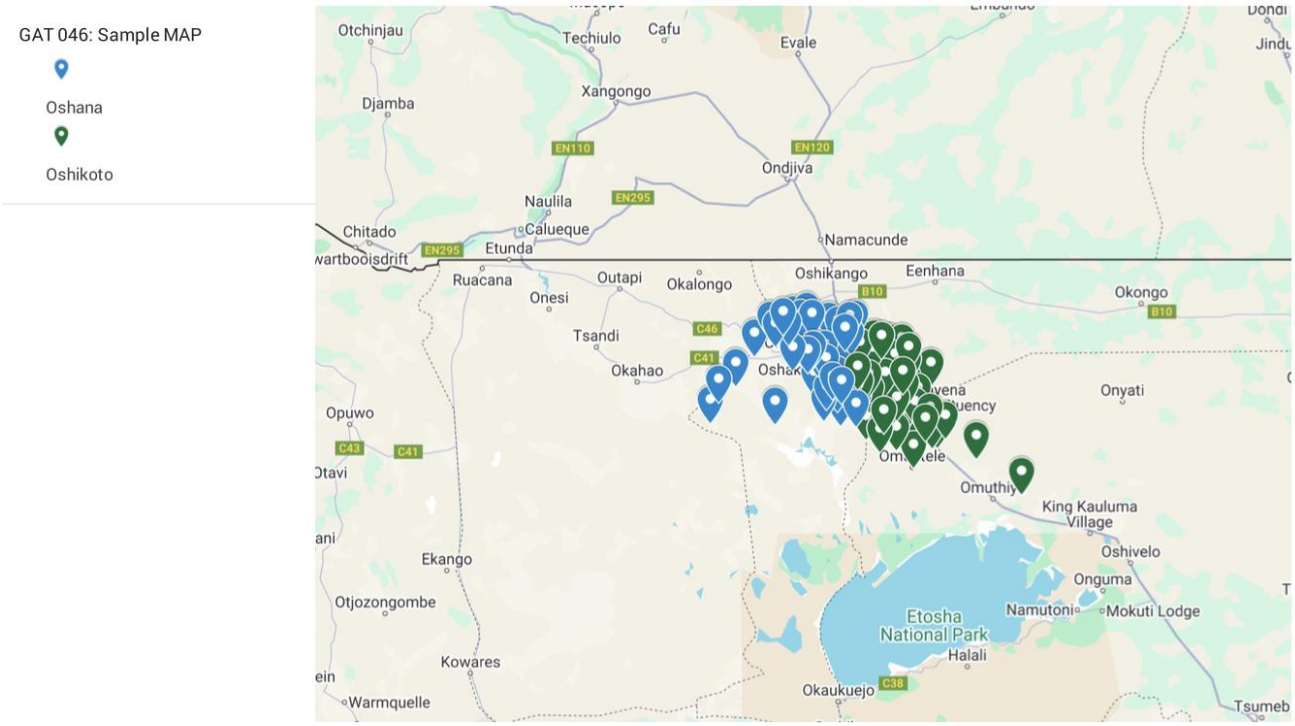
3.1.2 Target Sample

School Sample

The study aimed to include a total of 100 schools, with an even distribution of 50 schools in the Oshana Region and 50 in the Oshikoto Region (See Figure 1). These regions were selected due to their demographic and educational similarities, allowing Oshana to serve as the treatment area and Oshikoto as the comparison area. Within each school, the target sample consisted of learners who were enrolled in Grade 3 during the 2024 academic year and had progressed to Grade 4 by 2025. In addition, one Grade 3 teacher per school was targeted for participation, specifically, the main teacher who taught the sampled learners in Grade 3 during 2024.

[34] Unfortunately, most schools were designated as rural and so stratification on geotype was not possible (as it resulted in too many empty strata)

Figure 1: Sampled Schools



Class Sample

To ensure fairness and consistency in learner selection, a structured classroom sampling approach was followed. In schools where only one Grade 4 class received consent forms, the team first checked whether at least 20 learners had parental consent. If 20 or more learners consented, the team proceeded to select 20 learners using a lottery method. If fewer than 20 learners consented, all consenting learners were assessed, and the matter was escalated to the UNAM team to notify the Genesis team.

In schools with two Grade 4 classes, the sampling process varied depending on the number of consenting learners. If only one class had 20 or more consenting learners, sampling was done within that class. If both classes had 20 or more, one class was selected at random (by drawing the class name from a container), and 20 learners were sampled from that class. If the combined total of consenting learners across the two classes was at least 20 but neither class met the threshold individually, the two classes were merged for sampling, and 20 learners were randomly selected from the combined list.

Student Sample

Before sampling learners, the research team implemented strict eligibility criteria. All children who did not receive informed parental or guardian consent were excluded, as were learners who were repeating Grade 4 in 2025. Furthermore, any child identified through school records and teacher consultation as having significant learning disabilities, including but not limited to difficulty seeing or hearing, or significant challenges with communication or understanding others, was excluded from the sample to ensure ethical and accurate data collection during the EGRA.

Once eligibility was established, learners were sampled using a lottery-based random sampling method to ensure fairness and reduce selection bias. The enumerator team then drew 20 names at random to form the final learner sample. If a selected learner declined to participate or withdrew during the EGRA, they were replaced with the next numbered learner from the replacement pool to ensure that the target sample size was maintained.

3.1.3 Realised Sample

A total of 11 schools were replaced from the original sample. Five of these were identified as private schools, an attribute not captured in the original dataset but later flagged by regional school representatives from UNAM during their review. Two schools were inaccessible due to severe weather conditions and flooding in the area, while four schools had fewer than 25 learners enrolled in Grade 4, which was the minimum required to implement the sampling and administer the EGRA. Replacements were made in accordance with the original randomisation order, ensuring consistency in the sampling approach. Fieldwork was successfully completed with 50 schools in each of the regions.

3.2 Data Collection Instruments

During fieldwork, enumerators administered both the EGRA and a teacher questionnaire at each school. This section provides an overview of the data collection instruments used for the study.

EGRA

The Namibian version of the EGRA was used to assess Grade 4 learners' foundational reading abilities (See Annexe 6). Five different EGRA sub-tests were conducted, which assessed:

1. **Phonemic Awareness**, with learners being asked 10 questions to identify specific phonemic sounds, e.g. "What is the first sound in "egg"?"
2. **Letter Sound Recognition**, with learners tasked to sound as many letters as possible, out of 100, in one minute.
3. **Word Reading,** with learners tasked to read as many non-words and words as possible, from a list of 50, in one minute. If the learner got the first 5 words incorrect, the task automatically timed out.
4. **Oral Reading Fluency (ORF),** which required learners to read a short passage, with 79 words, in three minutes.
5. **Comprehension,** with learners asked 5 comprehension questions on the passage they had just read.

Teacher

The teacher questionnaire was administered to educators who taught Grade 3 during 2024. It included questions on their teaching experience, training in the Jolly methodology, and the types of resources they used in the classroom to support Grade 3 reading instruction.

3.3 Ethical Consideration

3.3.1 Research Ethics Approval

Ethical approval for the impact evaluation was obtained from the Ministry of Education, Arts, and Culture of Namibia. This clearance authorised the research team to commence data collection from research participants and to determine the effect of the Jolly programmes. The Ministry also approved the listed methods for the evaluation. A copy of the permission letter can be found here (See link to the permission letter).

3.3.2 Informed Consent Process

For this study, we designed consent processes to inform both teachers and primary caregivers of selected schools. The procedures ensured transparency about the study objectives, the data collection methods, and participants' rights to accept or decline participation. We outline the consent protocols for

teachers and primary caregivers below, detailing the steps taken to secure their approval and the measures implemented to safeguard the interests of all involved.

Teacher Consent

Before data collection, drop-off agents actively distributed consent forms to the selected schools. These forms provided detailed information about the study, including active contact information for participants with questions. The teacher who was selected indicated that they were willing to participate in the study, and this was recorded on the tablets.

Primary Caregiver Consent for Child Participation (EGRA Assessment)

Consent forms were distributed to each school for primary caregivers of eligible children. These forms explained the assessments to be conducted at the school with their child. Caregivers were given the option to indicate their willingness for their child and themselves to participate in the Index or to opt out.

Primary caregivers were instructed to complete the form and return it to their child's teacher, who would then pass it on to the research team. The consent form indicated that if the form were not returned, the research team would assume consent for the child's participation.

The field teams collected the consent forms, provided them to the regional coordinators, and subsequently sent them to the UNAM team for record purposes.

3.3.3 Inclusion/Exclusion Criteria

The following criteria were established to ensure that assessments were conducted only under appropriate conditions and with eligible students at the school. Criteria for inclusion in assessments:

- The child was within the eligible birth date range.
- The primary caregiver consented to participation.
- The child did not have hearing, vision, or mobility difficulties.
- The child was able to understand the assessment tasks.
- The child willingly participated.
- The child was well enough to perform the EGRA assessment.

4 | Data and Analysis Methodology

4.1 Final Dataset and EGRA Outcome Variables

The final dataset used for the analysis is at the learner level. It includes learners' EGRA performance, learners' characteristics and their school and teachers' characteristics. The final sample consists of 2,189 unique learners from 100 schools and 97 teachers' information (48 from Oshana and 49 from Oshikoto).

Table 1: Final Sample Size

Sample	Oshana	Oshikoto	Total
Schools	50	50	**100**
Learners	1078	1111	**2189**
Teachers	48	49	**99**

The scores for the five EGRA sub-tests are transformed in the final dataset for comparability across studies. Scores for Letter Sounds, Word Reading and ORF were converted to the number of correctly read letters or words per minute, hereafter referred to as correct words per minute (cwpm). As such, if a learner attempted to read all of the provided words/letter sounds in less than one minute, their score was proportionately adjusted upward, under the assumption that the learner would have continued to read words correctly at the same rate had there been more words to read.

Table 2: Description of the EGRA Outcome Variables

Component	Variable Description	Maximum
Phonemic Awareness	Number of phonemic sounds correctly identified.	10
Letters	Number of letter sounds correctly sounded per minute.	100+
Word Reading	Number of words correctly read per minute. Unit: cwpm	50+
Oral Reading Fluency	Number of words correctly and fluently read per minute. Unit: cwpm	26,33+
Comprehension	Number of comprehension questions correctly answered.	5

In addition to each individual EGRA outcome measure, two types of composite EGRA scores, outlined in Box 2 below, were constructed to analyse learners' overall reading skills. The first was calculated as a simple summation of the standardised scores for each of the five sub-tests, while the second was calculated using Principal Component Analysis (PCA)[35]. Effectively, PCA examines patterns in how scores

[35] For those unfamiliar with PCA, PCA is a statistical technique that reduces the dimensionality of a dataset while retaining as much variance as possible. It does this by transforming a set of possibly

across reading subskills tend to vary together and regards the first principal component in this variation as reflective of the underlying construct, which in our case is early grade literacy.

For ease of interpretation, we standardise the individual sub-tests and both composite indices to have a mean value of 0 and a standard deviation of 1. Throughout the impact evaluation, we estimate and interpret results for both the composite scores and on each of the sub-tasks separately.

Box 2: Composite EGRA Scores

The five EGRA sub-tests are combined to produce two types of composite indicators of learners' early grade reading skills:

1. **Simple Composite Score** (Composite Sum): Calculated as the unweighted sum of the five standardised components. This method treats each subskill equally and provides a straightforward aggregate measure of reading ability.
    - Each sub-test's score was rescaled to a 0–1 range using min-max normalisation to ensure comparability across measures with different ranges.
2. **Principal Component Analysis Score** (PCA Score): To account for potential correlations between components and to generate a data-driven weighting scheme, we conducted a PCA on the original (non-standardised) sub-test scores. The first principal component was extracted to serve as an alternative composite measure, representing the largest source of variance in the dataset.

4.2 Analysis Methodology

Ideally, the impact of the Jolly programmes would be assessed using a longitudinal study in which the EGRA outcomes would be compared before and after programme rollout. However, the study commenced after the programme had already been rolled out in Oshana for Grade 3 teachers. To estimate the programme's effect, we compared learners in Oshana (treatment group) with those in Oshikoto (comparison group). Since these groups may differ in background characteristics, we first estimate multivariate regressions before using Propensity Score Matching (PSM) to create comparable groups and better isolate the effect of the programmes.

Box 3: A Note on Treatment Assignment

In analysing the impact of the Jolly programmes, it is important to distinguish between two common approaches to estimating effects: the Intent-to-Treat (ITT) effect and the Treatment-on-the-Treated (TOT) effect. ITT analysis estimates the impact of being assigned to the treatment group, regardless of whether the participants actually received or complied with the intervention. By contrast, the TOT approach estimates the effect only among those who actually received the treatment.

This evaluation estimates the ITT effect. This approach recognises that some surveyed teachers in Oshana may not have received Jolly Grammar training in 2021, or that some who received training may have never implemented the programme, but this is expected in real-world scenarios and should be reflected in the estimated effects of the programme. The decision to estimate the ITT effect is consistent with standard practice in impact evaluation, where ITT is often preferred

correlated variables (like reading subskills) into a new set of uncorrelated variables called principal components.

because it preserves the original assignment structure and recognises that not all individuals assigned to treatment will always fully comply.

4.2.1 Multivariate Regression Analysis

Before applying the PSM, we ran multivariate regressions for each EGRA outcome to explore the relationship between the Jolly programmes and learner performance, while controlling for key differences in background characteristics. To maximise statistical power and account for baseline differences, we included control variables such as school location and size, learner age, gender, and language, as well as teachers' demographics, language of instruction, experience and training.

4.2.2 Propensity Score Matching Analysis

To estimate the impact of the Jolly programmes on Grade 4 learners' EGRA outcomes, we used PSM. Since learners were not randomly assigned to the 'treatment' or 'comparison' group, PSM helps reduce bias by creating more comparable groups based on background characteristics.

The PSM analysis is conducted as follows:

1. **Estimation of Propensity Scores:** We used a probit regression to calculate the learners probability (propensity) of receiving the intervention based on characteristics like learner age, gender, language, and teachers or school level variables. These variables were selected for their relevance to predicting EGRA outcomes.
2. **Matching:** Learners in the treatment group (Oshana) were matched to learners in the comparison group (Oshikoto) using the Nearest Neighbour (NN) matching method. To improve match quality and robustness, we toggled four different matching parameters, as outlined in the table below, to produce various matching models.
3. **Covariate Balance Checks:** We assess the quality of the match by comparing standardised mean differences and variance ratios of covariates before and after matching. Where balance was poor, we adjusted the matching approach by changing the number of nearest neighbours, and applying a caliper, toggling replacement, or enforcing common support to improve comparability between groups.

Table 3: Matching Method Parameters

Parameter	Explanation	Options
Stata Matching Command	Stata commands that create comparable groups by matching treated individuals with untreated individuals who are similar on observable characteristics. • *teffects psmatch* is Stata's official built-in command for PSM. • *psmatch2* is a user-written Stata command that offers a greater variety of matching options Both commands are explored due to their different input options and functionalities. For example, the *psmatch2* command allows one to enforce common support or specify no replacement of comparison	*teffects psmatch* or, *psmatch2*

Parameter	Explanation	Options
	individuals when matching. However, the *teffects psmatch* command is generally preferred as it accounts for the estimation of propensity scores when calculating standard errors.	
Number of Nearest Neighbours (NN)	Refers to how many comparison group individuals are selected to match with each treated group individual based on their propensity scores. • 1:1 NN Matching: For each treated individual, you find one comparison individual with the most similar propensity score. • 1:N NN Matching: For each treated individual, you find N (e.g. 3, 5) comparison individuals with the most similar propensity scores.	1:1 NN or, 1:3 NN or, 1:5 NN
Common Support	Enforcing common support refers to retaining only the individuals (both treated and comparison) whose propensity scores fall within the overlapping range of both groups' propensity score distributions.	No common support or, common support is enforced
Replacement Strategy	The replacement strategy defines whether a comparison individual can be used as a match for more than one treated individual.	With replacement or, without replacement
Caliper Restriction	A caliper restriction (often called a "caliper") defines the maximum allowable difference in propensity scores between a treated individual and a comparison individual for them to be considered a valid match. When a caliper restriction is put in place, it is set at 0.06, which is 0.2 times the pooled-sample standard deviation of the propensity score as recommended by Rosenbaum & Rubin (1985)[36].	No caliper or, a caliper of 0.06

4.3 Study Limitations

While every effort has been made to ensure the robustness of these results, this study still has important limitations which must be addressed. The study design and analytical approach are appropriate for the context but still constrain the robustness of the estimated effect sizes and the generalisability of the results.

First, the non-randomised design of the study introduces the potential for selection bias, even though the Jolly programmes were assigned at the regional level. The Oshana and Oshikoto regions in Namibia may differ systematically in terms of socioeconomic conditions, school quality, teacher capacity, and other contextual factors that influence learning outcomes. If such differences existed between the

[36] Rosenbaum, P. R., & Rubin, D. B. (1985). Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. *The American Statistician*, 39(1), 33-38.

treatment and comparison regions, they could bias the estimated effects of the Jolly programmes. While PSM and multivariate regression were used to control for observable covariates, these methods cannot adjust for unmeasured confounders such as learner motivation, classroom practices, or the local fidelity of programme implementation. As such, caution is warranted in interpreting causal claims.

Relatedly, the absence of baseline data substantially limits the study's ability to accurately estimate effect sizes and attribute gains solely to the Jolly programmes. Without pre-intervention measures of learners' reading skills, it is not possible to account for initial differences in reading proficiency between treatment and comparison groups. This makes it difficult to determine whether observed post-intervention differences are due to the programme itself or to pre-existing, unobservable disparities.

Third, each analytical method used has inherent limitations. Unconditional mean comparisons do not account for confounding variables, which can distort effect estimates. Multivariate regressions improve precision but rely on assumptions such as linearity and correct model specification. PSM enhances causal inference but is sensitive to the quality and completeness of covariates used in the matching process. As will be seen, the variability in significance and the magnitude of effects across methods highlights the need for cautious interpretation.

Fourth, while Oshana is considered the treatment group due to its earlier exposure to Jolly Phonics and Jolly Grammar, the actual use of Jolly resources between 2021 and 2024 remains unclear. Moreover, the initial training received by Grade 2 and 3 teachers in 2021 consisted of only a 2-day session, with no systematic monitoring of uptake or usage in classrooms. Given that the learner outcome data were collected in 2025, this introduces a significant time lag between the initial intervention/programme rollout and measurement. It is therefore difficult to establish whether the learners assessed were meaningfully exposed to the programme during their time in Grade 2 and 3, and thus whether any observed differences in outcomes can be attributed to early Jolly Phonics and Grammar intervention.

Taken together, these limitations should discourage us from making firm causal claims from this analysis, and rather, lead us to interpret the results with caution.

5 | Descriptive Statistics

5.1 Learner

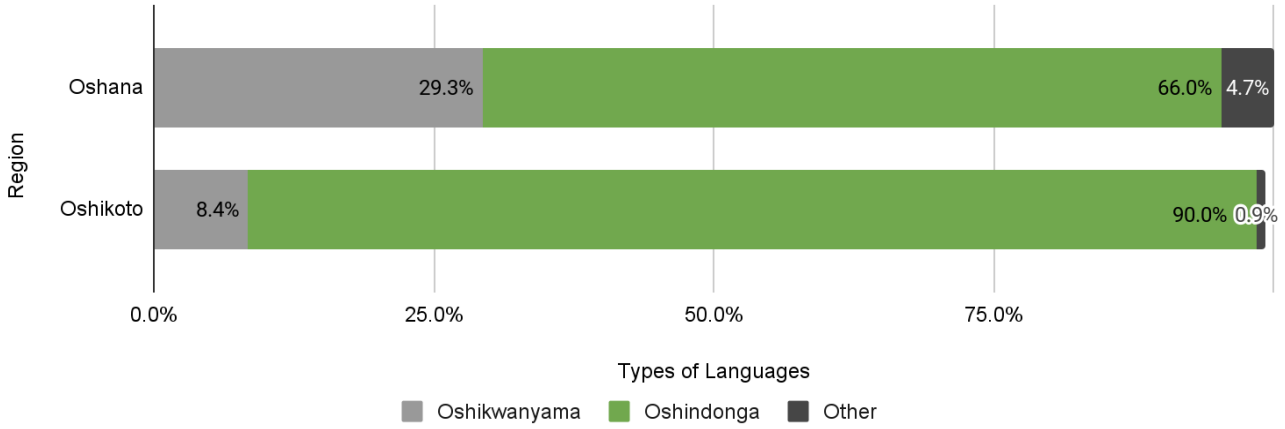
5.1.1 Learner Demographics

A total of 2,189 learner EGRA submissions were completed across the two regions, with 1,111 from Oshana and 1,078 from Oshikoto. The average age of students varies slightly across regions. In Oshana, the mean age is 9.6 years, with the youngest student being 7 years old and the oldest 14 years old. In Oshikoto, the mean is slightly higher at 9.7 years, with the youngest student being 7 years old and the oldest being 13 years old. Across both regions, the overall average age is 9.6 years.

The gender distribution of students is relatively balanced across both the Oshana and Oshikoto regions. In Oshana, just over half of the students are female (50.8%), and just under half are male (49.2%), while in Oshikoto, females make up just over half of the sample (52.4%), and males slightly less than half (47.6%). Overall, across both regions, just over half of the students are female (51.6%) and just under half are male (48.4%), indicating a slight predominance of girls in the total learner population.

In terms of home language, most students speak either Oshindonga or Oshikwanyama, with only a small fraction reporting other languages (2.8%). Oshindonga is the most widely spoken home language, used by approximately two-thirds of students in Oshana (66%) and the vast majority in Oshikoto (90%), accounting for just over three-quarters of the total sample (78.1%). Oshikwanyama is more common in Oshana, where nearly a third of students reported it as their home language (29.3%), compared to less than one in ten in Oshikoto (8.4%), totalling just under one-fifth overall (19%).

Figure 2: Student Home Languages by Region



5.2 Teachers

5.2.1 Teacher Demographics

A total of 97 teacher interviews were conducted, with 48 in Oshana and 49 in Oshikoto. The youngest teacher in Oshana is 26 years old, while in Oshikoto, the youngest is around 25 years old. The oldest teachers in both regions are about 59 years old. The average age of teachers across both regions is approximately 38 years, reflecting a wide age range among Grade 3 teachers, from early career professionals to those nearing retirement age. The gender distribution was similar across both regions: in Oshana, just over a quarter of the teachers were male (27.1%), and nearly three-quarters were female (72.9%), while in Oshikoto, almost a third of the teachers were male (30.6%) and just over two-thirds were female (69.4%).

In Oshana, most teachers reported speaking Oshindonga at home (62.5%), while just over a third spoke Oshikwanyama (35.4%) and a tiny proportion spoke other languages (2.1%). In Oshikoto, nearly all teachers reported speaking Oshindonga at home (91.8%), with a very small number speaking Oshikwanyama (8.2%). When teaching students, most teachers in Oshana used Oshindonga (70.8%), followed by smaller proportions using Oshikwanyama (18.8%) and English (10.4%). In Oshikoto, the majority also taught in Oshindonga (89.8%), with fewer using English (6.1%) and the least using Oshikwanyama (4.1%).

Figure 3: Teacher Home Languages by Region

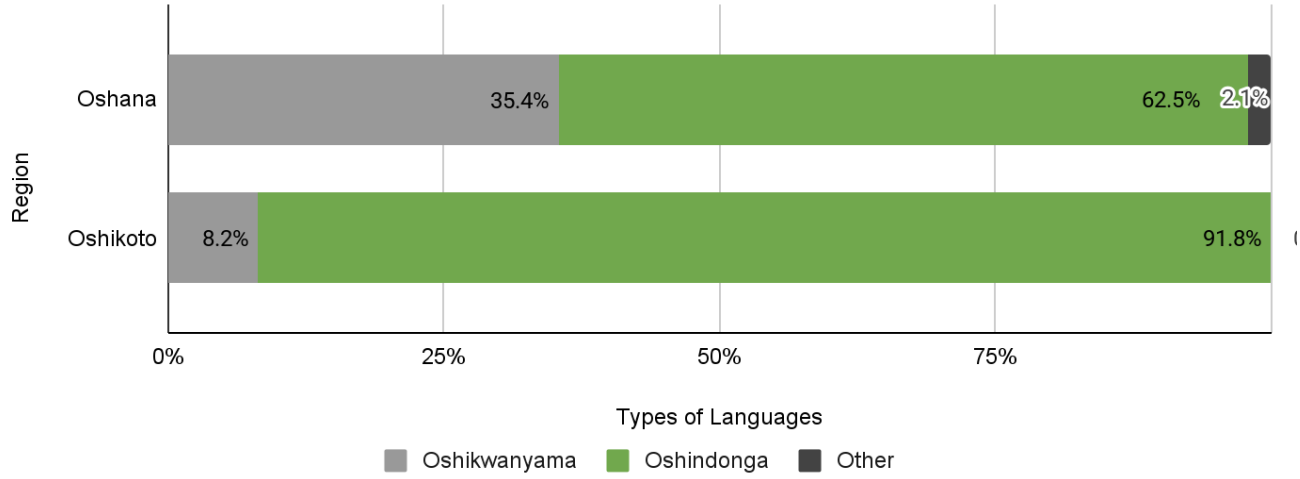
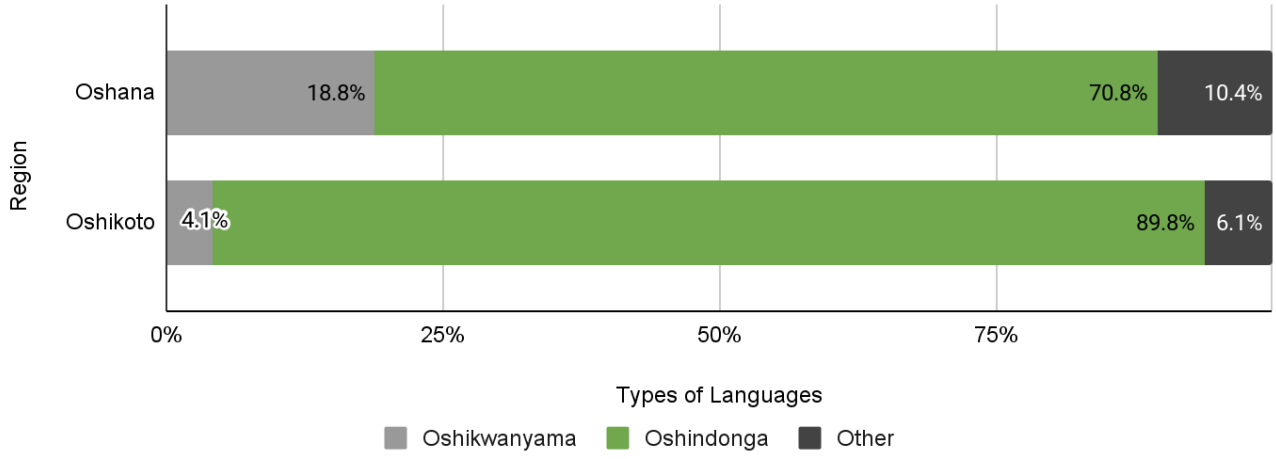


Figure 4: Teacher Teaching Languages by Region



5.2.2 Education and Qualifications

In both the Oshana and Oshikoto regions, most teachers reported having completed Grade 12. In Oshana, most completed Grade 12 under the old curriculum (93.8%), with only a small portion completing the new NSSCAS system (6.3%). In Oshikoto, while most also completed Grade 12 under the old system (75.5%), a higher proportion than in Oshana completed the new NSSCAS curriculum (24.5%).

Nearly all teachers across both regions reported having a tertiary education. In Oshana, almost all teachers held tertiary qualifications (97.9%), and in Oshikoto, every teacher reported having tertiary qualifications (100%). Overall, 99% of teachers interviewed (96 out of 97) had completed some form of tertiary education. The most common qualification in both regions was a Bachelor's degree, held by over half of teachers in Oshana (55.3%) and nearly half in Oshikoto (46.9%). Diplomas were also frequently reported, particularly in Oshikoto, where 42.9% of teachers held this qualification compared to 31.9% in Oshana. Postgraduate degrees were held by a small number of teachers, with slightly more reported in

Oshikoto. Qualifications such as Higher Certificates, Postgraduate Certificates, and other types were mentioned by fewer than 10% of teachers in either region and were therefore less common.

Figure 5: Post Graduate Qualifications Oshana

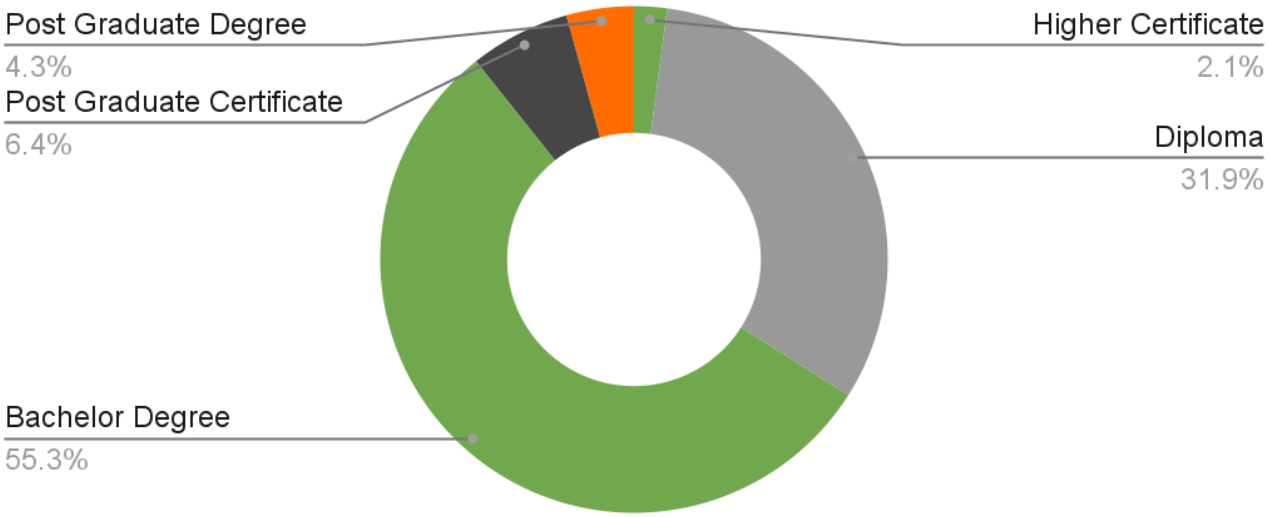
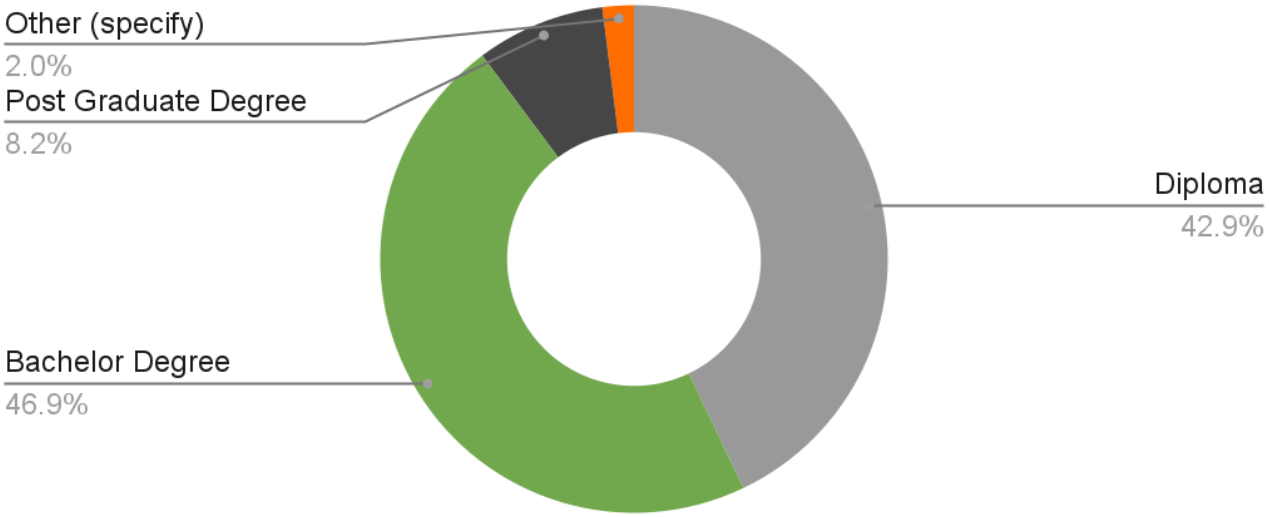


Figure 6: Post Graduate Qualifications Oshikito



The figure below illustrates the years of experience of teachers in the Oshana and Oshikoto regions. In Oshana, most teachers (46%) have between 0 and 5 years of experience, while 25% have 6-10 years, and less than a third (29%) have over 10 years of experience. In contrast, Oshikoto has a slightly different distribution, with less than half (38%) of teachers having 0-5 years of experience, around a quarter (26%) with 6–10 years, and a greater proportion (36%) with more than 10 years of experience. This suggests that Oshikoto has a more experienced teaching workforce than Oshana. In the Oshana region, most teachers have 0-5 years of experience (58.3%). Similarly, in Oshikoto, many teachers fall into this category (57.1%). Notably, Oshikoto has a slightly higher proportion of teachers with over 10 years of experience (18.4%) compared to Oshana (14.6%).

Figure 7: Teachers' Years of Experience by Region

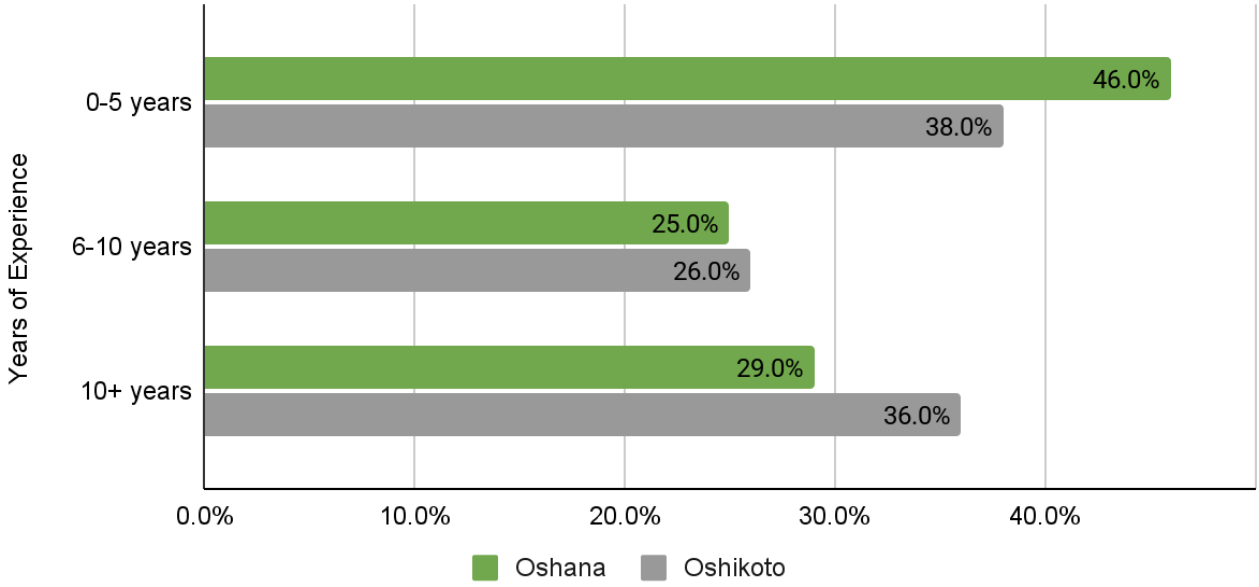
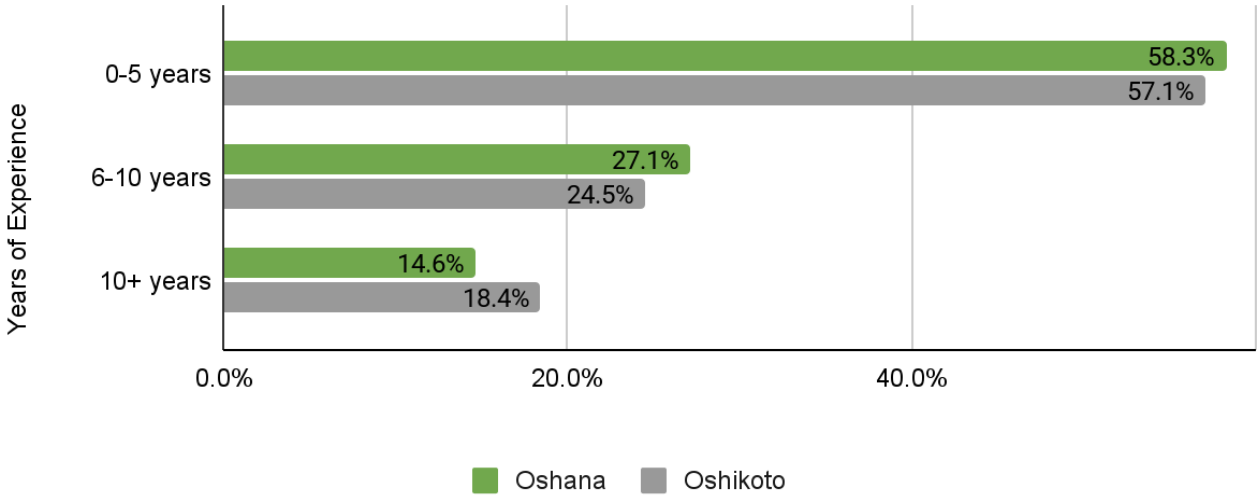


Figure 8: Teachers' Experience at Grade 3 Level by Region



5.3 Jolly Phonics and Grammar

5.3.1 Jolly Training and Support

Most teachers have benefited from Jolly Phonics or Grammar training, with the vast majority of surveyed teachers in both Oshana (93.1%) and Oshikoto (87.5) reporting attending some form of Jolly Phonics or Grammar training. Of those teachers that reported attending a training session, most felt that the training prepared them adequately for classroom implementation (84.4%), with strong confidence reported in Oshana (91.3%) and a solid majority in Oshikoto (77.3%).

While teachers responded positively to the training, most reported not receiving support from their regional office or HoD to implement the Jolly Phonics or Grammar programme after training (66%). In Oshana, just under half said they received support (47.9%), while in Oshikoto, far fewer (20.4%) reported receiving support from the regional office or their HoD.

Of those who did report receiving support, the frequency of support varied. In Oshana, most teachers reported never receiving support in 2024 (50.0%), while some received support at least once in 2024 (36.4%), and a few between 2 and 4 times (13.6%). In Oshikoto, while fewer teachers received support overall, some still received it once (30%), and others between 2 and 4 times (30%) in 2024.

Table 4: Frequency of Support from Regional Office or HoDs in 2024

	Oshana		Oshikoto		Totals	
Frequency	N	%	N	%	N	%
Never	11	50.0%	4	40.0%	15	46.9%
1 time	8	36.4%	3	30.0%	11	34.4%
2 to 4 times	3	13.6%	3	30.0%	6	18.8%
Total	22	100.0%	10	100.0%	32	100.0%

Amongst the teachers who received support, the most common types were regular classroom visits and feedback sessions, each selected by 36.4% of respondents[37]. These were more frequent in Oshana (43.5%) than in Oshikoto (20.0%). Assistance with lesson planning (27.3%) and one-on-one mentoring or coaching (21.2%) were also common. Mentoring was reported more often in Oshikoto (30.0%) than in Oshana (17.4%). Fewer teachers selected reporting on teaching progress (15.2%) and troubleshooting challenges (3.0%) as forms of support offered by the regional office of their HoDs.

Table 5: Type of Support from Regional Office or HoDs in 2024

Support Type	Oshana	Oshikoto	Total
Regular Classroom Visits	43.5%	20.0%	36.4%
Feedback Sessions	43.5%	20.0%	36.4%
Reports on Teaching Progress	17.4%	10.0%	15.2%
Assist with Lesson Planning	30.4%	20.0%	27.3%
One-on-one Mentoring	17.4%	30.0%	21.2%
Troubleshooting Challenges	4.3%	0.0%	3.0%
Other	26.1%	50.0%	33.3%

5.3.2 Jolly Materials

In Oshana, the Grade 3 teachers reported higher access to Jolly Phonics or Grammar resources (70.8%) compared to Oshikoto (30.6%, respectively). This is to be expected, given that in 2025, Jolly Grammar 2 had not yet been rolled out in Oshikoto, but had been introduced in Oshana in 2021.

5.4 Classroom

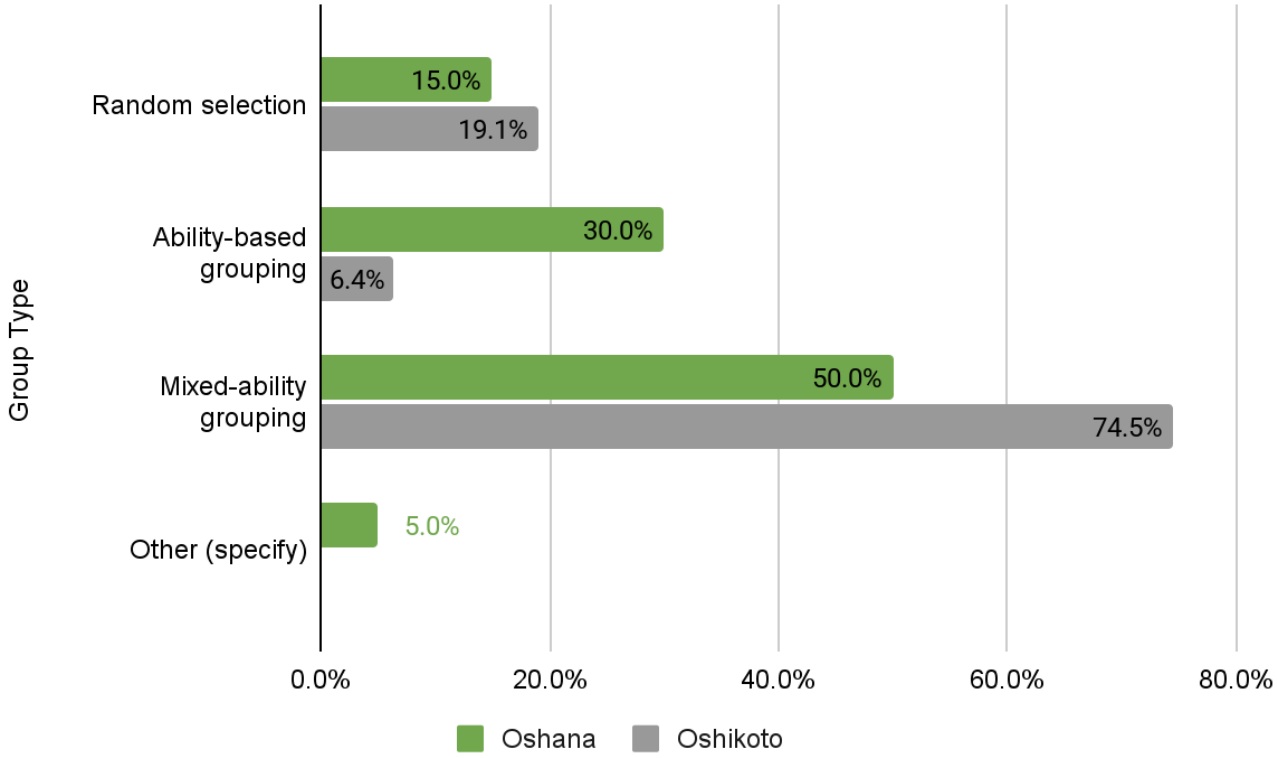
5.4.1 Classroom Overview

Class sizes were similar across both regions, with an average of 34 learners per class. In Oshana, class sizes ranged from 22 to 45 learners, while in Oshikoto the range was wider, from 21 to 66 learners. Most teachers reported grouping learners for activities, with 85.1% doing so in Oshana and an even higher

[37] Teachers could select multiple types of classroom support.

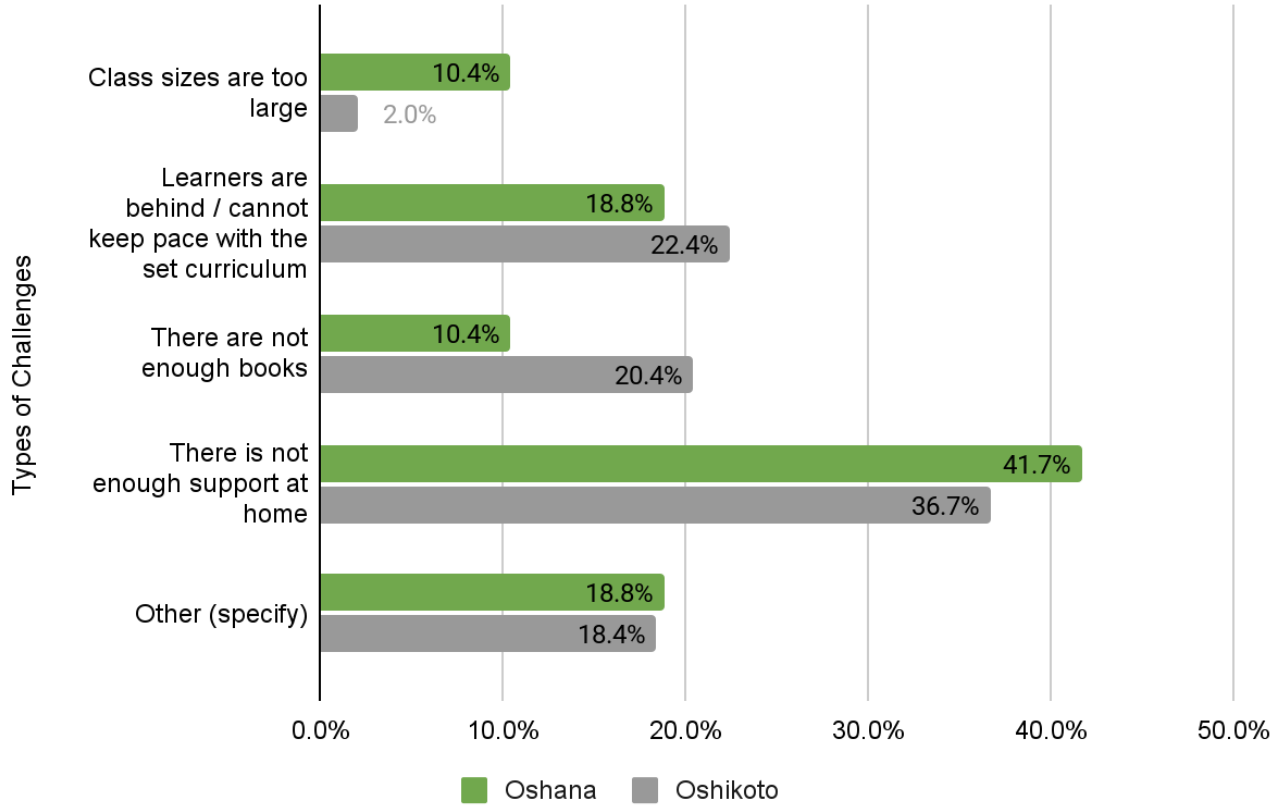
95.9% in Oshikoto. Teachers used various strategies to group learners during activities. In both regions, mixed-ability grouping was the most common approach, reported by half of the teachers in Oshana (50%) and three-quarters in Oshikoto (74.5%). Ability-based grouping was more commonly used in Oshana (30%) than in Oshikoto (6.4%), while a smaller share used random selection - 15% in Oshana and 19.1% in Oshikoto. Only a few teachers in Oshana (5%) reported using other grouping methods, while no alternative approaches were noted in Oshikoto.

Figure 9: Activity Group Types by Region



Teachers in both regions reported several challenges affecting classroom learning. The most common concern was a lack of support at home, cited by just over two-fifths of teachers in Oshana (41.7%) and over a third in Oshikoto (36.7%). In Oshikoto, a notable proportion also highlighted that learners were behind or struggling to keep pace with the curriculum (22.4%), slightly higher than in Oshana (18.8%). A lack of books was more frequently reported in Oshikoto (20.4%) than Oshana (10.4%), while large class sizes were a concern for some teachers in Oshana (10.4%) but were rarely raised in Oshikoto (2.0%) and around one in five teachers in both regions also identified other challenges not captured in the standard categories.

Figure 10: Classroom Challenges by Region



5.4.2 Activities in the Classroom

The frequency with which teachers conduct various English literacy activities in their classrooms was also analysed. For group-guided reading, the data reveals that the majority of teachers reported implementing this activity either once a week or 2–4 times per week. Notably, Oshikoto exhibited a higher proportion (47%) engaging in group-guided reading once weekly compared to Oshana (27%). Listening and speaking activities were predominantly conducted once a week. A significant proportion of teachers in Oshikoto (74%) selected this frequency, compared to 48% in Oshana. Daily implementation was more frequent in Oshana (29%) than in Oshikoto (18%). Phonics instruction demonstrated a different pattern. A majority of teachers in Oshana (56%) reported conducting phonics daily, whereas Oshikoto's daily implementation was considerably lower at 27%. Nevertheless, phonics remains one of the more frequently used activities overall, with only a small proportion reporting rare or no use.

Writing emerged as the most consistently practised daily activity, particularly in Oshana, where most teachers reported daily implementation (52%), compared to a lower but still significant proportion in Oshikoto (41%). Significantly fewer teachers reported rarely conducting writing activities. Shared reading was most commonly implemented once a week, especially in Oshikoto (53%), followed by a smaller share in Oshana (38%). Daily or twice-daily implementation of shared reading remained relatively low, particularly in Oshikoto (8%). Paired reading exhibited more variation in frequency. The most common implementation was once a week in both regions (31% in Oshana and 35% in Oshikoto), followed by implementation 2–4 times per week. Daily use of paired reading was relatively low across both regions. Notably, a small percentage of respondents reported never using paired reading (9%), while others used it only a few times per month or less (17%).

Figure 11: Classroom Activities Frequency in Oshana

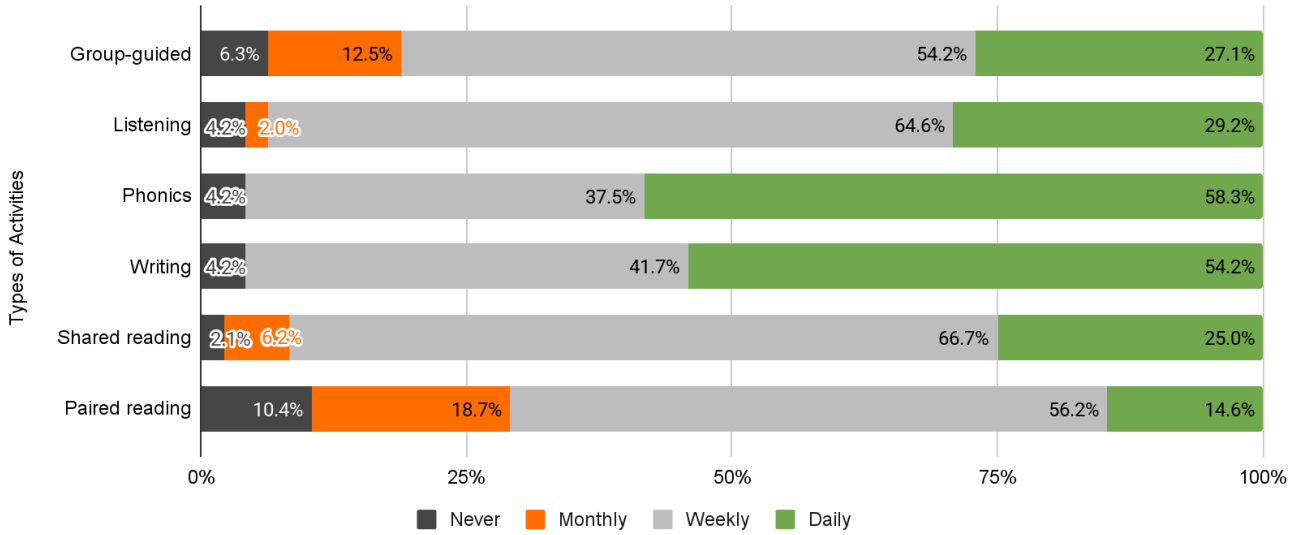
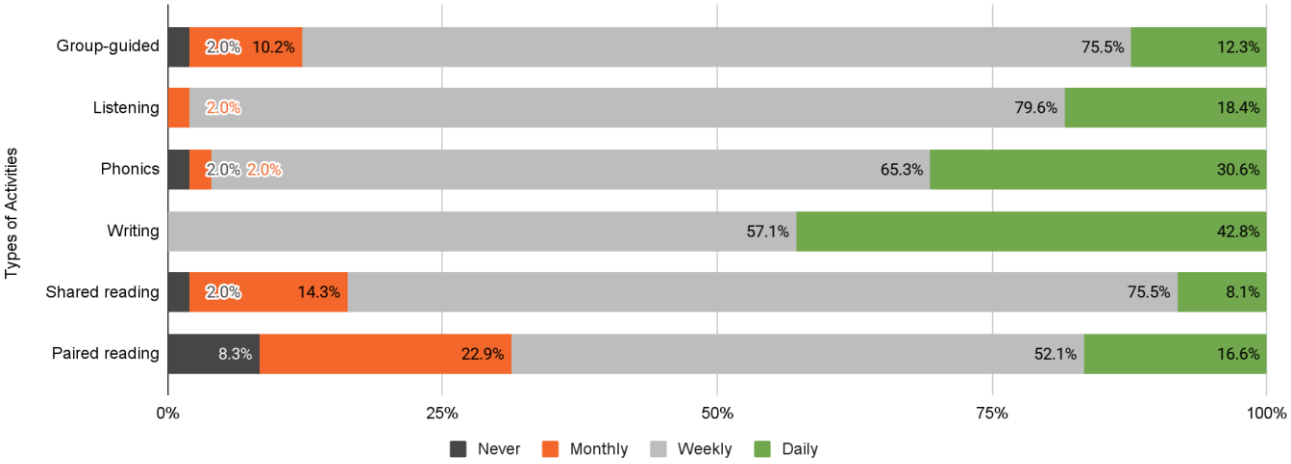


Figure 12: Classroom Activities Frequency in Oshikoto



5.4.3 Classroom Resources and Materials

In Oshana, the Grade 3 teachers reported higher access to Jolly Phonics or Grammar resources (70.8%) and multimedia materials (54.2%) compared to Oshikoto (30.6% and 36.7%, respectively). Conversely, Oshikoto had greater availability of lesson plans (93.9%), curriculum documents (93.9%), and textbooks/workbooks (81.6%) than Oshana (83.3%, 85.4%, and 60.4%, respectively). Access to online platforms, reading materials, interactive activities, and assessment tools was relatively similar across both regions.

In 2024, the availability of teaching materials varied between the Oshana and Oshikoto regions. In Oshana, most teachers reported having classbooks/exercise books (57.4%) and readers (38.3%). However, fewer teachers had access to MoEAC workbooks (31.9%) and interactive posters for synthetic phonics (51.1%). A notable portion of teachers in Oshana reported having no materials at all (14.6%). In Oshikoto, many teachers had classbooks/exercise books (60.4%) and readers (56.2%). However, fewer teachers had access to MoEAC workbooks (20.8%) and interactive posters for synthetic phonics (35.4%). Similar to Oshana, a portion of teachers in Oshikoto reported having no materials at all (14.3%). Across both regions, some teachers were unsure about material availability (21.0%).

Table 6: Classroom Resources by Region

Resources	Oshana	Oshikoto	Total
Lesson Plans	83.3%	93.9%	**88.7%**
Curriculum, Syllabi & IPM	85.4%	93.9%	**89.7%**
Jolly Phonics Resources	70.8%	30.6%	**50.5%**
Textbooks & Workbooks	60.4%	81.6%	**71.1%**
Online Platforms	45.8%	49.0%	**47.4%**
Reading Materials	62.5%	73.5%	**68.0%**
Multimedia Resources	54.2%	36.7%	**45.4%**
Interactive Activities	37.5%	34.7%	**36.1%**
Assessment Tools	45.8%	40.8%	**43.3%**
Other	12.5%	22.4%	**17.5%**

Table 7: Classroom Materials by Region

Materials	Oshana	Oshikoto	Total
Readers	38.3%	56.2%	**47.4%**
Classbooks / Exercise Books	57.4%	60.4%	**58.9%**
MoEAC Workbooks	31.9%	20.8%	**26.3%**
Photocopied Worksheets	42.6%	58.3%	**50.5%**
Interactive Posters (Synthetic Phonics)	51.1%	35.4%	**43.2%**
Other (specify)	31.9%	25.0%	**28.4%**
None (no materials at all)	14.6%	14.3%	**14.4%**
Don't Know	21.0%	20.0%	**21.0%**

5.5 Reading Outcomes: Unconditional Comparison of Means

Table 6 below presents a simple comparison of mean EGRA scores between learners taught by Jolly-trained teachers (Oshana) and those in the comparison group (Oshikoto). The final column shows mean differences (T-C), with asterisks indicating statistical significance. Full score ranges are available in the Annexe.

For Phonemic Awareness, Letter Sounds, ORF, Comprehension, and both composite scores (PCA and Sum), the mean differences between treatment and comparison groups are not statistically significant. The only statistically significant difference is in Word Reading, where the treatment group outperforms the comparison group by an average of 0.97 words per minute ($p < 0.05$).

Table 8: Comparative EGRA Outcomes

EGRA Outcome	Full Sample		Oshana (T)		Oshikoto (C)		T-C	
	Mean	**Std Dev.**	**Mean**	**Std Dev.**	**Mean**	**Std Dev.**	**Diff.**	**Diff. in SD**
Composite PCA	0.00	1.65	0.02	1.72	-0.02	1.58	0.037	0.022
Composite Sum	1.72	0.74	1.73	0.76	1.70	0.71	0.028	0.038
Phonemic Awareness	6.41	3.09	6.48	3.04	6.34	3.14	0.145	0.047
Letter Sounds	22.44	12.10	22.32	12.51	22.55	11.68	-0.233	-0.019
Word Reading	**14.26**	**9.99**	**14.73**	**10.63**	**13.78**	**9.26**	**0.951****	**0.097****
ORF	45.83	24.31	45.01	24.67	46.68	23.91	-1.673	-0.069
Comprehension	1.64	1.30	1.66	1.36	1.62	1.22	0.040	0.031
N	2189		1111		1078			

Notes: The significance of the difference in outcomes between treatment and comparison groups is calculated by regressing each outcome variable on the treatment variable. Significance is denoted as * p<0.1; ** p<0.05; *** p<0.01.

Figure 14 shows the proportion of learners per group achieving a zero score across sub-tests. The treatment group had more zero scores in the Word Reading, ORF and Comprehension sub-tests, while the comparison group had more zero scores in the Phonemic Awareness test. The two groups had a similar proportion of learners achieving a zero score in the Letter Sounds test.

Figure 13: Proportion of Zero Scores by EGRA Sub-test

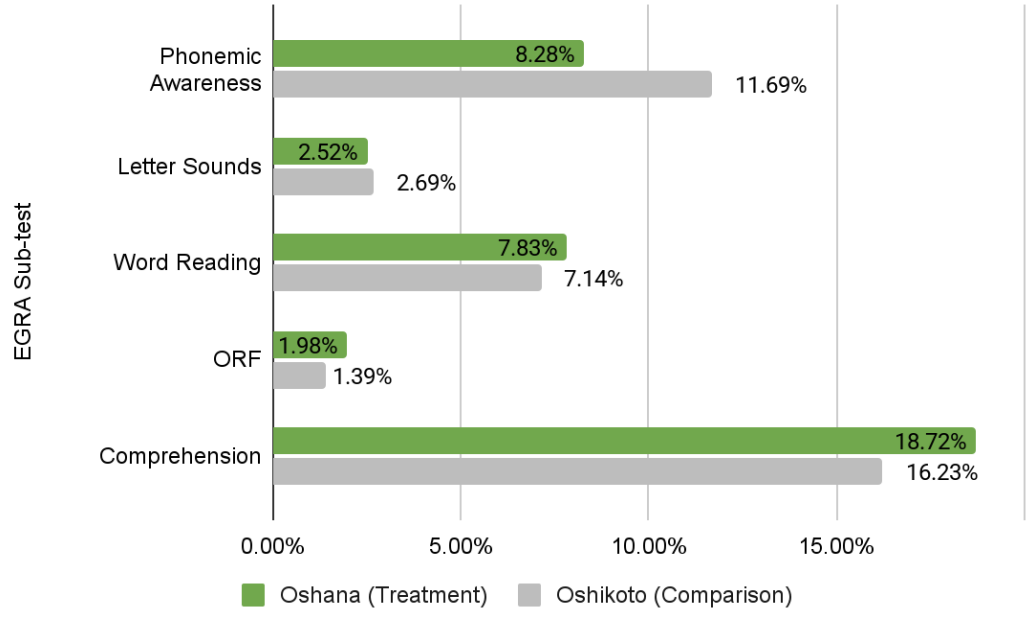


Figure 15 shows the percentage of children, in the treatment and comparison group, achieving above particular thresholds of correct words per minute during the Word Reading sub-test. The scores at the 25th percentile, median (50th percentile) and 75th percentile of the distribution for the full sample of 100 schools are indicated on the graph. Roughly 92% of learners in both groups correctly read at least one word, while more learners in the treatment group reached higher correct-word-per-minute thresholds. This suggests that the Jolly programmes may have had a greater effect on higher-performing students.

Figure 14: Word Reading for Treatment and Comparison Groups

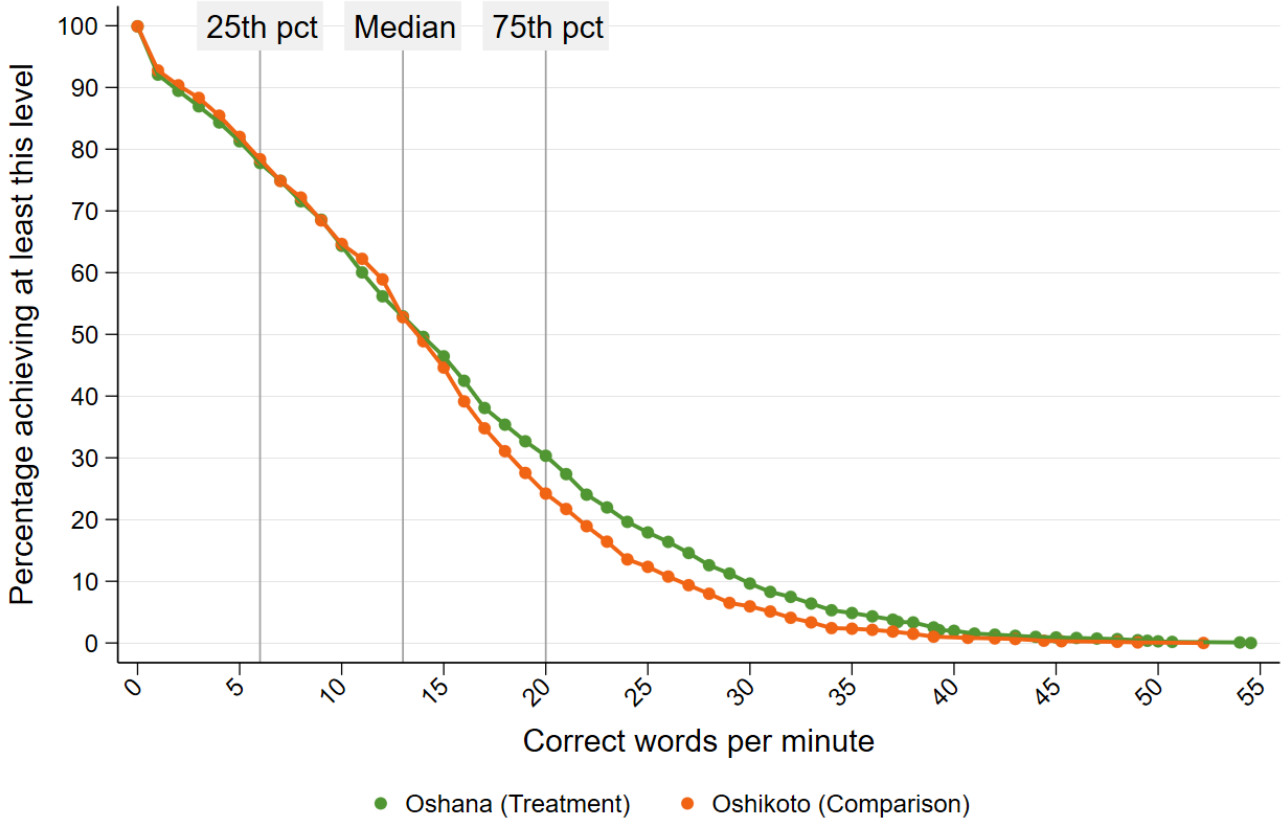
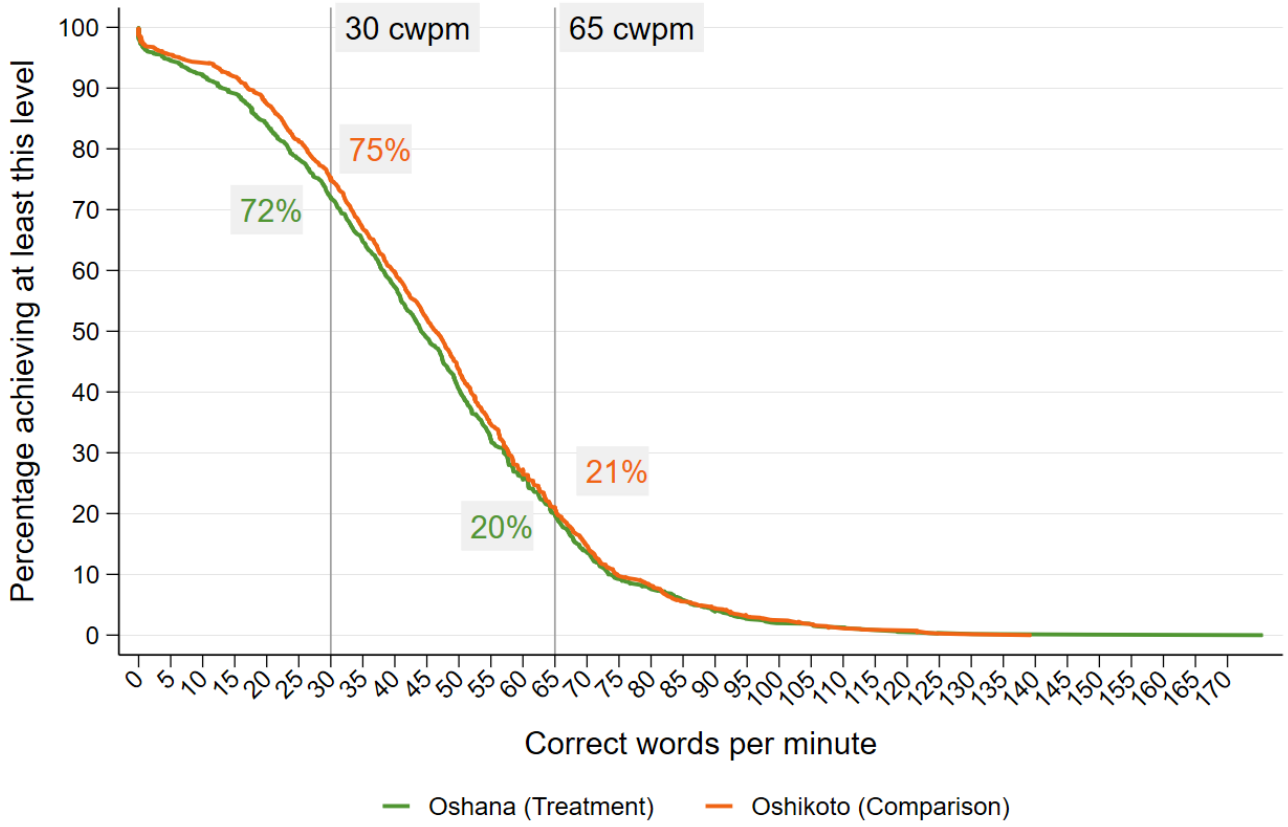


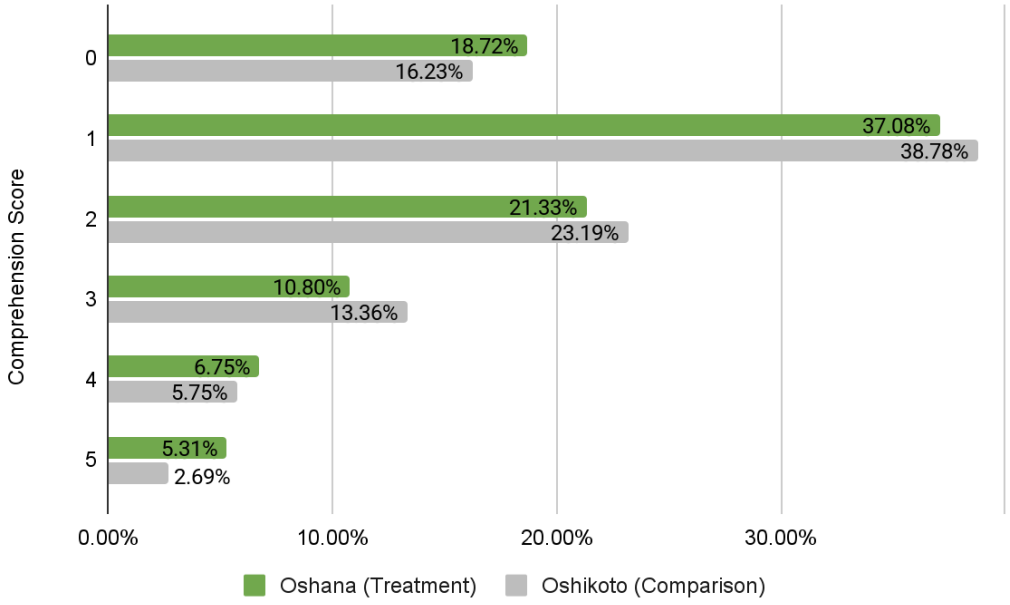
Figure 16 compares ORF between groups. The comparison group slightly outperforms the treatment group across most of the distribution. About 75% of learners in the comparison group could read at least 30 words correctly in a minute, compared to 72% in treatment. At the 65 cwpm benchmark, differences were minimal (21% comparison vs. 20% treatment). These small gaps suggest a possible limited effect on reading fluency.

Figure 15: ORF for Treatment and Comparison Groups



Comprehension results are presented in Figure 17. A larger share of learners in the treatment group scored either 4 or 5 out of 5 compared to the comparison group. Foundational reading-skill (Phonemic Awareness and Letter Sounds) sub-test distributions are shown in the Annex and reveal mixed patterns, with no consistent treatment effect.

Figure 16: Comprehension Scores for Treatment and Comparison Groups



5.5.1 Correlations between EGRA components

To better understand the relationships between reading sub-skills, Table 7 presents pairwise correlations across all learners. This analysis further supports the validity of the EGRA sub-tests by confirming that the expected relationships between components are present in the data.

The strongest correlation is between Word Reading and ORF, suggesting that the ability to read individual words accurately and efficiently is closely linked to fluent passage reading. ORF is moderately correlated with Comprehension, highlighting the role of fluency in understanding text. Word Reading also shows a moderate relationship with Comprehension, indicating that decoding skills play a key role in supporting comprehension.

Phonemic Awareness is more weakly correlated with Comprehension and ORF, which is consistent with its role as a foundational skill that contributes indirectly to more complex reading tasks. Letter knowledge is moderately associated with Word Reading and to a lesser extent with Comprehension, reinforcing the idea that strong letter-sound knowledge supports early reading development.

These correlations align with the programme's theory of change, that strengthening basic reading skills supports progress towards fluency and comprehension.

Table 9: Pairwise Correlations between EGRA Components

	Phonemic Awareness	Letters	Words	ORF	Comprehension
Phonemic Awareness	1.000	0.402***	0.322***	0.266***	0.280***
Letters		1.000	0.505***	0.351***	0.342***
Words			1.000	0.729***	0.488***
ORF				1.000	0.541***
Comprehension					1.000

6 | Impact Estimation

In this section we present the results of the impact estimation analysis. We start by running Ordinary Least Squares (OLS) multivariate regressions on learning outcomes with the treatment condition (Oshana or Oshikoto) as the variable of interest. We present the coefficient on the treatment assignment variable which represents the estimated impact of the Jolly programme. These results are run on the full sample, prior to applying PSM. The section following the multivariate regression analysis presents the same results, but after applying PSM. Comparing results from both the multivariate regression and PSM methods helps assess the sensitivity of our findings to the analytical approach.

6.1 Multivariate Regressions

Table 8 and Figure 18 present estimates of the treatment effects from the series of multivariate regressions on standardised EGRA outcomes, including two composite EGRA scores and five sub-tests. All outcome variables are expressed as z-scores to allow for comparison across sub-tests with different scales. The models control for a comprehensive set of covariates, as shown in the table, to reduce confounding and improve estimate precision.

These results indicate that the Jolly programmes did not have a statistically significant impact on most EGRA outcomes. Effects of the treatment on EGRA PCA Score (-0.008, p=0.861) and Composite Sum Score (0.010, p=0.826), Phonemic Awareness (0.069 SD, p=0.173), Letter Sounds (-0.014 SD, p=0.786), and Comprehension (-0.047 SD, p=0.323) were not statistically significant.

The results suggest that treatment had a positive effect at the 10% significance level (0.079 SD, p<0.10) on Word Reading, while a negative effect (-0.105 SD, p<0.05) was observed on ORF at the traditional 5% significance level. The negative effect on ORF is an unexpected finding, particularly given the possible positive effect on Word Reading.

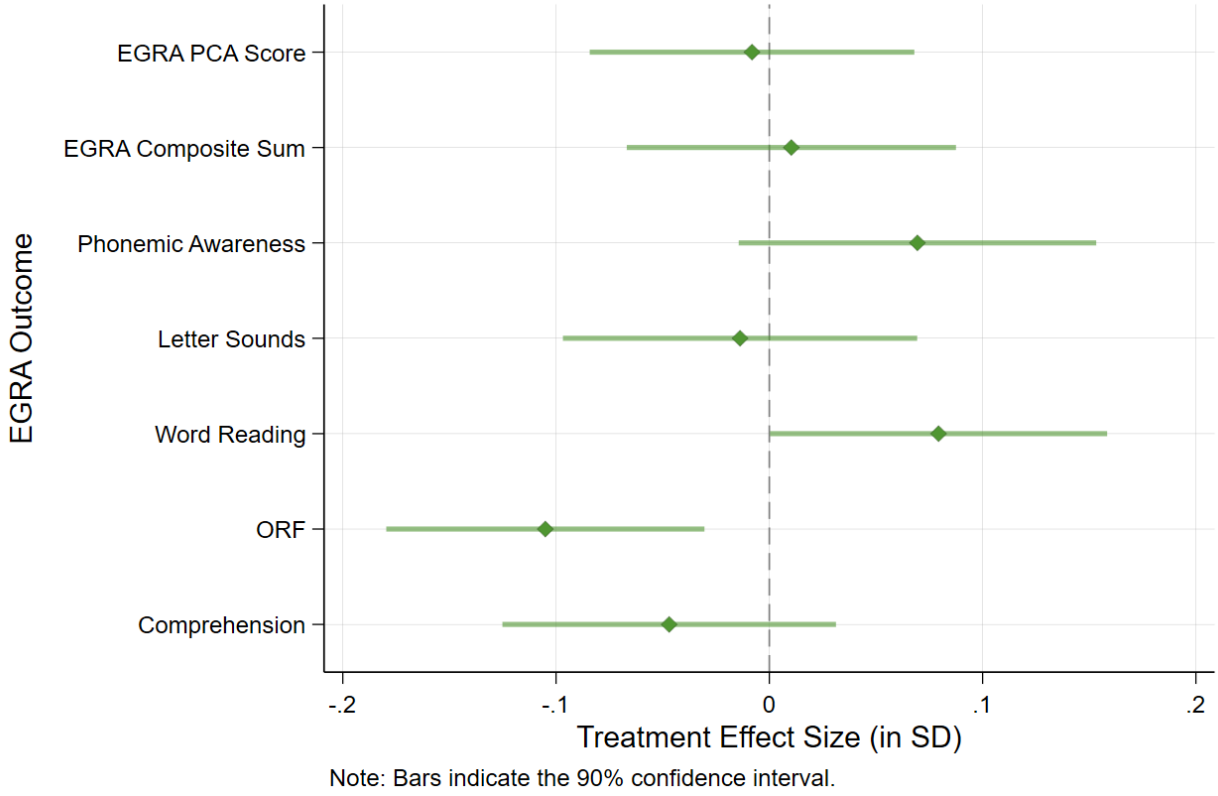
While OLS controls for observable differences between treatment and comparison groups through regression adjustment, it relies heavily on the correct specification of the functional form. To reduce this model dependence and improve covariate balance, we apply PSM as an alternative that pairs treated and comparison units with similar propensity scores. This matching approach allows us to construct a more comparable comparison group and re-estimate the treatment effect in a setting that more closely resembles a randomised experiment.

Table 10: Multivariate Regression Results

Outcome	Estimated Effect Size (in SD)	P-Value
EGRA PCA Score	-0.008	0.861
EGRA Composite Sum	0.010	0.826
Phonemic Awareness	0.069	0.173
Letter Sounds	-0.014	0.786
Word Reading	0.079*	0.099
ORF	-0.105**	0.021

Outcome	Estimated Effect Size (in SD)	P-Value
Comprehension	-0.047	0.323
Note: Significance denoted as *** p<.01, ** p<.05, * p<.1. Controls in each regression model include: distance to regional office, total learners in Grade 4, urban/rural, learner's age, learner's gender, learner's home language, teacher's language, teacher's age, teacher's gender, teacher's years of experience, teacher's training and schooling.		

Figure 17: Graphical Representation of Estimated Treatment Effects from Multivariate Regressions



Note: Bars indicate the 90% confidence interval.

6.2 Propensity Score Matching

6.2.1 Propensity Score Estimation

Following best practice in PSM, only covariates measured prior to treatment assignment, or those that remain fixed over time, were considered in the matching process. This approach helps to avoid the inclusion of variables that may be influenced by participation in the treatment or the anticipation of it, which could compromise the causal interpretation of the analysis[38]. In order to improve the precision of the propensity score model, priority was given to covariates known or hypothesised to be associated with the EGRA scores. Variables associated solely with treatment assignment, but not with the outcome, were excluded to prevent reducing precision.

Where subject-specific knowledge was insufficient to inform variable selection, an empirical method was applied. A series of linear regressions were conducted, regressing individual covariates on the EGRA PCA

[38] Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of economic surveys*, 22(1), 31-72.

score, to identify pre-treatment variables associated with the outcome. Covariates with a p-value of less than 0.10 were retained for inclusion in the propensity score estimation[39].

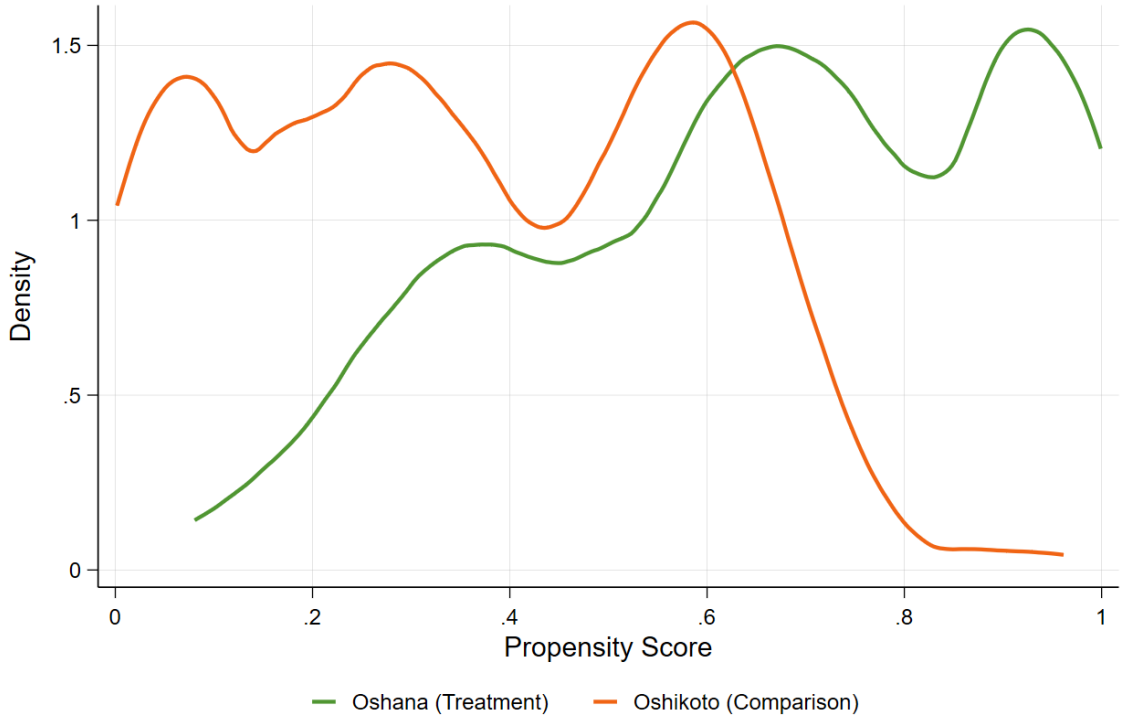
The final set of pre-treatment covariates included the following:

- *School characteristics:* distance to the regional office, urban/rural locality, total number of learners registered in the Grade 4 class
- *Teacher characteristics:* teacher's age, language most often spoken to teach, teacher's highest school year, teacher trained for junior primary, teacher trained for senior primary, years of primary school teaching experience
- *Learner characteristics:* learner's age, learner's gender, learner's home language

Using these covariates, propensity scores were estimated via a probit regression model (mean: 0.503, SD: 0.280). Figure 19 presents the kernel density estimates of the propensity scores for the unmatched sample by treatment status. A notable imbalance is observed between the treatment group (Oshana) and the comparison group (Oshikoto). The treatment group shows a concentration of propensity scores in the higher range (approximately 0.4 to 1), whereas the comparison group displays a greater density of scores in the lower range (approximately 0.05 to 0.6). This limited overlap between the two groups suggests that many learners in the treatment group do not have comparable counterparts in the comparison group prior to matching. If left unaddressed, this could result in biased estimates of the treatment effect.

These differences highlight the need for matching to improve covariate balance and reduce selection bias. Therefore, in the next stage of the analysis, matching techniques were applied to construct a more comparable sample by aligning individuals with similar propensity scores.

Figure 18: Kernel Density Estimates of the Propensity Scores for the Unmatched Sample



[39] Bergstra, S. A., Sepriano, A., Ramiro, S., & Landewé, R. (2019). Three handy tips and a practical guide to improve your propensity score models. *RMD open*, 5(1), e000953. Available at: https://rmdopen.bmj.com/content/5/1/e000953#DC1

6.2.2 Matching Methods

The NN matching method was used to estimate the Average Treatment Effect on the Treated (ATT). We explored several variations of this method, including matching with or without replacement, one-to-one or one-to-many matching, matching only those units with propensity score overlap (known as being on common support), and incorporating caliper restrictions. Results for four models, which vary based on their underlying matching parameters, are presented in the following sections. These four models are outlined in the table below.

Table 11: The PSM Models' Parameters

Model	Stata Command	Nearest Neighbours	Common Support	Replacement	Caliper
Model 1	*teffects*	1 NN	No	Yes	None
Model 2	*teffects*	3 NN	No	Yes	None
Model 3	*psmatch2*	3 NN	Yes	Yes	None
Model 4	*psmatch2*	1 NN	Yes	No	Set to 0.06

6.2.3 Checking the Quality of Matching

Before assessing outcomes, matching quality was tested using balance statistics, standardised biases, and propensity score overlap. Full diagnostic outputs are presented in the Annex, and summarised for each model below.

- **Model 1 (1NN, *teffects*):** Our baseline model employed the *teffects* command with 1-nearest neighbour (1NN) matching. This approach improved balance, with better propensity score overlap and standardised differences for key variables (eg. teacher's age, years of experience, NSSCAS) being near-zero. However, variables such as rural status, language spoken, and class size remained slightly imbalanced with standardised differences exceeding the commonly accepted 10% threshold. Additionally, variance ratios for several variables fell outside the ideal 0.8 to 1.25 range.
- **Model 2: (3NN, *teffects*)**: To address these limitations, we ran a second model using 3-nearest neighbour (3NN) matching. This method allows for the inclusion of a slightly wider set of similar comparison units, which can reduce the variance of treatment effect estimates. Using this method, balance further improved, with better propensity score overlap, stable or improved standardised differences for most covariates, and acceptable variance ratios. This model refines Model 1.
- **Model 3 (3NN with common support, *psmatch2*):** Model 3 applied 3NN matching with a common support condition using the psmatch2 command. Propensity score densities between groups were almost identical. Biases were generally below 10% and variance ratios were within acceptable limits, showing strong covariate balance.
- **Model 4 (1NN with no replacement, common support, and caliper=0.06, *psmatch2*):** stricter matching without replacement, with common support enforced and a caliper of 0.06. The caliper restricts matches to those within a specified propensity score distance, further ensuring high-quality matches. Propensity score overlap improved, although the densities were not as close as in Model 3. Standardised bias and variance ratios confirmed significant reduction in covariate imbalances across most variables. While a few variables retained modest imbalance, their post-matching bias was substantially reduced using this matching algorithm relative to Models 1 - 3.

In summary, across all four models, we observed progressive improvements in covariate balance. Model 4 showed the best overall performance, combining strong overlap, minimal bias, and consistent variance ratios. This sequence of refinements strengthens the credibility of our treatment effect estimates by ensuring that comparisons between treatment and comparison groups are based on well-matched observations.

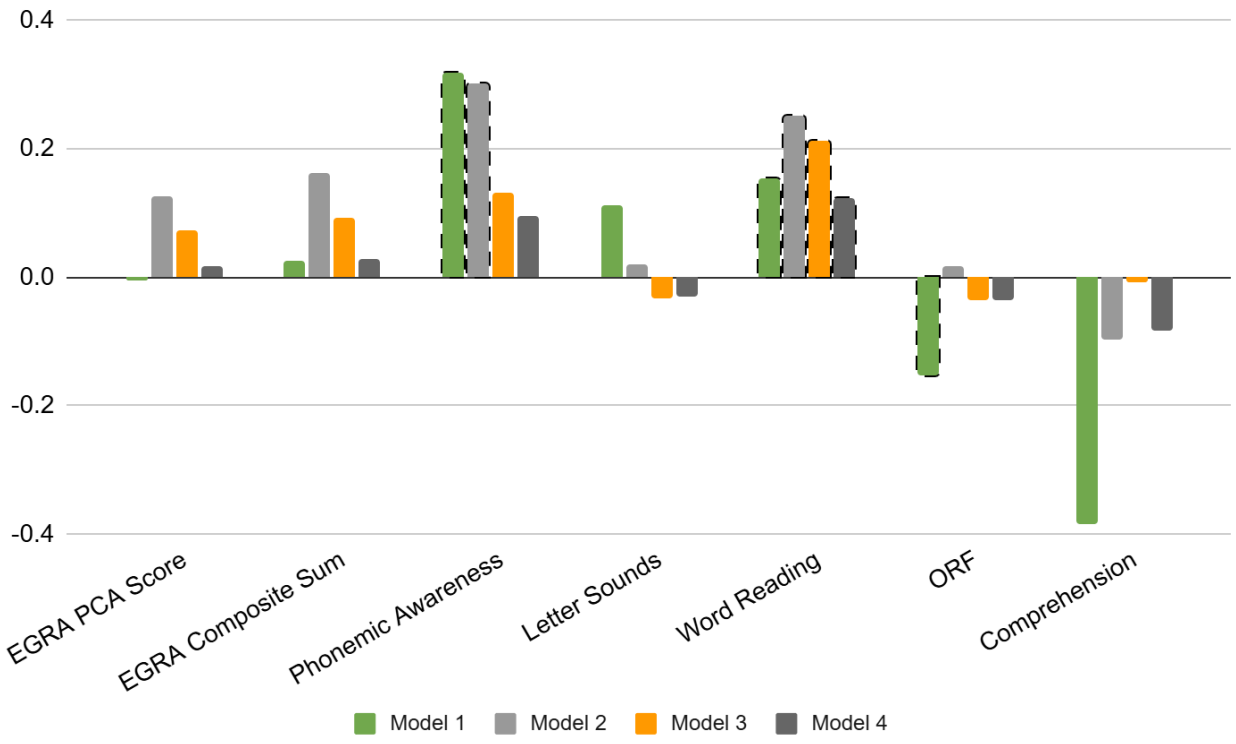
6.2.4 Treatment Effect Estimates

Table 10 and Figure 20 show the ATT estimates across models, expressed in standard deviations. Bars outlined with dashes in the figure reflect treatment effects significant at the 10% level, while the table highlights whether observed impacts were significant at traditional levels.

Across all models, the Jolly programmes consistently show a positive, statistically significant effect on Word Reading, with estimates ranging from 0.122 (Model 4) to 0.250 (Model 2) standard deviations. This provides strong statistical evidence to suggest that the Jolly programmes effectively enhanced the Word Reading skills of learners in the Oshana region.

The estimates of the treatment effect on the other EGRA outcomes are less robust across models. Model 1 and 2 show that treatment had a large and significant positive effect on Phonemic Awareness (0.317 and 0.301, respectively), but this is not the case in Models 3 and 4. Conversely, treatment appears to have an inconsistent effect on Comprehension: Model 1 shows a statistically significant negative effect (-0.153), while the other models estimate small, non-significant negative effects, or a small positive effect (Model 2). For EGRA PCA Score, EGRA Composite Sum, Letter Sounds, and ORF, effects are small, inconsistent across models, and largely not statistically significant.

Figure 19: Graphical Representation of Estimated Treatment Effects from PSM Models



Note: Dashed lines around the bar indicate statistical significance at the 10% level.

Table 12: Estimated Treatment Effects from PSM Models

Outcome	(1)	(2)	(3)	(4)
EGRA PCA Score	-0.006	0.125	0.072	0.018
EGRA Composite Sum	0.025	0.161	0.093	0.029
Phonemic Awareness	**0.317*** **	**0.301*** **	0.131	0.095
Letter Sounds	0.113	0.021	-0.034	-0.029
Word Reading	**0.153** **	**0.250*** **	**0.212** **	**0.122** **
ORF	**-0.153** **	0.018	-0.037	-0.035
Comprehension	-0.385	-0.097	-0.007	-0.084
Stata Command?	teffects	teffects	psmatch2	psmatch2
Nearest Neighbours?	1	3	3	1
Common Support?	No	No	Yes	Yes
With Replacement?	Yes	Yes	Yes	No
Caliper?	None	None	None	0.06
N	2,111	2,111	1,978	1,617

Note: Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1. The standard errors in the psmatch2 models do not take into account that the propensity score is estimated. Where common support is imposed, treatment observations are dropped whose pscore is higher than the maximum or less than the minimum pscore of the comparisons.

The comparison across models highlights the sensitivity of treatment effect estimates to PSM specification. Models 1 and 2, which did not explicitly enforce common support and used *teffects*, generally yielded larger and more frequently significant effects compared to Models 3 and 4, which utilised *psmatch2*, enforced common support, and in Model 4, used no replacement and a caliper. This variation underscores the importance of sensitivity analyses to assess the robustness of findings under different matching algorithms and criteria. The reduced sample size in Model 4 (N=1,617) compared to Model 1 (N=2,111) reflects the trade-off between strict covariate balance and the number of observations retained for analysis.

To further interrogate the robustness of the results, an additional set of matching models is explored in the Annex. The results show that, amongst the models whose matching process improved covariate balance and had strong propensity score overlap, treatment had a positive and significant effect on Word Reading. Where treatment had significant effects on other EGRA outcomes, including Phonemic Awareness and ORF, the covariate balance from matching was usually poor. This sensitivity analysis therefore reinforces the findings presented here on Word Reading (which suggest that the Jolly programmes had a statistically significant and positive effect on learners' Word Reading skills), while refuting any suggestive evidence of an effect on Phonemic Awareness and ORF.

Overall, the findings suggest a targeted positive impact on Word Reading, with less conclusive or no positive effects on other measured outcomes.

7 | Discussion

The analysis of EGRA outcomes presented thus far shows a nuanced picture of the effectiveness of the Jolly Phonics and Grammar programmes. Based on both unconditional means comparisons and more rigorous multivariate and PSM techniques, the programme's impact on learners' reading performance appears limited in scope and largely concentrated on one specific sub-skill: Word Reading.

Unconditional mean comparisons revealed no statistically significant differences between treatment and comparison groups for most EGRA sub-tests, except for the Word Reading sub-test. In the Word Reading sub-test, the treatment group outperformed the comparison group by a modest margin (0.951 cwpm), with this difference in means being statistically significant at the 5% level. Descriptive distributional analyses further suggested that this effect was more pronounced among higher-performing learners, indicating differential impact by learner ability.

Results from multivariate regressions, which controlled for a comprehensive set of learner, teacher, and school-level covariates, largely corroborated the unconditional findings. The Jolly programmes showed a statistically significant, albeit small, positive effect on Word Reading (0.079 SD, $p<0.10$), but a negative and statistically significant effect on ORF (-0.105 SD, $p<0.05$). No significant effects were observed on other outcomes. The negative result on fluency is unexpected and somewhat counterintuitive, given the observed improvement in Word Reading.

PSM provided a more refined estimate of treatment effects, in an attempt to account for potential selection bias in the samples. Across all four matching models, Word Reading consistently showed statistically significant and positive treatment effects, ranging from 0.122 to 0.250 SD. This lends strong support to the hypothesis that the Jolly Grammar training contributed to improved Word Reading skills. However, the inconsistency in results for other EGRA components raises questions about the programme's ability to influence broader literacy outcomes. For Phonemic Awareness, the treatment effect was positive and statistically significant in simpler matching models, but not robust across more stringent specifications. This suggests that the observed effect may be sensitive to matching parameters and not generalisable. Meanwhile, Comprehension outcomes showed no consistent improvement and in some models even declined slightly, hinting at possible trade-offs between decoding skills and comprehension when instructional time or emphasis is not adequately balanced. Fluency once again came out as negative, but inconsistently and less so in more stringently matched models.

An important contextual detail in better understanding the moderate effects sizes is the training that underpinned the Jolly Phonics and Grammar programmes. Although the training was rated highly by participants, many commented that it was too short, with insufficient time to consolidate complex phonics and grammar concepts. This constraint could have influenced the consistency of programme delivery in the classroom and, by extension, learner outcomes. It is thus worth acknowledging that the intervention's mode of delivery may be driving the surprising and inconsistent treatment effects observed in the analysis. Moreover, although teacher enthusiasm and engagement were high, gaps in teachers' understanding of English spelling conventions and grammar rules were noted, especially among those less confident in teaching English. These limitations may have contributed to the moderate-to-null effect sizes observed in the study. Additionally, logistical challenges, such as suboptimal technical equipment and a lack of hard-copy Jolly Learning materials, may have further diluted the programme's impact.

While the study provides useful insights into the effectiveness of the Jolly programmes in Namibia, several design and analytical limitations restrict the robustness and generalisability of the findings. The absence of baseline data hampers the ability to accurately estimate effect sizes and attribute observed gains to the Jolly programmes, as initial differences between groups cannot be accounted for. The non-random treatment and control assignment additionally raises concerns about selection bias due to unmeasured differences between regions, despite attempts to control for observable variables using

regression and PSM. Lastly, a lack of monitoring on the implementation of the Jolly programmes in Oshana between the 2021 teacher training and the 2025 evaluation limits the ability to determine whether any observed effects are due to sustained, high-fidelity delivery of the programmes or other contextual factors.

Taken together, these findings suggest that, although the observed effect on Word Reading is encouraging, it should be viewed as indicative rather than conclusive evidence of programme impact. Future research would benefit from a more rigorous design, including baseline measurement, random assignment where feasible and mixed-method approaches to strengthen the interpretation of quantitative results with qualitative insights.

8 | Recommendations on Effectiveness and Scalability

The Jolly programmes show a clear and consistent effect on learners' Word Reading skills. However, the lack of measurable impact on other reading sub-skills, such as comprehension and fluency, limits the overall effectiveness of the Jolly programmes. This suggests that while the programme's training components may have succeeded in strengthening teachers' focus on decoding instruction, it may not have sufficiently addressed the broader instructional competencies required to develop fluent and comprehending readers. Given the foundational nature of word reading in the reading acquisition process, these gains are meaningful. However, their limited translation into improvements in reading fluency and comprehension indicates a need for better integration of instructional strategies that support progression along the full continuum of reading development.

The Jolly Phonics and Grammar programmes, given their observable impact on a foundational skill, show promise for scale-up, particularly in contexts where learners have weak decoding abilities and where basic instructional quality is uneven. However, scalability should not proceed without refinements. Any expansion of the programme should include enhancements to the teacher training and ongoing teacher support to ensure greater commitment to implementation and a more balanced emphasis across the full range of reading competencies. In particular, strategies for improving oral reading fluency and comprehension should be prioritised. Moreover, the effectiveness of the programme appears to vary across the learner performance distribution, benefiting higher-performing learners more consistently. To ensure equitable impact, future iterations could include differentiated instructional strategies that support struggling readers and mitigate the risk of widening achievement gaps.

Recommendations for Improved Effectiveness and Scaling-Up:

1. **Extend Training Duration:** Future training should be extended to at least three days. This would allow teachers to consolidate key principles and feel more confident in their classroom delivery of the Jolly programmes.
2. **Curriculum Rebalancing:** Revise the training content to include greater focus on fluency and comprehension instruction, in addition to decoding.
3. **Targeted Support:** Include differentiated learning and formative assessment strategies in the training to ensure the Jolly programmes reach learners at all ability levels.
4. **Sustainability Planning and Teacher Support:** Ensure institutional support through ongoing coaching and refresher training, resource provision, and mentoring to reinforce teachers' understanding of the content and provide opportunities for peer learning and classroom-based support. HoDs should be actively involved and incentivised to support teachers, both in the practical implementation of the programme and in the dissemination and monitoring of hard-copy Jolly Learning resources.
5. **Strengthen Monitoring and System-Level Support:** Introduce structured monitoring systems to track implementation fidelity in schools, including classroom use of materials, lesson delivery, and learner engagement. Monitoring should also assess the quality and consistency of support provided by HoDs, advisory teachers, and Ministry officials, to ensure school-level implementation is being actively guided and reinforced.
6. **Further Evaluation:** Embed continuous feedback loops and conduct longitudinal studies to capture longer-term impacts beyond immediate decoding skills. Future studies should further include baseline measures and qualitative data to deepen understanding of the programmes' mechanisms and contextual influences.

Box 4: Recommendations for Further Evaluation

While this evaluation provides important early insights, the findings are also constrained by timing, limited exposure to the 2021 training, and unclear implementation fidelity in schools. We would thus strongly recommend that **the data collected through this study be used as baseline data for a longitudinal evaluation, with a follow-up assessment planned for 2026/27.** This would allow for a more rigorous study design that captures the full effects of the Jolly Grammar programme following its nationwide rollout in 2024 and 2025.

9 | Conclusion

This study sought to evaluate the effectiveness of the Jolly Phonics and Grammar programmes in improving reading outcomes among early grade learners in Namibia. The use of different methods of analysis, including unconditional mean comparisons, multivariate regression, and PSM, was instrumental in testing the robustness of the observed treatment effect.

The findings provide early evidence that the Jolly programmes may support foundational reading skills, particularly word reading, in Namibia, but currently lack the ability to impact reading skills more broadly. Across multiple analytical methods, the Jolly programmes showed a statistically significant positive effect on learners' ability to read words, although the estimated effect sizes vary. No impacts were observed on other sub-tasks, including Phonemic Awareness, ORF, or Comprehension. This is a promising result given the pressing need to improve early grade literacy outcomes in Namibia and the foundational nature of word reading as a literacy skill. However, more work is needed to ensure the programme provides teachers with adequate training and support to improve literacy in other domains.

The analysis also underscores the importance of rigorous evaluation designs. The absence of baseline data combined with non-random treatment assignment means that these findings should be interpreted with caution. Future studies should incorporate baseline assessments and consider randomised implementation where feasible.

Overall, this research contributes meaningfully to the emerging evidence base on interventions aimed at strengthening foundational literacy in Namibia. The study highlights the value of applying multiple analytical techniques to assess programme impact, and the need for ongoing efforts to improve both programme implementation and evaluation rigour. Furthermore, the results highlight the potential of Jolly Phonics and Grammar to strengthen literacy outcomes when supported by adequate training, curriculum alignment, and ongoing professional development. With sustained investment in teacher training and implementation support, this programme has the potential to deliver lasting benefits to learners across the country.

Annexe 1 | Additional Methodology Information

Box A1.1: Matching Algorithms

Nearest Neighbour (NN) Matching

Nearest Neighbour (NN) matching, one of the more commonly used estimators, has each treated individual matched to the comparison individual with the closest propensity score. This can be done with or without replacement.

- When matching **with replacement**, a comparison individual can be used as a match more than once. This typically improves the quality of matches and reduces bias, especially when the distribution of propensity scores differs between treatment and comparison groups. However, it can increase the variance of the estimates, since fewer unique individuals are used to construct the counterfactual.
- Matching **without replacement** avoids reusing comparison individuals but may lead to poorer matches and is sensitive to the order in which observations are matched. To address this, matches should be ordered randomly.

An alternative NN matching approach is to use **multiple nearest neighbours** for each treated individual (i.e. oversampling). This can reduce variance by incorporating more information, but may introduce bias due to weaker matches.

Caliper and Radius Matching

NN matching can result in poor matches when the closest available match has a substantially different propensity score. To address this, caliper matching sets a maximum allowable distance (caliper) between propensity scores, ensuring that only treated and comparison individuals within this range are matched. This improves match quality and reduces bias, but may limit the number of matches, potentially increasing the variance of the estimates. Selecting an appropriate caliper width can be challenging and typically requires judgement or sensitivity testing.

A variation of this approach is radius matching. Instead of selecting only the single nearest neighbour within the caliper, radius matching includes all comparison individuals within the caliper. This allows the number of matched comparisons to vary depending on how many suitable matches are available, combining the benefits of caliper matching and oversampling while reducing the risk of poor matches.

Box A1.2: The Bias-Variance Trade-off in Matching

Bias refers to systematic error, or how far, on average, your estimates are from the true value. When one refers to a highly accurate estimate, this means low bias: the estimate is correct on average i.e. the estimate is close to the true value.

- Increased bias means your estimates are consistently wrong in one direction (either too high or too low).
- It affects the precision of your estimates.
- In matching, bias increases when poor-quality matches are used (e.g., matching treated individuals to comparisons that differ substantially on important variables).

High bias leads to **misleading or inaccurate estimates** of the treatment effect, even if the effect looks precise.

Variance refers to the spread or variability of the estimate across repeated samples. When one refers to a highly precise estimate, this means low variance: estimates are tightly clustered, with narrow confidence intervals.

- Increased variance means your estimates may vary widely across samples, even if the average is correct.
- It affects the stability or reliability of your estimate.
- In matching, variance increases when fewer or duplicated comparisons are used (e.g., with replacement or calipers that exclude many potential matches).

High variance leads to unstable treatment effect estimates, with wider confidence intervals and **less certainty or precision.**

The following bias-variance trade-off exists when conducting propensity score matching:

- Allowing more (poorer) matches → reduces variance (more data points leading to greater precision) but increases bias (the estimates are of worse quality or are less accurate).
- Restricting matches to better ones (e.g., with calipers or tight criteria) → reduces bias (better quality or more accurate estimates) but increases variance (fewer data points leading to less certainty/precision).

Annexe 2 | Descriptive Statistics

Table A2.1: Detailed Descriptive Statistics for each EGRA Sub-test

Full Sample					
Outcome	N	Mean	SD	Min	Max
Composite PCA	2189	0.000	1.652	-3.716	7.053
Composite Sum	2189	1.716	0.736	0.000	4.450
Phonemic Awareness	2189	6.413	3.089	0.000	10.000
Letter Sounds	2189	22.436	12.104	0.000	100.000
Word Reading	2189	14.263	9.986	0.000	54.545
ORF	2189	45.832	24.307	0.000	175.556
Comprehension	2189	1.637	1.297	0.000	5.000
Treatment					
Outcome	N	Mean	SD	Min	Max
Composite PCA	1111	0.018	1.722	-3.716	7.053
Composite Sum	1111	1.730	0.764	0.000	4.450
Phonemic Awareness	1111	6.484	3.040	0.000	10.000
Letter Sounds	1111	22.321	12.510	0.000	100.000
Word Reading	1111	14.732	10.626	0.000	54.545
ORF	1111	45.008	24.667	0.000	175.556
Comprehension	1111	1.657	1.364	0.000	5.000
comparison					
Outcome	N	Mean	SD	Min	Max
Composite PCA	1078	-0.019	1.577	-3.709	5.592
Composite Sum	1078	1.701	0.706	0.002	4.019
Phonemic Awareness	1078	6.340	3.138	0.000	10.000
Letter Sounds	1078	22.555	11.676	0.000	67.000
Word Reading	1078	13.781	9.259	0.000	52.222
ORF	1078	46.681	23.912	0.000	139.412
Comprehension	1078	1.617	1.224	0.000	5.000

Figure A2.1: Phonemic Awareness Scores by Treatment and Comparison Group

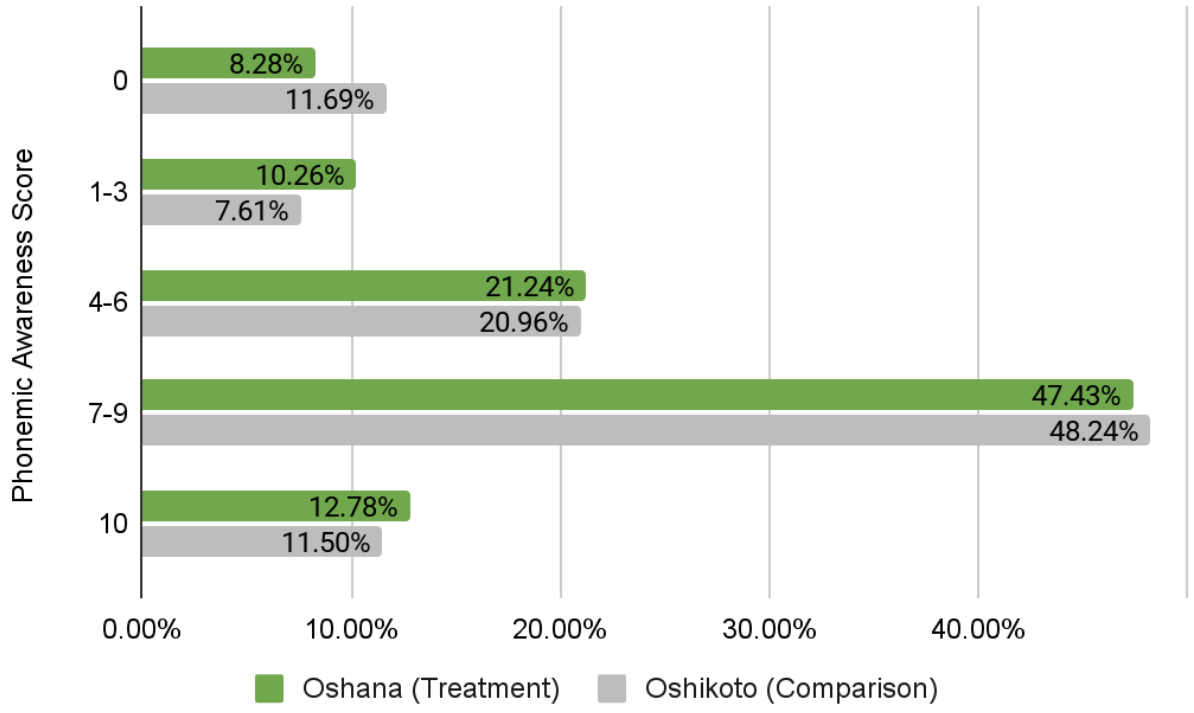


Figure A2.2. Letter Reading by Treatment and Comparison Group

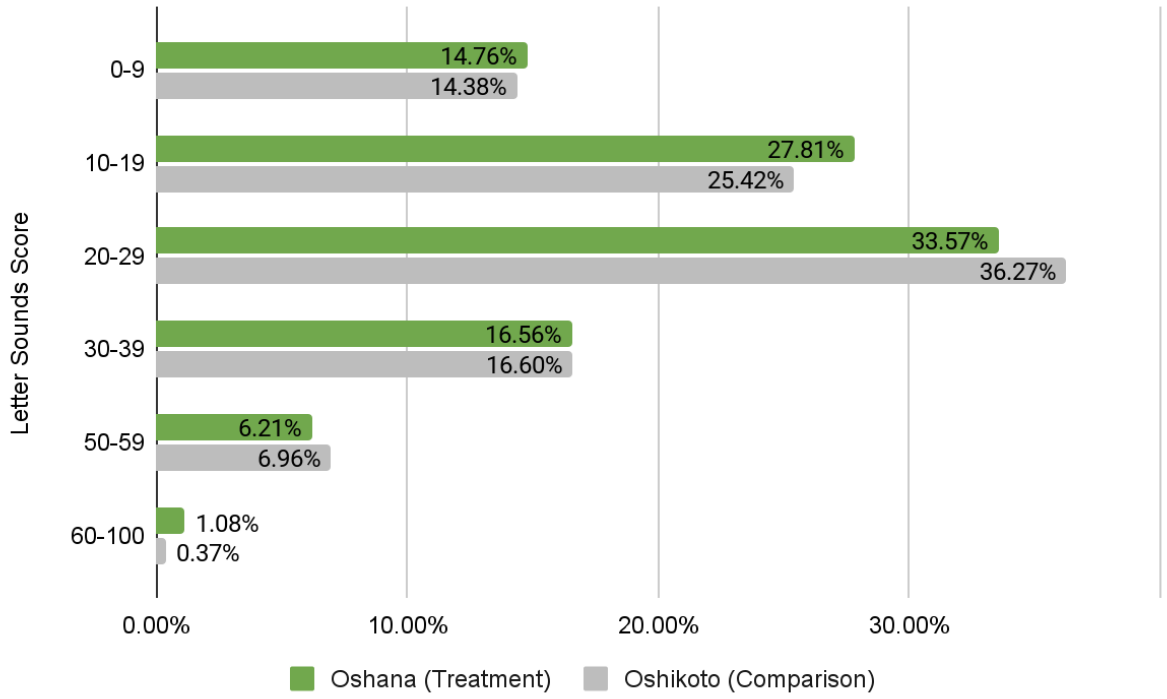


Figure A2.3: Word Reading by Treatment and Comparison Group

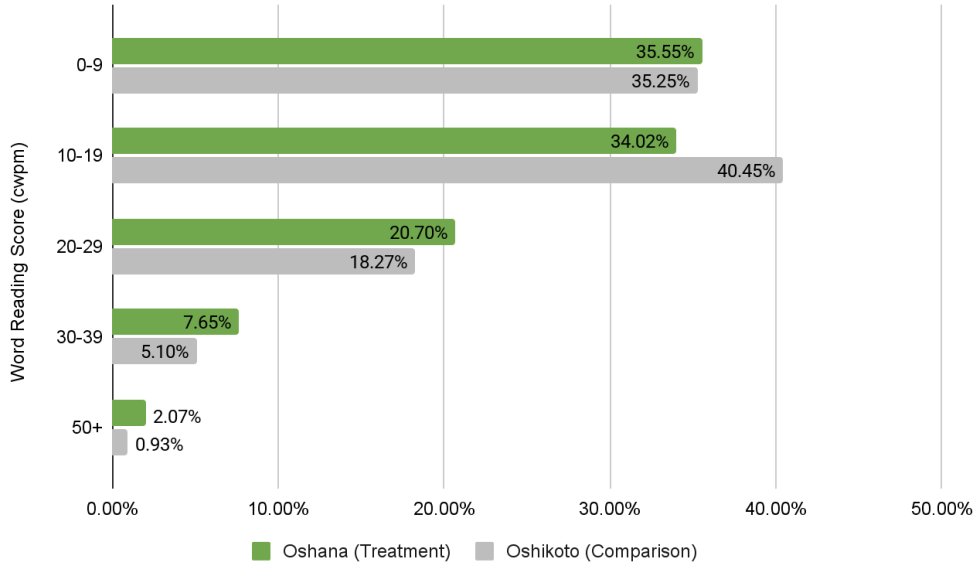
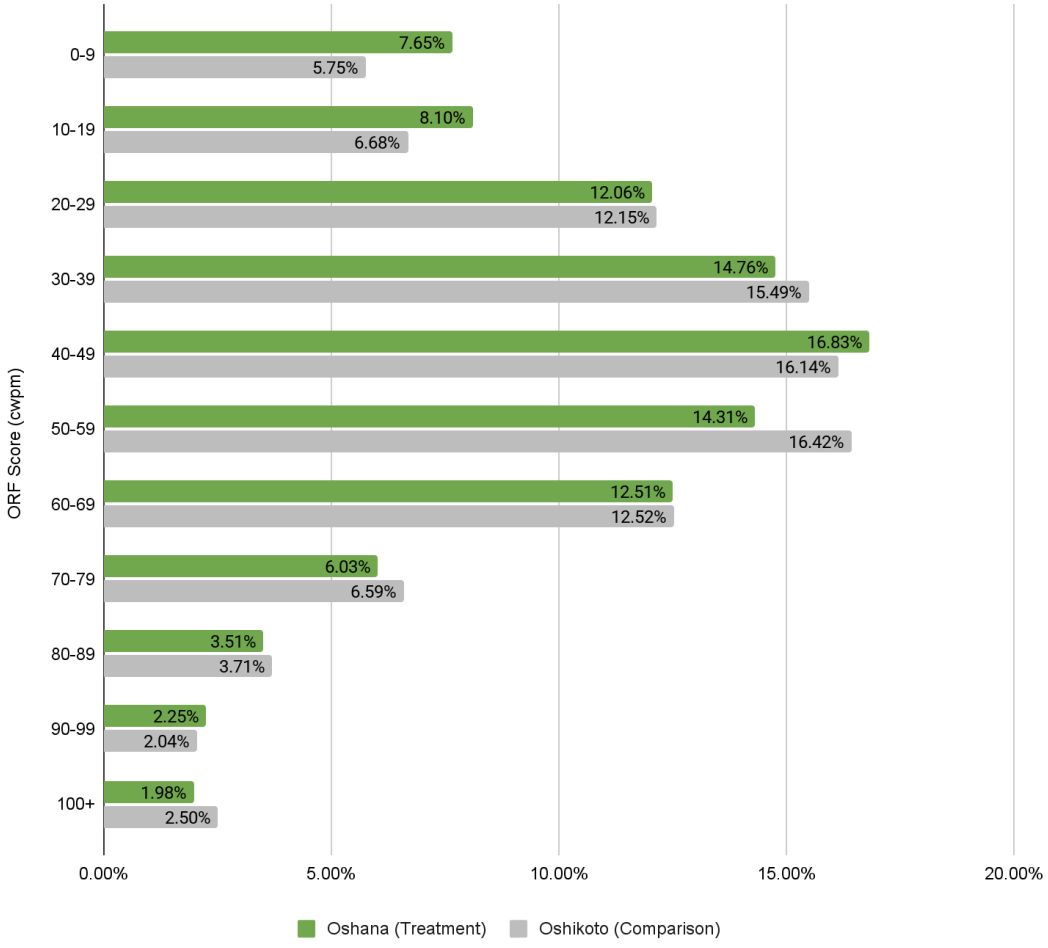


Figure A2.4: ORF by Treatment and Comparison Group



Annexe 3 | Regression Results

Table A3.1: Multivariate Regression Results

	EGRA PCA Score		EGRA Composite Sum		Phonemic Awareness		Letter Sounds		Word Reading		ORF		Comprehension	
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
Treatment	-0.008		0.010		0.069		-0.014		0.079	*	-0.105	**	-0.047	
School Characteristics														
Distance to Regional Office	-0.006	***	-0.005	***	0.000		0.000		-0.003		-0.007	***	-0.009	***
Rural	0.005	***	0.004	***	0.002		0.002	*	0.005	***	0.005	***	0.002	*
Number of Learners	-0.729	***	-0.716	***	-0.206		-0.328	**	-0.374	**	-0.764	***	-0.963	***
Learner Characteristics														
Learner's age	-0.295	***	-0.276	***	-0.136	***	-0.156	***	-0.308	***	-0.268	***	-0.189	***
Learner Gender	0.326	***	0.283	***	0.132	***	0.205	***	0.234	***	0.450	***	0.146	***
Learner Language: Oshikwanyama (base: Eng)	-0.522	**	-0.416	*	0.066		-0.508	**	-0.102		-0.773	***	-0.538	**
Learner Language: Oshindonga (base: Eng)	-0.609	***	-0.525	**	-0.003		-0.665	***	-0.154		-0.68	***	-0.704	***
Learner Language: Other (base: Eng)	-0.476	*	-0.371		0.144		-0.442		-0.155		-0.653	**	-0.561	**
Teacher Characteristics														
Teaching language: Oshikwanyama (base: English)^	-0.298	***	-0.340	***	-0.242	**	-0.044		-0.173		-0.254	**	-0.400	***
Teaching language: Oshindonga (base: English)^	0.117		0.110		0.072		0.243	**	0.085		0.001		0.053	
Teacher completed NSSCAS (base: Gr 12)^	0.103	*	0.105	*	0.066		0.041		0.109	*	0.077		0.081	
Trained for Junior Primary^	0.162	*	0.181	**	0.120		0.083		0.002		0.173	**	0.239	***

Trained for Senior Primary^	0.423	***	0.406	***	0.280	***	0.466	***	0.301	***	0.308	***	0.226	**
Years of Experience^	0.017	***	0.015	***	0.003		0.011	**	0.013	***	0.015	***	0.018	***
Teacher's Age^	-0.026	***	-0.026	***	-0.014	***	-0.017	***	-0.017	***	-0.021	***	-0.027	***
Intercept	4.601	***	4.360	***	1.707	***	2.550	***	3.722	***	4.409	***	4.171	***
Number of observations	2111		2111		2111		2111		2111		2111		2111	
Adjusted R-squared	0.2		0.18		0.04		0.06		0.13		0.23		0.16	
Note: Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1. Variables that are selected are denoted with a "^".														

Table A3.2: Multivariate Regression Results for Selecting the Matching Variables

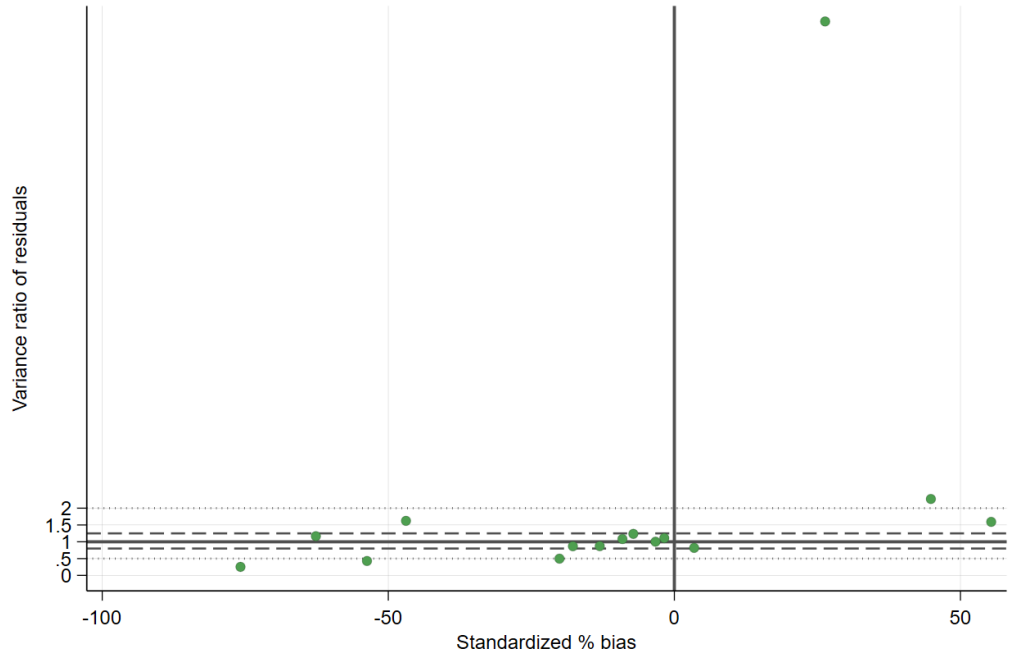
	EGRA PCA Score		EGRA Composite Sum		Phonemic Awareness		Letter Sounds		Word Reading		ORF		Comprehension	
	(1)		(2)		(3)		(4)		(5)		(6)		(7)	
School Characteristics														
Distance to Regional Office^	-0.005	***	-0.005	***	-0.001		0.000		-0.003	**	-0.006	***	-0.008	***
Rural^	-0.791	***	-0.777	***	-0.251		-0.390	**	-0.386	**	-0.838	***	-1.006	***
Number of Learners^	0.004	***	0.004	***	0.001		0.002		0.005	***	0.004	***	0.002	
Teacher Characteristics														
Teacher's Age^	-0.027	***	-0.027	***	-0.015	***	-0.018	***	-0.017	***	-0.022	***	-0.027	***
Teacher's Gender	-0.075		-0.079	*	-0.077		-0.071		-0.041		-0.058		-0.039	
Teaching language: Oshikwanyama (base: English)^	-0.292	***	-0.335	***	-0.251	**	-0.056		-0.151		-0.254	**	-0.383	***
Teaching language: Oshindonga	0.127		0.122		0.085		0.248	**	0.103		-0.002		0.058	

(base: English)^													
Teacher completed NSSCAS (base: Gr 12)^	0.097	*	0.090		0.022		0.028		0.083		0.106	*	0.101	*
Trained for Early Childhood Development	0.073		0.082		0.182		0.213		-0.131		0.127		-0.058	
Trained for Junior Primary^	0.187	**	0.210	**	0.172	*	0.135		-0.012		0.200	**	0.231	**
Trained for Senior Primary^	0.447	***	0.434	***	0.329	***	0.510	***	0.296	***	0.328	***	0.219	**
Years of Experience^	0.018	***	0.017	***	0.004		0.012	***	0.013	***	0.017	***	0.019	***
Learner Characteristics														
Learner Gender^	0.325	***	0.281	***	0.129	***	0.204	***	0.230	***	0.453	***	0.146	***
Learner Language: Oshikwanyama (base: English)^	-0.511	**	-0.398	*	0.111		-0.490	**	-0.080		-0.792	***	-0.553	**
Learner Language: Oshindonga (base: English)^	-0.603	***	-0.518	**	0.007		-0.654	***	-0.157		-0.672	***	-0.704	***
Learner Language: Other (base: English)^	-0.449	*	-0.334		0.215		-0.403		-0.125		-0.664	***	-0.575	**
Learner's age^	-0.294	***	-0.276	***	-0.137	***	-0.154	***	-0.313	***	-0.262	***	-0.187	***
Intercept	4.687	***	4.467	***	1.844	***	2.590	***	3.894	***	4.363	***	4.198	***
Number of observations	2111		2111		2111		2111		2111		2111		2111	
Adjusted R-squared	0.2		0.18		0.04		0.06		0.13		0.22		0.16	
Note: Significance denoted as: *** p<0.01, ** p<0.05, * p<0.1. Variables that are selected are denoted with a "^".														

Annexe 4 | Covariate Balance Checks

A4.1 Unmatched Sample's Covariate Balance

Figure A4.1: Unmatched Sample's Variance Ratio of Residuals vs. Standardized % Bias



A4.2 Matching Quality Checks for Models 1 - 4

Model 1: *teffects*, 1NN

Figure A4.2: PS Density Graphs for Model 1

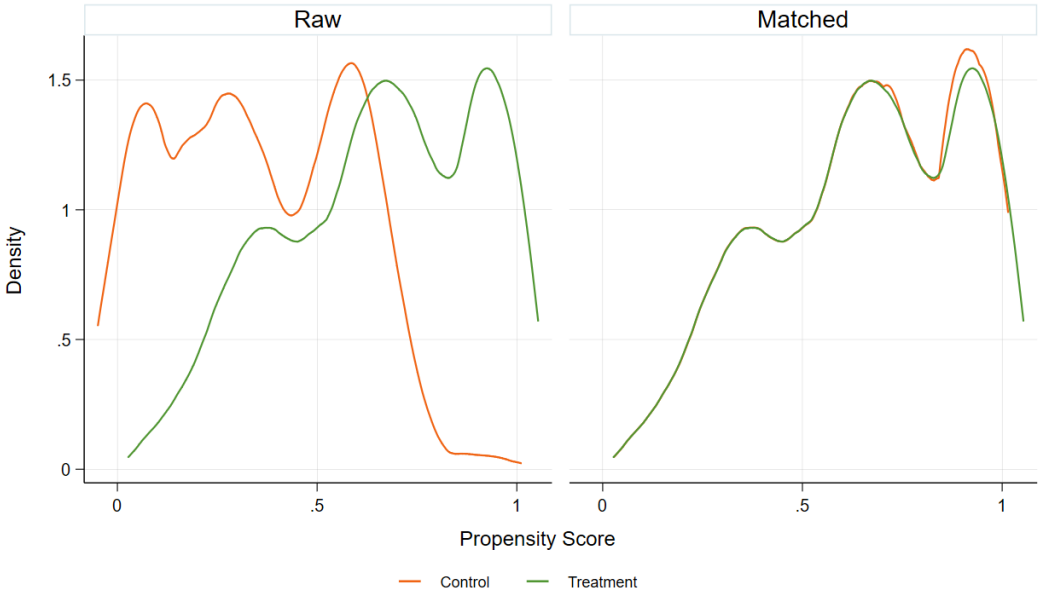


Table A4.1: Model 1 Balance Checks

	Standardised Difference		Variance Ratio	
Covariates	Raw	Matched	Raw	Matched
Distance to Regional Office	-0.762	-0.290	0.374	1.131
Rural	-0.024	-0.186	1.116	3.355
Number of Learners	0.059	0.129	0.843	2.091
Teacher's Age	-0.130	0.043	1.173	1.331
Teacher language: Oshikwanyama (base: Eng)	0.412	0.416	3.182	3.252
Teacher language: Oshindonga (base: Eng)	-0.440	0.134	2.121	0.891
Teacher completed NSSCAS (base: Gr 12)	-0.533	0.038	0.294	1.166
Trained for Junior Primary	-0.072	-0.197	1.246	2.045
Trained for Senior Primary	-0.201	0.053	0.496	1.283
Years of Experience	-0.176	0.020	0.933	1.089
Learner Gender	-0.054	-0.212	1.003	1.047
Learner: Oshikwanyama (base: Eng)	0.540	-0.238	2.613	0.857
Learner: Oshindonga (base: Eng)	-0.610	0.232	2.574	0.902
Learner: Other Language (base: Eng)	0.256	0.123	18.763	2.123
Learner's age	-0.193	0.107	0.859	0.868

Model 2: *teffects*, 3NN

Figure A4.2: PS Density Graphs for Model 2

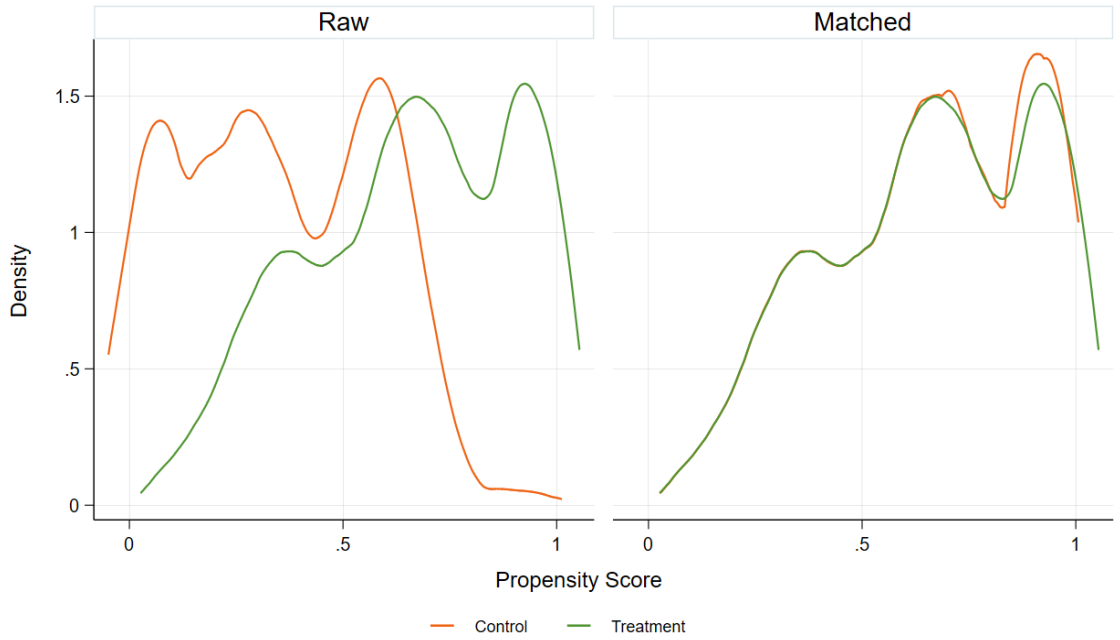


Table A4.2: Model 2 Balance Checks

	Standardised Difference		Variance Ratio	
Covariates	Raw	Matched	Raw	Matched
Distance to Regional Office	-0.762	-0.401	0.374	1.078
Rural	-0.024	-0.160	1.116	2.626
Number of Learners	0.059	0.014	0.843	1.701
Teacher's Age	-0.130	-0.106	1.173	1.247
Teacher language: Oshikwanyama (base: Eng)	0.412	0.441	3.182	3.677
Teacher language: Oshindonga (base: Eng)	-0.440	-0.128	2.121	1.164
Teacher completed NSSCAS (base: Gr 12)	-0.533	0.015	0.294	1.061
Trained for Junior Primary	-0.072	-0.194	1.246	2.016
Trained for Senior Primary	-0.201	0.063	0.496	1.354
Years of Experience	-0.176	-0.151	0.933	1.006
Learner Gender	-0.054	-0.020	1.003	1.000
Learner: Oshikwanyama (base: Eng)	0.540	-0.120	2.613	0.910
Learner: Oshindonga (base: Eng)	-0.610	0.218	2.574	0.905
Learner: Other Language (base: Eng)	0.256	-0.175	18.763	0.498
Learner's age	-0.193	-0.027	0.859	0.838

Model 3: *psmatch2*, 3NN, common support

Figure A4.4: PS Density Graph for Model 3

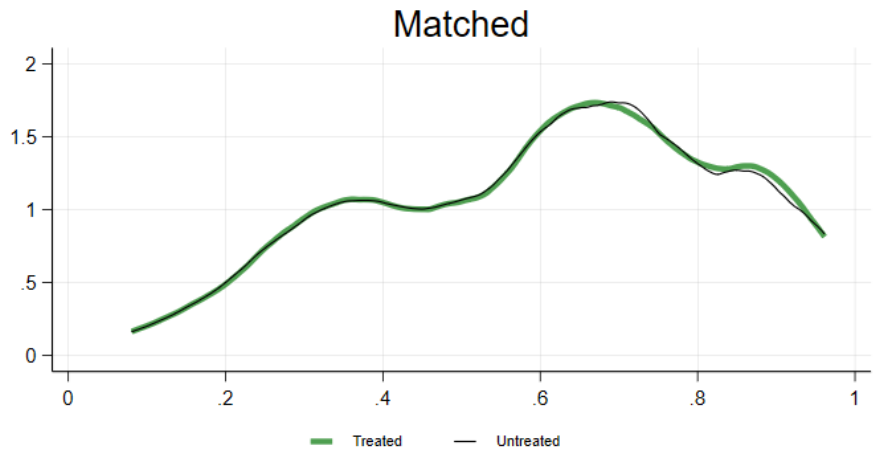


Figure A4.5: Model 1's Standardised % Bias across Covariates

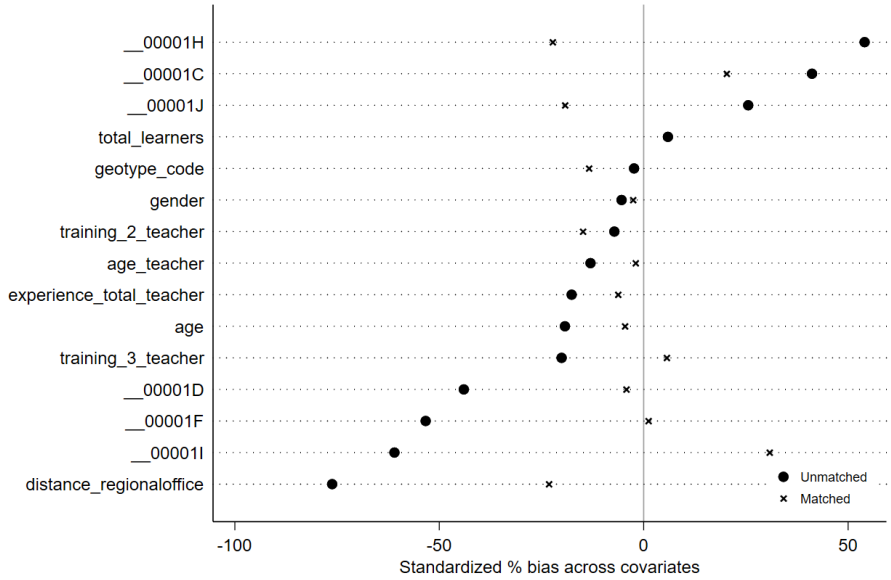
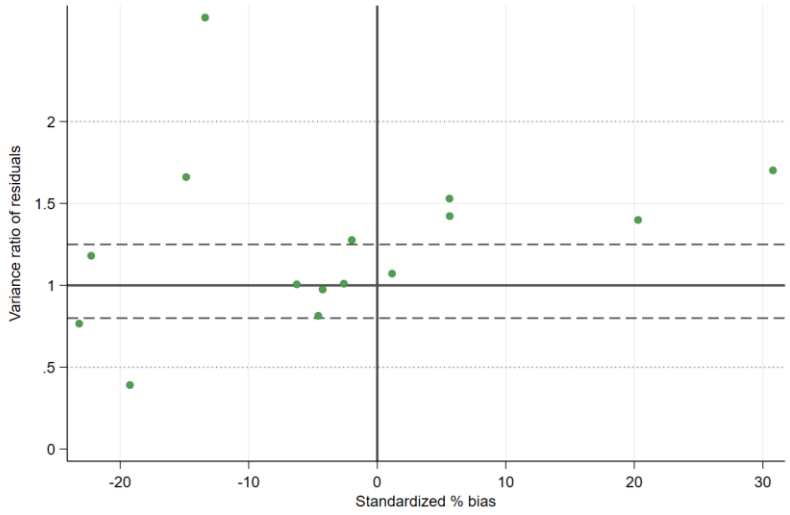


Figure A4.6: Model 3's Variance Ratio of Residuals vs. Standardized % Bias



Model 4: *psmatch2*, 1NN, common support, no replacement and caliper

Figure A4.7: PS Density Graph for Model 4

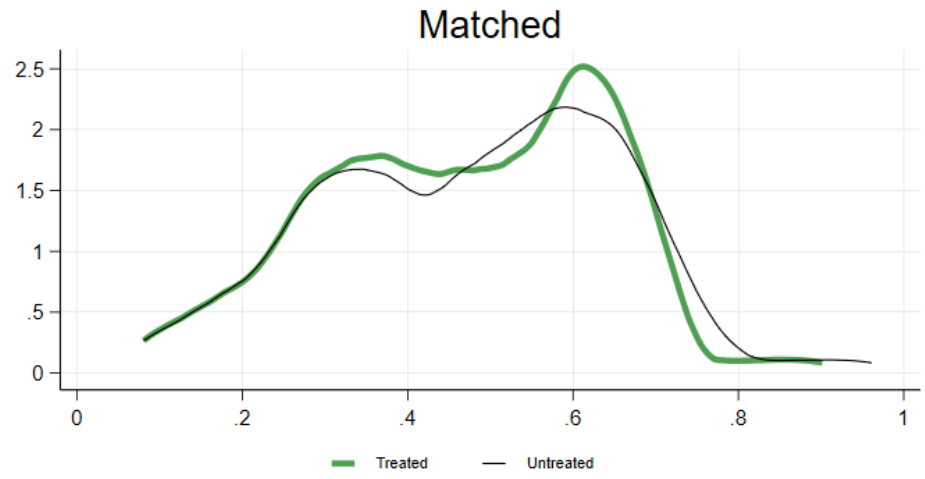


Figure A4.8: Model 4's Standardised % Bias across Covariates

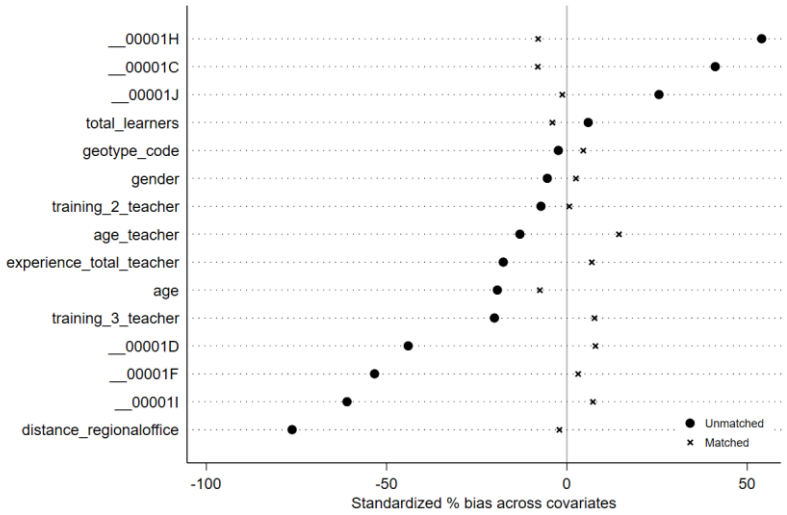
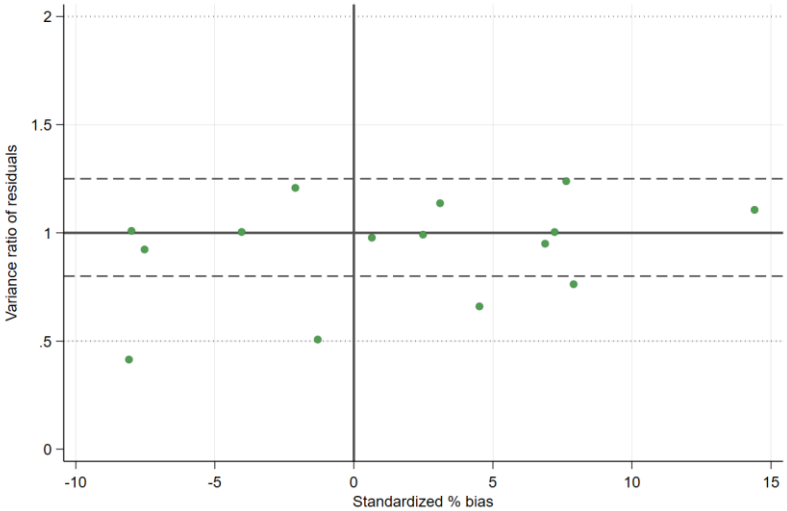


Figure A4.9: Model 3's Variance Ratio of Residuals vs. Standardized % Bias



Annexe 5 | Sensitivity Analysis

The covariate balance and propensity score overlap checks are presented for Models 5 - 8 below. Model 7, while having similar propensity score overlap to Models 5 and 6, does not demonstrate the same improvements in covariate balance as that in Model 5 and 6. Specifically, Model 7 has larger variance ratios of the covariates' residuals. Model 8 shows particularly weak propensity score overlap and little improvement to the covariate balance statistics. **These results suggest that Models 5 and 6, which both use five nearest neighbours during matching, produced a sample most similar in terms of covariates, and thus most reliable for estimating treatment effects.**

A5.1 Matching Quality Checks

Model 5: teffects, 5NN

Figure A5.1.10: PS Density Graphs for Model 5

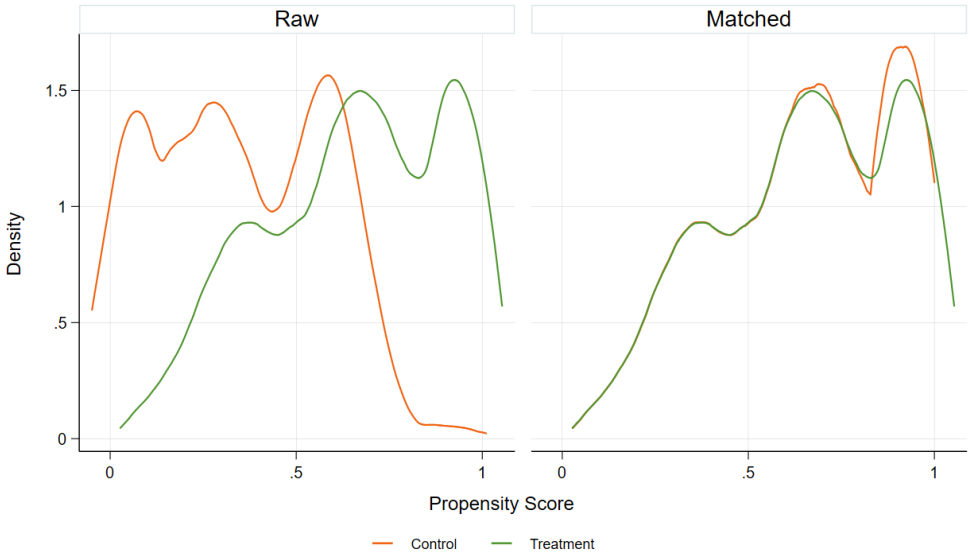


Table A5.1.3: Model 5 Balance Checks

	Standardised Difference		Variance Ratio	
Covariate	Raw	Matched	Raw	Matched
Distance to Regional Office	-0.762	-0.429	0.374	1.189
Rural	-0.024	-0.161	1.116	2.647
Number of Learners	0.059	0.106	0.843	1.710
Teacher's Age	-0.130	-0.057	1.173	1.271
Teacher language: Oshikwanyama (base: Eng)	0.412	0.442	3.182	3.695
Teacher language: Oshindonga (base: Eng)	-0.440	-0.176	2.121	1.248
Teacher completed NSSCAS (base: Gr 12)	-0.533	-0.003	0.294	0.990
Trained for Junior Primary	-0.072	-0.195	1.246	2.023
Trained for Senior Primary	-0.201	0.073	0.496	1.429
Years of Experience	-0.176	-0.089	0.933	1.032
Learner Gender	-0.054	0.034	1.003	1.001
Learner: Oshikwanyama (base: Eng)	0.540	-0.155	2.613	0.892
Learner: Oshindonga (base: Eng)	-0.610	0.211	2.574	0.906
Learner: Other Language (base: Eng)	0.256	-0.105	18.763	0.636
Learner's age	-0.193	-0.044	0.859	0.866

Model 6: psmatch2, 5NN, common support

Figure A5.1.11: PS Density Graph for Model 6

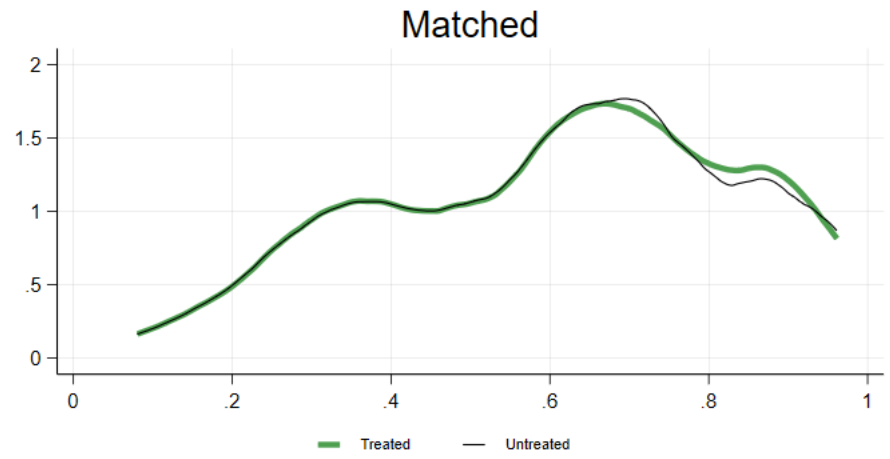


Figure A5.1.12: Model 6's Variance Ratio of Residuals vs. Standardized % Bias

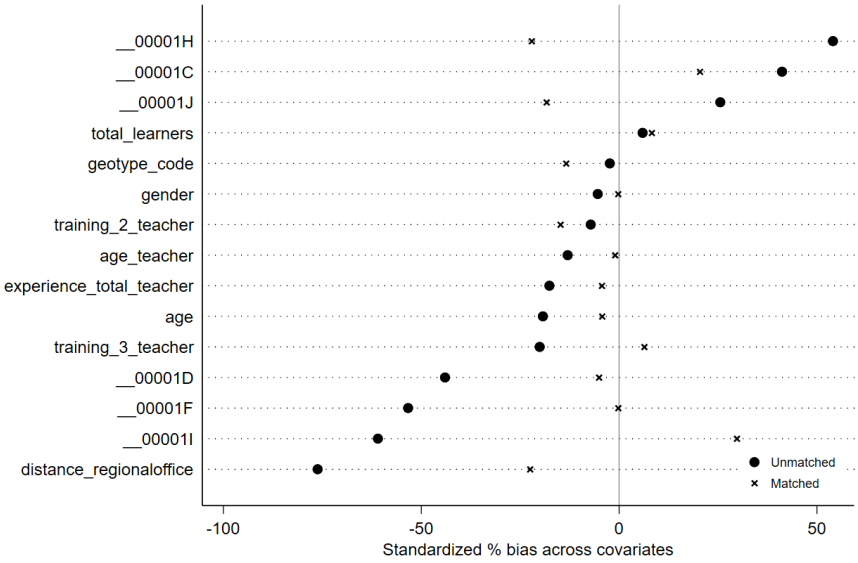
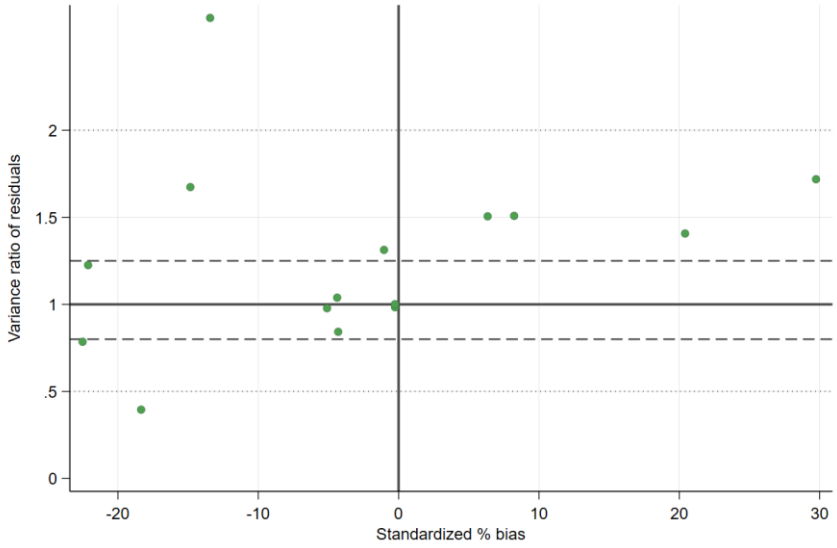


Figure A5.1.13: Model 6's Standardised % Bias across Covariates



Model 7: psmatch2, 1NN, common support

Figure A5.1.14: PS Density Graph for Model 7

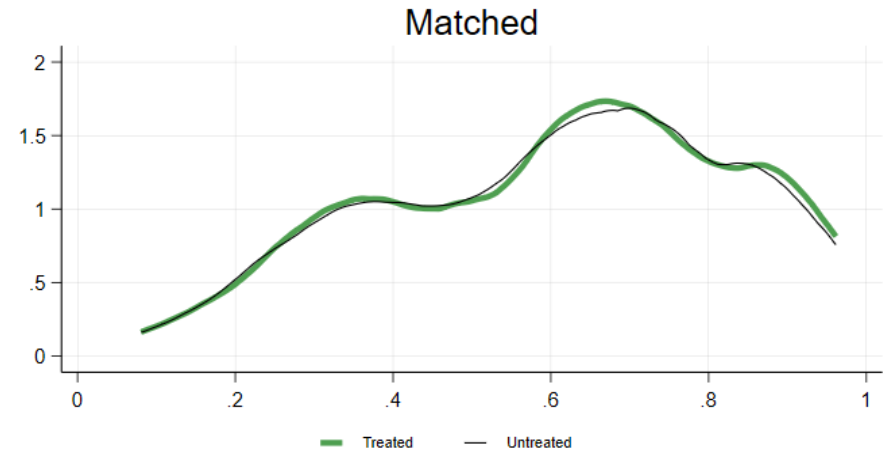


Figure A5.1.15: Model 7's Variance Ratio of Residuals vs. Standardized % Bias

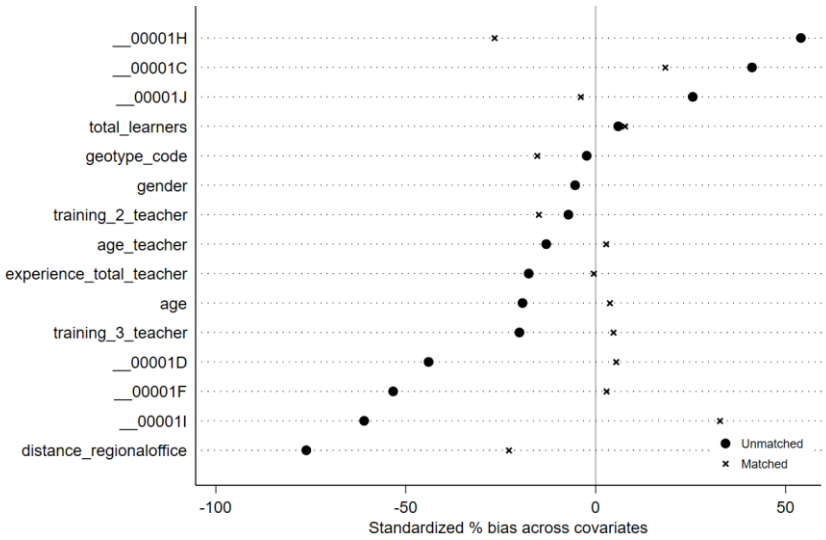
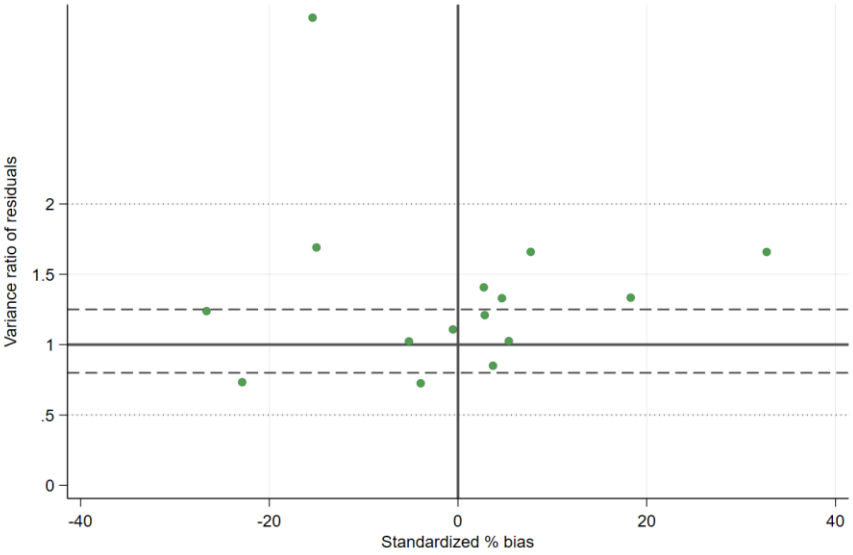


Figure A5.1.16: Model 7's Standardised % Bias across Covariates



Model 8: psmatch2, 1NN, common support, no replacement

Figure A5.1.17: PS Density Graph for Model 8

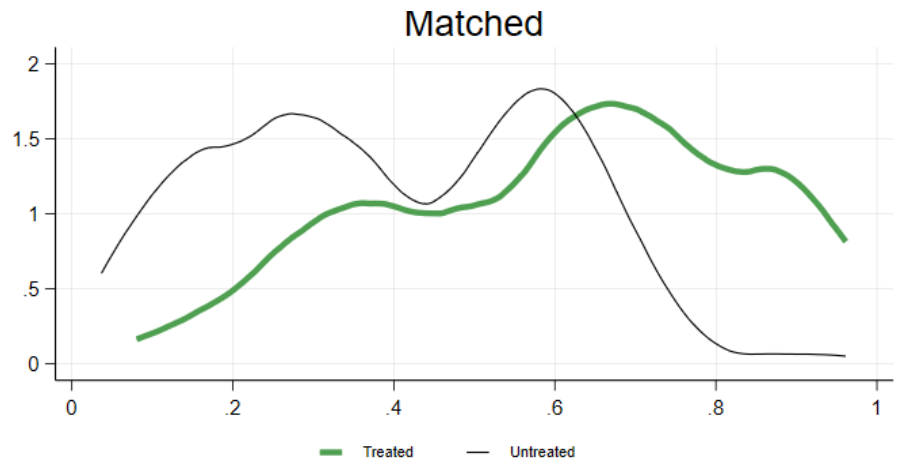


Figure A5.1.18: Model 8's Variance Ratio of Residuals vs. Standardized % Bias

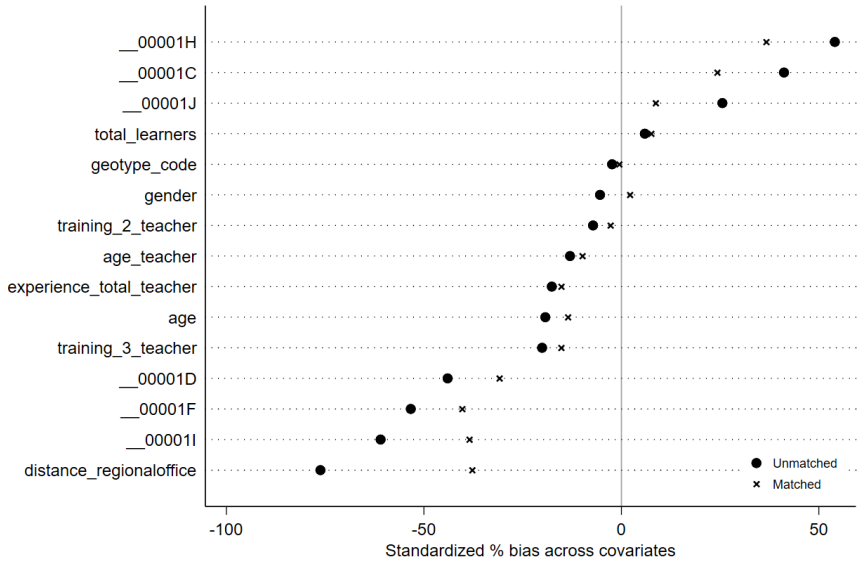
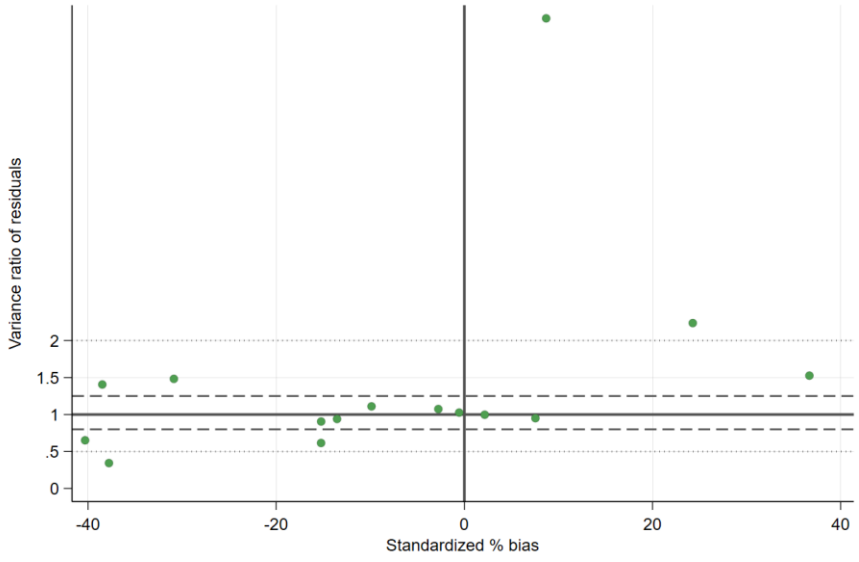


Figure A5.1.19: Model 8's Standardised % Bias across Covariates



A5.2 PSM Model Results

The results for all eight permutations of the matching approach are presented in the table below. Given the covariate balance checks, Model 5 and 6's estimates are likely the most reliable. These models find that treatment has a positive and significant effect on Word Reading scores, of 0.216 and 0.191 standard deviations, respectively. This result corroborates the findings presented in the main analysis.

Table A5.2.1: PSM Sensitivity Analysis Results

Outcome	(1)°	(2)°	(3)°	(4)°	(5)	(6)	(7)	(8)
EGRA PCA Score	-0.006	0.125	0.072	0.018	0.074	0.057	-0.013	0.012
EGRA Composite Sum	0.025	0.161	0.093	0.029	0.088	0.077	0.022	0.031
Phonemic Awareness	0.317***	0.301***	0.131	0.095	0.167	0.119	0.200*	0.052
Letter Sounds	0.113	0.021	-0.034	-0.029	-0.009	-0.038	0.003	-0.013
Word Reading	0.153**	0.250***	0.212**	0.122**	0.216***	0.191**	0.152	0.091**
ORF	-0.153**	0.018	-0.037	-0.035	0.001	-0.048	-0.176*	-0.098**
Comprehension	-0.385	-0.097	-0.007	-0.084	-0.098	-0.014	-0.189*	0.021
Stata Command?	*teffects*	*teffects*	*psmatch2*	*psmatch2*	*teffects*	*psmatch2*	*psmatch2*	*psmatch2*
Nearest Neighbours?	1	3	3	1	5	5	1	1
Common Support?	No	No	Yes	Yes	No	Yes	Yes	Yes
With Replacement?	Yes	Yes	Yes	No	Yes	Yes	Yes	No
Caliper?	None	None	None	0.06	None	None	None	None
N	2,111	2,111	1,978	1,617	2,111	1,978	1,978	1,978

Note: ° denotes whether the model is one of the main models presented in Section 6.2.4. The standard errors in the psmatch2 models do not take into account that the propensity score is estimated.

Annexe 5 | Permission Letter

REPUBLIC OF NAMIBIA

MINISTRY OF EDUCATION, ARTS AND CULTURE

Tel: *(061) 226 979*
Enquiries: *Ms. R Bosch*
Email: *Ronel.Bosch@moe.gov.na*
Ref 12/2/1

Luther Street, Govt. Office Park
P/Bag 13186, WINDHOEK

TO: **THE CHIEF REGIONAL OFFICERS**
OSHANA AND OSHIKOTO REGIONS

ATTENTION: **DIRECTORS OF EDUCATION, ARTS AND CULTURE**

Dear Directors,

SUBJECT: **PERMISSION GRANTED TO IMPLEMENT A PROPENSITY SCORE MATCHING IMPACT EVALUATION OF THE JOLLY PHONICS INITIATIVE IN THE OSHANA AND OSHIKOTO REGIONS.**

The Ministry of Education, Arts and Culture (MoEAC) is dedicated to making decisions based on evidence and has made Foundational Learning a major national goal. In line with the resolutions of the 2022 National Conference, the Namibian government has initiated a nationwide training programme for teachers of Grades 1 to 3, Junior Primary Heads of Departments, and Senior Educational Officers in the application of the Jolly Phonics/ Jolly Grammar method. This programme is highly respected on a global scale and is being adopted by numerous education systems, both in developed and developing countries. The methodology aligns with the most recent findings in the science of reading and play-based learning. The initiative has been met with great enthusiasm within Namibia. However, stakeholders have expressed concerns regarding the lack of independent evidence for the effectiveness of the intervention in Namibia.

In this context, the Ministry has authorised the implementation of an independent impact evaluation to determine the effect of the Jolly Phonics/ Jolly Grammar programme in Namibia. The Ministry, with continued support from UNICEF and in collaboration with the Bill and Melinda Gates Foundation, is committed to the evaluation process.

The impact evaluation will utilise a propensity score matching method. Capitalising on the earlier implementation of Jolly Phonics in the Oshana region, the evaluation will measure letter knowledge, as well as word, sentence, and paragraph reading and comprehension, using the Early Grade Reading Assessment (EGRA) instrument. This method will facilitate a valid and reliable comparison of reading outcomes between schools with extended exposure to Jolly Phonics/ Jolly Grammar and those in another region that uses the same medium of instruction and where Grade 4 learners have not yet been exposed to it.

The evaluation will test a random sample of 20 Grade 4 learners in each of the 100 schools, with 50 schools from the Oshana region and 50 comparable schools from the Oshikoto region. The Oshikoto region was selected due to the identical medium of instruction (Oshindonga and Oshikwanyama). Testing is scheduled to begin on 27 January 2025 and will last for approximately three weeks, employing the one-on-one EGRA instrument, which will be administered by trained graduates from the University of Namibia (UNAM) and overseen by impact evaluation specialists. Genesis Analytics (Pty) Ltd, in partnership with the UNAM Faculty of Education and Human Sciences, Department of Early Childhood Education and Care, have been contracted as impact evaluation specialists.

Regional Offices will receive notifications about the selected schools early in January 2025. Each region should designate a focal person, ideally a Senior Education Officer (SEO) for Senior Primary, to facilitate coordination. Please submit these names to Ms Ronel Bosch at Ronel.Bosch@moe.gov.na or ronelbosch@yahoo.co.uk. The appointed focal person will connect the UNAM evaluation team with a primary contact at each sampled school, aiding in the gathering of class lists (2024 lists if 2025 lists are not yet finalised) and distribution of parental consent forms. The evaluation will involve ten UNAM assessor teams, each comprising of three members, overseen by two UNAM coordinators who will travel between sites. Following a predetermined schedule, each team will assess one school per day, confirming each visit with the school's main contact to ensure the readiness and availability of the learners on the scheduled date.

Your cooperation and support in this matter are greatly appreciated.

Yours sincerely,

04.12.2024

Sanet L. Steenkamp
EXECUTIVE DIRECTOR

All official correspondences should be addressed to the Executive Director **Page 2 of 2**

Annexe 6 | EGRA Namibia

REPUBLIC OF NAMIBIA

MINISTRY OF EDUCATION, ARTS AND CULTURE

EARLY GRADE READING ASSESSMENT (EGRA) TOOL

ENGLISH SECOND LANGUAGE

This test is in English second language, so some learners may not understand the instructions you give them verbally. To be sure that they get an equal chance to perform their best, get all the learners together before conducting the test. Explain to them clearly (if possible in mother tongue) what is expected of each task. Do not give the details of the content of the sub-tasks.

General Instructions:
It is important to establish a relaxed atmosphere through some simple initial conversation of interest to the learner. The learner should perceive the assessment more as a game than a formal assessment. After you have finished, thank the learner.

Read the text in the box to the learner:

My name is _____. I'm working in the Ministry of Education.

- **We are trying to understand how learners learn to read. We would like your help in this.**
- **I'm going to ask you to listen to words, to sound out or read letter-sounds, read words and a short story out loud. Then I may ask you a few questions about the story you read.**
- **Using this stopwatch, I will see how long it takes you to do these things.**
- **This is NOT a test and it will not affect your marks at school.**
- **I will NOT write down your name so no one will know that these are your answers.**
- **Can we start?**

A. Date of assessment:	_____	E. Learner's gender:	○ Girl ☐	○ Boy ☐
B. Administrator's name:	_____	F. Learner's age:	Age: _____	
C. School name:	_____	G. Pre-school attendance:	○ Yes ☐	○ No ☐
D. Grade:	_____	H. Home language:	_____	

Comments: _____

_____

_____

_____

_____

_____

1

EGRA English Second Language © 2023

1. PHONEMIC AWARENESS: IDENTIFICATION OF INITIAL SOUNDS

Subtask 1

*This is NOT a timed exercise and **THERE IS NO LEARNER CHART**. Read aloud each word twice, and have the learner say the initial sound. Remember to model the "pure" sounds: /p/, not "pii" or "peh" or "puh", etc*

Say:

This is a listening exercise. I want you to tell me the first sound of each word. For example, in the word "pot", the first sound is "/p/". In this exercise, I would like you to tell me the first sound you hear in each word. I will say each word two times. Listen to the word, then tell me the very first sound in that word. If you don't know the sound, just leave it and listen to the next sound.

Let's practice. What is the first sound in "mouse"? "mouse."
[If the learner responds correctly, say]: **Very good, the first sound in "mouse" is /mmm/.**
[If the learner does not respond correctly, say]: **Listen again: "mouse." The first sound in "mouse" is /mmm/."**

Now let's try another one: What is the first sound in "day"? "day."
[If the learner responds correctly, say]: **Very good, the first sound in "day" is /d/.**
[If the learner does not respond correctly, say]: **Listen again: "day." The first sound in "day" is /d/.**

Do you understand what you should do?
[If the learner says no, say]: **Just try your best.**

Read the question and then pronounce the target word a second time. Accept only as correct the isolated sound (example /mmm/ not /muh/ or /meh/). Circle the appropriate response. If the learner does not respond after 3 seconds, mark as "No response" and say the next prompt.

Pronounce the words clearly, but do not overemphasize the beginning sound of each word.

***Early stop rule:** If the learner responds incorrectly or does not respond to the first five words, say "**Thank you!**". Discontinue this exercise; tick the box at the bottom of the page, and go on to the next exercise.*

What is the first sound in "_____"? "_____"? (Say the word twice)							
1.	**Egg**	/e/	Correct	Incorrect	Don't know	No Response	
2.	**Say**	/sss/	Correct	Incorrect	Don't know	No Response	
3.	**up**	/uh/	Correct	Incorrect	Don't know	No Response	
4.	**go**	/g/	Correct	Incorrect	Don't know	No Response	
5.	**now**	/nnn/	Correct	Incorrect	Don't know	No Response	**5 Words**
6.	**can**	/k/	Correct	Incorrect	Don't know	No Response	
7.	**unit**	/y/	Correct	Incorrect	Don't know	No Response	
8.	**eat**	/ii/	Correct	Incorrect	Don't know	No Response	
9.	**run**	/rrr/	Correct	Incorrect	Don't know	No Response	
10.	**look**	/lll/	Correct	Incorrect	Don't know	No Response	
TOTAL							

Tick in this box if the exercise was discontinued early because the learner had no correct answers in the first five words: ☐

2

EGRA English Second Language © 2023

2. LETTER SOUNDS

Subtask 2

Place the chart of letter-sounds (Chart 1) in front of the learner. Say:

Here is a page of letter-sounds . I would like you to say the sound of as many letters as you can. Some of them are two letters making one sound. You will start here and move across the page. (Point to the leftmost letter on the top row of the exercise, moving from left to right.) **When I say, 'Begin', you will sound the letters as fast as you can. Point to each letter as you sound it. If you don't know the sound of a letter, just leave it and read the next one..**

Let's practise first. Say the SOUND of the letter but NOT the letter name. (Point to the first letter and say, this is sound /b/, /b/ Do the same with the other two example letters on the chart, moving from left to right, to practise the instructions given above.)

Ok, now we're ready to begin.

Put your finger on the first letter. Ready? Begin….

- Start the timer when the learner starts.
- Strike a line through a letter-sound that the learner sounds incorrectly or cannot sound at all. For example: ~~b~~
- If the learner stops for more than 3 seconds, tell him/her to go on and strike a line through the letter that s/he did not read. For example: ~~b~~
- If the learner corrects himself/herself, accept it as correct. (If a strike has already been made on the letter, circle it to mark it correct.)
- If a learner skips a line, draw a continuous line through all the letters in that line and count them as incorrect. ~~[d][r][t][e][c][v].~~ DO NOT, however, discontinue the task because of this skipped line.
- If any entire line has strike-through lines across all the letters, ~~[e][j][o][d][h][x][k][l][n]~~ stop the assessment, place a bracket (]) after the last letter on that line and make a tick (✓) in the box at the bottom of the exercise to record that the exercise was discontinued.
- If the learner has not been able to say the sounds before the dark line, the exercise can also be discontinued.
- After one (1) minute, say "**Stop**". Place a bracket (]) after the last letter that the learner has attempted to sound.
- Count and record the number of letters that the learner sounded correctly.
- If the learner sounds all the letters in less than one (1) minute, stop the stopwatch and record the number of seconds taken to complete the exercise in the box below.

NOTE: /c/ and /k/ sound the same. There are two correct possible pronunciations for /th/ and /oo/.

LETTER SOUNDS: CHART 1

Examples: b S n

1.	**s**	**e**	**m**	**t**	**b**	**g**	**r**	**ai**	**a**	**h**	/10
2.	**c**	**j**	**i**	**k**	**n**	**u**	**p**	**l**	**d**	**f**	/10
3.	**o**	**y**	**ee**	**v**	**ng**	**ch**	**z**	or	**oo**	**x**	/10
4.	**ie**	**sh**	**er**	**qu**	**w**	**th**	**oi**	**ar**	**oa**	**ue**	/10
5.	**oa**	**S**	**E**	**M**	**T**	**B**	**G**	**R**	**A**	**H**	/10
6.	**C**	**J**	**I**	**K**	**N**	**U**	**P**	**L**	**D**	**F**	/10
7.	**O**	**Y**	**V**	**Z**	**X**	**W**	**Q**	**H**	**n**	**E**	/10
8.	**i**	**t**	**C**	**o**	**E**	**H**	**t**	**e**	**A**	**O**	/10
9.	**N**	**a**	**T**	**I**	**h**	**E**	**o**	**A**	**i**	**e**	/10
10.	**t**	**E**	**S**	**e**	**I**	**T**	**n**	**r**	**O**	**A**	/10
										Total number of letters sounded correctly:	***/100***
										If completed in less than 60 seconds, record the number of seconds it took to complete here:	
Tick in this box if the exercise was discontinued early:											
Did you notice any letter confusions in the learner's response? For example: b / d; p / q / g ; m / n; s / z; E/F										*Yes* ☐	*No* ☐

3

3. NON-WORD and WORD READING

Subtask 3

Place the words chart (Chart 2) in front of the learner. Say:

Here is a page of words. Some are made-up words and others are real. I would like you to read aloud as many made-up words as you can
(do not spell the words, but read them). You can blend the sounds together if needed or just say the word.
For example, this made-up word is "ubt."
1. Now you try: [point to the next word: "pog' and say] **Please read this word**
[If correct]: **"Very good: pog"**
[If incorrect]: **This made-up word is "pog."**
2. Now try another one: [point to the next word: mab and say] **Please read this word**.
[If correct]: **"Very good: mab"**
[If incorrect]: **This made-up word is "mab."**
Do you understand what you are supposed to do? When I say "begin," read the words as fast as you can. If you can't read a word, just leave it and read the next word.

Ok, now we're ready to begin.
Put your finger on the first word. Ready? Begin….

- Start the timer when the learner starts.
- Strike a line through a word that the learner reads incorrectly or cannot read at all. For example: ~~sug~~
- If the learner stops for more than three (3) seconds, tell the learner to go on, and strike a line through the word that s/he did not read. For example: ~~tob~~
- If the learner corrects himself/herself, accept it as correct. (If a strike has already been made on the word, circle it to mark it correct.)
- If a learner skips a line, draw a continuous line through all the words in that line and count them as incorrect. ~~[d][r][t][e][e][v].~~ DO NOT, however, discontinue the task because of this skipped line.
- If any entire line has strike-through lines across all the letters, ~~[e][j][o][d][h][x][k][l][n]~~ stop the assessment, place a bracket (]) after the last letter on that line and make a tick (✓) in the box at the bottom of the exercise to record that the exercise was discontinued.
- If the learner has not been able to say the sounds before the dark line, the exercise can also be discontinued.

- After one (1) minute, say: "**Stop**". Place a bracket (]) after the last word that the learner has read correctly.
- Count and record the number of words that the learner read correctly.
- If the learner reads all the words in less than one (1) minute, stop the stopwatch and record the number of seconds taken to complete the exercise in the box below.

NON-WORD and WORD READING: CHART 2

Examples:	ubt	pog	mab			
1.	**reb**	**wup**	**jub**	**eps**	**vuss**	/5
2.	**quoop**	**zook**	**chack**	**skap**	**blorn**	/5
3.	**meft**	**weems**	**chop**	**sing**	**dart**	/5
4.	**shock**	**flat**	**skill**	**gift**	**coins**	/5
5.	**var**	**slirt**	**weaf**	**pobe**	**flisp**	/5
6.	**braint**	**scrid**	**splote**	**twice**	**gloom**	/5
7.	**turn**	**mode**	**blast**	**groans**	**spray**	/5
8.	**strike**	**delay**	**modern**	**saucers**	**charming**	/5
9.	**I**	**the**	**he**	**she**	**me**	/5
10.	**we**	**be**	**was**	**to**	**do**	/5

	Total number of words read correctly:	**/50**
	If completed in less than 60 seconds, record the number of seconds it took to complete here:	
	Tick this box if the exercise was discontinued early:	

4. PASSAGE READING, Subtask 4 (if Pre-Primary substitute for oral vocabulary test)

Place the passage reading chart (Chart 3) in front of the learner. Say:

Now I'm going to ask you to read this story out loud. If you get stuck on a word, just leave it and keep on reading. When I say, 'Stop', stop reading the story. Then I will ask you some questions about what you have just read – so try to remember the story you're reading. You will start here. (Point to the first word of the passage.)

Ready? Begin.

- Start the timer when the learner starts.
- Strike a line through words that the learner reads incorrectly or cannot read at all. For example: ~~soccer~~
- If the learner stops for more than three (3) seconds, tell the learner to go on and strike a line through the word that s/he did not read. For example: ~~soccer~~
- If the learner corrects himself/herself, accept it as correct. (If a strike has already been made on the word, circle it to mark it correct.)
- If a learner skips a line, draw a continuous line through all the words in that line [~~Sara must stay in the house. She reads a book~~] and count them as incorrect. DO NOT, however, discontinue the task because of this skipped line.
- If any entire line has strike through lines across all the words, [~~Sara must stay in the house~~. ~~She reads a book~~] stop the assessment, place a bracket (|) after the last word on that line and make a tick (✓) in the box below to record that the exercise was discontinued.
- After one (1) minute, say: "**Stop**". Place a bracket (|) after the last word that the learner has read correctly.
- Count and record the number of words that the learner read correctly.
- If the learner reads the passage in less than one (1) minute, stop the stopwatch and record the number of seconds taken to complete the exercise in the box below.

PASSAGE READING: CHART 3

1.	**Sara (Q1) and Jim go to school on the red bus.**	**/10**
2.	**They take their bags on the bus (Q2). The bus stops.**	**/10**
3.	**'What now?' Sara looks at Jim. The bus stops for**	**/10**
4.	**the tall man and his dog (Q3). The man and his dog get**	**/12**
5.	**on the bus. They want to go to school too.**	**/10**
6.	**It is raining and the bus goes slowly (Q4).**	**/8**
7.	**When they get to school, the sun is shining. (Q5)**	**/9**
8.	**Sara, Jim, the man and his dog get off the bus. (Q5)**	**/11**
	Total number of words read correctly:	**/80**
If completed in less than 60 seconds, record the number of seconds it took to complete here:		
	Tick this box if the exercise was discontinued early:	

5

5. COMPREHENSION QUESTIONS

Subtask 5

Say:

Now I am going to ask you a few questions about the story you have just read. Try to answer all the questions I ask you.

- If the learner has read only part of the story, just ask the questions related to the part that s/he has read. Enter a dash (-) in the boxes for questions not covered.
- Enter a tick in a box (✓) for each question answered correctly.
- Put a cross in a box (✗) for each question answered incorrectly.
- If the learner corrects himself/herself, accept the answer as correct and put a tick (✓) in the box.
- Count and record the number of questions that the learner answered correctly at the bottom of the exercise.

	Questions	Answers	Signs
1.	**What was the name of the girl in the story?**	***Sara***	☐
2.	**What did Sara and Jim take on the bus with them?**	***Their bags***	☐
3.	**Why did the bus stop?**	***For the tall man and his dog (to get on the bus - accept any relevant answer)***	☐
4.	**Why did the bus move slowly?**	***Because it is raining (accept any relevant answer)***	☐
5.	**What happened when they got to school?**	***Sara, Jim, the man and his dog got off/ the sun was shining (either acceptable)***	☐
		Total number of questions asked:	
		Total number of questions answered correctly:	

4. Oral Vocabulary (alternative to passage reading) Place the oral vocabulary picture chart (Chart 4) in front of the learner. Say:

Now I'm going to say a word out loud and you will point to the picture. I will give you an example /sun/ and then you would point to this picture of the sun.
Ready? Begin.

- Start the timer when the learner starts.
- Strike a line through words that the learner incorrectly indicated the wrong picture or did not point to any picture. For example: ~~sun~~
- If the learner corrects himself/herself, accept it as correct. (If a strike has already been made on the word, circle it to mark it correct.)
- After one (1) minute, say: "**Stop**". Place a bracket (**|**) after the last word that the learner has read correctly.
- Count and record the number of pictures that the learner identified correctly.
- If the learner identified all the pictures in less than one (1) minute, stop the stopwatch and record the number of seconds taken to complete the exercise in the box below.

Oral Vocabulary (Chart 4)
Example sun

1	tree	hen	dog	flag	ant	/5
2	bed	ball	pencil	car	shoe	/5
3	book	chair	bag	fish	goat	/5
4	cup	banana	apple	cat	tree	/5

	Total number of pictures identified correctly:	/20
	If completed in less than 60 seconds, record the number of seconds it took to complete here:	
	Tick this box if the exercise was discontinued early:	

End of assessment. Make sure you have properly recorded all information on each page of the assessment before letting the learner go. Once everything is properly recorded and complete, thank the learner.

Chart 1: Letter Sounds

Example: b S n

s	e	m	t	b	g	r	ai	a	h
c	j	i	k	n	u	p	l	d	f
o	y	ee	v	ng	ch	z	or	oo	x
ie	sh	er	qu	w	th	oi	ar	oa	ue
oa	S	E	M	T	B	G	R	A	H
C	J	I	K	N	U	P	L	D	F
O	Y	V	Z	X	W	Q	H	n	E
i	t	C	o	E	H	t	e	A	O
N	a	T	I	h	E	o	A	i	e
t	E	S	e	I	T	n	r	O	A

8

EGRA English Second Language © 2023

Chart 2 : Word Reading – Section 1

Example: upt pog mab

Can also be sounded out then blended u-p-t p-o-g m-a-b

reb	wup	jub	eps	vuss
quoop	zook	chack	skap	blorn
meft	weems	chop	sing	dart
shock	flat	skill	gift	coins

Word Reading – Section 2

var	slirt	weaf	pobe	flisp
brait	scrid	splote	twice	gloom
turn	mode	blast	groans	spray
strike	delay	modern	saucers	charming

EGRA English Second Language

Word Reading – Section 3

I	the	he	she	me
we	be	was	to	do

Chart 3: Passage Reading

Sara and Jim go to school on the red bus. They take their bags on the bus. The bus stops. 'What now?'. Sara looks at Jim. The bus stops for the tall man and his dog. The tall man and his dog get on the bus.

They want to go to school too. It is raining and the bus goes slowly. When they get to school, the sun is shining. Sara, Jim, the tall man and his dog get off the bus.

12

Analysis Report

July 2025

G:ENESIS
UNLOCKING VALUE

UNAM
UNIVERSITY OF NAMIBIA