The impact of a story-based intervention on language, literacy, and cognitive development in South African pre-schoolers: randomised controlled trials for two language groups

Kate Cain, Department of Psychology, Lancaster University, Lancaster, LA1 4YF. UK

Jane Oakhill, University of Sussex

Shelley O'Carroll, Wordworks

Daleen Klop, University of Stellenbosch

Monique Visser, University of Stellenbosch

James Jackson, Lancaster University

Brian Francis, Lancaster University

Corresponding author: Kate Cain, k.cain@lancaster.ac.uk

Acknowledgements: This work was funded by British Academy grant ECE 190079 awarded to Oakhill (PI), Cain, O'Carroll, Klop (Co-Is). We would like to thank Andy Dawes for his comments during the project, and Bethany Pearson for her help with preparing the results tables.

Accepted for publication in International Journal of Educational Research 18/12/2025.

Abstract

Literacy rates in South Africa are low and many children start school without the requisite levels of emergent language and literacy skills needed to succeed. We report two RCTs of a story-based intervention delivered by preschool teachers to two language groups of children from low income backgrounds (isiXhosa: $N_{children}$=82, $N_{teachers}$=20; Afrikaans, $N_{children}$=118, $N_{teachers}$=24). The story-based intervention involved a 36-week programme, of 2-week cycles, each using a different culturally-appropriate story in the target language with activities designed to foster emergent language and literacy skills. Training for the teachers before and during the intervention was provided. The post-intervention assessment took place after 26 weeks. For both language groups (compared with the corresponding control group), the intervention had a positive impact on vocabulary taught in the programme and also developmental status across key learning domains. For early language and emergent literacy measures, baseline ability was the most consistent predictor for all outcome measures, with additional important contributions of initial vocabulary for some measures. This study demonstrates the feasibility of conducting gold-standard randomised controlled trials in low-resource settings. We draw on the data to set out practice and policy recommendations, critically the need to support school and literacy-learning readiness in homes and preschools, to enhance practice and children's outcomes.

The impact of a story-based intervention on language, literacy, and cognitive development in South African pre-schoolers: randomised controlled trials for two language groups

1. Introduction

Improving literacy is a key first step in overcoming the obstacles that lock individuals into a cycle of poverty and disadvantage (World Literacy Foundation, 2015). The standard of literacy in South Africa is exceptionally poor for a middle-income country: South Africa was placed last out of the 57 countries who participated in the Progress in International Reading Literacy Study (PIRLS) of Grade 4 schoolchildren in 2021, with 81% not able to read for meaning in any South African language (Department of Basic Education, 2023). Literacy skills build on early oral language skills: Preschool oral language skills such as phonological awareness, vocabulary and narrative predict later word reading and reading comprehension (Duff et al., 2015; Jago et al., 2025; Sénéchal, 2006; Silva & Cain, 2015), and preschool oral language interventions delivered by classroom teachers boost these foundational oral language skills (Bianco et al., 2010; Fricke et al., 2013; Language and Reading Research Consortium et al., 2022). Such findings informed the story-based intervention (called *Little Stars*) designed to develop the precursors of literacy in South African preschoolers, which we evaluated in this study.

Good reading comprehension requires good word reading (decoding) skills and good language skills (Garcia & Cain, 2014). Children with more accurate and fluent word reading, and those with better listening comprehension skills, obtain higher scores on standardised measures of reading comprehension from the earliest stages of reading instruction (Language and Reading Research Consortium (LARRC), 2015). Accurate and efficient word decoding enables effective retrieval of word meanings, frees up cognitive resources for the construction of meaning (Perfetti & Stafura, 2014), and predicts reading comprehension

concurrently and longitudinally (LARRC, 2015; Jago et al., 2025; Kendeou et al., 2009; Oakhill & Cain, 2012). Similarly, understanding of words, sentences, and prose read aloud to children predicts reading comprehension concurrently and longitudinally (Language and Reading Research Consortium (LARRC) & Logan, 2017; Jago et al., 2025; Kendeou et al., 2009; Oakhill & Cain, 2012).

In an attempt to improve the low levels of literacy in South Africa, a number of reading programmes have been introduced, the majority targeted at children in Grade 1 and above (see Grigg et al., 2016, for an overview). Programmes targeting basic literacy skills show positive, but small, impacts (Grigg et al., 2016; Meiklejohn et al., 2021). A systematic review of interventions targeting reading comprehension noted good quality of some, but was not able to draw firm conclusions on their efficacy due to lack of rigour either in study design, implementation, or reporting (Carter et al., 2024). Despite the promise of these evidence-based interventions, they have had limited impact to date on the significant challenge faced by South Africa, where the percentage of Grade 4 children not attaining basic standards of literacy has increased in recent years from 72% in PIRLS 2016 (Howie et al., 2017) to 81% in 2021.

1.1. Reading development: the South African context

Two pertinent contextual factors in South Africa provide particular challenges to the implementation of good quality instruction and its success. The provision of early years education is characterised by a lack of qualified teachers and poorly resourced classrooms (Bertram et al., 2013; Naidoo et al., 2014). This means that the achievement gap becomes entrenched from the earliest years. In high-income countries, high-quality instruction and teacher talk in preschool classrooms is associated with language gains for 4-year-olds over the academic year (Yeomans-Maldonado et al., 2019), and professional development to

enhance the quality of the classroom is related to improved language and literacy outcomes (Dickinson & Caswell, 2007; Egert et al., 2018). While the research is clear that there is a relation between quality of instruction and child outcomes, the current evidence is largely drawn from studies in high-income countries. There is a need for greater understanding of the influence of teacher experience and the role of professional development to support effective programme delivery in lower income settings, both of which were considered in this study (and explored in detail in separate reports: Oakhill et al., in preparation; Visser et al., in press).

Variation in the effectiveness of educational interventions is related to both the fidelity of implementation and adaptation of the programme due to the local context (Lendrum & Humphrey, 2012). These may be influenced by the educator's experience and style, as well as classroom factors such as resources and class size (Pedder, 2006). We provided professional development to support reliable implementation of the intervention and collected independent assessments of the quality of implementation. We also took into account teachers' qualifications and years of experience as potential mediators of any effects, and also considered the influence of independent measures of the classroom resources and teacher-led interactions outside of the intervention. In addition, we drew on the theoretical framework of situated learning (Brown et al., 1989) and designed the activities in the intervention in light of the specific cultural context of young children in South Africa.

A second critical challenge is that children start school in South Africa without sufficient proficiency in the foundational skills that support reading instruction (Wills et al., 2022) and educational attainment and learning in general (Tredoux et al., 2024). Recent work reports that, depending on the region, up to half of children have no alphabetic

knowledge at the start of Grade One (Wills et al., 2022); other work reports a developmental lag in verbal abilities of 15 months in 5-year-old preschoolers (Naudé et al., 2003). These stark findings may be explained (in part) by the fact that very few children in low income South African homes have access to books and language-rich early learning experiences at home (Naudé et al., 2003; O'Carroll, 2011).

Both theory and empirical research indicate the potential of preschool support to mitigate the risk of poor reading development, and learning in general. The influence of preschool oral language skills and foundational code skills on literacy outcomes is well established (Dickinson & McCabe, 2001). Teacher-delivered pre-school oral language interventions have been shown to benefit the oral language skills that were targeted, such as phonological awareness, vocabulary and narrative (Bianco et al., 2010; Fricke et al., 2017; Haley et al., 2017; Language and Reading Research Consortium et al., 2022). While the research is clear on the gains that can result from high quality language-rich learning environments, this evidence, like the evidence-base for the impact of classroom variables, is largely drawn from studies in high-income countries. Given the challenges in the South African context, there is a pressing need to determine if a high-quality early learning programme can make a significant impact on early oral language skills which, in turn, can enhance readiness to learn to read.

1.2. Cognitive development in South African preschoolers

Recent national surveys in South Africa find low performance measures of cognitive development that underpin learning, indicating poor school readiness (Giese et al., 2025; Tredoux et al., 2024). In South Africa's most comprehensive national surveys of preschool child outcomes conducted in 2021 and 2024, fewer than 50% of children met the standard of being 'on track' in critical domains including gross and fine motor skills, emergent

numeracy, emergent language and literacy and cognition and executive functioning. These were assessed using the Early Learning Objectives Measure (ELOM), designed for population-level monitoring of the developmental status of children aged 50–69 months and evaluation of early learning programmes in South Africa (Snelling et al., 2019). We included the ELOM subscale on emergent language and literacy development as a key outcome measure, because the intervention was designed to foster these skills. The other four ELOM domains were examined separately to determine if attending a preschool setting in general improved performance on these other key developmental domains and, in addition, to determine if exposure to the structured language-rich intervention programme resulted in better outcomes relative to the control group. The impact on a total score calculated across all five ELOM domains was also examined.

Several additional childhood variables known to influence early cognitive development are pertinent to the setting of our study, and were considered as potential influences on both primary language and reading-related measures and the secondary broader domains of learning (assessed with the ELOM). The first was a child's height. Stunting is related to childhood nutrition and a major concern in low- and middle-income countries, such as South Africa, because it is a significant predictor not only of health, but also cognitive performance in middle childhood (Mendez & Adair, 1999). Gender is another potential influence on language and literacy performance. In the PIRLS 2021 data, all countries had girls achieving higher mean reading scores compared to boys, but South Africa had the highest gender gap (Department of Basic Education, 2023; Mullis et al., 2023). However, studies using the ELOM as a performance measure find an influence of stunting and gender on scores (Tredoux et al., 2024). We also examined the influence of early vocabulary and the home learning environment on the effectiveness of the programme. In

high-income contexts, both preschool vocabulary and the home literacy environment influence language and literacy development in school (Sénéchal, 2006). In the South African context, out-of-school literacy practices may involve less parental involvement and fewer book-based experiences than was the case in those settings that informed the development of existing models of the home literacy environment (Kajee & Sibanda, 2019). For that reason, we assessed vocabulary and the home *learning* environment, which captures a broader set of variables related to early learning and is therefore more appropriate than an assessment of the home literacy environment in our study context.

1.3. The current study: approach and aims

We report two randomised controlled trials (RCTs) of the Little Stars story-based programme for literacy development in South African preschools located in the Western Cape, with separate trials for two different language groups. The programme was designed to target instruction of key foundational skills for literacy development, which have been shown to be weak in school starters in South Africa (Wills et al., 2022). The story-based approach drew on the theoretical framework of situated learning (Brown et al., 1989) and is particularly appropriate in the South African context where very few children have access to books and language-rich early learning experiences at home (Naudé et al., 2003; O'Carroll, 2011). Stories have the potential not only to support language development, but to provide a bridge between oral and written language, as demonstrated in studies on the relation between parental storybook reading style and later literacy (Haden et al., 1996; Silva & Cain, 2015). There is a lack of research evaluating preschool interventions for foundational skills in low-income settings. We sought to establish whether our theoretical and research-informed approach would benefit preschoolers, in this context, to develop the foundational literacy skills and broader readiness to learn skills associated with good progress in literacy.

We compared performance of an experimental group that received the classroom intervention to that of a (wait list) control group that received normal classroom teaching. There were two language groups: isiXhosa and Afrikaans. Their data are modelled separately because there are substantive differences in the phonological and morphosyntactic structure between these languages. The pre-registration of the study design and aims, and also the dataset and analytic code, can be found at LINK IN HERE.

The programme involved oral storytelling, book sharing, listening and speaking and writing activities designed to teach print awareness, phonological awareness, vocabulary, and narrative skills, with the aim to support the development of code-related and meaning-related skills that underpin later word reading and listening (and reading) comprehension, respectively. To address resourcing concerns and the sustainability of this project, teachers received professional development and support before and during the intervention to enhance their knowledge, and hard copies of the opensource resources needed to deliver the programme.

1.3.1. Research question 1: Relative to a non-intervention control group, does the Little Stars story-based intervention result in better outcomes in children's early language and emergent literacy? To address our primary research question our analyses examined whether or not the intervention resulted in gains on measures of children's emergent language and literacy, print and phonological awareness, and narrative skills relative to a wait-list control group. We examined how the programme influenced learning of target vocabulary items, and whether child or classroom characteristics influenced this learning outcome. Target vocabulary was not assessed prior to the intervention, so we were not able to look at gain scores. We examined whether outcome performance on each of these measures was influenced by child characteristics (initial vocabulary, gender, and age),

teacher characteristics (teacher's age, qualification level, and experience of teaching this age group), and classroom quality (quality of the teaching and learning environment). Class fees were used as an indicator of socio-economic status (SES) (Henry & Giese, 2023) to control for this variable. For all measures, we predicted that children who received the intervention would show stronger language skills than those in the control group at post-test.

1.3.2. Research question 2: Relative to a non-intervention control group, does the Little Stars story-based intervention result in gains in children's broader cognitive and learning readiness skills?

To address this second, exploratory, research question we examined the influence of the intervention on more general areas of cognitive development known to influence educational attainment: gross motor development, fine motor and visual integration, emergent numeracy and mathematics, and cognition and executive functioning (Snelling et al., 2019; Tredoux et al., 2024). Whilst the intervention was not designed to foster these skills directly, they may be enhanced by specific activities: fine motor coordination and visual integration were practiced through drawing and fine-motor tasks in the intervention, number, position, and measurement vocabulary featured in some stories, and tasks were developed to support children's ability to attend to auditory information. Evidence for an effect of the intervention would be stronger outcome scores for the intervention group, relative to the control group.

2. Method

2.1. Study design

We conducted two separate RCTs: one comprising isiXhosa-speaking children and their educators, the other comprising Afrikaans-speaking children and their educators. Early

Childhood Development (ECD) centres were randomly assigned either to the treatment (Little Stars) condition or to a business-as-usual (existing curriculum) wait control condition. One ECD centre in each language group had two classrooms and these classrooms were assigned to the same group.

Recruitment and consent took place before randomisation. A larger group of teachers and principals were invited to an information session about the project and trainers explained the consent form in the participants' home language. They were given an option of completing the form in their home language or English. Teachers who consented to participate explained the project to parents, and only children whose parents had completed consent forms were included in the study. The study was approved by the Health Research Ethics Committee, Stellenbosch University, in May 2021 [N21/05/047].

2.2. Study setting

All ECD centres were in the Western Cape province of South Africa, located in low income suburbs of rural towns and a township NGOs (Inceba Trust, Sikhula Sonke, Ikamva Labantu) supported recruitment and training.

2.3. Child participants

The final sample for data analysis comprised: 82 (36 girls) isiXhosa-speaking and 118 (54 girls) Afrikaans-speaking participants across 20 isiXhosa (10 intervention, 10 control) and 24 Afrikaans ECD centres (12 intervention, 12 control). These children had complete data on key measures. Reasons for exclusion of data detailed in Figures 1a and 1b.

Figures 1a and 1b around here

Participant characteristics for the final sample by group are reported in Table 1. Values were compared (by t-test) for the intervention and control groups within each language group (see Table 1 for details). The majority of these differences did not reach

statistical significance at the 5% significance level. For participants with available data (59 isiXhosa and 100 Afrikaans), standardised height for age was compared between intervention and control groups; differences were not statistically significant. There was very little evidence of stunted growth: 1 isiXhosa child in the intervention group and 5 Afrikaans children in the intervention group fell into the moderate stunting category, and 1 Afrikaans child in the control group fell into the severe stunting category.

Table 1 around here.

2.4. The intervention

The programme aligns with the South African National Curriculum Framework (Department of Basic Education, 2015), with stories linked to each of the six Early Learning and Development Areas. It has a play-based approach to teaching and learning and follows a similar structure to the widely used Wordworks Grade R story-based home language teaching programme. It was informed by pedagogical theories, such as situated learning theory (Hedegaard, 1998) and, therefore, designed around key principles of early learning including: nurturing and responsive relationships are crucial for learning and language development; children learn best when new learning has meaning and is connected to something familiar; children learn by being active and using all their senses; children make meaning through stories and play; children learn best when they are encouraged to interact, share ideas and ask and answer questions (O'Carroll et al., 2023).

The programme includes 18 story packs providing a total of 36 weeks of teaching. The two language programmes follow the same structure and are based on the same stories (which are locally written or selected to be appropriate for the South African context). Versions of the teacher guides and classroom materials were translated into Afrikaans and isiXhosa by mother tongue speakers who have experience of materials for young children.

The programme comprises five activities per week with one daily teaching activity (in a two-week cycle/routine), which requires only 15–30 minutes per day and could be integrated into an existing daily teaching programme. The activities were designed to develop the foundational skills, reviewed in section 1, which have been shown to support early literacy development. Each two-week cycle begins with telling a new story, followed by retelling the story, focusing on target vocabulary. Children are then engaged in activities related to each story: a song or rhyme to reinforce vocabulary introduced in the story; listening for beginning sounds in words used in the story; role playing the story to provide opportunities to use new vocabulary and phrases; retelling the story with sequence picture cards to build understanding of narrative structure; shared writing and reading a big book based on the oral story to teach print concepts; children drawing their favourite part of the story and creating 3D objects related to the story; little books for children to take home to retell the story. The materials are Creative Commons licensed (Wordworks, 2023).

2.4.1. Training of teachers

The teachers in the intervention group attended an orientation workshop to find out about the programme and the research project and to clarify expectations. Following this, they participated in two full-day training workshops and four further workshops (2.5 hours each, held monthly between April and August 2022). The training ended with a graduation event that included opportunities for feedback and reflection (Visser et al., in press). Attendance was excellent (average attendance at 5.2 out of 6 training workshops). The control group received the same training after the conclusion of the trial (from August to December 2022).

2.4.2. Quality of programme delivery

Trainers visited the teachers in both language intervention groups in July 2022. As trainers could not observe teachers doing all 10 activities in the two-week cycle, we asked teachers to do one of the main teacher-led activities: Storytelling, Sequence pictures or Reading a Big Book and a child-led activity: Drawing and emergent writing. The trainers rated the teachers across a range of items (see Supplemental Materials Methods for details). Delivery of the programme was generally very competent: 75% of teachers achieved an overall score of 4-5 (fairly competent/very competent), 17% achieved a score of 3, and 8% a score of 1-2. These data were used to create an implementation quality score.

2.5. Child language and reading-related measures

In February 2022 (Time 1) trained and accredited assessors conducted classroom observations and assessments of the children. The programme was designed to last for the full school year (36 weeks). Covid delayed the programme start. Since ethics approval required the control group to access the intervention in the same calendar year as the intervention group, the evaluation was completed at 26 weeks in August 2022 (Time 2), rather than at 36 weeks on completion of the full programme, as originally planned. Validated processes to measure age, height, gender, and language spoken were followed.

Few standardised instruments for the assessment of general ability, language and literacy are available in South Africa, particularly in African languages. The assessments used are described below (see Supplemental Materials Methods, for more detail) and were selected or developed specifically for this study, because they were available (or deemed suitable for translation and use in this context) in isiXhosa and Afrikaans. All assessments were administered and responses transcribed and scored by native speakers of isiXhosa and Afrikaans, as appropriate. Sessions including assessments that required post-scoring were recorded. Children were assessed individually in a space in the ECD Centre and all responses

were written down on the test forms and voice-recorded on tablets. Different to our pre-registration, different vocabulary assessments were administered at pre- and post-test, as detailed below.

2.5.1. Screening measures

2.5.1.1. Hearing. Before the study commenced, a hearing screening was conducted by a qualified audiologist on all children. This revealed that six children (4 isiXhosa and 2 Afrikaans) had hearing loss; they were referred for diagnostic audiology and excluded from the final sample. Twenty three percent of the isiXhosa sample and 26% of the Afrikaans sample had middle ear problems and were referred for medical treatment of middle ear problems and re-screening. Their data were not excluded from the final study sample.

2.5.1.2. Disability screen. The World Health Organisation (WHO) Disability Assessment Schedule (Durkin et al., 1995) was administered at both times of assessment. Children who were failed on the WHO screening on 2 or more items were excluded, with the following exception. Item 3 ("when you told this child to do something, did he/she seem to have difficulty understanding what you are saying?") is subject to interpretation by assessors. If a child failed on this item, but scored above 25 on the ELOM total score, their data were not excluded.

2.5.2. Child language covariate: initial vocabulary

The Cross-linguistic Lexical Task (CLT) was administered before the start of the intervention. It was developed as a cross-linguistically and cross-culturally comparable tool for the lexical assessment of children (Haman et al., 2017). It is available in isiXhosa and Afrikaans and has been used in studies in South Africa in mono- and multilingual populations (Perold Potgieter & Southwood, 2016).

The CLT comprises four sections: noun and, verb comprehension, assessed with picture recognition tasks, and noun and verb production, assessed with picture naming tasks. Each section has 30 test items and 2 practice items, with a total maximum score of 120 points. The assessment takes approximately 15 minutes per participant. Only target responses were accepted as correct in the comprehension sections. We followed published guidelines (Bohnacker et al., 2016) to score the production sections. Inter-scorer reliability was good for both language groups (94-100%). Internal consistency (Cronbach's alpha) was good for both language groups: isiXhosa, α=0.77; Afrikaans, α=0.88.

2.5.3. Environmental covariates: teacher experience, classroom quality, and home learning environment

2.5.3.1. Teacher experience. We used three metrics as indicators of teacher experience and training: practitioner age, practitioner qualification, and years spent teaching.

2.5.3.2. Classroom quality. Observers rated the quality of language, literacy, and learning activities in each classroom using three subscales from versions of the Early Childhood Environment Rating Scale (Sylva et al., 2006); the ECERS-R has been used previously in South Africa (Biersteker et al., 2016). We administered the Language and Literacy and the Learning Activities subscales of the ECERS-3 (Harms et al., 2014) and the Literacy subscale from the ECERS-E (Siraj-Blatchford et al., 2010). To obtain a more culturally-sensitive measure of quality classroom practice, evidence for adults engaging in storytelling was included in the book reading code for both ECERS-3 and ECERS-E, because the former is not resource-dependent and the intervention targets oral storytelling. Observations were conducted near to the start of the study (Time 1) and at post-test after 26 weeks of the intervention (Time 2). We used the Time 2 scores as predictors in our analyses as a proxy for the highest quality of support provided to children. To avoid a reduction in the number complete cases,

we used the Time 1 values where the Time 2 scores were not available (2 isiXhosa control, 1 isiXhosa intervention, and 1 Afrikaans intervention teachers, whose classrooms involved 17 isiXhosa and 4 Afrikaans children in the study).

2.5.3.3. Home learning environment. The ELOM Home Learning Environment Tool (HLE) (Dawes et al., 2023) was used to measure the time a caregiver spends with the child during the week and at weekends, and early learning resources and activities in the home. This measure has been rigorously tested and validated for the South African context. The HLE interview was completed by telephone with caregivers for 52 of the final isiXhosa sample and 59 of the final Afrikaans sample. We did not use this measure as a covariate in the main analyses because there were no related measures to use for data imputation. We report correlational analyses between the HLE and baseline performance on our child measures to determine their associations.

2.5.4. Child language and reading-related outcome measures

2.5.4.1. Emergent language and literacy (including print and phonological awareness). This was assessed using two measures: the Emergent Language and Literacy (ELL) subscale from ELOM, and the Early Literacy Protocol (ELP), both administered at pre-test (Time 1) and post-test (Time 2). The ELL subscale of the ELOM assesses how well children are able to communicate effectively and use language. Direct assessment includes description of feelings and actions (from picture prompts), naming of common objects, description of everyday tasks, story comprehension, and identification of initial sounds in words. This assessment was administered and scored according to the manual (Dawes et al., 2020). Psychometric analysis shows good internal consistency for this (and all other) ELOM subscales (Snelling et al., 2019).

The ELP was designed in the Stellenbosch University Division of Speech, Language and Hearing Therapy to assess print and phonological awareness. We selected tasks of print awareness (environmental print, concepts of print) and phonological awareness (syllable segmentation, syllable synthesis, and identification of phonemes at the start of words) relevant for this age group. Items were scored 1 point for correct (0 for incorrect) and summed to produce a print awareness score (max=10) and a phonological awareness score (max=12). Internal consistency was calculated using Cronbach's alpha for Time 1 performance. This was weak for the print awareness scale (α = .61 isiXhosa; α = .60 Afrikaans) and good for phonological awareness (α = .77 isiXhosa; α = .72 Afrikaans).

2.5.4.2. Child language. This was assessed using two measures: the Multilingual Assessment Instrument for Narratives (MAIN) (Gagarina et al., 2019) at pre- and post-test and a measure of vocabulary taught in the programme, the Proximal Vocabulary Test (PVT) at post-test only. The MAIN was developed as a tool for the assessment of narrative abilities of children aged 4 to 9 years in multilingual populations and from diverse cultural backgrounds, and has been used previously in South Africa (Klop & Visser, 2020). There were two stories with parallel structures: three distinct episodes portrayed across a sequence of six colourful pictures. Children completed one story at pre-test (Time 1) and one at post-test (Time 2) with order counterbalanced.

Children were first presented with the six picture sequence depicting the whole story, and then asked to tell the story, seeing two pictures (representing one episode of the story) at a time. Two scores were derived from the retells. A score for the number of story grammar elements included (macrostructure) (maximum possible score = 17) and a score for story structure complexity (maximum score = 9: (Maviş et al., 2016)). After the retell, comprehension was assessed with 10 open-ended questions (maximum score = 10). Inter-

rater agreement was assessed for the scoring of structural complexity and found to be high: isiXhosa = 90%; Afrikaans = 91%.

The PVT was developed to assess learning of vocabulary taught in the programme. Seventeen target words (7 nouns, 7 verbs and 3 adjectives) were identified from the wordlists for each picture book used in the programme. After two training items (a noun and a verb), children completed an expressive definition task (using picture prompts) and then a comprehension task (picture selection for spoken word prompt) (maximum score across both = 45). Inter-rater reliability for the expressive task was high: isiXhosa = 98.4%; Afrikaans = 98.1%. This was not administered at baseline because

2.5.5. Child general cognitive outcome measures

The Early Learning Outcomes Measure (ELOM) is a standardised tool suitable for measuring the effects of early learning programmes and children's readiness to learn in children aged 50-69 months (Snelling et al., 2019). It assesses five domains: gross motor development, fine motor coordination and visual integration, emergent numeracy and mathematics, cognition and executive functioning, and emergent literacy and language (ELL). Content and construct validity, reliability and cross-cultural fairness have been established (Anderson et al., 2021; Dawes et al., 2018). The total scores correlate well with the WPPSI (Anderson et al., 2021; Wechsler, 2012) indicating that this measure is a good indicator of general ability and readiness to learn.

2.5.6 Data reduction and analysis plan

When tests are translated and adapted for use in ethnolinguistic samples that are different from those on which the source test was developed, it is necessary to undertake psychometric analyses to establish their conceptual, construct, and metric equivalence to the source (Hambleton & Zenisky, 2010; van de Vijver & Tanzer, 2004). This is a lengthy and

costly procedure, particularly in a country like South Africa with 12 official languages and within-language variations. At the time this study was conducted, the equivalence of the CLT, the MAIN, and the ELP across languages had not been established. In addition, IRT corrected (standardised) scores were not available (Bortolotti et al., 2012) and, in any event, IRT scores are only comparable within (e.g., longitudinally) and not across languages (Leon & Singh, 2017). With these limitations in mind, we decided to use participants' total scores on each instrument for our analyses. The same procedure was used in analyses of receptive vocabulary data from the Young Lives longitudinal study of the development of cognitive skills from five to sixteen years of age in four countries and multiple languages (Tredoux & Dawes, 2018).

2.5.6.1. Data reduction. Some of our child language and reading-related outcome measures yielded several performance indicators: two for the CLT and ELP, and 3 for the MAIN. In addition, there were multiple indicators for the teachers' experience, the classroom environment, and the home learning environment. To avoid having multiple outcomes for each construct, principal components analysis (PCA) was conducted to reduce each construct to a single variable. In PCA, by taking linear combinations of the original variables, a smaller set of uncorrelated variables is created (Jolliffe & Cadima, 2016). The PCA was undertaken using the R base function princomp. Note that, for each construct we combined the baseline and follow-up data when applying the PCA and we applied PCA to the combined IsiXhosa and Afrikaans data.

For the various constructs, the proportion of variance explained by the first component ranged from 0.58 to 0.84 (see Supplemental Materials Data Reduction, Table SM.DR.1). Moreover, all loadings on the first component were positive and > 0.5, all first component eigenvalues > 1 and all subsequent eigenvalues < 1.0. These first components

therefore explained most of the component variability and were used as outcome variables, created to have means of zero, in the subsequent analyses. For the subscales of the ELOM (the language and emergent literacy-related subscale ELL, and the other general cognitive, fine and gross motor and numeracy attainment subscales) we used the ELOM subscale scores. We adopted a complete-case approach because missingness was typically due to child absenteeism (i.e. MCAR), and imputation methods are not likely to be effective, owing to the lack of predictive observed information.

2.5.6.2. Implementation Quality. Instead of using a treatment/control variable, we used a more finely tuned variable "Implementation Quality" which measured the degree of implementation for those children in the treatment group and had the value zero for those children in the control group. Trainers visited each centre in the intervention group before the last workshop. The trainers rated practitioners on overall classroom management and planning and also on different aspects of the quality of implementation of the activities in the Little Stars programme (scores ranged from 10 to 40, M = 32.63, SD = 8.02).

2.5.6.3. Model fitting. We fit a series of models to the n = 200 version of the data (complete cases: isiXhosa = 82, Afrikaans = 118), ranging in complexity from a linear model to a linear mixed-effects regression model. In our simplest model (Model 1), we modelled the outcome follow-up construct, with just two predictors - the outcome baseline measure and the Implementation Quality variable (the intervention classroom-level variable, for which the control group were awarded 0). This allowed us to estimate the average change in the follow-up score per unit increase in implementation quality, after adjusting for baseline scores (only). Note that we did not administer the PVT outcome measurement variable at baseline, so this was not included in the model.

We then built on this model by adding the three individual-level covariates of initial vocabulary, age, and gender (Model 2). We added these using forward stepwise selection, based on the Bayesian Information Criterion (BIC; Schwarz, 1978).This model allowed us to estimate the average change in the follow-up score per unit increase in implementation quality, after adjusting for baseline scores and individual covariates. Gender was not selected in any of the analyses at this stage. Child age was selected as a significant covariate in some of the analyses of ELOM domains, as reported below.

Next, we added the three classroom-level covariates, teacher experience, classroom quality, and class fees to the model (Model 3a); the first two variables were derived through PCA. This allowed us to estimate the average change in the follow-up score per unit increase in implementation quality, after adjusting for baseline scores and for both individual- and classroom-level covariates. Finally, we fit a random-intercept mixed effects model (Model 3b). Including classroom-level random effects allowed us to capture the remaining variability that exists between classrooms.

The models were fit in R (R Core Team, 2024); specifically, the linear models were fit using the lm function, the mixed effects models were fit using the lmer function from the lme4 package (Bates et al., 2015), the forward selection procedure was undertaken using the stepAIC function in the MASS package (Venables and Ripley, 2002) and the R^2 values for the mixed-effects models were computed using the r2mlm function from the r2mlm package (Shaw et al., 2023). The differences between models 1, 2 and 3a were assessed by likelihood ratio tests, comparing the difference of minus twice the log-likelihoods to a chi-squared distribution. The significance of the random effect comparing Model 3a to 3b was assessed using the rand function in the lmerTest package (Kuznetsova et al., 2017). Our dataset is available at BLINDED FOR REVIEW.

3. Results

3.1. Descriptives

Children's performance on the covariates (age, gender, and initial vocabulary (CLT)) was reported by language group and condition in Table 1 above. Children's performance on the outcome measures is reported in Table 2. Initial values were compared (by t-test) for the intervention and control groups within each language group (see Table 2 for details and Supplemental Materials Results, Tables SM.R.1 and SM.R.2 for individual subtest means). As might be expected given the random assignment of classrooms to condition, there were few statistically significant differences between intervention and control groups within the two language groups. The exception was for the MAIN scores: the Afrikaans intervention group obtained higher scores than their controls, $t(116) = 2.57$, $p = .011$.

Table 2 around here

Teacher experience, classroom environment quality factor scores, and class fees (as a proxy for SES) and comparisons are reported, by language group and condition in Table 3 (means and comparisons on all individual tests reported in the Supplemental Materials Results, Table SM.R.3 and SM.R.4). Despite random assignment of classrooms to condition, the teachers (classrooms) in the isiXhosa intervention group obtained a statistically significantly higher teacher experience factor score than controls ($t(20) = -2.20$, $p = .039$) and the Afrikaans intervention group teachers (classrooms) obtained a significantly lower score ($t(22) = 2.09$, $p = .048$). No other comparisons reached conventional levels of statistical significance.

Table 3 around here.

3.2. Research question 1: Relative to a non-intervention control group, does the Little Stars story-based intervention result in gains in children's early language and emergent literacy?

3.2.1. isiXhosa children

Table 4 summarises the four models for the isiXhosa children with the first principal components for the ELL, ELP, MAIN, and the single indicator PVT as outcome variables. For the ELL measure, which assesses how well children are able to communicate effectively and use language, there was no significant effect of implementation quality, but initial scores on the ELL were a strong and significant predictor of outcome scores (Model 1). For this measure, Model 2 was identical to Model 1 because none of the child covariates (age, gender, CLT) were selected when applying the BIC model selection process. Models 3a (with classroom covariates) and 3b (with additional classroom level random effects) did not provide a statistically better fit to the data, and explained little additional variance. Thus, Model 1 is the preferred model. This model explained a sizeable proportion of variation in the data (R^2 = .422). These analyses demonstrate that children with stronger initial language and literacy scores obtained higher scores on this measure after 26 weeks, regardless of intervention status, and child-level and teacher-specific and classroom-quality covariates.

Table 4 around here.

There was a different pattern of results in the analysis of the ELP scores, which assesses print and phonological awareness. Similar to the analysis of ELL, there was no significant effect of implementation quality, and initial scores on the ELP were a strong and significant predictor of ELP outcome scores (Model 1). Further, none of the child-level covariates were selected (Model 2); thus Model 2 was identical to Model 1. However, in contrast to the analysis of the ELL data, Model 3a (with classroom-level covariates) was a

significantly better fit than Model 2: Initial ELP scores and also the measure of SES (class fees) were each significant predictors of outcomes. Model 3b (including classroom random effects) was the best fitting model and explained an additional 10.5% of variance in scores relative to Model 3a. Model 3b explained a sizeable proportion of variation in the data (R^2 = .523). The estimated effects of initial ELP scores explained similar and significant variance to earlier models, but the estimated effects of classroom variables and SES were no longer significant. These analyses indicate that differences between the classrooms not captured by our study measures explained variation in outcomes in addition to initial ELP scores.

The pattern of results for the MAIN assessment of narrative skills was broadly similar to that reported for ELL. In Model 1, initial MAIN performance was a significant predictor of MAIN outcome, and successive models did not provide a significantly better fit to the data. The best-fitting model (Model 1) explained very little variation in outcome performance (R^2 = .067), and initial narrative skills (MAIN) were the only substantive predictor of outcome performance.

The final set of analyses concerned the bespoke measure of vocabulary taught in the programme (PVT). This measure was not administered at baseline. In Model 1, implementation quality was found to be a significant predictor of performance, demonstrating that those who received the intervention outperformed those in control classrooms. Specifically, there was an estimated benefit to outcome scores of .08 per unit increase in implementation quality. In Model 2, initial vocabulary (CLT) was included as a strong and significant predictor, and implementation quality remained a significant predictor. The inclusion of classroom covariates (Model 3a) and random effects (Model 3b) did not significantly improve the fit of the model. Thus, Model 2 was selected as the best fitting model. It explained a moderate proportion of variation in PVT scores (R^2 = .373). This

set of analyses indicates that the child-level covariate of initial vocabulary and implementation quality each explained outcomes on this measure, regardless of other child-level covariates (age and gender), and teacher-specific and classroom-quality covariates.

In addition, we examined the correlations between the HLE and Time 1 scores for ELL, ELP, and MAIN, and also the CLT, combined across intervention and control groups. The correlations were weak (range 0.033-0.153) indicating that the HLE did not significantly predict initial performance on these measures. The correlations are reported in Supplemental Materials Results, Table SM.R.6, see also Table SM.R.5 for breakdown of means for HLE components).

3.2.2. Afrikaans children

Table 5 summarises the four models for the Afrikaans children that have the first principal components for the ELL, ELP, MAIN, and the single indicator PVT as outcome variables. For the ELL measure, similar to the findings for the isiXhosa sample, there was no significant effect of implementation quality, but initial scores on the ELL were a strong and significant predictor of outcome scores (Model 1). Different to the findings for the isiXhosa group, the CLT was a strong and significant predictor of outcome, and initial ELL scores did not make a statistically significant contribution once CLT was added (Model 2). Model 3a (with classroom covariates) was a significantly better fit, with both CLT and the classroom environment predicting significant variance in outcomes. Although Model 3b (with random effects) explained slightly more variance, it was not a better fit to the data than Model 3a. Thus, the best fitting model was Model 3a. This model explained a moderate proportion of variance in outcome scores (R^2 = .315) and indicates that initial vocabulary and measures of the classroom environment were the most important predictors of final ELL scores.

Table 5 around here.

The pattern of results for the ELP data differed from that of the ELL data. In Model 1, both initial performance and implementation quality were significant predictors of performance; after adjusting for initial ELP performance, there was an estimated 0.13 increase in the outcome ELP scores per unit increase in implementation quality, demonstrating that those who received the intervention outperformed those in control classrooms. Model 2 did not explain additional variance. Model 3a (with classroom covariates) was a better fit to the data; implementation quality was no longer a statistically significant predictor, but teacher experience and classroom environment each explained significant variance in outcomes. Model 3b (additional classroom level random effects) provided a statistically better fit to the data, but the estimated effects of teacher experience and classroom variables were no longer significant. Model 3b was the preferred model and explained sizeable variance in scores (R^2 = .468). These analyses indicate that differences between the classrooms not captured by our study measures explained variation in outcomes, in addition to pre-test ELP scores.

The pattern of results for the MAIN also confirmed Model 3b as the best fitting model. Significant predictors of performance were initial MAIN scores and initial vocabulary (CLT). This model explained a moderate proportion of variance in scores (R^2 = .388). As for the ELP analyses, the analyses indicate that differences between the classrooms not captured by our study measures explained additional variation in outcomes. The final set of analyses concerned the bespoke measure of vocabulary taught in the programme (PVT), which was not administered at baseline. As for the isiXhosa sample, in Model 1, implementation quality was found to be a significant predictor of performance, demonstrating that those who received the intervention outperformed those in control classrooms with an estimated benefit to outcome scores of .09 per unit increase in

implementation quality. In Model 2, initial vocabulary (CLT) was an additional significant predictor. The inclusion of classroom covariates (Model 3a) and random effects (Model 3b) did not significantly improve the fit of the model. Thus, Model 2 was the best fitting model; it explained sizeable variance in PVT scores (R^2 = .587). As for the isiXhosa sample, this set of analyses indicates that the child-level covariate of initial vocabulary and implementation quality each explained outcomes on this measure, regardless of other child-level covariates (age and gender) and teacher-specific and classroom-quality covariates.

We examined the correlations between the HLE and Time 1 scores for ELL, ELP, and MAIN, and also the CLT, combined across intervention and control groups. Similar to the isiXhosa group, the correlations were weak (range -0.040-0.215) indicating that the HLE did not strongly influence performance on these measures.

3.2.3. Summary

These analyses show that initial ability is an important predictor of outcomes for the isiXhosa sample. Initial ability on our measures of emergent language and literacy (ELL), narrative (MAIN), and print and phonological awareness measure (Early Literacy Protocol: ELP), was a significant predictor in the best fitting model for each. Similarly, for the Afrikaans sample, initial performance on the ELP and MAIN predicted outcomes. The home learning environment was not a significant predictor of children's initial ability for either language group.

We were able to explain a reasonable proportion of variance in outcome scores across all measures for both groups, with the exception of the narrative scores for the isiXhosa sample. In addition to children's initial performance on each measure, we found that additional variance in these measures was explained by classroom factors for the Afrikaans sample; the classroom environment (assessed by the ECERS subscales) predicted

ELL outcome scores, and random variation between classrooms not captured by this measure and the teacher experience factor predicted ELP and MAIN outcomes. Furthermore, for both language groups, implementation quality, in addition to initial vocabulary (Cross-linguistic Lexical Task: CLT), predicted outcomes on the vocabulary taught in the programme (PVT).

3.3. Research question 2: Relative to a non-intervention control group, does the Little Stars story-based intervention result in gains in children's broader cognitive and learning readiness skills?

3.3.1. isiXhosa children

Table 6 summarises the four models for the isiXhosa children with the scores for the four ELOM domains: Gross Motor Development, Fine Motor Coordination and Visual Integration, Emergent Numeracy and Mathematics, Cognition and Executive Functioning, and ELOM total scores. For the domain of Gross Motor Development, there was no significant effect of implementation quality, but initial scores were a significant predictor of outcome scores (Model 1). For this measure, Model 2 was identical to Model 1 because none of the child covariates (age, gender, CLT) were selected when applying the BIC model selection process. Models 3a (with classroom covariates) and 3b (with additional classroom level random effects) did not provide a statistically better fit to the data, and explained little additional variance. Thus, Model 1 is the preferred model, because the other models did not provide a statistically better fit to the data. This model explained a small proportion of variation in the data (R^2 = .090). These analyses demonstrate that children with stronger initial Gross Motor Development scores obtained higher scores on this measure after 26 weeks, regardless of intervention status, child-level, teacher-specific and classroom-level covariates.

Table 6 around here.

There was a different pattern for domain 2 - Fine Motor Coordination and Visual Integration, for which Model 3a was the best fit (R^2 = .251). Pre-test performance made a positive contribution to post-test scores, and class fees made a small and negative contribution. For the two other domains and the ELOM total scores, Model 1 was the best fitting model: Emergent Numeracy and Mathematics (R^2 = .329), Cognition and Executive Functioning (R^2 = .444), and ELOM total score (R^2 = .555) indicating that initial scores were a significant predictor of outcome scores.

The ELOM also provides cut scores, which can be used to classify children as On Track, Falling Behind, or Falling Far Behind the expected developmental standards on all domains. Table 7 reports the number and proportion of children in the intervention and control groups whose performance was categorised as On Track at Time 1 and Time 2 for all five domains. Before the start of the study, no more than 60% of children were On Track across the different ELOM domains; by Time 2 the range of those On Track was 50 – 83%. A full break down of categories is reported in Supplemental Materials Results SM.R.7.

Table 7 around here

We used McNemar's change test to determine if a significant proportion of children changed status between 'falling behind/falling far behind' and 'on track' between Times 1 and 2. We conducted separate analyses for intervention and control groups. Due to the number of comparisons (5 domains and also total scores) we used a conservative significance level (p = .008; .05/6 comparisons). For the isiXhosa intervention group, significant improvement was evident for domain 2 (Fine Motor Coordination and Visual Motor Integration) with 14 children moving from 'falling behind/falling far behind' to 'on track' status and only 3 in the other direction, and for the ELOM total scores with 18

children showing an improvement in category and none changing to a poorer developmental status. For the control group, there was an improvement in performance for domain 3 (Emergent and Mathematics) with 11 children moving to 'on track' status. A summary of findings is reported in Supplemental Materials Results Table SM.R.8.

3.3.2. Afrikaans children

Table 8 summarises the four models for the Afrikaans children with the scores for the four ELOM domains reported above. The pattern of results differed to that found for the isiXhosa sample and also across ELOM subscales. For Gross Motor Development, initial scores and implementation quality were each significant predictors of outcome scores in Model 1. Children in intervention classrooms outperformed those in control classrooms with an estimated benefit to outcome scores of .07 per unit increase in implementation quality. None of the child-level covariates were included in Model 2 (using the forward selection procedure). Model 3a, which included the teacher, classroom, and SES covariates explained additional variance and Model 3b, with random effects, was the best fitting model (R^2 = .536). In this model, initial scores remained a significant predictor of outcome, but implementation quality was no longer a significant predictor.

Table 8 around here.

For Fine Motor and Visual Integration, Model 3b was the best fitting model (R^2 = .595). As for Gross Motor Development, implementation quality was a unique predictor in Models 1 and 2. In addition, age was included as a significant child-level covariate in Model 2. However, like Gross Motor Development, the only unique predictor in Model 3b was initial performance and also random variation between classrooms not captured by our measures. For Emergent Numeracy and Mathematics, Model 3b was also the best fitting model (R^2 = .511), with initial performance, initial vocabulary (CLT), and random effects

making significant contributions. Cognition and Executive Function showed the same pattern. Model 3b was the best fitting model (R^2 = .552) with initial subscale performance, initial vocabulary (CLT), and random effects making significant contributions. The same was true for the ELOM total score, which includes all five domains (these four and ELL) and is used as an indicator of readiness to learn. The best-fitting model was Model 3b in which initial ELOM total score and initial vocabulary (CLT), each made significant positive and unique contributions to outcomes, in addition to random effects (R^2 = .749).

We used the same procedure, as for the isiXhosa sample, to calculate the number and proportion of children in the intervention and control groups whose performance was categorised as On Track at Time 1 and Time 2 for each of the five ELOM domains (including domain 5 - Emergent Language and Literacy) and to test for improvement (suing McNemar's change test, as described above). A summary is provided in Table 9.

Table 9 around here.

Before the start of the study, between 38 and 77% of children were On Track across the different ELOM domains, which increased to between 52 and 80% by Time 2. A full break down of categories is reported in Supplemental Materials Results Table SM.R.7. For the Afrikaans intervention group, a significant proportion of students improved status from 'falling behind/falling far behind' to 'on track' status on three domains (Gross Motor, Fine Motor, and Emergent Numeracy and Mathematics) and also the ELOM total scores. There was no evidence for significant improvement on any scores for the control group. A summary of findings is reported in Supplemental Materials Results SM.R.8.

3.3.3. Summary

These analyses show that initial ability on the ELOM subdomains and total scores was the sole predictor of outcomes for the isiXhosa sample. There was a different pattern

for the Afrikaans sample. Initial ability was an important predictor for all measures. In addition, initial vocabulary and also variation between classrooms not captured by our measures explained outcomes scores for several ELOM subdomains, as well as ELOM total scores. Exposure to the intervention resulted in categorical improvement on ELOM total scores for both language groups.

4. Discussion

We conducted two RCTs to evaluate the impact of a story-based intervention on the language, emergent literacy, and cognitive ability of two groups of South African preschoolers, one group was isiXhosa speakers, the other Afrikaans. The intervention was designed to be culturally-sensitive and sustainable in a low-income context by developing and providing open access culture- and language-specific resources and professional development before and during the study. The intervention directly influenced learning of target vocabulary in the programme, for both language groups. In addition, a significant proportion of children exposed to the intervention improved to the category of 'on track' status for learning on our indicator of cognitive ability. For both language groups, baseline performance was a strong indicator of outcomes for both language groups, with initial vocabulary and variation between classrooms also predicting outcomes for the Afrikaans group.

There was no evidence that implementation quality predicted significant unique variance in outcomes on measures of early language and emergent literacy once controlling for the influence of initial performance, and child-, teacher-, and classroom-level covariates. Thus, the intervention evaluated at week 26 of the full 36 week programme did not result in greater gains on these measures for the intervention group in relation to controls, with these stringent controls. However, the finding of a positive impact on our proximal measure

of taught language is promising, particularly given that post-test assessment was conducted earlier than our original study design (due to conditions of the ethical review board). The ELOM improvements to On Track status seen for both intervention groups indicate broader benefits of a structured intervention that might develop children's general readiness to learn and enhance future learning. Future studies should include follow-up assessments to determine longer-term impacts.

There was a very strong influence of pre-test (baseline) performance on most measures for both language groups; the exception was the ELL measure for the Afrikaans group. This finding underscores the need for evaluation studies to control for baseline performance, which is not always the case (Carter et al., 2024). One interpretation is that children with stronger initial skills benefitted more from the intervention than their peers. Another possibility could be that there is a certain threshold of cognitive ability (or maturity) for the intervention to work effectively. Relative to controls, children in the intervention classrooms in both language groups were more likely to improve in categorical status on the ELOM total scores to being 'on track'. These findings identify a strong need to understand the factors associated with individual differences in early language and literacy skills, which may have an impact on the effectiveness of early years and classroom instruction. Research in high-income contexts reports a strong association between early language and literacy performance and the home literacy environment (Sénéchal, 2006). The homes of the children in our study contained few books: caregivers of Afrikaans children reported an average of fewer than 6, and the isiXhosa sample reported fewer than 1 on average. We used a measure of the environment that was developed for the South African context ELOM Home Learning Environment Tool (HLE) (Dawes et al., 2023), in order to survey the broader home learning context and not bias toward more affluent homes. Despite variability in

those scores, they were not strongly associated with baseline scores on our outcome measures. A limitation of this tool is that it is self-report and does not measure frequency or quality of interactions between caregivers and children, which may be more predictive of early language and literacy development. Further, responses were obtained for just over half of our sample. We found little evidence for stunting in our sample, which is noted as a predictor of educational and cognitive outcomes (Mendez & Adair, 1999). Future research should explore a range of culturally sensitive measures of the early learning environment in homes and other influences on development from birth to three years to determine the most influential factors that inform knowledge and ability at the start of preschool.

In line with other research, our measure of the broader classroom environment and use of resources (three subscales from the ECERS) was not a strong predictor of children's performance (Giese et al., 2025). One interpretation is that this variable has little influence on child development. Another interpretation is that our measure was not appropriate to capture meaningful variation between classrooms. Indeed, for the Afrikaans sample, variation between classrooms that was not captured by any of our measures improved model fit for the measures of print and phonological awareness (ELP) and narrative skills (MAIN) and all ELOM domains and total scores. Other work has questioned the suitability of the ECERS-R in low-income contexts (Betancur et al., 2021). Furthermore, others have commented that standardised measures of ELP quality do not capture the aspects of quality that have a positive impact, such as the quality and richness of adult-child interactions (UNICEF, 2017). Together with these current findings, we see an urgent need to develop sensitive measures of the preschool and early school environment for future research that capture *how* early years practitioners interact with children. This may better account for the

influence of variability between classrooms that we were not able to measure directly in our study.

We selected and developed culturally-appropriate language and literacy materials for this study, because there are no standardised instruments for the language and literacy in South Africa, across both languages. These captured variability in our sample, but average performance on some measures was low even on post-test after 26 weeks. For example, whilst all groups obtained 45% or higher average scores for the MAIN story comprehension task, production scores remained low. In comparison to other studies, our materials appear suitable to study change. For example, whilst some studies report that 49-68% of children in Grade 1 are unable to demonstrate phonemic awareness (Wills et al., 2022), our measures included other code-related skills such as print knowledge and syllable awareness and we found that very few (<5% isiXhosa and <2% Afrikaans) were unable to successfully complete any items at pre-test. Similarly, for the early vocabulary measure (CLT) only one child (isiXhosa) could not produce a single item, and all of our sample demonstrated comprehension of several words. Narrative skills are identified as weak elsewhere in the literature with between 9 and 12% of Grade 1 children in South Africa unable to answer comprehension in a listening comprehension task (Wills et al., 2022). In our study, only 4 children (isiXhosa) were not able to answer any questions correctly. Thus our tasks, while challenging for preschoolers in this context, were suitable.

4.1. Limitations

In addition to limitations discussed above, we discuss five that are also pertinent. First, we conducted our post-test assessment at 26 weeks for a programme designed to last for 36 weeks. This was a condition for ethical approval, because the ethics review panel required that the control group (both teachers and students) had the opportunity to benefit

from the training in the same school year. Evaluation after completion of the entire programme as intended may have seen stronger impacts of the intervention. However, we note that all the classrooms, including the 'control' classrooms were selected from ECD Centres that were receiving support from local NGOs, including support with administrative systems, leadership and management, health and nutrition of children, compliance with municipal regulations to enable registration, training to support quality teaching. Second, there were missing data across a range of measures, particularly in the isiXhosa sample, where children left ECD centres during the study period, or not been at school because parents were unable to afford fees. We adopted a complete-case approach, due to concerns about data not being missing at random, and there being insufficient observed information to reliably predict the missing values. Although this may have introduced bias in the estimates obtained, other approaches such as imputation would not have been effective in this scenario. We recommend that future studies build in additional data collection points to mitigate such risks, and record reasons (where known) for missing datapoints (absenteeism, assessor error, child refusal, etc) to inform analysis choices. Third, we did not administer a pre-test of the target vocabulary so were not able to examine gain scores. As discussed in our method section, standardised assessments are not available for the range of language groups in South Africa. We recommend that future studies include baseline assessments of all experimental outcome measures. Fourth, we were not able to capture reliable data about attendance and, therefore dosage, for individual children. It was reported that within the isiXhosa sample, some parents saved on school fees by taking their children out of the ECD centres to be looked after by older siblings during ordinary schools closure periods. This was confirmed by sampling attendance records. We recommend that future studies integrate monitoring systems to record dosage. Fifth, we note the limitation that our study

findings are directly relevant to our specific context (Western Cape, South Africa). However, we believe that these findings provide critical evidence to inform future intervention work in this region, and other regions facing similar resource and performance challenges; future work is required to test the generalisation of these findings to other low-income settings.

4.2. Conclusions

We have demonstrated the feasibility of a model to conduct a gold-standard RCT evaluation of a preschool intervention in a low-income setting. Our model of delivery is sustainable: it was delivered by teachers, who received training before and after the study, and who received free resources to support programme delivery. The intervention had positive impacts on taught vocabulary and early cognitive development, but initial performance and variation between classrooms were the strongest predictors of emergent literacy outcomes. These findings indicate a need for policy to support school and literacy-learning readiness in homes and preschools, to enhance practice and children's outcomes. Future research to assess the longer-term benefits of this programme on formal literacy instruction are warranted.

References

Anderson, K. J., Henning, T. J., Moonsamy, J. R., Scott, M., du Plooy, C., & Dawes, A. R. (2021). Test-retest reliability and concurrent validity of the South African Early Learning Outcomes Measure (ELOM). *South African Journal of Childhood Education*, *11*(1), 1-9. https://doi.org/ https://doi.org/ 10.4102/sajce.v11i1.881

Bertram, C., Mthiyane, N., & Mukeredzi, T. (2013). 'It will make me a real teacher': Learning experiences of part time PGCE students in South Africa. *International Journal of Educational Development*, *33*(5), 448-456. https://doi.org/https://doi.org/10.1016/j.ijedudev.2012.10.001

Betancur, L., Maldonado-Carreño, C., Votruba-Drzal, E., & Bernal, R. (2021). Measuring preschool quality in low-and middle-income countries: Validity of the ECERS-R in Colombia. *Early Childhood Research Quarterly, 54, 86-98.*, *54*, 86-94. https://doi.org/10.1016/j.ecresq.2020.08.001

Bianco, M., Bressoux, P., Doyen, A. L., Lambert, E., Lima, L., Pellenq, C., & Zorman, M. (2010). Early training in oral comprehension and phonological skills: Results of a three-year longitudinal study. *Scientific Studies of Reading*, *14*(3), 211-246. https://doi.org/DOI: 10.1080/10888430903117518

Biersteker, L., Dawes, A., Hendricks, L., & Tredoux, C. G. (2016). Center-based early childhood care and education program quality: A South African study. *Early Childhood Research Quarterly*, *36*, 334-344. https://doi.org/10.1016/j.ecresq.2016.01.004

Bohnacker, U., Lindgren, J., & Öztekin, B. (2016). Turkish-and German-speaking bilingual 4-to-6-year-olds living in Sweden: Effects of age, SES and home language input on vocabulary production. *Journal of Home Language Research 1*, 17-41. https://doi.org/http://doi.org/10.16993/jhlr.26

Bortolotti, S. L., Tezza, R., Andrade, D. F., Bornia, A. C., & A.F., J. (2012). Relevance and advantages of using the item response theory. . *Quality and Quantity*, *47*(4), 2341-2360. https://doi.org/10.1007/s11135-012-9684-5

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32-42. https://doi.org/10.3102/0013189X018001032

Carter, J., Podpadec, T., Pillay, P., Babayiğit, S., & Gazu, K. A. (2024). A systematic review of the effectiveness of reading comprehension interventions in the South African multilingual context. *Educational Research and Evaluation*, *29*(1-2). https://doi.org/10.1080/13803611.2024.2314522

Dawes, A., Biersteker, L., Girdwood, E., Snelling, M., & Tredoux, C. G. (2018). Early learning assessment innovation in South Africa: A locally appropriate monitoring tool. *Childhood Education*, *94*(1), 12-16. https://doi.org/10.1080/00094056.2018.1420358

Dawes, A., Biersteker, L., Girdwood, E., Snelling, M. J. T. L., & Tredoux, C. G. (2020). *Early Learning Outcomes Measure Technical Manual.* The Innovation Edge. https://datadrive2030.co.za/wp-content/uploads/2022/09/ELOM-Technical-Manual_2020-1.pdf

Dawes, A., Snelling, M., & Biersteker, L. (2023). *Home Learning Environment Tool Technical Manual*. Datadrive2030

Dickinson, D. K., & Caswell, L. (2007). Building support for language and early literacy in preschool classrooms through in-service professional development: Effects of the

Literacy Environment Enrichment Program (LEEP). *Early Childhood Research Quarterly*, *22*, 243-260.

Dickinson, D. K., & McCabe, A. (2001). Bringing it all together: The multiple origins, skills, and environmental supports of early literacy. *Learning Disabilities Research & Practice*, *16*, 186-202.

Duff, F. J., Reen, G., Plunkett, K., & Nation, K. (2015). Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry*, *56*, 848-856. https://doi.org/10.1111/jcpp.12378

Durkin, M. S., Wang, W., Shrout, P. E., Zaman, S. S., Hasan, Z. M., Desai, P., & Davidson, L. L. (1995). Evaluating a ten questions screen for childhood disability: reliability and internal structure in different cultures. *Journal of Clinical Epidemiology*, *48*(5), 657-666. https://doi.org/https://doi.org/10.1016/0895-4356(94)00163-K

Department of Basic Education. (2023). *PIRLS 2021: South African Preliminary Highlights Report*.

Egert, F., Fukkink, R. G., & Eckhardt, A. G. (2018). Impact of in-service professional development programs for early childhood teachers on quality ratings and child outcomes: A meta-analysis. *Review of Educational Research*, *88*, 401-433.

Fricke, S., Bowyer-Crane, C., Haley, A. J., Hulme, C., & Snowling, M. J. (2013). Efficacy of language intervention in the early years. *Journal of Child Psychology and Psychiatry*, *54*, 280-290. https://doi.org/10.1111/jcpp.12010

Fricke, S., Burgoyne, K., Bowyer-Crane, C., Kyriacou, M., Zosimidou, A., Maxwell, L., Lervag, A., Snowling, M. J., & Hulme, C. (2017). The efficacy of early language intervention in mainstream school settings: A randomized controlled trial. *Journal of Child Psychology and Psychiatry*, *58*(10), 1141-1151. https://doi.org/doi:10.1111/jcpp.12737

Gagarina, N., Klop, D., Kunnari, S., Tantele, K., Välimaa, T., Bohnacker, U., & Walters, J. (2019). MAIN: Multilingual assessment instrument for narratives–Revised. *ZAS Papers in Linguistics*, *63*, 1-36.

Garcia, J. R., & Cain, K. (2014). Decoding and reading comprehension: A meta-analysis to identify which reader and assessment characteristics influence the strength of the relationship in English. *Review of Educational Research*, *84*(1), 74-111. https://doi.org/10.3102/0034654313499616

Giese, S., Pettersson Gelander, G., Cook, C., Carnegie, T., Tredoux, C., Dawes, A., Biersteker, L., Beukes, S., Pahla, Z., & Kika-Mistry, J. (2025). *Thrive by Five Index 2024: National Findings*.

Grigg, D., Joffe, J., Okeyo, A., Schkolne, D., van der Merwe, N., Zuma, M., Mulenga, C., Boodhoo, A., & Louw-Potgieter, J. (2016). The role of monitoring and evaluation in six South African reading programmes. *Southern African Linguistics and Applied Language Studies*, *34*, 359-370. https://doi.org/10.2989/16073614.2016.1262270

Haden, C. A., Reese, E., & Fivush, R. (1996). Mothers' extratextual comments during storybook reading: Stylistic differences over time and across texts. *Discourse Processes*, *21*, 135-169.

Haley, A., Hulme, C., Bowyer-Crane, C., Snowling, M. J., & Fricke, S. (2017). Oral language skills intervention in pre-school—A cautionary tale. *International Journal of Language andCommunication Disorders*, *52*(1), 71-79. https://doi.org/DOI: 10.1111/1460-6984.12257

Haman, E., Łuniewska, M., Hansen, P., Simonsen, H. G., Chiat, S., Bjekić, J., Blažienė, A., Chyl, K., Dabašinskienė, I., Engel de Abreu, P., Gagarina, N., Gavarró, A., Håkansson, G., Harel, E., Holm, E., Kapalková, S., Kunnari, S., Levorato, C., Lindgren, J.,…Armon-Lotem, S. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics and Phonetics*, *31*(11-12), 818-843. https://doi.org/10.1080/02699206.2017.1308553

Hambleton, R. K., & Zenisky, A. (2010). Translating and adapting tests for cross-cultural assessment. In D. Matsumoto & F. van de Vijver (Eds.), *Cross-cultural Research Methods* (pp. 46-74). Cambridge University Press.

Harms, T., Clifford, R. M., & Cryer, D. (2014). *Early Childhood Environment Rating Scale, third edition (ECERS-3)*. Teachers College Press.

Hedegaard, M. (1998). Situated learning and cognition: Theoretical learning and cognition. *Mind, Culture, and Activity*, *5*(2), 114–126. https://doi.org/10.1207/s15327884mca0502_5

Henry, J., & Giese, S. (2023). Reviewing the socio-economic gradient in learning outcomes for children who participated in the Thrive by Five Index. https://datadrive2030.co.za/wp-content/uploads/2023/08/Data-insights-Reviewing-the-Socio-Economic-Gradient-TECHNICAL-Aug-23.pdf

Howie, S. J., Combrinck, C., Tshele, M., Roux, K., McLeod Palane, N., & Mokoena, G. (2017). *PIRLS 2016: South African highlights report*. https://repository.up.ac.za/bitstream/handle/2263/66183/PIRLS%202016%20HL%20Report_12.12.2017_FINAL.pdf?sequence=1

Jago, L. S., Monaghan, P., Alcock, K., & Cain, K. (2025). The effect of preschool vocabulary and grammar on early reading comprehension and word reading: A systematic review and meta-analysis. *Educational Research Review*, *47*, 1000680. https://doi.org/10.1016/j.edurev.2025.100680

Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: a review and recent developments. *Philosophical transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150202. https://doi.org/10.1098/rsta.2015.0202

Kajee, L., & Sibanda, R. (2019). Home as a primary space: Exploring out-of-school literacy practices in early childhood education in a township in South Africa. *South African Journal of Childhood Education*, *9*(1), 1-10. https://doi.org/https://doi.org/10.4102/sajce.v9i1.686

Kendeou, P., van den Broek, P., White, M., & Lynch, J. S. (2009). Predicting reading comprehension in early elementary school: the independent contributions of oral language and decoding skills. *Journal of Educational Psychology*, *101*, 765-778. https://doi.org/10.1037/a0015956

Klop, D., & Visser, M. (2020). Using MAIN in South Africa. *ZAS Papers in Linguistics*, *64*, 207-210. https://doi.org/10.21248/zaspil.64.2020.575

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Language and Reading Research Consortium (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, *50*, 151-169. https://doi.org/10.1002/rrq.99

Language and Reading Research Consortium & Logan, J. (2017). Pressure points in reading comprehension: a quantile multiple regression analysis. *Journal of Educational Psychology*, *109(4),* 451-464. https://doi.org/10.1037/edu0000744: 10.1037/edu0000150

Language and Reading Research Consortium, Lo, M., & Xu, M. (2022). Impacts of the Let's Know! curriculum on the language and comprehension-related skills of prekindergarten and kindergarten children. *Journal of Educational Psychology*, *114*, 1205–1224. https://doi.org/10.1037/edu0000744

Lendrum, A., & Humphrey, N. (2012). The importance of studying the implementation of interventions in school settings. *Oxford Review of Education*, *38*(5), 635-652. https://doi.org/https://doi.org/10.1080/03054985.2012.734800

Leon, J., & Singh, A. (2017). Equating test scores for receptive vocabulary across rounds and cohorts in Ethiopia, India and Vietnam. *Young LIves Technical Report*, *40*, 1-116.

Maviş, İ., Tunçer, M., & Gagarina, N. (2016). Macrostructure components in narrations of Turkish–German bilingual children. Applied Psycholinguistics, 37(1), 69-89. https://doi.org/10.1017/S0142716415000429

Meiklejohn, C., Westaway, L., Westaway, A. F., & Long, K. A. (2021). A review of South African primary school literacy interventions from 2005 to 2020. *South African Journal of Childhood Education*, *11*(1), 1-11. https://doi.org/http://dx.doi.org/10.4102/sajce.v11i1.919

Mendez, M. A., & Adair, L. S. (1999). Severity and timing of stunting in the first two years of life affect performance on cognitive tests in late childhood. *The Journal of Nutrition*, *129*(8), 1555-1562. https://doi.org/https://doi.org/10.1093/jn/129.8.1555

Mullis, I. V. S., von Davier, M., Foy, P., Fishbein, B., Reynolds, K. A., & Wry, E. (2023). *PIRLS 2021 International Results in Reading*. Boston College, TIMSS & PIRLS International Study Center. https://doi.org/10.6017/lse.tpisc.tr2103.kb5342

Naidoo, U., Reddy, K., & Dorasamy, N. (2014). Reading literacy in primary schools in South Africa: Educator perspectives on factors affecting reading literacy and strategies for improvement. *International Journal of Educational Sciences*, *7*(1), 155-167. https://doi.org/10.1080/09751122.2014.11890179

Naudé, H., Pretorius, E., & Viljoen, J. (2003). The impact of impoverished language development on preschoolers' readiness-to-learn during the foundation phase. *Early Child Development and Care*, *173*(2-3), 271-291. https://doi.org/DOI: 10.1080/03004430303098

O'Carroll, S. (2011). An exploratory study of early letter-sound knowledge in a low socio-economic context in South Africa. *Reading and Writing*, *2*(1), 7-26. https://doi.org/doi.org/10.4102/rw.v2i1.10

O'Carroll, S., Oakhill, J., Cain, K., Klop, D., & Visser, M. (2023). *Research brief 1. Little Stars: A low-cost, open-source language and literacy professional development programme for teachers of 4- and 5-year-olds*. https://www.wordworks.org.za/wp-content/uploads/2023/10/BA-LS-research-brief-1_E_low-res.pdf

Oakhill, J., & Cain, K. (2012). The precursors of reading ability in young readers: Evidence from a four-year longitudinal study. *Scientific Studies of Reading*, *16*(2), 91-121. https://doi.org/10.1080/10888438.2010.529219

Oakhill, J., Cain, K., O'Carroll, S., Visser, M., & Klop, D. (in preparation). The positive impact of professional development and implementation of an intervention on pre-school teachers' language and literacy knowledge and practice.

Pedder, D. (2006). Are small classes better? Understanding relationships between class size, classroom processes and pupils' learning. *Oxford Review of Education*, *32*(2), 213-234. https://doi.org/10.1080/03054980600645396

Perfetti, C. A., & Stafura, J. (2014). Reading comprehension: Including word knowledge in a theoretical framework. *Scientific Studies of Reading*, *18*(1), 22-37. https://doi.org/10.1080/10888438.2013.827687

Perold Potgieter, A., & Southwood, F. (2016). A comparison of proficiency levels in 4-year-old monolingual and trilingual speakers of Afrikaans, isiXhosa and South African English across SES boundaries, using LITMUS-CLT. *Clinical Linguistics and Phonetics*, *30*, 87-100. https://doi.org/ 10.3109/02699206.2015.1110715

Sénéchal, M. (2006). Testing the home literacy model: Parent involvement in Kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific Studies of Reading*, *10*, 59-87. 10.1207/s1532799xssr1001_4

Silva, M. T., & Cain, K. (2015). The relations between lower- and higher-level oral language skills and their role in prediction of early reading comprehension *Journal of Educational Psychology*, *107*(2), 321-331. https://doi.org/10.1037/a0037769

Siraj-Blatchford, I., Taggart, B., & Sylva, K. (2010). *ECERS-E: The Early Childhood Environment Rating Scale Curricular Extension to ECERS-R* (4 ed.). Trentham Books Ltd.

Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. G. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning program outcomes. *Child: Care, Health and Development*, *45*(2), 257-270. https://doi.org/10.1111/cch.12641

Sylva, K., Siraj-Blatchford, I., Taggart, B., Sammons, P., Melhuish, E., Elliot, K., & Totsika, V. (2006). Capturing quality in early childhood through environmental rating scales. *Early Childhood Research Quarterly*, *21*(1), 76-92. https://doi.org/10.1016/j.ecresq.2006.01.003

Tredoux, C., Dawes, A., Mattes, F., Schenk, J. C., Giese, S., Leach, G., van der Berg, S., & Horler, J. (2024). Are South African children on track for early learning? Findings from the South African Thrive By Five Index 2021 Survey. *Child Indicators Research*, *17*(2), 601-636. https://doi.org/10.1007/s12187-023-10093-3

Tredoux, C. G., & Dawes, A. (2018). *Predictors of mathematics and literacy skills at fifteen years in Ethiopia, India, Peru and Vietnam*. Young Lives https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-WP179.pdf

UNICEF. (2017). *Overview: MELQO: Measuring early learning quality and outcomes*. UNESCO Publishing.

van de Vijver, F., & Tanzer, N. K. (2004). Bias and equivalence in cross-cultural assessment: An overview. *European Review of Applied Psychology / Revue Européenne de Psychologie Appliquée*, *54*(2), 119–135. https://doi.org/https://doi.org/10.1016/j.erap.2003.12.004

Visser, M., O'Carroll, S., Oakhill, J., Cain, K., & de Wet, K. (in press). Teachers at the core: Exploring the usefulness, feasibility, and implementation of an early literacy and language programme. In *Reimagining Early Childhood Care and Education (ECCE): Bridging Global Practices with 21st Century Innovations*.

Wechsler, D. (2012). *WPPSI-IV: Wechsler preschool and primary scale of intelligence*. NCS Pearson, Inc.

Wills, G., Ardington, C., & Sebaeng, L. (2022). Foundational skills in home language reading in South Africa: Empirical evidence from 2015–2021. In N. Spaull & E. Pretorius (Eds.), *Early grade reading in South Africa* (pp. 37-63). Oxford University Press.

Wordworks. (2023). *Little Stars: Supporting early language and literacy teaching in pre-Grade R*. https://www.wordworks.org.za/wp-content/uploads/2023/08/Programme-information-sheets-2023_Little-stars.pdf

World Literacy Foundation. (2015). *The economic and social cost of illlliteracy: A snapshot of illiteracy in a global context*. W. L. Foundation. https://worldliteracyfoundation.org/wp-content/uploads/2015/02/WLF-FINAL-ECONOMIC-REPORT.pdf

Yeomans-Maldonado, G., Justice, L. M., & Logan, J. A. (2019). The mediating role of classroom quality on peer effects and language gain in pre-kindergarten ECSE classrooms. *Applied Developmental Science*, *23*, 90-103. https://doi.org/10.1080/10888691.2017.1321484

Table 1.

Participant characteristics at Time 1 (pre-test)

		isiXhosa	**isiXhosa**		**Afrikaans**	**Afrikaans**	
		Intervention	**Control**	t-test	**Intervention**	**Control**	t-test
		(N=42; 20 girls)	(N=40; 16 girls)	(df=80)	(N=60; 29 girls)	(N=58; 25 girls)	(df=116)
Height for age*	Mean (SD)	-.321 (.817)	.002 (1.04)	*t* = 1.338	-.429 (1.77)	-.115 (1.06)	*t* = 1.294
(standardised score)	*min - max*	*-2.61 - .972*	*-1.64 – 1.99*	*p* = .186	*-2.99 – 2.86*	*-3.09 – 1.98*	*p* = .167
Age in months	Mean (SD)	54.67 (3.44)	56.08 (3.14)	*t* = 1.933	55.35 (3.62)	54.67 (3.70)	*t* = 1.007
	min - max	*50 - 61*	*50 - 61*	*p* = .0568	*50 - 64*	*50 - 61*	*p* = .316
Cross-linguistic	Mean (SD)	-0.56 (0.92)	-0.69 (1.18)	*t* = 0.559	-0.56 (0.92)	0.39 (1.02)	*t* = 0.365
Lexical Task (factor	*min - max*	*-2.12 – 1.61*	*-3.61 – 1.52*	*p* = .577	*-2.12 – 1.61*	*-2.23 – 2.02*	*p* = .716
score)							

*Missing data. Means calculated for 59 isiXhosa (34 intervention, 25 control; df = 57) and 100 Afrikaans (54 intervention, 46 control; df = 98).

Table 2.

Descriptives for child language and reading-related measures at Time 1 (pre-test) and Time 2 (post-test after 26 weeks of intervention)

		Time 1			Time 2		
		isiXhosa					
		Intervention	**Control**	**t-test (df=80)**	**Intervention**	**Control**	**t-test (df=80)**
ELL raw score	Mean (SD)	10.11 (3.45)	10.41 (3.42)	*t* = -0.390	12.07 (3.83)	12.95 (3.64)	*t* = -1.059
	min-max	*3.19 - 18.28*	*3.72 - 19.27*	*p* = 0.697	*4.44 - 19.27*	*5.59 - 18.49*	*p* = 0.293
ELP factor score	Mean (SD)	-0.44 (1.09)	-0.25 (0.95)	*t* = -0.874	0.47 (1.16)	0.61 (0.90)	*t* = -0.603
	min-max	*-2.77 - 1.80*	*-2.51 - 2.40*	*p* = 0.385	*-2.21 - 2.65*	*-1.46 - 2.65*	*p* = 0.548
MAIN factor score	Mean (SD)	-1.16 (0.98)	-0.99 (1.02)	*t* = -0.795	0.37 (1.36)	0.34 (1.25)	*t* = 0.100
	min-max	*-2.36 - 2.22*	*-2.56 - 0.82*	*p* = 0.429	*-2.36 - 4.03*	*-1.89 - 2.69*	*p* =.920
PVT raw score	Mean (SD)	-	-	-	15.57 (5.90)	11.93 (4.39)	*t* = 3.163
	min-max	-	-	-	*7.00 - 28.00*	*4.00 - 23.00*	*p* =.002
		Afrikaans					
Afrikaans		**Intervention**	**Control**	**t-test (df=116)**	**Intervention**	**Control**	**t-test (df=116)**
ELL raw score	Mean (SD)	12.02 (4.74)	12.50 (4.62)	*t* = -0.558	14.75 (4.22)	13.87 (3.67)	*t* = 1.209
	min-max	*2.25 - 20.00*	*0.00 - 20.00*	*p* = 0.578	*2.25 - 20.00*	*4.17 - 20.00*	*p* = 0.229
ELP factor score	Mean (SD)	-0.43 (1.11)	-0.74 (1.02)	*t* = 1.553	0.80 (1.33)	0.09 (1.07)	*t* = 3.162
	min-max	*-2.82 - 2.20*	*-2.57 - 1.90*	*p* = 0.123	*-2.26 - 2.95*	*-1.51 - 2.70*	*p* = 0.002
MAIN factor score	Mean (SD)	-0.11 (1.36)	-0.76 (1.37)	*t* = 2.573	1.25 (1.37)	0.61 (1.28)	*t* = 2.586
	min-max	*-2.36 - 2.89*	*-2.56 - 3.01*	*p* = 0.011	-2.36 - 3.76	-1.56 - 3.76	*p* = 0.011
PVT raw score	Mean (SD)	-	-	-	27.78 (5.91)	25.64 (4.75)	*t* = 2.168
	min-max	-	-	-	*14.00 - 38.00*	*15.00 - 38.00*	*p* = 0.032

Note. SD = standard deviation; ELL = Emergent Language and Literacy; ELP = Early Literacy Protocol Factor Score; MAIN = Multilingual Assessment Instrument for Narrative; PVT = Proximal Vocabulary Test.

Table 3.

Teacher experience and classroom quality factor scores and class fees

		isiXhosa	isiXhosa		Afrikaans	Afrikaans	
		Intervention	**Control**	**t-test**	**Intervention**	**Control**	**t-test**
		(N=11)	(N=11)	(df=20)	(N=12)	(N=12)	(df=22)
Teacher experience	Mean (SD)	-1.06 (0.79)	-0.01 (1.37)	*t* = -2.200	0.84 (1.13)	-0.20 (1.30)	*t* = 2.091
factor score	*min - max*	*-2.56 – 0.02*	*-2.51 – 1.45*	*p* = .039	*-1.32 – 3.28*	*-2.60 – 2.16*	*p* = .048
Classroom quality	Mean (SD)	-0.01 (1.94)	-.48 (1.12)	*t* = 0.686	0.33 (1.45)	-0.16 (1.67)	*t* = 0.777
factor score	*min - max*	*-2.51 – 4.02*	*-2.30 – 1.42*	*p* = .500	*-1.56 – 3.64*	-2.88 – 3.27	*p* = .446
Class fees	Mean (SD)	275.45 (40.34)	288.18 (62.74)	*t* = -.566	447.50 (125.70)	400.83 (166.81)	*t* = 0.774
	min - max	*200.00 - 300.00*	*200.00 - 400.00*	*p* = .578	*120.00 – 650.00*	*0 – 600.00*	*p* = .447

Note. SD = standard deviation

Table 4.

isiXhosa language group: Summary of models to determine predictors for early language and emergent literacy measures.

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Intervention (and baseline) only (N=82)		**Individual covariates only (N=82)**		**Adding classroom covariates (no RE) (N=82)**		**Adding classroom covariates (with RE) (N=82)**	
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
ELL post-test								
Implementation quality	-0.01 (0.02)	.46	-0.01 (0.02)	.46	-0.00 (0.02)	.83	-0.00 (0.02)	.84
ELL pre-test	0.71 (0.09)	<.001	0.71 (0.09)	<.001	0.64 (0.10)	<.001	0.64 (0.10)	<.001
Teacher experience					0.26 (0.29)	37	0.26 (0.29)	.37
Classroom quality					-0.02 (0.22)	.92	-0.02 (0.22)	.92
Class fees					0.01 (0.01)	.12	0.01 (0.01)	.14
SD of random effects							0.00	
Log-likelihood (ML)	-201.62		-201.62		-199.91		-199.91	
Likelihood ratio test χ^2					3.42	.33	0.00	1.00
R^2	0.42		0.42		0.45		0.43	
ELP post-test								
Implementation quality	0.00 (0.01)	.53	0.00 (0.01)	.53	0.01 (0.01)	.35	0.01 (0.01)	.57
ELP pre-test	0.60 (0.09)	<.001	0.60 (0.09)	<.001	0.55 (0.10)	<.001	0.54 (0.10)	<.001
Teacher experience					-0.08 (0.09)	.38	-0.07 (0.11)	.53
Classroom quality					-0.04 (0.06)	.58	-0.03 (0.09)	.72
Class fees					0.01 (0.00)	.01	0.01 (0.00)	.07
SD of random effects							0.40	
Log-likelihood (ML)	-101.11		-101.11		-96.60		-96.07	

Likelihood ratio test χ^2				9.02	.03	3.99	.05
R^2	0.35		0.35	0.42		0.52	
MAIN post-test							
Implementation quality	0.00 (0.01)	.63	0.00 (0.01)	.63	0.01 (0.01)	.38	0.01 (0.01)	.39
MAIN pre-test	0.34 (0.14)	.02	0.34 (0.14)	.02	0.25 (0.15)	.09	0.25 (0.15)	.09
Teacher experience					0.24 (0.13)	.07	0.24 (0.13)	.06
Classroom quality					0.04 (0.10)	.70	0.04 (0.10)	.70
Class fees					0.00 (0.00)	.32	0.00 (0.00)	.33
SD of random effects							0.00	
Log-likelihood (ML)	-134.49		-134.49		-131.91		-131.91	
Likelihood ratio test χ^2					5.16	.16	0.00	1.00
R^2	0.07		0.07		0.12		0.12	
PVT post-test							
Implementation quality	0.08 (0.03)	.02	0.08 (0.03)	.01	0.07 (0.03)	.04	0.07 (0.04)	.07
CLT			2.72 (0.48)	<.001	2.79 (0.50)	<.001	2.62 (0.50)	<.001
Teacher experience					-0.26 (0.47)	.57	-0.26 (0.54)	.63
Classroom quality					0.18 (0.35)	.60	0.16 (0.41)	.70
Class fees					-0.00 (0.01)	.84	-0.00 (0.01)	.93
SD of random effects							1.45	
Log-likelihood (ML)	-252.76		-238.68		-238.38		-238.71	
Likelihood ratio test χ^2			28.16	<.001	0.60	.90	0.84	.36
R^2	0.07		0.34		0.34		0.37	

Note. ELL = Emergent Language and Literacy; ELP = Early Literacy Protocol Factor Score; MAIN = Multilingual Assessment Instrument for Narrative; PVT = Proximal Vocabulary Test; CLT = Cross-linguistic Lexical Task. RE = random effects; SE = standard error; SD = standard deviation; ML = Maximum Likelihood.

Table 5.

Afrikaans language group: Summary of models to determine predictors for early language and emergent literacy measures.

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Intervention (and baseline) only (N=118)		**Individual covariates only (N=118)**		**Adding classroom covariates (no RE) (N=118)**		**Adding classroom covariates (with RE) (N=118)**	
	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value	Estimate (SE)	p-value
ELL post-test								
Implementation quality	0.03 (0.02)	.17	0.02 (0.02)	.38	0.03 (0.02)	.11	0.03 (0.03)	.20
ELL pre-test	0.30 (0.07)	<.001	0.14 (0.08)	.07	0.09 (0.08)	.28	0.12 (0.08)	.12
CLT			1.29 (0.30)	<.001	1.55 (0.31)	<.001	1.41 (0.32)	<.001
Teacher experience					0.03 (0.29)	.92	0.03 (0.36)	.94
Classroom quality					-0.68 (0.24)	.01	-0.63 (0.30)	.05
Class fees					0.00 (0.00)	.82	0.00 (0.00)	.92
SD of random effects							1.26	
Log-likelihood (ML)	-320.55		-311.88		-307.22		-306.97	
Likelihood ratio test χ^2			17.33	.00	9.33	.03	2.89	.09
R^2	0.14		0.26		0.32		0.39	
ELP post-test								
Implementation quality	0.01 (0.01)	.02	0.01 (0.01)	.02	0.01 (0.01)	.07	0.01 (0.01)	.16
ELP pre-test	0.57 (0.09)	.00	0.57 (0.09)	<.001	0.60 (0.09)	<.001	0.59 (0.10)	<.001
Teacher experience					0.21 (0.09)	.02	0.21 (0.12)	.10
Classroom quality					-0.15 (0.07)	.03	-0.15 (0.10)	.13
Class fees					-0.00 (0.00)	.75	-0.00 (0.00)	.77

SD of random effects						0.46	
Log-likelihood (ML)	-172.12		-172.12		-167.47	-166.34	
Likelihood ratio test χ^2					9.30	.03	5.47	.02
R^2	0.31		0.31		0.36	0.47	
MAIN post-test							
Implementation quality	0.01 (0.01)	.08	0.01 (0.01)	.09	0.01 (0.01)	.20	0.01 (0.01)	.27
MAIN pre-test	0.39 (0.08)	.00	0.25 (0.09)	.01	0.27 (0.10)	.01	0.24 (0.09)	.01
CLT			0.35 (0.10)	<.001	0.32 (0.11)	<.001	0.32 (0.11)	<.001
Teacher experience					-0.03 (0.10)	.75	-0.03 (0.13)	.83
Classroom quality					0.09 (0.08)	.31	0.09 (0.11)	.41
Class fees					-0.00 (0,00)	.95	-0.00 (0.00)	.95
SD of random effects						0.48	
Log-likelihood (ML)	-188.65		-182.76		-182.16	-181.29	
Likelihood ratio test χ^2			11.78	<.001	1.21	.75	4.20	.04
R^2	0.22		0.30		0.30	0.39	
PVT post-test							
Implementation quality	0.09 (0.03)	<.001	0.06 (0.02)	<.001	0.06 (0.02)	.02	0.05 (0.02)	.02
CLT			3.25 (0.28)	<.001	3.15 (0.29)	<.001	3.15 (0.29)	<.001
Teacher experience					-0.06 (0.30)	.83	-0.06 (0.30)	.83
Classroom quality					0.05 (0.25)	.83	0.05 (0.25)	.84
Class fees					0.00 (0.00)	.14	0.00 (0.00)	.15
SD of random effects						0.00	
Log-likelihood (ML)	-361.85		-315.07		-313.62	-313.62	

Likelihood ratio test χ^2			93.57	<.001		2.90	.41		0.00	1.00
R^2	0.09	0.59			0.60			0.59		

Note. ELL = Emergent Language and Literacy; ELP = Early Literacy Protocol Factor Score; MAIN = Multilingual Assessment Instrument for Narrative; PVT = Proximal Vocabulary Test; CLT = Cross-linguistic Lexical Test; RE = random effects; SE = standard error; SD = standard deviation; ML = Maximum Likelihood.

Table 6.

isiXhosa language group: Summary of models to determine predictors for each ELOM domain and ELOM total scores

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Intervention (and baseline) only (*N*=82)		**Individual covariates only (N=82)**		**Adding classroom covariates (no RE) (*N*=82)**		**Adding classroom covariates (with RE) (*N*=82)**	
	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value
ELOM GM								
Implementation quality	0.01 (0.02)	.92	0.01 (0.02)	.92	-0.01(0.02)	.90	-0.01 (0.02)	.90
ELOM GM pre-test	0.22 (0.08)	.01	0.22 (0.08)	.01	0.18 (0.08)	.04	0.18 (0.08)	.04
Teacher experience					-0.45 (0.32)	.17	-0.45 (0.32)	.17
Classroom quality					0.19 (0.25)	.46	0.19 (0.25)	.47
Class fees					0.01 (0.01)	.08	0.02 (0.01)	.10
SD of random effects							0.00	
Log-likelihood (ML)	-212.55		-212.55		-210.01		-210.01	
Likelihood ratio test χ^2			0.00	1.00	5.07	.17	0.00	1.00
R^2	0.09		0.09		0.15		0.14	
ELOM FM&VI								
Implementation quality	0.01 (0.02)	.84	0.01 (0.02)	.84	-0.02 (0.02)	.35	-0.02 (0.02)	.37
ELOM FM&VI pre-test	0.42 (0.11)	< .001	0.42 (0.11)	< .001	0.43 (0.10)	< .001	0.42 (0.10)	< .001
Teacher experience					-0.44 (0.26)	.10	-0.44 (0.27)	.11
Classroom quality					0.25 (0.21)	.24	0.25 (0.21)	.26
Class fees					-0.01 (0.01)	.05	-0.01 (0.01)	.07

SD of random effects						0.18	
Log-likelihood (ML)	-197.92		-197.92		-193.53	-193.57	
Likelihood ratio test χ^2			0.00	1.00	8.77	.03	0.01	.97
R^2	0.17		0.17		0.25	0.24	
ELOM EMN							
Implementation quality	-0.01 (0.02)	.68	-0.01 (0.02)	.68	-0.01(0.02)	.70	-0.01 (0.03)	.88
ELOM EMN pre-test	0.54 (0.09)	< .001	0.54 (0.09)	< .001	0.54 (0.09)	< .001	0.52 (0.09)	< .001
Teacher experience					0.17 (0.32)	.60	0.21 (0.40)	.60
Classroom quality					0.19 (0.25)	.44	0.18 (0.31)	.58
Class fees					0.01 (0.01)	.47	0.01(0.01)	.46
SD of random effects						1.28	
Log-likelihood (ML)	-208.75		-208.75		-207.97	-208.04	
Likelihood ratio test χ^2			0.00	1.00	1.55	.67	2.06	.15
R^2	0.32		0.32		0.33	0.41	
ELOM C&EF							
Implementation quality	0.02 (0.02)	.21	0.02 (0.02)	.21	0.04 (0.02)	.09	0.04 (0.02)	.10
ELOM C&EF pre-test	0.67 (0.08)	< .001	0.67 (0.08)	< .001	0.65 (0.09)	< .001	0.65 (0.09)	< .001
Teacher experience					0.13 (0.29)	.66	0.13 (0.29)	.66
Classroom quality					-0.29 (0.22)	.20	-0.29 (0.22)	.22
Class fees					0.01 (0.01)	.55	0.01 (0.01)	.56
SD of random effects						0.00	
Log-likelihood (ML)	-200.83		-200.83		-199.58	-199.58	
Likelihood ratio test χ^2			0.00	1.00	2.50	.48	0.00	1.00
R^2	0.44		0.44		0.46	0.45	

ELOM total								
Implementation quality	-0.01 (0.05)	.81	-0.01 (0.05)	.81	-0.02 (0.06)	.71	-0.02 (0.06)	.71
ELOM total pre-test	0.70 (0.07)	< .001	0.70 (0.07)	< .001	0.70 (0.07)	< .001	0.70 (0.07)	< .001
Teacher experience					-0.30 (0.78)	.70	-0.30 (0.78)	.70
Classroom quality					0.25 (0.60)	.69	0.25 (0.60)	.69
Class fees					0.01(0.02)	.72	0.01 (0.02)	.72
SD of random effects							0.01	
Log-likelihood (ML)	-282.11		-282.11		-281.91		-281.80	
Likelihood ratio test χ^2			0.00	1.00	0.39	.94	0.00	1.00
R^2	0.56		0.56		0.56		0.54	

Notes: ELOM = Early Learning Objectives Measure; GM = gross motor development; FM&VI = fine motor coordination and visual integration; EMN = emergent numeracy and mathematics; C&EF = cognition and executive functioning; RE = random effects; SE = standard error; SD = standard deviation; ML = Maximum Likelihood.

Table 7.

isiXhosa language group: Number of children (and proportion) in intervention and control group who were classified as On Track at Time 1 (pre-test) and Time 2 (post-test) on ELOM domains

	Intervention (N=42)		**Control (N=40)**	
	Time 1	**Time 2**	**Time 1**	**Time 2**
ELOM Total Time 1	16 (.38)	34 (.81)	18 (.45)	24 (.60)
ELOM GM Time 1	25 (.60)	35 (.83)	24 (.60)	30 (.75)
ELOM FM&VI Time 1	8 (.19)	21 (.50)	6 (.15)	15 (.38)
ELOM EMN Time 1	14 (.33)	25 (.60)	13 (.33)	23 (.58)
ELOM C&EF Time 1	19 (.45)	26 (.62)	18 (.45)	19 (.48)
ELOM ELL Time 1	23 (.55)	27 (.64)	20 (.50)	26 (.65)

Note. ELOM = Early Learning Objectives Measure; GM = gross motor development; FM&VI = fine motor coordination and visual integration; EMN = emergent numeracy and mathematics; C&EF = cognition and executive functioning; ELL = emergent language and literacy.

Table 8.

Afrikaans language group: Summary of models to determine predictors for each ELOM domain and ELOM total scores

	Model 1		**Model 2**		**Model 3**		**Model 4**	
	Intervention (and baseline) only (*N*=118)		**Individual covariates only (N=118)**		**Adding classroom covariates (no RE) (*N*=118)**		**Adding classroom covariates (with RE) (*N*=118)**	
	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value	Estimate (SE)	*p*-value
ELOM GM								
Implementation quality	0.07 (0.02)	.01	0.07	0.02	0.06 (0.02)	.01	0.06 (0.04)	.12
ELOM GM pre-test	0.46 (0.07)	<.001	0.46	0.07	0.47 (0.07)	<.001	0.43 (0.08)	<.001
Teacher experience					0.66 (0.33)	.05	0.65 (0.53)	.24
Classroom quality					-0.32 (0.27)	.25	-0.31 (0.44)	.49
Class fees					-0.01 (0.01)	.06	-0.01 (0.01)	.20
SD of random effects							2.39	
Log-likelihood (ML)	-330.85		-330.85		-326.09		-320.10	
Likelihood ratio test χ^2			0.00	1.00	9.51	.12	16.84	<.001
R^2	0.283		0.28		0.34		0.54	
ELOM FM&VI								
Implementation quality	0.03 (0.02)	.04	0.04 (0.02)	.03	0.03 (0.02)	.17	0.03 (0.03)	.35
ELOM FM&VI pre-test	0.50 (0.07)	<.001	0.45 (0.07)	<.001	0.51 (0.08)	<.001	0.48 (0.08)	<.001
Child age			0.18 (0.08)	.02	0.15 (0.08)	.07	0.11 (0.07)	.14
Teacher experience					0.44 (0.25)	.09	0.47 (0.39)	.25
Classroom quality					-0.01 (0.20)	.99	0.04 (0.32)	.92

Class fees					-0.01 (0.01)	.13	-0.01 (0.00)	.36
SD of random effects							1.74	
Log-likelihood (ML)	-294.20		-291.54		-288.97		-284.07	
Likelihood ratio test χ^2			5.32	0.02	5.14	.16	14.20	<.001
R^2	0.38		0.41		0.43		0.59	
ELOM EMN								
Implementation quality	0.05 (0.02)	.02	0.04 (0.02)	.03	0.02 (0.02)	.36	0.02 (0.03)	.50
ELOM EMN pre-test	0.47 (0.07)	<.001	0.30 (0.09)	.01	0.32 (0.10)	<.001	0.21 (0.10)	.04
CLT			1.03 (0.36)	.01	1.04 (0.37)	.01	1.25 (0.35)	<.001
Teacher experience					0.77 (0.30)	.01	0.82 (0.44)	.08
Classroom quality					0.01 (0.25)	.99	-0.04 (0.37)	.91
Class fees					-0.01 (0.00)	.49	-0.01 (0.01)	.70
SD of random effects							1.87	
Log-likelihood (ML)	-319.63		-315.58		-311.77		-309.07	
Likelihood ratio test χ^2			8.11	0.01	7.61	.06	9.37	.01
R^2	0.30		0.35		0.39		0.51	
ELOM C&EF								
Implementation quality	0.05 (0.02)	.01	0.04 (0.02)	.02	0.03 (0.02)	.15	0.03 (0.03)	.34
ELOM C&EF pre-test	0.62 (0.09)	<.001	0.43 (0.10)	<.001	0.41 (0.10)	<.001	0.44 (0.11)	<.001
Teacher experience			1.15 (0.30)	<.001	1.24 (0.31)	<.001	1.24 (0.31)	. <.001
CLT					0.56 (0.27)	.04	0.60 (0.37)	.12
Classroom quality					-0.25 (0.24)	.29	-0.22 (0.31)	.49
Class fees					0.01 (0.00)	.68	0.01 (0.03)	.86
SD of random effects							1.44	

Log-likelihood (ML)	-311.04		-303.65		-301.20		-299.84	
Likelihood ratio test χ^2			14.78	.00	4.89	.18	5.88	.02
R^2	0.36		0.43		0.46		0.56	
ELOM total								
Implementation quality	0.22 (0.06)	.01	0.20 (0.06)	.01	0.16 (0.07)	.02	0.16 (0.12)	.21
ELOM total pre-test	0.59 (0.07)	<.001	0.44 (0.09)	<.001	0.44 (0.09)	<.001	0.49 (0.09)	<.001
CLT			3.14 (1.19)	.01	3.62 (1.22)	.01	3.17 (1.02)	.01
Teacher experience					2.41 (0.95)	.01	2.53 (1.70)	.15
Classroom quality					-0.95 (0.81)	.24	-0.73 (1.43)	.62
Class fees					-0.01 (0.01)	.25	-0.01 (0.01)	.38
SD of random effects							8.27	
Log-likelihood (ML)	-456.62		-453.12		-448.73		-433.49	
Likelihood ratio test χ^2			7.01	.01	8.79	.03	36.07	<.001
R^2	0.44		0.48		0.51		0.75	

Notes: ELOM = Early Learning Objectives Measure; GM = gross motor development; FM&VI = fine motor coordination and visual integration; EMN = emergent numeracy and mathematics; C&EF = cognition and executive functioning; CLT = Cross-Linguistic Vocabulary Test; RE = random effects; SE = standard error; SD = standard deviation; ML = Maximum Likelihood.

Table 9.

Afrikaans language group: Number of children (and proportion) in each group who were classified as On Track at Time 1 (pre-test) and Time 2 (post-test) on ELOM domains

	Intervention (N=59)		**Control (N=58)**	
	Time 1	**Time 2**	**Time 1**	**Time 2**
ELOM Total Time 1	33 (.55)	46 (.77)	30 (.52)	41 (.71)
ELOM GM Time 1	23 (.38)	36 (.60)	31 (.53)	31 (.53)
ELOM FM&VI Time 1	28 (.47)	42 (.70)	24 (.41)	30 (.52)
ELOM EMN Time 1	29 (.48)	48 (.80)	27 (.47)	38 (.66)
ELOM C&EF Time 1	28 (.47)	37 (.62)	26 (.45)	29 (.50)
ELOM ELL Time 1	39 (.65)	48 (.80)	41 (.71)	45 (.78)

Note. ELOM = Early Learning Objectives Measure; GM = gross motor development; FM&VI = fine motor coordination and visual integration; EMN = emergent numeracy and mathematics; C&EF = cognition and executive functioning; ELL = emergent language and literacy.

Highlights

Teacher-delivered randomised controlled trial of South African preschoolers.

Story-based intervention designed to improve foundational skills for reading.

Intervention implementation quality influenced learning of target vocabulary.

Post-intervention, more treatment children on track in key developmental domains.

Baseline performance predicted performance gains for treatment and control groups.

See separate file for Figures 1a and 1b