

Thrive by Five Index 2024: Technical Report

September 2025





About the Thrive by Five Index

The *Thrive by Five Index*, first launched in 2021, is the largest survey of preschool child outcomes in South Africa. It was established as a collaborative effort to track progress and promote positive change in Early Childhood Development.

The *Thrive by Five Index 2024* is the second in a planned series of cross-sectional, nationally and provincially representative studies conducted every three years. The 2024 survey was led by the Department of Basic Education and coordinated by DataDrive2030, with support from the FirstRand Foundation, the LEGO Foundation, This Day Foundation, and Yellowwoods.

Suggested citation

Pettersson Gelande, G., Giese, S., Dawes, A., Ardington, C., Tredoux, C., Cook, C., Carnegie, T., Brophy, T., Zastrau, E. (2025). *Thrive by Five Index 2024: Technical Report*. DataDrive2030, Cape Town.

www.thrivebyfive.co.za

Published: September 2025

Acknowledgements

The *Thrive by Five Index* is a testament to the power of collaboration. This report, and the broader effort it represents, would not have been possible without the dedication, expertise, and generosity of many individuals and organisations.

Our Funding Partners - This work was generously supported by: FirstRand Foundation, LEGO Foundation, This Day Foundation and Yellowwoods. These partners not only provide financial support but also share a deep commitment to advancing Early Childhood Development (ECD) in South Africa.

Project Oversight - We acknowledge the leadership of the national Department of Basic Education (DBE), under whose guidance the *Thrive by Five Index* is implemented, and the strategic oversight provided by the 2024 *Index's* Steering and Communications Committees.

Project Management - The *Thrive by Five Index* is coordinated by DataDrive2030, a non-profit organisation committed to unlocking the power of data to improve outcomes for young children. We would especially like to recognise the extraordinary dedication of every member of the DataDrive2030 team, who demonstrated, time and again, that this work is so much more than a job.

Technical Contributors - This study was designed, implemented and quality assured with expert input from a multidisciplinary team with deep technical expertise in ECD, survey design, assessor training, fieldwork management, data processing, analysis, and evidence communication. Their collective knowledge ensured that the *Thrive by Five Index* measures what truly matters and that the findings are technically robust and policy relevant.

Field Work Teams - We extend our heartfelt thanks to the listing teams, the fieldwork managers, regional coordinators, call centre operators, assessors and auditors whose tireless efforts made this study possible. Working under extremely challenging conditions, often in remote and resource-constrained settings, they demonstrated professionalism, patience, and perseverance.

Early Learning Programmes (ELPs) - We are deeply grateful to the many ELP principals and practitioners who participated in the 2024 round of the *Thrive by Five Index*. They form the foundation of South Africa's early learning landscape.

Primary Caregivers - We also wish to recognise the primary caregivers who allowed us into their homes, and trusted us with their children and their information, enabling us to understand the contextual realities of their lives more deeply.

Acknowledgements

Finally, we extend our deepest thanks to the thousands of 4-year-old children who took part in assessments across the country. Their curiosity, energy, and willingness to engage are what gave life and meaning to this study, and their contributions are invaluable.

Contributors

We gratefully acknowledge the following individuals for their contributions to this study. Names are listed alphabetically by organisation and surname.

Core team: Simone Beukes, Tyryn Carnegie, Caylee Cook, Andrew Dawes, Sonja Giese, Gunilla Pettersson Gelandar, Colin Tredoux.

DataDrive2030 (extended team and consultants): Joshua Arendse, Linda Biersteker, Brinthia Dawood, Svetlana Doneva, Melicia Fernandez, Alasdair Fraser, Elizabeth Girdwood, Aimee Hare, Callum Hollywood, Sibongile Khumalo, Jesal Kika-Mistry, Tawanda Madamombe, Sobongile Moyo, Tahmeed Omar, Zenzi Pahla, Cailyn Sonderup, Liezl Worship.

Dataquest: Shani Cooper, Zisanda Dladla, Sizani Mthembu, Shepard Muleya, Nonkululeko Seloba. Regional coordinators: Johannes Bodiba, Aziza Cassiem, Phumlani Mathe, Gugulethu Mbana, Fatima Mgiti, Athi Mqakelana, Thandeka Msebenzi, Selby Nkambule, Isaac Phukubje, Letlhogonolo Ramosepele.

Department of Basic Education: Janeli Kotze, Stephen Taylor.

Genesis Analytics: Sue Aitken, Morongwa Marutha, Isabella Rose Miller, Amaal Salie, Jan Schenk, Elzeth Zastrau.

Geospace: Leboke Bapela, Francois Bezuidenhout, Ruan Botha, Etienne de Fortier, Nthabiseng Mafoko, Lucas Mokwena, Thulani Nhlapo, Adriaan Voors.

SALDRU: Zuri Adonis, Victoria Basopo, Paidamoyo Bodzo, Timothy Brophy, Morne Hoffman, Kim Ingle, Alansa Klaasen, Lulama Mabaleka, Theophilus Mathoma, Andrew Partridge, Nobubele Tyembile.

University of Cape Town (UCT): Cally Ardington.

Other contributors: Megan Borole, Nicholas Dowdall, Kath Hall, Gregory Haugan, Grace Leach, Alicia Menendez, Mark Tomlinson, Servaas van der Berg, Gabrielle Wills.

As a team, we recognise the profound responsibility that comes with collecting and interpreting data about children's lives. We are deeply committed to ensuring that the Thrive by Five Index is fully leveraged to accelerate progress for young children across South Africa.

Table of Contents

List of Boxes, Figures, and Tables	01
1. Introduction	02
1.1 Study Objectives and Scope	03
1.2 Key Differences Between the 2021 and 2024 Index Rounds	05
1.3 2024 Index Responsibilities and High-Level Timeline	06
1.4 Structure of This Report	07
2. Survey Forms, Sampling And Sample Sizes	08
2.1 Survey Forms	08
Sub-Study: Survey Forms	12
2.2 Sampling Strategy and Weights	12
Sub-Study: Selection of Wards, Small Area Layers and Households	17
2.3 Sample Sizes	17
Sub-Study: Sample Sizes	21
3. Fieldwork Overview	22
3.1 ELP Listing and Sampling	23
3.2 ELP Data Collection	24
3.3 Collection of Data from Enrolled Children’s Primary Caregivers	29
Sub-Study: Household Listing and Non-Enrolled Data Collection	31
4. Data Preparation	34
4.1 The 2024 Index Datasets	34
4.2 Data Quality Checks and Cleaning	35
5. Some Considerations for the Next Index Round	39
Annex 1: LPQA Revisions, Scoring, Ratings and Factor Structure	40
Annex 2: Derivation of the New ELP Fee Variable	44
Annex 3: Elom Score Adjustments (Enrolled Children)	45
Annex 4: Non-Enrolled Sub-Study Supplemental Information	48
Annex 5: Ethics Overview	50
References	51

List of Boxes, Figures and Tables

Boxes

Box 1:	2024 Index Major Achievements	02
Box 2:	Developing the 2024 Index Fee Groups	36

Figures

Figure 1:	2024 Index High Level Timeline	07
Figure 2:	Map of ELPs Visited for the 2024 Index	18
Figure 3:	Data Handshakes Between the Two Service Providers	22
Figure 4:	Day at the ELP	27
Figure 5:	Example of an End of Day Report for the Child Sampling Form	29

Tables

Table 1:	Overview of 2024 Index Study Components	04
Table 2:	ELOM 4&5 Total and Domain Score Bands for Children Aged 50-59 Months	09
Table 3:	SEF Score Bands for Children Aged 50-59 Months	10
Table 4:	Allocation of Strata, Wards, ELPs and Children Across Provinces	13
Table 5:	Extract from Child Sampling Form: Child Eligibility Ladder	15
Table 6:	Expected Confidence Interval Widths for Provincial ELOM On Track Estimates	16
Table 7:	2024 Index Survey Forms and Sample Sizes (Enrolled Children)	19
Table 8:	ELP Verification Outcomes and Sampling Frame	23
Table 9:	Roll-Out of Fieldwork Across the Nine Provinces (Enrolled Children)	26
Table 10:	Primary Caregiver Interview Outcomes (Enrolled Children)	30
Table 11:	Household Listing and Screening Outcomes	31
Table 12:	Roll-Out of Fieldwork Across the Three Provinces (Sub-Study)	32
Table 13:	Datasets Received from the Service Providers	34
Table 14:	ELOM Data Investigation Areas, Checks and Outcomes	37

01 Introduction

The *Thrive by Five Index* (Index) produces indicators on key child outcomes – early learning levels, social-emotional functioning and physical growth – for children enrolled in Early Learning Programmes (ELPs) in South Africa. These indicators are used to assess and track child performance nationally. The study was designed in partnership with the Departments of Basic Education, Health, Planning, Monitoring and Evaluation, Social Development, and Statistics South Africa. It also included extensive consultations with various stakeholders engaged in the early childhood development (ECD) sector.

The *Index* was launched in 2021 to provide data that informs policy; to support the delivery of quality ECD services to preschool children; and to monitor South Africa's progress towards the realisation of local and global development goals, including UN Sustainable Development Goal (SDG) 4.2: all girls and boys have access to quality early childhood development, care and pre-primary education so that they are ready for primary education.

The *2024 Index* constitutes the second in this series of cross-sectional nationally and provincially representative surveys of the developmental status of children aged 50-59 months, to be conducted three-yearly until 2030. It follows the first round, conducted in 2021 amid the COVID-19 pandemic.

The process of generating the key child outcome and accompanying indicators has been long and complex, spanning from the study design phase to fieldwork preparation, data collection, and the subsequent checking and cleaning of the collected data. This Technical Report aims to describe the key features of the *2024 Index* study to ensure transparency, help guide analysis using the *2024 Index* data, and allow for lessons to be shared.

Findings based on the *2024 Index* data and insights are presented in a separate report, the *Thrive by Five Index 2024: National Findings*.

Box 1: 2024 Index Major Achievements



6000+

Early Learning Programmes listed in **432 wards** across South Africa



5000+

children aged **50-59 months** enrolled in ELPs assessed



3800+

primary caregivers of enrolled children interviewed



1300+

Early Learning Programmes observed and assessed



250+

non-enrolled children aged **50-59 months** assessed and their primary caregivers interviewed

1.1 Study Objectives and Scope

It is well established that the first five years of life are particularly crucial in laying the foundation for healthy physical, neurological, and psychological development. Children who are nourished and nurtured in the early years and have opportunities for learning in the home and in Early Learning Programmes (ELPs), are more likely to start school On Track in key areas of development such as emergent language and literacy, emergent numeracy and mathematics and social-emotional functioning (Duncan et al., 2007; Tran et al., 2017). A good early start also has positive effects on subsequent learning. Importantly, the benefits are typically greater for children from socioeconomically disadvantaged backgrounds (Watts et al., 2023).

The objectives of the Index are to:

- Determine the proportion of young children in South Africa who attend Early Learning Programmes (ELPs) who are On Track for their age in three key areas of development, early learning, social-emotional functioning (SEF) and physical growth
- Monitor trends in performance gaps and gains in this population over time and across socio-economic groups
- Explore a range of potential predictors of key child outcomes
- Track progress in the attainment of local and global development goals, including the Sustainable Development Goals (SDGs); and
- Inform decision-making and actions to improve key development outcomes for young children.

The *2024 Index* has two study components: children enrolled in ELPs, referred to as **enrolled children**, and those not enrolled in ELPs, referred to as the **sub-study of non-enrolled children**.

The main activities under the enrolled children component were the listing of ELPs in the selected study areas; data collection from the sample ELPs in these areas, including child assessments; and collection of data from the assessed children's primary caregivers (Table 1). The reason for the ELP listing was two-fold. First, to develop the sampling frame from which to select ELPs for the study on a rolling basis, which was necessary given the size of the fieldwork window (the 3rd of September to the 21st of November 2024), to accommodate end-of-year ELP closures. Second, to update the Department of Basic Education's database of ELPs in the selected study areas. The ELP listing was followed by the collection of data from the ELPs selected for the sample, assessment of children aged 50-59 months enrolled in these ELPs, and interviews with their primary caregivers. To be eligible for inclusion in the sample, ELPs had to be open at least eight hours per week and have at least one enrolled child aged 50-59 months. This component is national in scope, covering all of South Africa's nine provinces.

The sub-study of the non-enrolled children component is a test sample to assess the feasibility of reaching non-enrolled children. It is intended to allow for an indicative comparison of key child outcomes between enrolled and non-enrolled children after careful matching to account for the different nature and features of the two samples. This component also covers children aged 50-59 months, but who were not enrolled in an ELP in 2024 (strictly speaking, between January 2024 and the time of the data collection in October-November 2024).

In contrast to the enrolled component, it only covers three provinces, Gauteng, KwaZulu-Natal and Western Cape. Within these provinces, it covers children living in a variety of informal, rural and urban low-income areas (see Annex 4 for details). The main activities were the listing of households to identify children meeting the eligibility criteria for the sub-study, followed by assessment of the identified children and interviews with their primary caregivers.

Table 1: Overview of 2024 Index Study Components	
Enrolled Children	Sub-study of Non-Enrolled Children
<p>Listing of ELPs</p> <ul style="list-style-type: none"> • Listing of eligible ELPs in 432 wards/combined wards selected for the main sample, across all 9 provinces. • Construction of ELP sampling frame based on the listing, and sampling of ELPs. 	<p>Listing of households</p> <ul style="list-style-type: none"> • Listing of households with eligible children in 45 low-income wards in Gauteng, KwaZulu-Natal and Western Cape provinces.
Data Collection	
<ul style="list-style-type: none"> • ELOM 4&5 assessment of children aged 50-59 months currently enrolled in ELPs. • Practitioner rating of assessed children’s social-emotional functioning (SEF). • Interviews with assessed children’s primary caregivers. • Enumerator observation of the class attended by assessed children. • Enumerator observation of ELP facilities. • Principal and practitioner interviews. 	<ul style="list-style-type: none"> • ELOM 4&5 assessment of children aged 50-59 months not enrolled in an ELP between January 2024 and the time of the data collection (October-November 2024). • Interviews with assessed children’s primary caregivers.

The *2024 Index* also provided an opportunity for the South African Medical Research Council to conduct a small separate study (not discussed further here), with financial support from DataDrive2030. This study covered 70 of the *2024 Index* sample ELPs in Gauteng and measured the concentration of metals (including arsenic, cadmium, lead and mercury within ELPs) as well as indoor temperatures, all of which have been shown to affect early childhood development negatively. Details of this component of the study, including key findings, are shared in a separate report.

1.2 Key Differences Between the 2021 and 2024 Index Rounds

There are two key differences between the 2021 and the *2024 Index* surveys.

The first is that the *2021 Index* was conducted amid the **COVID-19 pandemic** in South Africa where lock-downs and pandemic restrictions lead to closures of many ELPs and/or affected the services they provided, and some families either choose or were forced to keep their children at home, which affected which ELPs data were collected from and which children were assessed.

The second is that a **new sampling strategy** was developed for the *2024 Index*. There were two main reasons for this. First, to also include smaller ELPs for the Index sample to be representative of children enrolled in ELPs with **one or more** enrolled children aged 50-59 months, whereas the *2021 Index* sample was representative of children enrolled in ELPs with **six or more** enrolled children aged 50-59 months. Second, at the time of the *2021 Index*, there was no national database of ELPs in South Africa. In contrast, the *2024 Index* was able to draw on the national 2021 ECD Census. However, an update and expansion were necessary. The census was conducted during the COVID-19 pandemic, when many ELPs had closed, at least temporarily. In addition, the passage of time and natural turnover in the sector made it essential to refresh the data to ensure a comprehensive and up-to-date sampling frame.

As a result, direct comparison of key child outcomes between 2021 and 2024 is not appropriate. The *2024 Index* will serve as the baseline for future rounds, enabling direct comparison of child outcomes from 2024 onward.

There were also three important additions to the *2024 Index*:

- collection of data for non-enrolled children (sub-study);
- interviews with primary caregivers of assessed children to capture data on the children's home environment and their caregivers (enrolled children and sub-study of non-enrolled children); and
- collection of data from all ELPs attended by the assessed children compared to one-third of ELPs for the *2021 Index*.

1.3 2024 Index Responsibilities and High-level Timeline

The DataDrive2030 technical team comprised specialists for each required technical area. They designed the study, provided technical guidance, co-led training events, quality assured all activities, supported incoming data checking, conducted final checking and cleaning of collected data, and prepared accompanying documentation.

For the ELP and household listings and data collection, two different national Service Providers were selected, given the complexity and magnitude of the *2024 Index*, as well as the fixed window for data collection due to the need to conduct all child assessments within the final term.

- The South African Labour and Development Research Unit (SALDRU) at the University of Cape Town - Service Provider 1 - was responsible for the listings of ELPs and households and the telephonic collection of data from enrolled children's primary caregivers, along with the development of all associated fieldwork protocols.
- Genesis Analytics - Service Provider 2 - was responsible for the pre-test of survey forms and development of fieldwork protocols and the collection of data for both enrolled and non-enrolled children, and non-enrolled children's primary caregivers.

The purposes of the pre-test in April-May 2024 were to test the survey forms, which were successfully done, and also to develop fieldwork protocols for use during the data collection. Due to capacity constraints, Service Provider 2 was unable to complete the latter part of the pre-test as planned, which subsequently impacted the first enumerator training.

The design phase of the study began in earnest in August 2023 and included development of the sampling strategy; major revision and extensions of the existing survey forms; development and digitisation of the new survey forms; preparation and submission of the Ethics Application to the Ethics Review Committee of the Faculty of Humanities, at the University of Cape Town¹; and revisions of the prepared survey forms after the pre-test in Eastern and Western Cape provinces in April-May 2024, and later, during the enumerator training in September 2024 (see Figure 1).

¹ Approval was granted on the 6th of August 2024 (reference number PSY2024-032).

The preparations for the ELP and household listings, and telephonic collection of data from enrolled children’s primary caregivers by Service Provider 1 were completed between June and August 2024, followed by the listing fieldwork between August and October 2024.

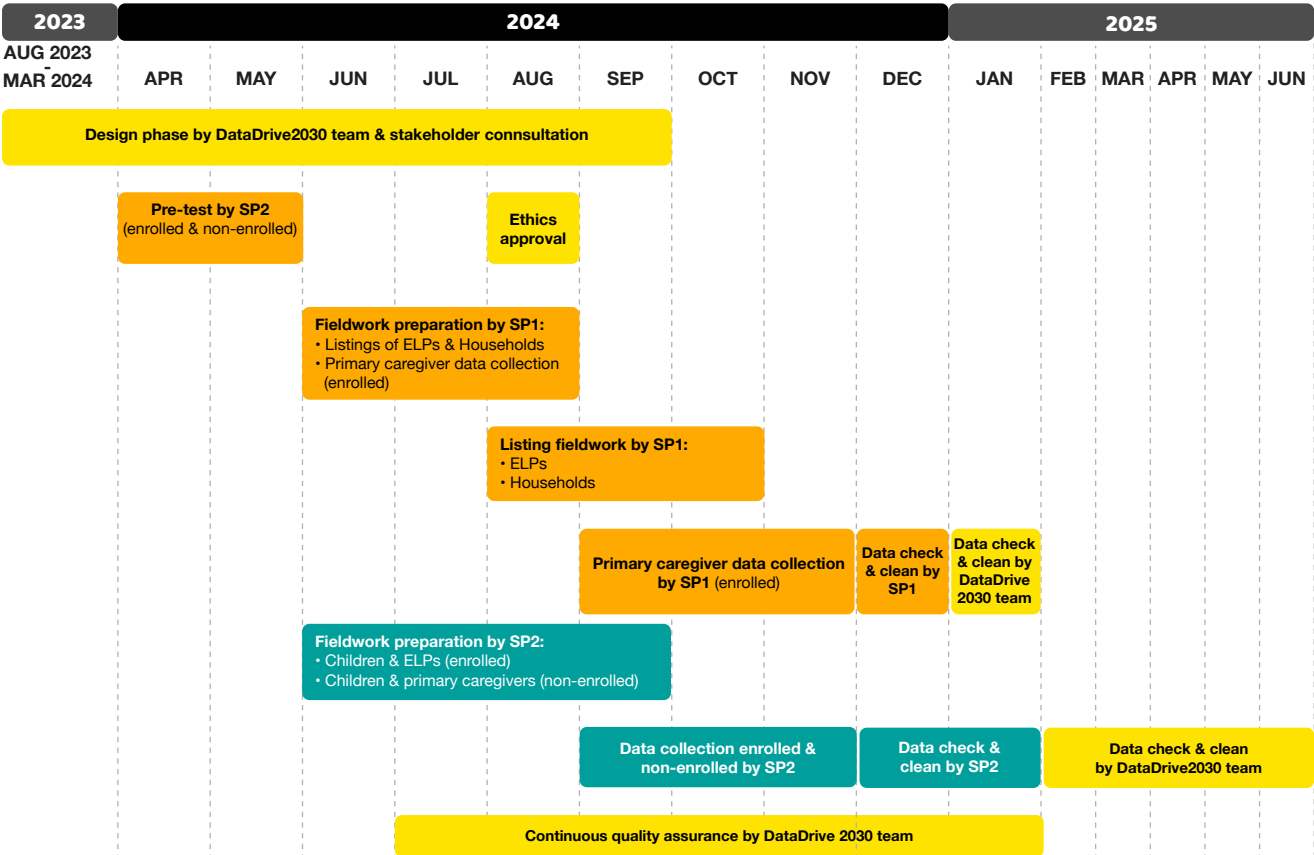
Service Provider 2 made the preparations for the collection of data for enrolled children and ELPs and from non-enrolled children and their primary caregivers, between June and September 2024. The data collection started in early September and was completed in late November 2024.

In addition to continuous checking of incoming data during the fieldwork, both the listings and the data collection were followed by data checking and cleaning first by the Service Providers and second by the DataDrive2030 team.

1.4 Structure of This Report

The remainder of this report begins by describing the 2024 Index study design and provides an overview of the fieldwork preparation and data collection. This is followed by an overview of the checking and cleaning process to produce the final datasets. Finally, the report considers two main areas for improvement for the next Index round. The Annexes provide additional details on selected topics.

Figure 1: 2024 Index High Level Timeline



02 Survey Forms, Sampling and Sample Sizes

This section describes the key design elements of the *2024 Index*: the survey forms used to collect the data, the sampling strategy, and the planned and realised sample sizes.

2.1 Survey Forms

The development of the *2024 Index* survey forms (questionnaires) was informed by the *Nurturing Care for Early Childhood Development Framework* and builds on the *2021 Index* survey forms and data analysis. The development of the *2024 Index* forms included consultations and workshops with over 50 national and international researchers from various fields of expertise, spanning roughly three months.

The survey form development also included extensive desk-testing of each of the electronic survey forms to ensure that skip logics (the path followed by each respondent depending on their answers), value constraints for questions and in-form calculations worked as intended in order to ensure data quality. The electronic survey forms were then pre-tested in the field in Eastern Cape and Western Cape provinces, which was followed by a limited number of revisions to increase form functionality. The final adjustments to the survey forms were made during the enumerator training, for a small number of items requiring additional refinements, mainly in terms of translation.

To achieve the *2024 Index* objectives, the following eight survey forms were used to collect data:

- 1 Child Sampling Form;
- 2 ELOM 4&5 Years Assessment Tool (ELOM 4&5), direct child assessment valid for children aged 50-69 months;
- 3 Social-Emotional Functioning (SEF) Rating Scale - practitioner rating of child's SEF;
- 4 Learning Programme Quality Assessment (LPQA v2) - classroom observation tool;
- 5 Facility Observation Form;
- 6 Principal interview;
- 7 Practitioner interview; and
- 8 Primary caregiver interview.

The main contents of each of these forms are set out in Table 7.

The Child Sampling Form is new for the *2024 Index*. It captures enrolment and attendance data, as well as child characteristics, to identify the group of eligible children from which to select children for the assessment randomly; to enable the construction of weights; and to collect contact information for the assessed children's primary caregivers for use in telephonic interviews.

The **ELOM 4&5 Assessment** is specifically created for South Africa, and is a psychometrically validated instrument with standards for overall child development and specific domains: Gross Motor Development (GMD), Fine Motor Coordination and Visual Motor Integration (FMC&VMI), Emergent Numeracy and Mathematics (ENM), Cognition and Executive Function (CEF), and Emergent Literacy and Language (ELL) (Dawes et al., 2020; Snelling et al., 2019).

The ELOM 4&5 aligns with the *South African National Curriculum Framework for Children from Birth to Four* and was developed and standardised for use with children aged 50-69 months, with performance bands for two groups: 50-59 months and 60-69 months. It provides a reliable and fair assessment of children regardless of their socio-economic circumstances or ethnolinguistic background and is available in all official languages of South Africa. Content, construct, age, and concurrent validity (with the WPSSI-IV) and test-retest reliability have been established (Anderson et al., 2021; Dawes et al., 2020; Snelling et al., 2019). ELOM 4&5 assessors require five to seven days of training and must meet strict accreditation requirements. The assessment is conducted in the child's home language, and ELOM assessors can only conduct assessments in the languages they are accredited in.

A child's performance on each item of the five ELOM 4&5 domains is assigned a raw score, which is then transformed into a scaled score. In each domain, item scaled scores are summed to produce a domain total score out of 20. The five domain scores are then summed to derive the ELOM 4&5 Total Score out of 100. Scores for each domain and the total fall within one of three performance bands: On Track (achieving the expected standard of performance), Falling Behind and Falling Far Behind. The cut-off points for On Track, Falling Behind and Falling Far Behind for children aged 50-59 months and 60-69 months were set empirically using standard score percentiles (see Table 2), and in consultation with key stakeholders in 2016.

The ELOM 4&5 also includes measurement of child height using a stadiometer and a board to ensure a level floor surface when measuring.

Table 2: ELOM 4&5 Total and Domain Score Bands for Children Aged 50-59 Months

ELOM 4&5 Score Bands			
	Falling Far Behind (<32 nd percentile)	Falling Behind (32 nd – 59 th percentile)	On Track (=>60 th percentile)
Domain 1: Gross Motor Development	0 – 5.40	5.41 – 8.59	8.60 – 20
Domain 2: Fine Motor Coordination & Visual Motor Integration	0 – 9.70	9.71 – 12.31	12.32 – 20
Domain 3: Emergent Numeracy & Mathematics	0 – 6.34	6.35 – 9.32	9.33 – 20
Domain 4: Cognition & Executive Functioning	0 – 4.07	4.08 – 7.16	7.17 – 20
Domain 5: Emergent Literacy & Language	0 – 6.53	6.54 – 10.25	10.26 – 20
ELOM Total	0 – 36.01	36.02 – 46.31	46.32 – 100

The **Social-Emotional Functioning (SEF) Rating Scale** is used alongside the ELOM 4&5 to measure a child’s social-emotional functioning – aspects of behaviour that can only be reliably assessed by someone who knows the child, such as a practitioner or primary caregiver. The SEF Rating Scale has 13 items across two key areas: social relations with peers and adults, and emotional readiness for school (see Table 3).

Scores for each scale fall within one of two performance bands – children either meet the expected standard, or they do not. For social relations, the score ranges from 0 to 18, and a child is defined as meeting the standard if their score is 12 or above. For emotional readiness for school, the score ranges from 0 to 12, and a child scoring 9 or above is defined as meeting the standard. Concurrent validity with the Strengths and Difficulties Questionnaire (Goodman, 1997) has been established (Dawes et al., 2020).

Table 3: SEF Score Bands for Children Aged 50-59 Months

SEF score bands			
	Items	Don't Meet the Standard	Meet the standard
Social relations with peers and adults	child able to cooperate without prompting; work with peers in group activities; resolve problems without aggression; seek support and information from familiar adults; take initiative to create cooperative activities with a familiar adult; and independent toilet use	0-11	12-18
Emotional readiness for school	child able to communicate clearly; express needs and feelings appropriately; adjust to changes in the classroom or home routine; and approach new experiences confidently; and likes to do things without help and is a self-starter.	0-8	9-12

The **Learning Programme Quality Assessment Tool (LPQA) (v2)** involves a two-hour observation of a class with some or all of the assessed children, depending on whether they are in the same class. The tool has 22 items spread across five domains: Materials and Equipment, Planning and Assessment, Learning Programme, Teaching Strategies, and Relationships and Interactions. The domain and total scores are classified into three rating categories: Inadequate, Basic and Good (see Annex 1 for details).

The **Practitioner and Principal Interview Forms** for the 2024 Index were significantly revised compared to the 2021 forms. They were extended to include new topics such as views on gender, growth mindset and play, and barriers to doing the job, in addition to expanded sections on in-service training.

The **Facility Observation Form** captures basic ELP features such as the existence of a fence around the premises, source of drinking water and type of toilets, which can influence child security and health.

The **Primary Caregiver Interview Form** was new to the *2024 Index*, as primary caregivers were not interviewed as part of the *2021 Index*. The form captures data on the child's socio-economic circumstances, home learning environment, and health and key characteristics of primary caregivers to allow for disaggregation of results by some of these factors, and for examination of predictors of the key child outcomes. For the home learning environment section, the form draws heavily on the *DataDrive2030 Home Learning Environment Tool* that has been developed and validated for use in South Africa (Dawes et al., 2023). The respondent for this form is the assessed child's primary caregiver, defined for the *2024 Index* as the person who cares for the child most of the time, assumes responsibility for the child, and knows the child well enough to answer questions about their development and home environment. The interviews with assessed children's primary caregivers (PCGs) were conducted telephonically, as it was not deemed feasible from either a fieldwork or a budget perspective to conduct face-to-face interviews.

Comparability of Survey Forms Between 2021 and 2024

The ELOM 4&5 and SEF forms used for the *2024 Index* are the same as those used for the *2021 Index*, except for the addition of a small number of complementary questions (asked of assessors after conducting the child assessment) and slight revisions to the child disability screening questions for the former; and the addition of two subscales namely: task focus and following instructions, each consisting of four questions, to the latter.

For the facility, practitioner and principal forms, key questions were retained for the *2024 Index* to allow for trend analysis of these.

The *2024 Index* LPQA tool was substantially revised compared to 2021, to add skip logic and simplify response choices in order to facilitate more consistent administration of the form in the field. By necessity, this involved some reformulation of questions. The LPQA factor structure was also revised for the *2024 Index*. For a detailed description of the revisions, see Annex 1. Given the revisions to the LPQA tool, comparability across the two surveys is very limited. The trade-off between leaving the LPQA tool unchanged to allow for comparison over time and revising it to facilitate its administration in the field to increase data quality for the *2024 Index* and future rounds was carefully considered before deciding the latter was of a higher priority.

The primary caregiver interview and the child sampling form were developed for the *2024 Index*; therefore, only comparison with future surveys will be possible.

Sub-Study: Survey Forms

Three survey forms were used to collect data for the purposively selected sample of children aged 50-59 months in low-income areas in Gauteng, KwaZulu-Natal and Western Cape, currently not attending ELPs: ELOM 4&5, SEF Rating Scale and the PCG interview.

These three forms for the sub-study of non-enrolled children are the same as the ones used for the 2024 enrolled sample, except that:

- The SEF Rating Scale form is administered to primary caregivers instead of to practitioners
- The PCG interview form section about ELP attendance has a different set of questions, including if the child has ever attended an ELP and the reason why the child currently does not attend an ELP.

2.2 Sampling Strategy and Weights

Three key principles guided the development of the 2024 *Index* sampling strategy (for more details, see Ardington 2024):

- The sample must be statistically representative of the underlying population of children aged 50-59 months enrolled in ELPs, at both national and provincial levels.
- Forward comparability to allow for comparison of key child outcomes over time, using future *Index* surveys.
- It must be possible to replicate the sampling strategy for future surveys.

A stratified multistage sampling design was used. Before sampling, 38 strata were formed using province and a weighted school quintile (see Table 4), where the weights were based on the number of Grade 3 learners in each school. The reason for there not being 45 strata (nine provinces x five weighted school quintiles) is that in all provinces except Gauteng and the Western Cape, weighted school quintiles 4 and 5 were collapsed into one due to the limited number of wards (unit selected at the first stage of the sampling) in the top two quintiles.

Table 4: Allocation of Strata, Wards, ELPs and Children Across Provinces

Province	Weighted school quintile					No. of allocated wards	No. of allocated ELPs	No. of allocated children
	q1	q2	q3	q4	q5			
EC	12	12	13	12		49	147	588
FS	9	9	9	8		35	105	420
GT	15	15	15	15	16	76	228	912
KZN	18	19	18	18		73	219	876
LIM	13	13	12	12		50	150	600
MP	9	10	9	10		38	114	456
NC	8	9	9	9		35	105	420
NW	9	9	9	8		35	105	420
WC	8	8	8	9	8	41	123	492
Total			38			432	1296	5184

Note: In all provinces other than Gauteng and Western Cape, weighted school quintiles 4 and 5 were collapsed into one, as there were a limited number of wards (unit selected at the first stage of the sampling) in the top two quintiles.

Within each stratum, three-stage sampling was implemented as follows.

Stage 1: Selection of wards, except for in a minimal number of cases where two or more wards were combined to create a feasible enumeration area (henceforth referred to as wards), defined as having 60 or more Grade 3 learners based on data provided by the Department of Basic Education. Wards were selected with probability proportional to size, where the measure of size was the number of Grade 3 learners in each ward.

Across the country, 432 wards were selected (equivalent to roughly 11% of all wards in the country), and each province was allocated a minimum of 35 wards, with the balance allocated to the other provinces in proportion to the number of two-year-olds in each province as per the 2022 National Census.

After the selection of wards, an ELP listing exercise (discussed below) was conducted in each selected ward to construct the sampling frame from which to select ELPs and to update the Department of Basic Education's ELP database for these wards.

Stage 2: Selection of ELPs. For an ELP to be eligible, it must have at least one enrolled child aged 50-59 months; be open for at least eight hours per week; and be operational.

Within each stratum, the target number of ELPs was three times the number of selected wards. If all selected wards in a stratum had at least three ELPs with registered 4-year-old children, three ELPs were randomly selected in each selected ward. For strata where some wards had fewer than three eligible ELPs, all available ELPs were selected and the remaining sample of ELPs was allocated across other wards in the same stratum that had more than three eligible ELPs. If an ELP was unavailable during data collection, the next highest-ranked ELP was used as a replacement.

This means the planned sample could be achieved without needing to replace wards. This is possible because the full ELP listing enables the calculation of probabilities of selection at each sampling stage. This is not only simpler for fieldwork but is also optimal for controlling sample bias (areas with very few ELPs will have some chance of being in the sample, rather than merely being replaced) and for the straightforward calculation of weights.

Stage 3: Selection of children attending the sample ELPs after checking their eligibility (see Table 5). A child was eligible for the ELOM assessment if they were 50-59 months old at the time of data collection; their primary caregiver had not refused consent for the child's participation in the study; the child was present on the day of the survey visit; they could be assessed in their home language (that is, one of 11 official South African languages); and they had passed the ELOM 4&5 screening for difficulties seeing, hearing, mobility and communication. Child assent was also required for participation.

Within each ward, the aim was to assess a number of 50 to 59-month-old children equal to four times the number of selected ELPs, and an equal number of boys and girls. All eligible children present at the ELP on the day of the survey were randomly assigned numbers, and the girls and boys with the highest numbers were selected for assessment (the random sampling was integrated into the survey form). If a child could not be assessed, then the child with the next highest number of the same sex was assessed.

Table 5: Extract from Child Sampling Form: Child Eligibility Ladder

Variable Name	Question	Eligibility Criterion No.
note_assess	Now we will capture some information about children who are attending today to determine who will participate in the assessment.	n.a.
count_children	How many children in total are attending the early learning programme today?	n.a.
count_year	How many of those $\${count_children}$ children are 4 years old?	n.a.
count_month	How many of those $\${count_year}$ children were born between $\${birthdate_min}$ and $\${birthdate_max}$? [Note: Birth date calculated automatically based on assessment date.]	1
count_consent	How many of those $\${count_month}$ children have their caregiver's consent to participate in the study?	2
count_language	How many of those $\${count_consent}$ children can be assessed in their home language?	3
count_able_seeing	How many of those $\${count_language}$ children do not have any difficulties seeing (with or without glasses)?	4
count_able_hearing	How many of those $\${count_able_seeing}$ children do not have any difficulties hearing (with or without a hearing aid)?	5
count_able_moving	How many of those $\${count_able_hearing}$ children do not have any difficulties walking or moving their arms, or have weakness or stiffness in the arms or legs?	6
count_able_understand	How many of those $\${count_able_moving}$ children do not have SIGNIFICANT difficulties understanding what people say?	7
note_gender	How many of those $\${count_able_understand}$ children are...?	n.a.
count_girls	Girls	n.a.
count_boys	Boys	n.a.

Note: $\${variable\ name}$ denotes an automatically populated value from a previous question in the survey form

During data collection, if fewer than the target of 12 children were assessed in the first three ELPs in a selected ward, an additional ELP was selected within the same ward, or if this was not possible, an additional ELP was selected in another ward within the same stratum. These were not replacements but additions to the ELP sample to reach the target of 12 assessed children per ward.

As the 2024 Index is a sample-based survey, to produce statistically representative estimates, weights need to be applied. Each weight (ward, ELP, child) is the inverse of the probability of a ward/ELP/child being selected at each of the three stages set out below. The probability of selection is the product of the following probabilities:

- Probability that a ward is selected = $(\text{number of wards selected in stratum}) \times (\text{Grade 3 learners in ward}) / (\text{Grade 3 learners in stratum})$.
- Probability that an ELP is selected = $(\text{number of ELPs selected in that ward}) / (\text{number of ELPs in ward})$.
- Probability that a child is selected = $(\text{number of children selected in that ELP}) / (\text{number of eligible children in ELP})$.

Confidence interval (CI) widths for each province depend both on the number of wards allocated to that province and the degree of homogeneity of child outcomes within wards and ELPs in that province. Using the national design effect for the percentage of children On Track for early learning from the *2021 Index*, expected CI widths were calculated for key indicators to inform the determination of sample sizes for the *2024 Index* (see Table 6). While a larger sample size would have been desirable to increase the precision of provincial-level estimates, this was not financially feasible given the already large scope and logistical complexity of the *2024 Index*.

Table 6: Expected Confidence Interval Widths for Provincial ELOM On Track Estimates

Province	Confidence interval width for Percentage of children on track for early learning
EC	10.8%
FS	12.7%
GT	8.6%
KZN	8.8%
LIM	10.6%
MP	12.2%
NC	12.7%
NW	12.7%
WC	11.8%

Sub-study: Selection of Wards, Small Area Layers and Households

The sub-study sample of non-enrolled children was purposively selected and is not statistically representative of the relevant underlying population. It captures key characteristics and outcomes for the non-enrolled assessed children and their primary caregivers and households. However, it can indicate differences with enrolled children using appropriate statistical methods.

Three provinces were selected for the sub-study – Gauteng, KwaZulu-Natal and Western Cape. Gauteng and KwaZulu-Natal were selected for the non-enrolled sample because they have the largest populations of 3-5-year-old children in South Africa, together accounting for nearly 45% of this age group nationally. These two provinces also have the highest absolute numbers of non-enrolled children, making them especially important for understanding the barriers to early learning participation at scale. The Western Cape was included for reasons of logistical convenience and to build on operational insights gained during the pre-test phase. Across these provinces, 45 wards in low-income areas (classified using weighted school quintiles) were selected pre-fieldwork to capture three area types: formal urban, informal and rural. (For further details see Annex 4y).

Within each of the selected wards, three small area layers (areas created by Statistics South Africa by combining enumeration areas (EAs) with a population of less than 500 with adjacent EAs) were purposively selected to maximise the distance from ELPs to increase the likelihood of identifying households with non-enrolled children.

Within each selected SAL, the fieldwork teams conducted a screening of all structures to identify dwellings. This was followed by door-to-door visits to collect the data required to determine if a household had children eligible for the sub-study. If a household did, they were included in the non-enrolled sample.

2.3 Sample Sizes

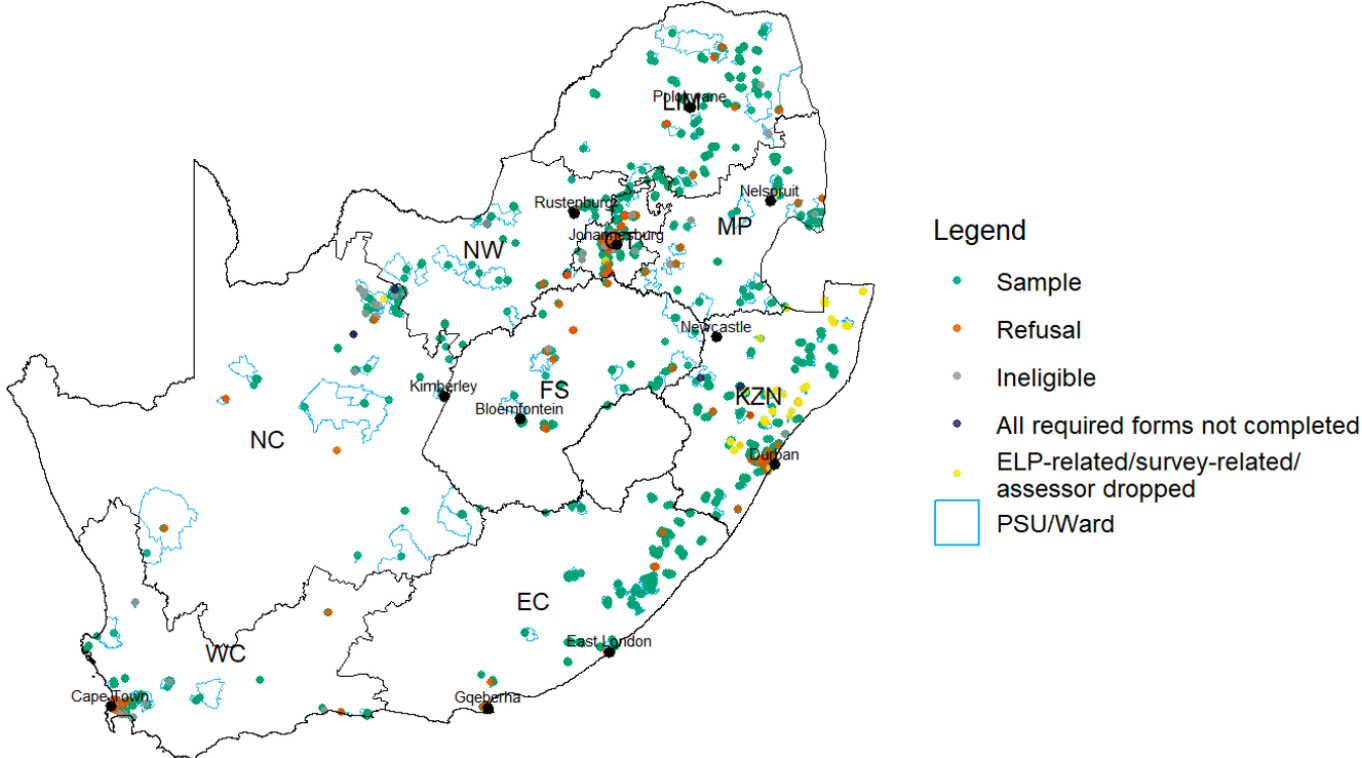
The *2024 Index* sample covers all nine provinces and was implemented with realised sample sizes as a percentage of planned sample sizes ranging from 94% to 107% (Table 7).

A total of 432 wards were selected for the *2024 Index*, and the realised number of wards is 414. The discrepancy is due to the removal of 18 wards. Five wards could not be visited because no ELP was identified during the listing. Three wards were not visited due to safety and security issues during the fieldwork; a type of risk that was not limited to these wards, but these were considered too dangerous for the fieldwork teams to visit after consultations with the local South African Police Service. Another 10 wards were removed during data checking and cleaning because the ELOM 4&5 Assessments were conducted by two assessors whose observations had to be removed (see section 4.2).

The planned sample size for ELPs was 1,296. This was exceeded for each ELP survey form except the practitioner interview. This was because in ELPs where the principal was the only practitioner, the practitioner form was not administered; only the principal form was (with some additional questions carried over from the practitioner form).

Daily monitoring was conducted to ensure the sample was implemented as planned. The 1,624 sample ELPs visited for the 2024 Index enrolled component are shown in Figure 2. Boundaries of the study wards/combined wards (in some cases, two or more wards were combined to ensure more than 60 Grade 3 learners, a proxy for children aged 50-59 months, to ensure workable enumeration areas) are drawn in blue. Sampled ELPs that had no age-eligible children on the day of data collection, either due to child absence or a listing error, in a small number of cases, or because they had closed down since being listed, are marked with grey markers (48). ELPs that had to be dropped for ELP-related (1) or survey-related (3) reasons, or because the observations of two ELOM assessors were removed from the sample (45), are marked with yellow markers (see section 4.2). ELPs that refused to participate in the study (125) are marked with orange markers and were mostly concentrated in the urban areas of Gauteng, KwaZulu-Natal and Western Cape provinces. A vast majority of refusals (78%) were by ELPs located in school quintiles 4 and 5 areas. Finally, some ELPs were dropped because not all required survey forms were completed (14), as indicated by the dark blue markers. This leaves the 1,388 sample ELPs where data was collected and retained, shown in green.

Figure 2: Map of ELPs Visited for the 2024 Index



The planned child sample size was 5,184. During the fieldwork, 5,358 ELOM 4&5 Assessments were conducted, resulting in a **realised sample size of 5,001 children** after removing the following:

- five observations at one ELP that were not randomly sampled
- 110 observations deemed invalid under the ELOM rules (assessment was not conducted in the child’s home language, child did not assent to the assessment, child scored zero on two or more ELOM domains, child exited the assessment early and/or child had a disability)
- 185 observations by two assessors, one who was not fully ELOM accredited and one for whom all 100+ children they assessed scored identically on one ELOM item (see section 4.2); and
- 57 observations that lacked one or more of the required corresponding survey forms (child sampling, SEF, facility, LPQA, principal).

For the PCG sample, the target was to interview PCGs for 70% of the assessed children. This rate was exceeded with a realised sample of 3,841 PCGs for the 5,001 assessed children included in the final sample, yielding a subsample of 77% of assessed children with a primary caregiver interview. Examination of the whole child sample and the subsample of children with primary caregiver interviews found good alignment across key child and ELP characteristics, including ELOM 4&5 total scores, child age, sex, growth status, total enrolment, registration status, ELP building type, and use of a curriculum. This suggests that the subsample is broadly representative of the full sample across these dimensions.

Table 7: 2024 Index Survey Forms and Sample Sizes (Enrolled Children)

Survey Form	Respondent	Main Purpose	Sample Size		Realised/ planned sample size
			Planned	Realised	
Child sampling form	Principal, includes checking of child register and child age records	Randomly select 4 children aged 50-59 months per ELP (half boys and half girls) for the ELOM 4&5 assessment and SEF rating. In ELPs with fewer than 4 eligible children, all eligible children assessed, and the shortfall made up in another randomly selected ELP, where the maximum number of assessed children per ELP is 6. Select the class to be observed using the LPQA (v2) tool. Collect contact information for assessed children’s primary caregivers.	1296	1388	107%
ELOM 4&5 Assessment (face-to-face) • gross motor development • emergent literacy and language • emergent numeracy and mathematics • fine motor coordination & visual motor integration • cognition and executive functioning • height measurement using a stadiometer	Child aged 50-59 months at the time of the survey visit	Capture early learning levels, overall and by domain Measurement of child height to capture physical growth status	5189	5001	96%

Table 7: 2024 Index Survey Forms and Sample Sizes (Enrolled Children)

Survey Form	Respondent	Main Purpose	Sample Size		Realised/ planned sample size
			Planned	Realised	
SEF Rating Scale (face-to-face)	Practitioner who knows the assessed child	Capture child social and emotional functioning levels	5189	5001	96%
Facility observation (face-to-face)	Enumerator observation	Capture key ELP facility features such as the existence of a fence, source of drinking water and type of toilets, that can influence child security and health and the learning experience	1296	1388	107%
LPQA (v2) Tool (face-to-face)	Enumerator observation of class with assessed children	Capture the quality of the learning programme across five areas: learning environment, learning and teaching, relationships and interactions, curriculum and teaching strategies	1296	1388	107%
Principal interview (face-to-face)	Principal/ ECD manager	Capture respondent's background characteristics, work environment, barriers to doing their job, views of gender, growth mindset and play, key ELP characteristics and funding	1296	1388	107%
Practitioner interview (face-to-face)	Practitioner/teacher who works with the assessed children	Capture respondent's background characteristics, work environment, barriers to doing their job and views of gender, growth mindset and play. For ELPs where the principal is the only practitioner, this form was not administered.	Max. 1296	1223	94%
Primary caregiver interview (telephonic)	Primary caregiver of assessed child	Capture child socio-economic circumstances, home learning environment, home stressors and primary caregiver and household characteristics that are expected to influence the key child development outcomes measured by the Index	5001	3841	77%

Sub-study: Sample Sizes

The planned sample size for non-enrolled children was 540, but it was extremely challenging to identify non-enrolled children in the field for a variety of reasons (SALDRU, 2025). For example, because primary caregivers feared the child support grant might be taken away if it became known that their child did not attend an ELP.

During the fieldwork, 334 non-enrolled children were identified but after removing 1 child who was not within the eligible age range; 1 child who did not have a PCG interview; and 60 ELOM 4&5 Assessments that were invalid according to the ELOM rules (no PCG consent, assessment not conducted in child's home language, no child assent, child scored zero on two or more ELOM domains, child exited assessment early, child had a disability) the realised sample size was 272. These children reside in 262 households, as the study design allowed for multi-family arrangements, for instance, two families living in the same household or compound.

Some of the many challenges associated with the household data collection are highlighted in this experience from a fieldwork spot check in PSU 154.

“As the area is a hotspot, high crime with gang violence, robbery and car hijacking, the SAPS Colonel allocated two police vehicles to patrol the area where the teams were working. It was recommended that no vehicle or clothing branding be used to blend in better.

The respondents for the non-enrolled household visits were quite hesitant to respond when approached. Some assisted and cooperated with information, even guiding the team to where they could find 4 and 5-year-old children. Others were not willing to give any information at all, asking if their grants were going to be taken away. Just hearing the words ‘Department of Education’ had some respondents withdrawing.” (SALDRU, 2024: p.29).

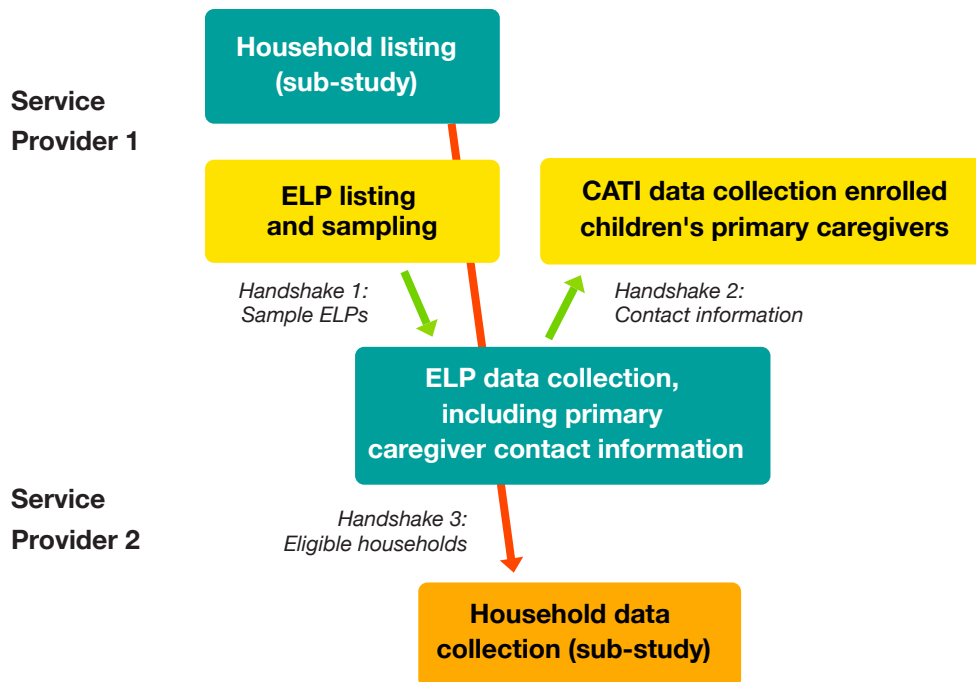
03 Fieldwork Overview

This section describes the main activities for the fieldwork preparation and implementation. It draws heavily on the fieldwork reports prepared by the two Service Providers and the continuous dialogue between the DataDrive2030 team and the Service Providers during the fieldwork period.

As discussed earlier, Service Provider 1 conducted the listings of ELPs and households and the telephonic primary caregiver interviews for the enrolled component. Service Provider 2 conducted the assessments of enrolled children, interviews and observations at the ELPs, as well as child assessments and PCG interviews for the non-enrolled component (sub-study of non-enrolled children).

This dual approach had several advantages but required close collaboration to ensure timely transfers of data and information (Handshakes) between the two providers during the fieldwork. The first Handshake involved the transfer of sample ELPs from Service Provider 1 to Service Provider 2 on a rolling basis after the listing of ELPs was completed in each sample ward. The second handshake involved transfer of primary caregiver contact information obtained during the ELP data collection from Service Provider 2 to Service Provider 1, also on a rolling basis, to enable the telephonic interviews with enrolled children’s primary caregivers (see Figure 3). The third Handshake involved the transfer of household records from Service Provider 1 after the completion of the household listing to Service Provider 2 for the non-enrolled data collection for the sub-study.

Figure 3: Data Handshakes Between the Two Service Providers



3.1 ELP Listing and Sampling

The preparations, including enumerator training, and listing and sampling of ELPs took place between June and October 2024. Each of the key activities is discussed below.

Desk-based compilation of an initial list of ELPs using multiple data sources, including

- The Department of Basic Education’s Education Management Information System (EMIS)
- Calls to schools and ELPs for referrals
- Aerial imagery and Google Maps

Fieldwork verification and expansion: In each selected ward, the fieldwork teams systematically visited each ELP identified through the desk-based exercise. During these visits, they sought to verify the ELP’s existence and collect key data for each ELP, including accurate GPS location; ELP name, address and contact information; operating hours; and the number of children aged 50–59 months enrolled (the eligible age range for the ELOM 4&5 Assessment). Newly established ELPs, not present in the initial desk-based listing, were identified through engagement with local stakeholders, which included community organisations, ECD Forums, health clinics, primary schools, police offices, social workers, traditional leaders, and ward councillors.

The desk research and site visits identified a total of 6019 possible ELPs in the 432 wards, of which 71% were able to be verified. The most common reasons for non-verification were that the ELP had closed down since the 2021 ECD Census; there was no ELP at the provided address or in the vicinity; or that the ELP was located outside the wards selected for the 2024 Index sample (Table 8).

Table 8: ELP Verification Outcomes and Sampling Frame

	Number of ELPs
Preliminary ELP list	6019
ELP closed down	558
No ELP at address & could not locate the ELP in the vicinity	461
ELP located outside study area	282
Refused to participate	228
Duplicate ELP	168
Not an ELP	55
Verified ELPs	4267
ELPs <u>not</u> eligible for the study (for example, no age eligible children, open less than 8 hours per week)	300
Sampling frame	3967

In preparation for data collection, these ELPs were mapped to the 432 selected wards via Geographical Information Systems (GIS) methods to visualise the location of the ELPs and develop an effective fieldwork model. The teams also distributed *2024 Index* information materials, such as posters and pamphlets, to raise awareness of the Index.

Data quality checks: In-field, this included regular checks and audits by field supervisors and spot checks by Service Provider 1's survey assistant team. Another key part of the quality assurance was back-check phone calls to 569 ELPs to verify collected data and assess the professionalism of the field teams. Unit-level quality checks ensured that each ELP was visited and assigned a verification outcome. The item-level quality checks were conducted using customised Stata scripts, focused on data validity, accuracy, and completeness.

ELP sample ranking: The 3,967 identified, eligible ELPs constitute the ELP sampling frame (Table 8). These ELPs were randomly ranked using an algorithm approved by the DataDrive2030 team. This ranking determined the selection of ELPs for the data collection.

Transfer of sample ELPs to Service Provider 2: Out of the 3,967 ELPs in the sampling frame, 2,218 ELPs were randomly selected and sent to Service Provider 2 (Handshake 1). The reason for sending more than the planned number of three sample ELPs per ward (up to six were sent) was to accommodate ELP refusals to participate, ELPs closing down between the ELP listing and data collection, etc. Towards the last part of the ELP data collection, replacement sample records beyond the top six ranked were pushed to Service Provider 2 to reach the ward target when required. Service Provider 1's pipeline included backups of both the incremental sample records sent with each push to Service Provider 2, as well as the full sample at the time of each data push.

Consent form distribution: The fieldwork teams distributed consent forms to give to primary caregivers of children in the eligible age range, at the top four ranked ELPs in each sample ward, in preparation for the data collection by Service Provider 2. Each selected ward received enough forms for all the children in the eligible age range, in their home language.

3.2 ELP Data Collection

The preparations and data collection for both the enrolled children and ELPs, as well as for non-enrolled children and their primary caregivers, took place between June and November 2024. The main activities are described below.

Training of enumerators

Service Provider 2 organised the training of enumerators across three hubs: Johannesburg (August 19–23 2024), Durban (August 26–30 2024), and Cape Town (September 2–6 2024). At each hub, Service Provider 1 and DataDrive2030 team members conducted the training sessions jointly, with some of the same team members overlapping across hubs to ensure as much consistency in training content and delivery as possible.

The development of the training materials and fieldwork manual for the enrolled and the non-enrolled data collection was led by Service Provider 2, working with the DataDrive2030 team. The fieldwork protocols were not final for the first enumerator training in Johannesburg. To help address this, the protocols were uploaded to the fieldworker platform in an easy-to-understand format, and additional support was provided by the regional coordinators at the start of the fieldwork.

All newly recruited ELOM 4&5 assessors (those who did not have prior ELOM accreditation) participated in a five-day training programme, including theory, observed direct assessments with children and inter-rater reliability testing. Core training followed over a further five-day period. During the first two days of this second training week, field workers were split into two groups – LPQA auditors and ELOM assessors. The 'auditors' were trained on the LPQA, and the ELOM assessors received two days of refresher training. This also provided an opportunity to ensure consistency between old and new ELOM assessors as they joined the refresher together. For the remaining three days, all field workers were trained on other instruments and protocols. The main topics were research ethics, fieldwork protocols, interview and observation techniques, the SurveyCTO survey forms, and fieldwork logistics.

During the LPQA training, fewer than expected candidate auditors met the accreditation requirements, which are set high to ensure the learning programme quality is captured as accurately as possible. Furthermore, during the data collection, some enumerators dropped out, and language gaps were identified across the fieldwork teams. This resulted in a shortage of LPQA auditors, and an additional LPQA training session was conducted in Cape Town from the 21st to the 25th of October 2024.

In-field support for ELOM assessors and LPQA auditors was provided from the outset and was ramped up in fieldwork week 4 based on the quality assurance of incoming data. This included the development and implementation of a standardised assessor observation and coaching protocol, and weekly online office hours organised by the DataDrive2030 team. The latter focused on areas where the incoming data monitoring indicated a need for further support, including ELOM domain-specific training, review of specific survey form sections, and Q&A.

Teams and rollout across provinces

Each province had a dedicated regional coordinator, except Gauteng, which had two due to its notably larger sample size. For the data collection, 55 ELOM assessors and 55 LPQA auditors worked in pairs for the ELP visits.

Data collection rollout was staggered across the provinces after training had been completed at each hub and based on the sequencing of information (ELP sample records and consent drop-offs) from Service Provider 1 to Service Provider 2, as the listing in each ward was completed (Handshake 1). As a result, the data collection started in Gauteng, Limpopo and North West, followed by Free State, KwaZulu-Natal and Mpumalanga, Eastern Cape, Western Cape and Northern Cape (Table 9).

There were some initial delays with the drop-off of consent forms, which meant that the start of data collection was delayed in some wards. Once this challenge was identified, measures were put in place by both Service Providers and the DataDrive2030 team to accelerate drop-offs. In addition, the consent period was shortened from 7 days to 4 days for a limited period.

Table 9: Rollout of Fieldwork Across the Nine Provinces (Enrolled Children)

Total fieldwork days		Fieldwork period
EC	43	18/09/24 - 18/11/24
FS	42	11/09/24 - 07/11/2024
GT	56	03/09/24 - 18/11/24
KZN	40	11/09/24 - 15/11/24
LIM	49	03/09/24 - 19/11/24
MP	47	09/09/24 - 22/11/24
NC	36	01/10/24 - 21/11/24
NW	46	03/09/24 - 22/11/24
WC	43	18/09/24 - 21/11/24

The regional coordinators were responsible for calling principals of the ELPs to schedule appointments, collect basic information about the ELP, and ask the principal to prepare the information required for the visit. This information included official registration documents (if available), returned consent forms, child registers and details on children’s dates of birth.

Upon arrival at the ELP, the auditor conducted the facility observation as it included one item that required observation when children were dropped off in the morning, while the ELOM 4&5 assessor ensured all the necessary information to complete the Child Sampling Form was available. This included collecting the received opt-out consent forms from primary caregivers and checking them following a specific protocol.

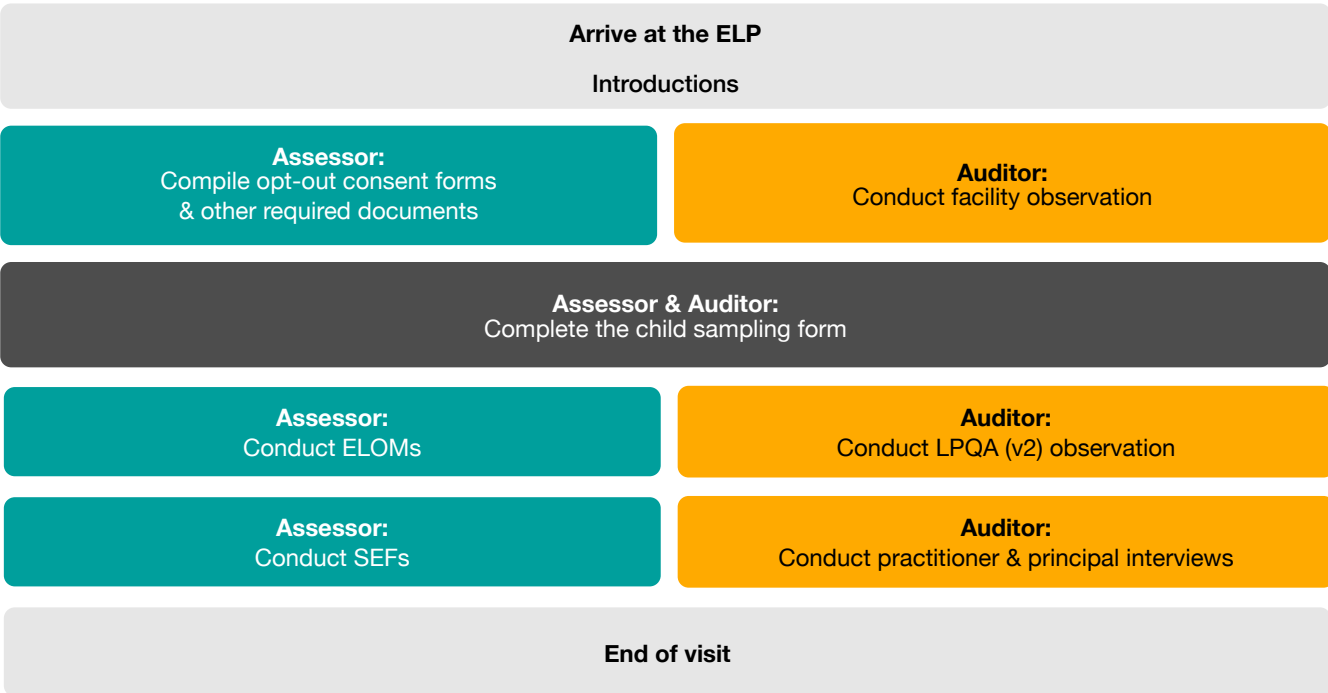
After the auditor had checked the opt-out consent forms, the random sampling of children was conducted using the Child Sampling Form, typically jointly by the ELOM assessor and the auditor, given the need to check records while using this form.

The child assessment took on average 45 minutes, and the majority of assessments were completed by 1:00 pm, given that young children typically tire later in the day, and some children may leave earlier. But in reality, many ELPs struggle with basic infrastructure, and ELOM assessors were asked to record the frequency of disruptions for consideration during the analysis of assessment data. The disruptions include background noise or interruptions by other children or adults. For the next Index round, more training will be provided specifically on strategies for managing common disruptions. After completing the ELOM 4&5 assessments, including height measurement, the assessor conducted the interviews with practitioners for them to rate each assessed child’s social and emotional functioning.

Based on the sampled children, the Child Sampling Form selected a class to be observed, ensuring it included at least some of the assessed children. The auditor began with the two-hour LPQA observation, typically by 9:00 am, to ensure the full duration could be observed before children started to leave for the day.

The practitioner and principal interviews were then conducted in either order, depending on availability. For the practitioner interviews, the respondent was automatically selected by the child sampling form to ensure a practitioner responsible for at least one of the assessed children was interviewed. In cases where the fieldwork team had unsuccessfully attempted an in-person interview with a practitioner or principal three times, a telephone interview was instead conducted, using a specific protocol. In total, six practitioner interviews, seven principal interviews, and four SEF ratings by practitioners were conducted telephonically.

Figure 4: Day at the ELP



Child height measurement. Child height and weight were both intended to be captured for the *2024 Index*. However, due to enumerators struggling to measure weight during the first week of fieldwork accurately, the difficult decision was made only to capture height to focus on the relatively more important measure for the *Index*. Every child's height was measured twice using a stadiometer; if the two measurements differed by more than 0.3 cm, height was measured a third time. During the first week of fieldwork, there were some problems measuring height accurately due to uneven floor surfaces in some ELPs. Therefore, from the second fieldwork week onwards, every enumerator was provided with a board to place under the stadiometer and a 1 metre spirit level to ensure the stadiometer was level when measuring. In addition, the DataDrive2030 team prepared an instructional video on how to accurately use the plank and stadiometer, which was shared with enumerators.

At the end of each visit, ELPs were provided with a certificate of thanks for their participation in the *Thrive by Five Index* and an educational resource (Finding Thabo posters) with accompanying instructional materials. ELPs were also provided with contact details for the Index team in the event of any concerns or questions.

Mop-ups. The fieldwork teams were not always able to complete all survey forms during the initial visit to an ELP. In these cases, mop-up visits were used.

ELP replacement protocol. A strict ELP replacement protocol was adhered to throughout the data collection to ensure the sample was implemented as planned, including following the ELP sample rank order (see section 3.1). All replacement requests were reviewed and approved by the DataDrive2030 team.

ELPs were replaced only if the request by the regional coordinator was made for one of the following reasons:

- ELP had no eligible children
- principal refused to allow the ELP to participate
- primary caregivers did not consent for children to participate
- ELP was no longer operational
- ELP was inaccessible due to safety concerns or road closures
- ELP had recently been visited by fieldwork teams from the Department of Basic Education's Quality Assurance and Support System (QASS) survey, which was ongoing at the same time as the *2024 Index* in some study areas.

Incoming data checks. During fieldwork week one, adequate systems to effectively monitor the quality of the incoming data were not yet in place due to Service Provider 1 capacity constraints. The DataDrive2030 team therefore provided support for incoming data monitoring until the capacity constraint had been resolved.

Airtable data quality controls, End-of-Day reports and Dashboard. Leveraging the linkage between SurveyCTO and Airtable, real-time quality control alerts were set up to flag potential data quality issues. When a potential issue was flagged, the enumerator responsible for the collection of the data was sent a query on Slack. If their response to the query indicated an error, a correction was made and logged. End-of-Day reports were made available via an Interface on Airtable for each of the survey forms (Figure 5). These reports were used to provide feedback on each of the quality control alerts that appeared on Slack, and the feedback received for each alert was displayed. The outcomes of the quality control investigations were also made visible on the End-of-Day report interface. The DataDrive2030 team monitored the End-of-Day reports daily and submitted queries in the report itself or on Slack for time-sensitive priority issues, to Service Provider 2.

Figure 5: Example of an End of Day report for the Child Sampling form

slack_msg_com	QC response	Description of Enumerator Feedback (from Data Correction: Quality C...
You have captured that there are 11 present at the ELP today, b...	No, no changes requir...	The fieldworker confirmed that this is correct and stated...
Your child sampling form was 6 min long. Please explain why it...	No, no changes requir...	The principal presented all of the necessary documents to...
You have captured that there are 13 present at the ELP today, b...	No, no changes requir...	The fieldworker confirmed that this is correct and added...
Your child sampling form was 6 min long. Please explain why it...	No, no changes requir...	The enumerator confirmed that the numbers were correct...

A digital dashboard linked to Airtable was developed for fieldwork monitoring. The dashboard provided near real-time tracking of ELP fieldwork performance indicators. Initially, the number of quality controls on the Dashboard was limited to core controls monitored for all survey forms, and over time as the fieldwork progressed, the number of quality controls was increased.

Stata scripts: Customised Stata scripts were prepared to conduct near-daily checks across all survey forms. The checks included:

- validation of unique identifiers
- skip logic checks
- checks for non-allowed missing values; and
- monitoring of Don't know and Refused response rates for all variables.

Back-check phone calls: Calls were made to 10% of the sample principals to verify that a visit had occurred (100% were visited) and to assess the professionalism of the fieldwork team (99% were rated as professional).

3.3 Collection of Data from Enrolled Children's Primary Caregivers

The fieldwork preparations and collection of data from primary caregivers of enrolled children were conducted between August and early December 2024 by Service Provider 1. Each of the key activities for this exercise is set out below.

Management of primary caregiver record. As Service Provider 1 received the primary caregiver records from Service Provider 2 (Handshake 2), Service Provider 1 developed a system to process and quality assure the received data.

Training. Service Provider 1 delivered comprehensive enumerator training that covered the survey form, data collection techniques, and using computer-assisted telephonic interviewing (CATI) software. DataDrive2030 collaborated on the training content and delivery to ensure correct administration of the primary caregiver survey form and consistency with the training provided by Service Provider 2 on the equivalent form used for non-enrolled children’s primary caregivers for the sub-study of non-enrolled children.

Receipt of sample records from Service Provider 2 and data collection. The interviews with primary caregivers were conducted following strict calling rules to ensure a sufficient level of effort to contact each respondent, and consent was sought at the start of each interview following a strict protocol. This included a minimum of six call attempts per telephone number over at least two days, at various times, including weekends. The average call duration was 20 minutes.

The primary caregiver sample records were received on a rolling basis, totalling 5,334 records (see Table 10), of which 5,041 were viable (95%), indicating that they had complete contact information. Among the viable records, 7% had incorrect or non-existent phone numbers; another 10% were not contactable; and 2% of respondents did not consent to be interviewed. This yielded 4,072 complete interviews (from almost 14,000 call attempts), equivalent to 81% of the 5,041 received sample records, an unusually high rate for a telephonic interview, and even more so as it includes sensitive questions.

Table 10: Primary Caregiver Interview Outcomes (Enrolled Children)

	Number of primary caregiver interviews (enrolled children)	Column 1 indicator as percent of viable sample records
Viable sample records	5,041	n.a.
Incorrect number/number does not exist	351	7%
Not contactable	499	10%
Respondent refusal	119	2%
Interview complete	4072	81%
Received sample records	5,334	n.a.
Attempted phone calls	13,766	n.a.

Data quality checks. Service Provider 1 performed high-frequency checks daily. This involved using specialised software for unit-level checks and ensuring compliance with calling rules. To conduct item-level checks of skip pattern adherence, valid ranges, variable completeness and non-response, and cross-question sense checks, customised Stata scripts were used. In addition, 446 successful back-check phone calls (equivalent to 11% of completed interviews), were made to verify data accuracy and enumerator professionalism during the interviews.

Sub-study: Household Listing and Non-Enrolled Data Collection

Household listing

The preparations for the listing of households with children eligible for the study, including enumerator training, and subsequent data collection took place between June and October 2024, with the key activities described below.

Mapping of structures and dwelling verification. Prior to fieldwork, using a geospatial information system (GIS), all dwelling units within the 135 selected SALs in the 45 study wards were identified to create a comprehensive listing of 34,151 dwellings. An enumerator visited each of these to confirm if they were a viable dwelling, that is, a residential inhabited structure. Contact was made with a resident 18 years or older for 22,359 of the listed dwellings (see Table 11). Reasons for not visiting certain identified dwellings were documented, and included safety and security concerns, non-viable dwellings (for instance, non-residential structures) and access issues (for example, gated communities).

Identification of eligible households. Service Provider 1 and the DataDrive2030 team developed a short screening form to be administered at households to capture data required to determine if they had children eligible for the study. Within each selected SAL, enumerators systematically visited the verified dwelling units, using the screening form. This resulted in 339 complete household rosters and 324 eligible children for the sub-study sample. These household records were transferred to Service Provider 2 for the collection of data for non-enrolled children and their primary caregivers (handshake 3).

Table 11: Household Listing and Screening Outcomes

	Number of dwellings
Successful contact (18+ years)	22,359
No eligible children	20,832
Refused to participate	37
Did not consent	1,151
Household roster complete	339
Household roster complete (% of successful contacts)	1.5%
Number of eligible children	324

Some of the study areas were too dangerous for the field teams to enter without police escort, or in some cases at all, which resulted in two selected wards in the Western Cape province being replaced.

Data quality checks. This included in-field spot checks by Service Provider 1’s survey assistant team and desk-based checks for completeness, accuracy and validity of the household listing data.

Non-enrolled data collection

The sub-study preparations started in June 2024, and the data collection concluded in the third week of November 2024.

Training of enumerators. The enumerator training for the sub-study was conducted jointly with the training for the enrolled children component section. Additional training sessions were conducted for the sub-study enumerators on the following:

- primary caregiver interview (the corresponding form for enrolled children’s caregivers was covered in Service Provider 1 training)
- how to conduct the ELOM 4&5 Assessment in a home environment with possible interference by caregivers, siblings, etc.
- how to administer the SEF Rating Form to primary caregivers (rather than to practitioners)
- how to accurately identify different water sources and types of buildings, and
- when and how to refer caregivers for support if needed.

Teams and rollout across provinces.

A regional coordinator oversaw each province. There were 15 fieldwork teams visiting households, each consisting of an ELOM 4&5 assessor paired up with an auditor. Assessors conducted the ELOM 4&5 Assessments while auditors conducted the primary caregiver interviews. The most experienced ELOM assessors had been chosen for the sub-study given the added challenges of assessing in homes as compared to at ELPs.

The data collection started in Gauteng and KwaZulu-Natal, followed by the Western Cape a few days later (Table 12).

Table 12: Rollout of Fieldwork Across the Three Provinces (Sub-Study)

	Total fieldwork days	Fieldwork period
GT	18	28/10/24 - 14/11/24
KZN	10	28/10/24 - 12/11/24
WC	17	04/11/24 - 21/11/24

Service Provider 1 transferred the household records to Service Provider 2 (Handshake 2) on a rolling basis. The regional coordinators were responsible for organising household appointments at least two weeks before arriving in the study ward.

If the primary caregiver was unavailable when the fieldwork team arrived, despite the regional coordinator making an appointment, the team attempted three visits to the household on three different days at three different times. If unsuccessful, the team moved on to the next household.

Typical visit to a home. The ELOM child eligibility criteria for non-enrolled children are the same as for enrolled children (see Table 5). However, for non-enrolled children, one additional criterion was applied, if a child lived in a home considered inappropriate to conduct the ELOM 4&5 Assessment in, for example, due to the presence of intoxicated adults, unsafe conditions or extreme distractions, the assessment was not conducted.

Consent from the primary caregiver to be interviewed and to assess the eligible child was sought in person on the day of the visit. If consent was not provided, the visit was terminated.

The fieldwork team was provided with a set of protocols specific to the home environment. For example, the enumerators were asked to try to find a quiet space, inside or outside the home, for the ELOM 4&5 Assessment to avoid disruptions as much as possible. They were also instructed how to explain the assessment to the child and the primary caregiver without evoking anxiety or expectations and were reminded that the child could take breaks or stop at any point.

The enumerators had also been trained on how to respond to queries from primary caregivers around the child's performance on the ELOM 4&5 Assessment. The protocol was first to remind them that the aim of the assessment was not to measure their individual child's performance but rather to evaluate children's performances across South Africa overall. Secondly, they were instructed to offer a few general comments such as "Your child managed to complete the assessment, which shows good attention skills", "your child knows their colours/shapes/numbers" etc.

As the fieldwork team did not know in advance what would be available in the home in terms of tables and chairs, they were provided with child lap desks or foldable tables to have a flat surface to use when conducting the ELOM 4&5 Assessment.

Height and weight measurement. For the sub-study, children's weight was measured in addition to their height. For this, the fieldwork teams used boards to have a level surface, and for the height measurement, they used a 1 metre spirit level to ensure that the stadiometer was well-positioned and calibrated.

At the end of the visit, a token of appreciation was given, including an airtime voucher, crayons and a colouring book for the child, as well as a handout with some ideas for fun ways to help your young child learn numbers.

Mop-ups. The fieldwork teams were not always able to complete all forms on the first household visit, and 3% of the sample households were re-visited.

Incoming data checks and back-check phone calls. As the same quality control system was used for the sub-study as for the ELP data collection, this is not discussed again here. Back-check phone calls were made to 41% of the interviewed primary caregivers of non-enrolled children to verify that the fieldwork team had visited and behaved professionally. These interviews also included a small number of questions about whether the assessed child had ever been enrolled in an ELP, and the reason for currently not being enrolled, which were not asked initially due to a skip-logic design error in the primary caregiver interview form.

04 Data Preparation

This section describes the post-fieldwork checking and cleaning undertaken by the DataDrive2030 team to generate the final 2024 Index datasets for both study components: enrolled children and the sub-study of non-enrolled children.

4.1 The 2024 Index Datasets

Each Service Provider prepared the datasets and documentation for the data they had collected. Service Provider 1 submitted datasets containing data from the ELP and HH listings and the telephonic interviews with primary care givers (PCGs) of enrolled children (see Table 13). Meanwhile, Service Provider 2 submitted datasets for the child sampling at ELPs, the child assessments conducted at ELPs and in households respectively, the ELP level interviews and observations, and the in-person interviews with non-enrolled children’s PCGs.

The datasets submitted by Service Provider 1 were less complex and were thoroughly checked and fully documented, requiring minimal data cleaning by the DataDrive2030 team in February 2025. For the datasets received from Service Provider 2, data checking and cleaning were conducted by the DataDrive2030 team over the period January to June 2025, as discussed below.

The 2024 Index datasets will be uploaded to the *DataFirst Open Data Portal* as public use files, together with accompanying documentation.

Table 13: Datasets Received From the Service Providers

Service Provider 1	Listing	<ul style="list-style-type: none"> • ELP listing data • Household listing data
	Enrolled children	<ul style="list-style-type: none"> • Primary caregiver data (telephonic interviews with 77% of the final sample of enrolled children)
Service Provider 2	Enrolled children	<ul style="list-style-type: none"> • ELOM-SEF-height data • ELP (child sampling, facility, LPQA, practitioner, principal) data
	Non-enrolled	<ul style="list-style-type: none"> • ELOM-SEF-height-weight data • Primary caregiver data

4.2 Data Quality Checks and Cleaning

General checks

A set of general checks was conducted for both the enrolled and non-enrolled datasets using customised, structured Stata scripts, one for each dataset. Checks for non-allowed missing values, adherence to value constraints, and Don't know/Refused rates for each variable were conducted in addition to those conducted by Service Provider 2 during the data collection, and minimal corrections were necessary.

Some inconsistencies were identified when matching children across the Child Sampling Form, ELOM 4&5, and practitioner SEF Rating Scale forms in terms of child age and sex; these were corrected to ensure cross-form consistency. The ELOM 4&5 form was treated as the source of correct values for these two variables.

In a small number of cases, within-form data inconsistencies were identified, and if other available data did not allow for resolution, the affected values were recoded as invalid responses. Some data inconsistencies of this type may remain in the public datasets, and data users who detect them are encouraged to report them to the DataDrive2030 Index team.

Child Sampling Form checks: Given the importance of collecting accurate data on the number of children aged 50-59 months present on the day of the survey, not only for correct sampling of children at the ELP, but also for the weight construction, this form received extra attention. This included the general checks discussed above as well as:

- Checking that the constraints built into the survey form to ensure the collection of accurate eligibility data had been followed for the eligibility ladder (see Table 5), and
- Checking that the number of children present on the day aged 50-59 months was less than or equal to the number of enrolled children aged 50-59 months.

The built-in checks were found to have worked as intended.

Checks of key variables

Three key variables were subjected to additional checking: (1) ELP fee amount, (2) total enrolment, and (3) ELP building type for the enrolled children's datasets. Additional checks were also run on ELOM data for both the enrolled and non-enrolled children's datasets. Each of these is discussed below.

ELP fee amount: The collected ELP fee data was used to develop the fee band variable. This is a key disaggregation variable for the 2024 Index. There were two sources for this variable: (1) principal reported fees charged and (2) PCG reported fees paid. (Primary care givers reported fees are available for 95% of ELPs). These data were combined to create a final, new fee variable, using the rules set out in Annex 2, where the general principle was to trust the principal's reported data unless there was clear evidence to the contrary.

Box 2: Developing the 2024 Index Fee Groups

One of the objectives of the *2024 Thrive by Five Index* is to monitor the disparity in early learning between children from less and more advantaged socio-economic circumstances. But the type of data required to construct an indicator for children's socio-economic circumstances is not collected for all sample children. Instead, an indicator that may be considered a proxy for household income and that is: available for all sample children; can provide valid comparisons over time; and can be replicated for future Index rounds needed to be developed.

Different options were explored and the pros and cons of each were examined, including consideration of the level at which data are available, collection frequency, and stability over time, as summarised below.

Weighted school quintile (ward level, annual, secondary data);

- South African Index of Multiple Deprivation (SAIMD) (ward level, 2011, secondary data);
- Priority wards based on SAIMD (ward level, 2011, secondary data);
- ECD subsidy (ELP level, binary, not stable over time due to policy changes, Index data);
- ELP structural quality indicator (ELP level, imprecise, Index data); and
- Fee amount charged by ELP (ELP level, Index data).

Based on this exploration, an Index fee variable consisting of seven groups was selected by an Advisory Group led by the DataDrive2030 team and comprising specialists from the Department of Basic Education, National Treasury and Research on Socioeconomic Policy (RESEP) at Stellenbosch University. For further details, see Kika-Mistry (2025).

Users of the *2024 Index* data may choose to use the *Index* fee group variable or to develop their own indicator, depending on what they consider more appropriate given their needs.

Total child enrolment: Child enrolment at ELPs is a key variable not only to assess the size of ELPs, but also, in combination with other variables, for example, deriving class size or the number of children per practitioner. Data on total enrolment was collected through both the principal interview and the Child Sampling Form. In the received principal and child sampling datasets, there were discrepancies in the total enrolment variable. A threshold was set, and only observations with discrepancies above it were addressed. For these observations, a third data source, the total enrolment data collected by the ELP listing team, was used to triangulate following a set of rules (available on request). This left 21 ELPs with discrepancies above the threshold and without total enrolment data from the ELP listing to use for triangulation. For these, 15 successful backcheck calls were made and 12 values were replaced while three were verified.

ELP building type: Data quality checks revealed unexpected results regarding the relationship between ELP building type and other variables, warranting further investigation. During field work, photos were taken of the front of each visited ELP, and the DataDrive2030 team systematically reviewed these photos and corrected 113 out of the 1,388 observations (8%). To reduce the number of misclassifications in the next Index round, additional training will be provided, and revisions to the response choice formulations in the survey form will be considered.

ELOM scores: The ELOM scores are the most important data collected by the Index, and the review of the ELOM dataset received from Service Provider 2 was extensive, taking place over three months.

For the enrolled sample, it was found that one of the ELOM assessors had not been fully accredited, and their 68 observations were removed from the sample to ensure adherence to the ELOM rules. For a second ELOM assessor, all children assessed scored identically (zero standard deviation) on ELOM item 13 (see Annex 3), which is not feasible, and their 117 observations were also removed from the sample.

Review of the photos taken by ELOM assessors of the drawings made by children for ELOM items 6 and 7 (drawing of a person, copy of a triangle), revealed some incorrect scoring by assessors. The DataDrive2030 team systematically reviewed all photos and rescored, as necessary. This process was applied equally to both the enrolled and non-enrolled samples.

Additional preliminary analysis of ELOM scores raised some queries, and the DataDrive2030 team systematically explored potential explanations. This included checking the functioning of the ELOM survey form; reviewing the weight calculations; examining key child characteristics; and exploring any fieldwork issues that could potentially affect scores (see Table 14). This method of elimination led to one remaining explanation, ELOM assessor bias (explicit and implicit).

Table 14: ELOM Data Investigation Areas, Checks and Outcomes

Area	Check conducted	Outcome
ELOM SURVEY FORM	Does the ELOM survey form work as intended?	Yes
	Does the ELOM survey form translate correctly?	Yes
SAMPLING STRATEGY IMPLEMENTATION & WEIGHTS	Sample design implemented as intended?	Yes
	Weights constructed correctly?	Yes
REALISED SAMPLE	Child age distribution similar across provinces?	Yes
	Child sex distribution similar across provinces?	Yes

Table 14: ELOM Data Investigation Areas, Checks and Outcomes

Area	Check conducted	Outcome
FIELDWORK ISSUES	Significant difference in ELOM scores related to the time of day the assessment was conducted?	No
	Significant differences in ELOM scores for children assessed at the start versus the end of term?	No
	Assessors conducted the permitted number of ELOM 4&5 assessments per day?	Yes
	ELOM maturation effect skewed?	No
ELOM ASSESSORS	Effect of assessors attending different training hubs?	No
	Influential assessors scoring too leniently on specific items	X

Exploratory summaries of the ELOM data revealed three assessors whose original domain mean scores were implausibly high. A closer item-by-item screen confirmed that these three assessors registered consistently higher scores than other assessors. Large positive Cohen’s d values backed these diagnoses, $d > 0.8$ (see Annex figure 1), which is conventionally accepted as a powerful effect (Cohen, 1977), and by post-survey interviews.

Before deciding on whether to adjust scores or drop these assessors’ observations, the DataDrive2030 team more accurately quantified bias. Modelling was conducted (see Annex 3 for details) and clarified the picture: the three previously identified lenient assessors topped the national bias ranking. As a result, for the three assessors flagged as too lenient, the estimate of assessor bias from the modelling was treated as a constant to be subtracted from every domain total score they had awarded. All domain scores for these three assessors were re-centred in this way so that internal rank order within a child’s profile would be maintained.

05 Some Considerations for the Next Index Round

There are two main areas for improvement for the next *Index* round.

- **Enhance and expand enumerator training and in-field quality assurance.** Extended training would provide additional practice in the field for assessors with limited prior experience in assessment or classroom observation. In addition, expanded in-field quality assurance would involve direct observation of child and LPQA assessments for all assessors during the initial fieldwork week to ensure compliance with protocols and consistency in assessment administration. Budget permitting, this could also involve systematic sampling of assessments throughout the fieldwork to confirm accuracy and consistency, to be done on a 'risk' basis to ensure efficient monitoring.
- **Streamline systems and communication across teams.** The scale and complexity of the 2024 *Index* required multiple organisations to collaborate on a large number of interdependent activities under very tight timelines. There is further scope to consolidate data platforms across teams and to facilitate better coordination, data transfers, communication and quality assurance of fieldwork implementation and incoming data.

Annex 1: LPQA Revisions, Scoring, Ratings and Factor Structure

A1.1 Background

The DataDrive2030 ELOM Learning Programme Quality Assessment (LPQA) is an observation-based assessment that looks at 22 items related to ELP quality. A review informed the selection of items, drawing on relevant literature and the key classroom focus areas of established observational measures to support construct validity. While definitions of high-quality early childhood care and education (ECCE) vary somewhat depending on context, and include many important factors beyond the classroom, there is general consensus that the learning environment set up and resourcing, curriculum, approaches to extending learning, teacher-child interactions and relationships are all critical contributors working together to influence programme quality and child learning outcomes. Items that are included in the LPQA speak to these factors and meet the following criteria:

- Easily observable (through direct observation or documentary records) and with differentiated levels for scoring
- Strongly associated with overall quality scores and child outcomes related to readiness to benefit from Grade R
- Aligned with the SA National Curriculum Framework, aims towards Grade R competencies and the ECD Programme sections of the Children’s Act ECD Norms and Standards

To conduct the LPQA, assessors must spend at least two hours in an ELP observing the various aspects of programme quality. In version 1 (v1) of the LPQA used for the *2021 Index*, assessors assigned a rating - inadequate, basic, or good - to each item based on a set of criteria, an approach that depends heavily on the assessors interpreting the criteria consistently. The LPQA (v1) was included in the *2021 Index* data collection and administered to a subsample of approximately 500 ELPs. However, concerns arose regarding assessors’ ability to interpret the criteria reliably and consistently at scale, which prompted a revision of the tool for use in the *2024 Index*. Additionally, there were concerns around the five domains that were identified in the confirmatory factor analysis of (v1). Specifically, the item ‘Gross Motor Materials’ did not contribute to any domain, and the ‘Assessment for learning and teaching’ domain only consisted of two items, limiting its reliability as a factor or domain. The LPQA (v1) was included in the pre-test ahead of the *2024 Thrive by Five Index* data collection. The pre-test confirmed the concerns around reliability and consistency, and the decision was made to revise the LPQA ahead of the 2024 data collection.

A1.2 Revisions to Create LPQA (v2)

To improve consistency and reliability, version 2 (v2) broke down each item into multiple sub-questions with simple response options (for example, Yes or No). While the content and criteria for each item remain the same as in v1, disaggregating the items into sub-questions simplifies administration for the assessor and helps yield more reliable data. Each sub-question is scored using a predefined numeric code (ranging from 0 to 2), and the total score determines the item's overall rating.

An example using the room arrangement item is shown in Annex table 1. This process was followed for each of the 22 items in the LPQA.

Annex table 1: Room Arrangement Item			
1. Room arrangement			
1a) How many organised learning or interest areas are there in the classroom?	0-2 learning areas (0)	3 learning areas (1)	4 or more learning areas (2)
1b) Are quiet learning areas separated from active and/or noisier ones?	Yes (1)		No (0)
1c) Is the space arranged so that children can carry out activities independently of the practitioner?	Yes (1)		No (0)

LPQA (v2) domain and total scores are also classified into three rating categories - Inadequate, Basic, and Good. The process for determining domain rating categories involves transforming raw item scores into a standardised percentage, as follows:

- 1. Convert item rating categories to numeric codes:** Each item is initially rated as inadequate, basic, or good based on criteria outlined in (v1) (Annex Table 2). These ratings are then coded numerically as 1 (Inadequate), 2 (Basic), or 3 (Good) to standardise the scale across items.
- 2. Sum the numeric codes within each domain:** For each of the five domains, the numeric codes of the items are added together.
- 3. Calculate a percentage score:** The sum of the numeric codes for each domain is divided by the maximum possible score for that domain – 12 for domains with four items ($4 \times 3 = 12$) or 15 for domains with five items ($5 \times 3 = 15$) – to yield a percentage.
- 4. Assign the rating category:** The percentage score determines the domain's rating category (Inadequate, Basic, or Good), as specified in Annex table 2.

For the total LPQA score, the numeric codes of all 22 items are summed and divided by the maximum possible score of 66 ($22 \times 3 = 66$) to obtain a percentage. Similarly, this percentage is then mapped to the Inadequate, Basic or Good category.

Annex table 2: Domain and Total Rating Categories

Rating category	Threshold	Interpretation
Inadequate	Less than 60% of domain total	Room for significant improvement
Basic	60% to <80% of domain total	Some room for improvement
Good	80% or more of domain total	Small enhancements possible

A1.3 Psychometrics for LPQA (v2)

Psychometric validation on LPQA (v2) was conducted on the *2024 Thrive by Five Index* ELP sample. In collaboration with content specialists, several adjustments were made to the initial factor structure, resulting in a revised structure comprising five domains with 22 items, as outlined in Annex table 3. The revised LPQA (v2) structure was developed through an iterative process combining statistical analysis and theoretical considerations.

Annex table 3: LPQA v2 Domains and Items

Domain	Item	
Materials and Equipment	Open-ended materials Gross motor materials Indoor materials	Developmentally appropriate materials Accessible materials
Planning and Assessment	NCF & ELDA's Programme planning	Child observation Recording progress
Learning Programme	Daily schedule Numeracy & mathematics Language & literacy	Free play Group times
Teaching Strategies	Room arrangement Staff engagement	Open-ended questions Opportunities for independence
Relationships and Interactions	Child-child interactions Staff-child interactions	Children's efforts acknowledged Behaviour management

There was strong evidence supporting the reliability of the LPQA (v2) as a measure of programme quality, demonstrated by high Cronbach's Alpha (0.905) and McDonald's Omega values (0.927) (Hayes & Coutts, 2020). Confirmatory factor analysis strongly supported the internal factor structure of the LPQA v2 domains (in other words, the analysis showed strong evidence that the items within each domain consistently measured the same underlying concept). Hierarchical linear regression analyses were used to investigate the ability of the LPQA (v2) domain and total scores to predict early school readiness, as assessed by the ELOM 4&5.

Although the LPQA (v2) Total Score shows a statistically significant relationship with ELOM 4&5 Total Score after controlling for ELP fee band, its predictive effect is minimal, explaining only a small proportion of the variance, underscoring the substantial influence of numerous other factors that were not controlled for.

A1.4 Considerations for the Next Index Round

Item adjustments: Based on feedback from trainers and assessors, as well as data exploration, changes to three of the items in the LPQA will be considered for the next survey round. These include removing the skip logic for one section, changing the framing of one question to reduce the use of the 'Not applicable' choice, and rewording one question to improve scoring accuracy.

Tool administration: Consider increasing the minimum observation duration to ensure the assessor observes a larger proportion of the ELP's daily activities.

Annex 2: Derivation of the New ELP Fee Variable

The collected fee data is important as it will be used to construct the *2024 Index* fee groups that serves as a key disaggregator for the *2024 Index*. There were two data sources for the new variable: (1) principal reported fees charged and (2) PCG reported fees paid (available for 95% of ELPs). Data from these sources were combined to create a new fee variable. Before creating this variable, five enumerator finger errors and five observations for one enumerator who mistakenly collected annual rather than monthly principal-reported fee data were corrected during the data checking and cleaning by the DataDrive2030 team.

The rules used to create the new fee variable are set out below; the general principle was to trust the principal's reported data unless there was clear evidence to the contrary.

- If no primary caregiver fee data, use principal-reported data
- If no principal fee data, use:
 - The primary caregiver value if there was one respondent
 - The matched value, if two or more primary caregiver values matched
 - The maximum primary caregiver value, if there was more than one respondent, but no matching value.
- If the principal value equals the maximum reported primary caregiver value, use principal-reported data
- If the difference between the principal and maximum primary caregiver value is less than R50 or 25%, use principal-reported data
- If the difference between the principal and maximum primary caregiver value is greater than R50 and 25% use:
 - The principal value if only one primary caregiver value, and it does not match
 - The principal value if at least one primary caregiver value matches
 - The principal value if there is no principal and primary caregiver match, and no match between two or more primary caregivers
 - The matching primary caregiver value, if at least two primary caregiver values match.
- If the principal value is zero and at least one primary caregiver value exists use:
 - The primary caregiver value if one respondent
 - The matched value if two or more primary caregiver values match
 - The maximum primary caregiver value, if more than one respondent, but no matching value.

Annex 3: ELOM Score Adjustments (Enrolled Children)

A3.1 Background

After receiving the ELOM dataset prepared by the survey company, the DataDrive2030 technical team, led by Professor Colin Tredoux, conducted extensive checking of the ELOM data using the statistical software R. Some of the preliminary results raised queries and recorded evidence and statistical anomalies suggested that a small number of ELOM assessors were potentially not assessing to the same standards and/or reliability as the other ELOM assessors. This meant there was a likelihood of bias in some of the assessor judgements, and this would impact both provincial and national ELOM results.

As the *2024 Index* will be used as a baseline for future survey rounds, it is crucial that it gives a correct picture of the current state to track changes over time accurately. It is not ideal to make post-data collection corrections to data, but it is preferable to allow knowingly incorrect data to be used. Established methodologies and procedures do exist to allow for this sort of correction. It was decided, in consultation with multiple specialists, to correct for assessor bias using objective evidence from within the collected data. Their recommendations were:

- To objectively identify which assessors, and which of their items, were systemically impacted through statistical means. The reason for identifying assessor-item pairings was to allow for adjustments targeted at only those cases where post data-collection alignment to standards was needed to give accurate and reliable results
- Rescale/recentre any erroneous judgments or scores rather than deleting all the domain or assessor scores, as this might introduce hidden biases to the data. This would also maintain as much statistical power in both the national and provincial results as possible; and
- Apply scaling/centering to the identified assessor-item pairings to align all the assessors' judgements to the standards. Easy methods such as those correlating with other more objective domain scores or averaging across non-impacted domain scores to give a best estimate of the erroneous score are preferred.

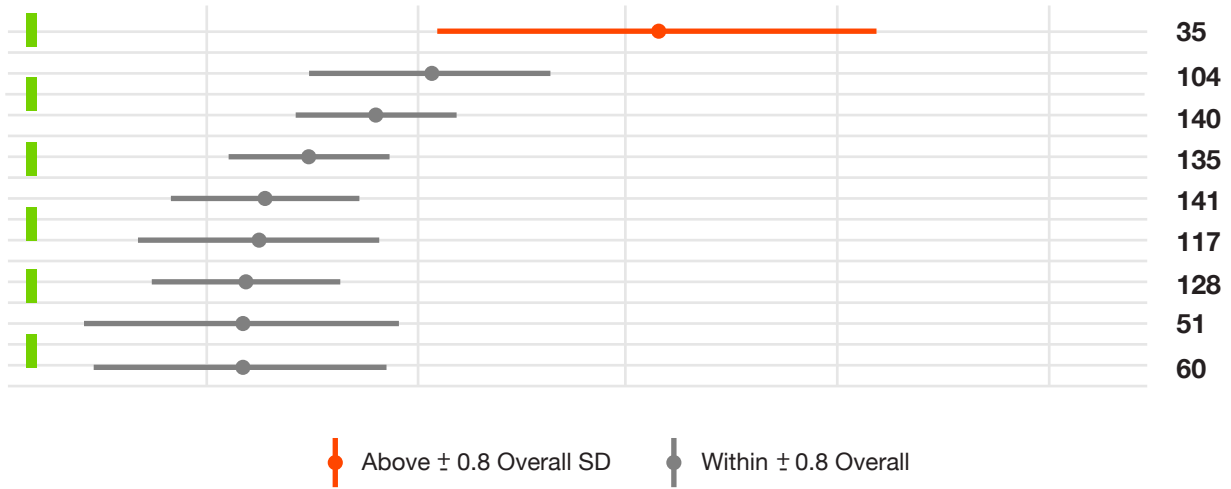
A3.2 Data Exploration

Exploratory summaries of the 2024 ELOM data revealed one assessor who recorded identical scores for over 100 children on a single item, an extremely unlikely outcome, and three assessors whose original domain mean scores were implausibly high. A closer item-by-item screen confirmed that these three assessors consistently gave higher scores than other assessors. These discrepancies were backed by large positive Cohen's *d* values ($d > 0.8$, conventionally accepted as a very strong effect), and by post-survey interviews in which, in several cases, procedural errors were confirmed.

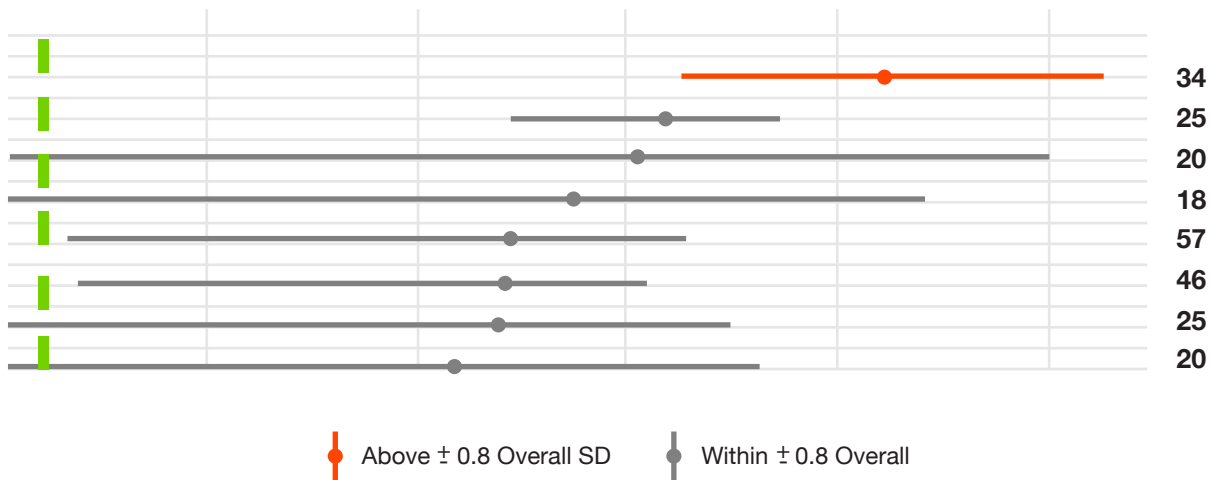
Annex Figure 1: Selected Cohen's d for assessors scoring too leniently and for assessor with zero standard deviation for ELOM score 13

In the panels below, the dots represent mean scores, and the lines represent the 95% confidence interval for each assessor's observations. The numbers on the right-hand side are the number of ELOMs conducted by each assessor.

Panel 1: Assessor 1 scoring too leniently, national all ELP fee quintiles, Total Score (red line)



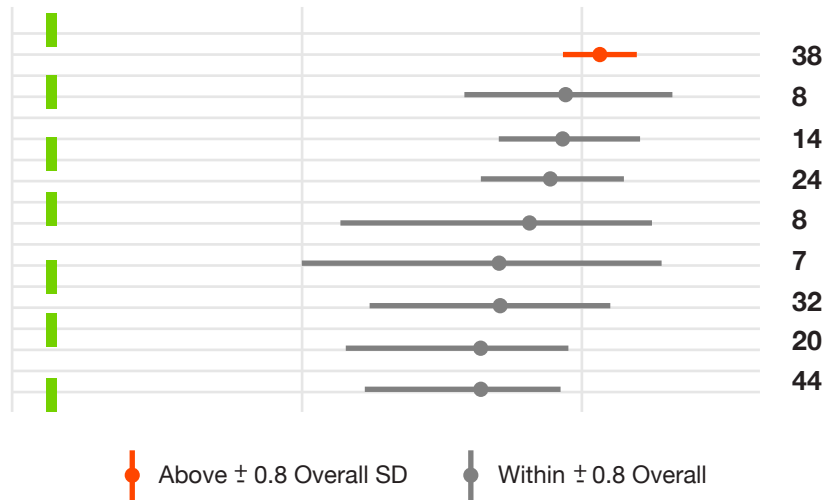
Panel 2: Assessor 2 scoring too leniently, national ELP fee quintile 4, domain 1 score (red line)



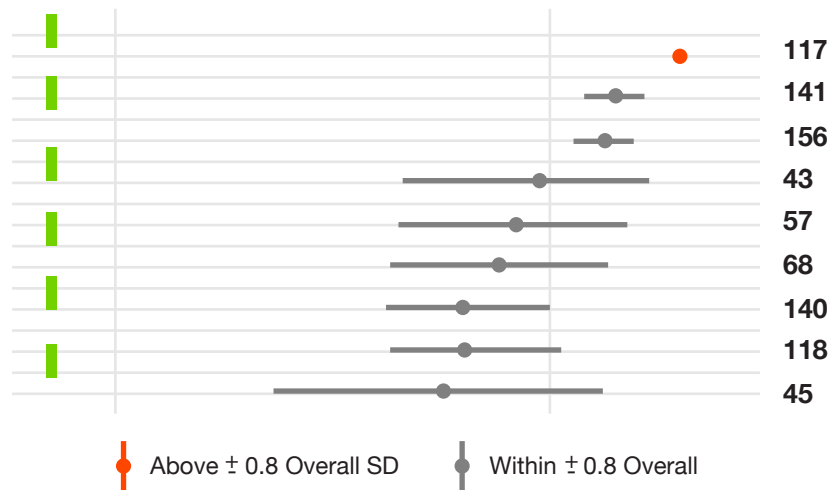
Annex Figure 1: Selected Cohen's d for assessors scoring too leniently and for assessor with zero standard deviation for ELOM score 13

In the panels below, the dots represent mean scores and the lines represent the 95% confidence interval for each assessors' observations. The numbers on the right-hand side are the number of ELOMs conducted by each assessor.

Panel 3: Assessor 3 scoring too leniently, national ELP fee quintile 1, score 22 (red line)



Panel 4: Assessor with standard deviation zero for score 13, national all ELP fee quintiles (red dot)



Before deciding on a remedy, whether to adjust scores or remove these assessors' observations, the team decided to quantify bias more precisely. A mixed-effects model was fitted in which child age (in months), sex and socio-economic circumstances (approximated by ELP fee quintile) were entered as fixed effects, while the hierarchical sampling units - child, ELOM domain, assessor - were treated as random intercepts. In the baseline linear specification, the assessor variance component was not large, yet substantial enough to distort many estimates if left uncorrected. Extraction of the resulting best-linear-unbiased predictors made the picture unambiguous: the three previously flagged lenient assessors topped the national bias ranking, with large random intercepts, while the one other assessor showed a different anomaly entirely—several flat item distributions (standard deviation zero)—confirming the earlier protocol concerns.

A3.3 Adjustment Rules

Based on these findings, the team agreed on a set of adjustment rules that would be transparent, replicable and minimally disruptive. For the three assessors flagged as too lenient, the mixed-effects-model bias estimate was treated as a constant to be subtracted from every domain total score they had awarded. In practice, the adjustment was implemented by removing those assessors' cases, recomputing age-by-sex-by-fee-quintile-specific means for the remaining pool of assessors, calculating each outlier's deviation from those benchmarks and then shifting their original scores accordingly. All domain scores for the three assessors who were judged to be outliers were recentred in this way so that internal rank order within a child's profile would be maintained. After the adjustment, three further observations received zero scores on two or more domains and were removed to ensure adherence to the ELOM rules.

The zero standard deviation assessor faced a different issue. After weighing two alternatives—imputing only the aberrant items versus discarding all their observations—the team decided to remove them because an error of this sort and magnitude is likely to be indicative of other issues in administering the assessment properly.

Once the adjustments had been scripted, another analyst independently replicated the full adjustment pipeline in a different statistical software to verify the procedure and coding accuracy. The logic of the procedure was also reviewed, endorsing the described approach as preferable to a wholesale re-weighting that might erase genuine score differences.

In its final form, the adjustments eliminated all ELOM assessments conducted by an unauthorised assessor and by one assessor who had zero standard deviation on one item for over 100 children and reclustered all domain scores awarded by the three systematically lenient assessors. This exercise is supported by a single reproducible R script that future Index survey rounds can re-run with minor parameter edits.

Annex 4: Non-Enrolled Sub-Study Supplemental Information

A4.1 Objectives

For the *2024 Index*, the non-enrolled child sample was a ‘test’ group to assess the feasibility of reaching non-enrolled children and, as such, was substantially smaller than the enrolled children sample. The objective was to learn more about the home circumstances of non-enrolled children and to be able to make indicative comparisons of the key child development outcomes measured by the Index: early learning (percentage On Track/average ELOM scores) and physical growth (percentage stunted), for non-enrolled and enrolled children aged 50-59 months.

A comparison of the key child development outcomes across statistically representative samples of non-enrolled and enrolled children from the lower end of the socio-economic spectrum was not financially feasible for the *2024 Index*. Instead, an approach that allows for comparison of the key child development outcomes across statistically non-representative but indicative samples of non-enrolled and enrolled children from the lower end of the socio-economic spectrum was selected.

A4.2 Eligibility Criteria

Ideally, the non-enrolled sample would consist only of children with no exposure to an ELP. However, based on experience (for instance, the 2022 Western Cape stunting baseline survey and the 2022 National Census) and the findings from the *2024 Index* pre-test, it was not deemed feasible to identify enough children with zero ELP exposure in practice within a feasible resource envelope. Therefore, the target population for the non-enrolled sample were children aged 50-59 months not currently enrolled in any early learning programme (ELP) and who come from the lower end of the socio-economic spectrum. An ELP was defined as: pre-school, nursery school, crèche, ECD centre, mobile ECD service, Grade R or playgroup.

There was also a disability eligibility criterion. Children were ineligible for the sample if they had difficulties (a) seeing (with or without glasses); (b) hearing (with or without a hearing aid); or (c) walking or moving their arms or have weakness or stiffness in the arms or legs; and/or (d) significant difficulties understanding what people say.

A4.3 Pre-test Findings

During the pre-test in April-May 2024, households generally welcomed and engaged with the enumerators. Nevertheless, there were two main challenges.

Challenge 1: Initially, an interval of two homes was used during the random walk, but this was changed to one because it was difficult to locate enough households with children aged 4 to 5 years (the Household visit screening form (first visit) asked for this age group rather than only 4-year-olds to avoid incorrectly excluding households with potentially age-eligible children).

Challenge 2: In households with 4-year-olds, the main challenge was to find children who met the original ELP attendance eligibility criterion defined as not having attended an ELP in 2024 or in 2023. Based on the initial ELP attendance criterion, 1.3% of households visited had eligible children, and the ELOM assessment and primary caregiver interview were administered. This proportion was considered too low for the sub-study, and it was deemed necessary to relax the eligibility criterion to “did not attend an ELP in 2024” (i.e. children who attended an ELP in 2023 were considered eligible). This meant adopting the assumption that for children who attended an ELP before 2024, any effects on key child development outcomes would have faded out by the time of the data collection.

A4.4 Ward Selection for the Sub-Study

To select the 45 wards in the three provinces – Gauteng, KwaZulu-Natal and Western Cape – for the sub-study of non-enrolled children, the following steps were used.

1. Across the three provinces, wards that had fewer than 15 enrolled Grade 3 learners (reliable, recurrent administrative data) per square kilometre were excluded
2. Three area types – informal, rural and urban formal – were defined using Statistics South Africa 2011 Census sub-area type: (a) urban formal = formal residential areas; (b) informal = informal residential areas; and (c) rural = farms + small holdings + traditional. (Statistics South Africa 2022 Census classifications not available in time)
3. Cutoffs for a ward to count as a given area type were as follows:
 - a) **Informal:** Consists of 10% or more informal residential areas. (Before the fieldwork, the survey firm examined up-to-date maps of each relevant ward to locate the informal areas and only cover those for the non-enrolled sample and the informal areas located furthest away from the ELPs listed for the enrolled children component)
 - b) **Rural:** Consists of 20% or more farms, small holdings and traditional areas and not already selected as informal (and covers the rural areas located furthest away from the ELPs listed for the enrolled children component)
 - c) **Urban formal:** Consists of 60% or more formal residential areas and not already selected as informal or rural (and covers the urban formal areas located furthest away from the ELPs listed for the enrolled children component)
4. More than 15 wards exist for a given area type (informal, rural, urban informal); the wards were to be ranked by Grade 3 enrolment per square kilometre and the wards with the higher cut-off values were to be selected.

Using these rules, 45 wards were chosen for the sub-study of non-enrolled children. After up-to-date maps showing structures had been prepared for each of the selected wards, these were visually inspected before data collection to ensure there had not been any significant changes since 2011. For instance, large outmigration from a ward results in very low population (and child) density. No such cases were identified.

Annex 5: Ethics Overview

Ethical approval for the *2024 Index* was obtained from the University of Cape Town's (UCT) Faculty of Humanities Ethics Committee (Reference Number: PSY2024-032).

The study was designed and implemented in strict accordance with national and institutional ethical guidelines that govern research involving human participants, especially vulnerable populations such as young children.

The research process prioritised informed consent, voluntary participation, and the protection of personal and sensitive information at every stage of the project. All individuals involved in the study, including children, primary caregivers, early learning practitioners, and principals, were fully informed about the nature and purpose of the research. Consent was obtained before participation, with additional provisions made for age-appropriate assent where applicable.

To ensure the safety and protection of all participants, all fieldwork teams, including enumerators and assessors, underwent comprehensive vetting procedures, including:

- South African Police Service (SAPS) clearance
- Sexual Offences Register clearance application, and
- Child Protection Register clearance application.

Only individuals who met all the required safeguarding criteria were permitted to engage directly with children and other participants.

Throughout the study, measures were taken to safeguard participant confidentiality and to uphold data integrity. Data was anonymised during analysis and securely stored, with access limited to authorised personnel. Fieldworkers were trained in ethical conduct and safeguarding protocols, and continuous oversight mechanisms were put in place to monitor adherence to ethical standards across all sites.

The *Thrive by Five Index* remains committed to upholding the highest ethical standards in research and recognises the importance of respecting the dignity, rights, and well-being of all participants involved.

The various *2024 Index* consent forms are available on request from the DataDrive2030 Index team.

References

- Anderson, K.J., Henning, T.J., Moonsamy, J.R., Scott, M., Du Plooy, C. & Dawes, A.R.L. (2021). 'Test-retest reliability and concurrent validity of the South African Early Learning Outcomes Measure (ELOM)', *South African Journal of Childhood Education* 11(1), a881. <https://doi.org/10.4102/sajce.v11i1.881>
- Ardington, C. (2024). *2024 Thrive by Five Index Sampling Strategy Enrolled Children*. Cape Town: DataDrive2030.
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*, Revised Edition. Academic Press: New York
- Dawes, A., Snelling, M., & Biersteker, L. (2023). *DataDrive2030 Home Learning Environment Tool Technical Manual*. Cape Town: Datadrive2030
- Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L., Tredoux, C.G. et al. (2020). *Early Learning Outcomes Measure (ELOM) Technical Manual*. Claremont, Cape Town: The Innovation Edge.
- Duncan, G. J., Claessens, A., Huston, A.C., Pagani, L.S., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 44, 1428-1446.
- Genesis Analytics. (2025). *Thrive by Five 2024 Fieldwork Report*. Cape Town: Genesis Analytics.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38 (5), 581–586. doi:10.1111/j.1469-7610.1997.tb01545.x
- Hayes, A. F., & Coutts, J. J. (2020). Use Omega instead of Cronbach's Alpha for Estimating Reliability. *Communication Methods and Measures*, 14(1), 1–24. <https://doi.org/10.1080/19312458.2020.1718629>
- Kika-Mistry, J. (2025). Indicator options for tracking the gap in early learning outcomes for children from different socioeconomic backgrounds. *Background Note for the 2024 South Africa Thrive by Five Index*. Cape Town: DataDrive2030.
- SALDRU. (2025). *Thrive by Five Closeout Report. Report 2. January*. Cape Town: Southern Africa Labour & Development Research Unit.
- Snelling, M., Dawes, A., Biersteker, L., Girdwood, E., & Tredoux, C. (2019). The development of a South African Early Learning Outcomes Measure: A South African instrument for measuring early learning programme outcomes. *Child: Care, Health and Development*, 45, 257-270. <https://doi.org/10.1111/cch.12641>

Tran, T. D., Luchters, S., & Fisher, J. (2017). Early childhood development: impact of national human development, family poverty, parenting practices, and access to early childhood education. *Child: Care, Health and Development*, 43(3), 415-426.

Watts, T. W., Jenkins, J. M., Dodge, K. A., Carr, R. C., Sauval, M., Bai, Y., Escueta, M., Duer, J., Ladd, H., Muschkin, C., Peisner-Feinberg, E., & Ananat, E. (2023). Understanding heterogeneity in the impact of public preschool programmes. *Monographs of the Society for Research in Child Development*, 87(4). <https://doi.org/10.1111/mono.12463>



2024 | **thrive⁵**
thrive by five | INDEX
BUILDING FOUNDATIONS FOR LIFELONG SUCCESS

The Thrive by Five Index 2024 is led by the Department of Basic Education and coordinated by DataDrive2030, with the support of the FirstRand Foundation, the LEGO Foundation, This Day, and Yellowwoods.

