



South Africa 2024 Thrive by Five Index

Datasets Summary: Enrolled Children

Pub Version 1

Prepared by DataDrive2030 Tb52024 Technical Team, 7th of May 2025

POPI Act

The 2024 Thrive by Five Index Pub Version 1 datasets must be stored in compliance with the POPI Act, and the data must not be used for purposes other than those agreed to in the NDA.

Study objectives¹

The objectives of the South Africa Thrive by Five Index 2024 (2024 Index) are to:

- determine the proportion of young children in South Africa who attend early learning programmes (ELPs) that are on track for their age in three key areas of development, early learning, social and emotional functioning (SEF) and physical growth;
- monitor trends in performance gaps and gains 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 ECD decision-making and actions at multiple levels.

Survey forms administered in the field

The development of the 2024 Index survey forms was guided by the [Nurturing Care for Early Childhood Development Framework](#) and informed by the 2021 Index survey forms and 2021 analysis. The 2024 Index form development included consultations with more than 50 national and international researchers from across the required fields of expertise.

To achieve the 2024 Index study objectives, eight survey forms were used to collect data for a statistically nationally and provincially representative sample of children aged 50-59 months attending ELPs across South Africa:

1. child sampling;
2. early learning outcomes measure for ages 4 and 5 (ELOM4&5);
3. practitioner rating of child social and emotional functioning (SEF);
4. facility observation;
5. learning programme quality assessment (LPQA);
6. principal interview;
7. practitioner interview; and
8. primary caregiver interview.

Comparison of 2021 and 2024 survey forms

The **child sampling form** (not public use data) is new for the 2024 Index. It captures enrolment and attendance data to identify the group of eligible children to randomly select children for the child

¹ The 2024 Index also include a *pilot sample of children aged 50-59 months not enrolled in ELPs who live in low-income areas in three provinces: Gauteng, KwaZulu-Natal and Western Cape*. Two survey forms were used to collect data for the pilot sample, a household version of the ELOM 4&5-SEF form and a household version of the primary caregiver interview. The 2024 Index also contains a *sub-study by the South Africa Medical Research Council (SAMRC) to measure environmental factors in ELPs to determine the presence and concentration of metals (including lead, arsenic, cadmium and mercury) and heat exposure within ELPs*. This sub-study covers 70 ELPs purposively selected from the main study sample of ELPs in Gauteng province.

South Africa Thrive by Five Index 2024

assessment, to enable construction of weights and collect contact information for the selected children's primary caregivers.

The **ELOM4&5 form** is the same as the one used for the 2021 Index except for the addition of a small number of complementary questions; slight revisions to the disability screening questions; and that the practitioner SEF section was extracted and administered as a separate form to facilitate the interviews with practitioner on the day of the data collection visit.

The **practitioner rating of children's social and emotional functioning (SEF) form** is the same as in 2021 with two additional subscales namely, task focus and following instructions, each consisting of four questions.

The **practitioner and principal interview forms** for the 2024 Index were substantially revised compared to the 2021 forms and were extended to include new topics such as views on gender, growth mindset and play and barriers to doing the job.

The 2024 **LPQA form** covers five learning areas as did the 2021 form, but the former was revised to add skip logics and simplify responses choices in order to facilitate administration of the form in the field. By necessity, this involved some reformulation of questions.

The **facility observation form** is a shortened version of the 2021 form, with some question revisions and limited additions.

The **primary caregiver interview form** is new for the 2024 Index. 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 with about 5,000 PCGs. It captures features of the child's socio-economic background, home learning environment and health and key characteristics of primary caregivers.

Sample sizes

The new sampling strategy developed for the 2024 Index (*available separately*) covers all nine provinces and was implemented as intended with response rates ranging from 94% to 107% (**see Table 1**).

The planned sample size for ELPs was 1,296 and this was exceeded for each ELP level form except the practitioner form. This was because for ELPs where the principal is the sole practitioner, the practitioner form was not administered, instead a principal form with some additional questions carried over from the practitioner form was administered.

The planned child sample size was 5,189 compared to a realised sample size of 5,001. This was after removing ELOM observations (a) that were invalid according to the ELOM rules (no consent, no assent, child scored zero on two or more ELOM domains, child exited assessment early, child has a disability, assessment not conducted in child's home language); (b) for two assessors identified as problematic based on extensive statistical analysis, in-field quality assurance and post-survey interviews; and (c) that did not have the required ELP level survey forms (facility, LPQA, principal).

For the PCG sample, the target was to interview PCGs for around 70% of the assessed children. This rate was exceeded with a realised sample of 3,841 PCGs for the 5,001 assessed children yielding a response rate of approximately 77%.

Variable naming conventions, IDs, weights, flags & missing values

Variable labels are intended to be self-explanatory. But the appropriate survey form(s) should always be consulted when using the dataset(s) to ensure correct interpretation and usage.

Each survey form and by extension dataset, has its own variable prefix to indicate (a) that it is a collected (as opposed to derived) variable and (b) to indicate which survey form collected it to facilitate when multiple datasets are merged. The variable prefix n_ indicates that a variable was derived by the DataDrive2030 Tb52024 technical team using collected data. For example, in the ELOM-SEF dataset, the collected variables are prefixed elom_ while derived variables are prefixed n_elom_. *The derived variables include calculated ELOM scores and cutoffs and LPQA scores and ratings.*

There are three identifiers for the 2024 Index: id_prov=province ID, id_eed=ELP ID and id_child = child ID. The province ID is included in all datasets while the child ID is included in the ELOM-SEF and the PCG datasets.

- For merges of the ELOM-SEF (child level) and PCG (child *subsample*) datasets, use id_child.
- For merges of the facility, LPQA, principal and practitioner datasets use id_eed.
- For merges of the child-ELP level datasets use id_eed.

There are three different weights to be applied to generate nationally and provincially representative estimates.

- For child level estimates (full sample) use weight_child.
- For PCG estimates (*child subsample*) use weight_pcg.
- For ELP level estimates (facility, LPQA, principal, practitioner) use weight_elp.

While provincial sample sizes were selected to allow for as precise estimates as possible given the available budget and capacity to ensure data quality, any disaggregation by subgroups at provincial level, will require examination of confidence intervals to determine feasibility for each specific province-subgroup pairing.

In the ELOM-SEF and the principal datasets there are flag variables for a small number of variables to indicate whether they were adjusted, corrected or imputed. For example, child height was imputed for 42 children in the ELOM-SEF dataset and the flag variable is called n_elom_flag_height.

Missing values are coded as follows: .=system missing (mostly due to skip logic), .a or -1=Don't know, .b or -2=Refused, .c or -3=Not applicable, .z=Invalid response (due to tablet error in the field or inconsistent reporting due to enumerator/respondent error).

Potential remaining data errors in the datasets

The Pub Version 1 datasets have been through a two-step data checking and cleaning process, nonetheless, some data errors are likely to remain. If a data user discovers a potential data error, please fill in the provided error reporting form and submit to Simone Beukes, 2024 Index Operations Manager at simone@datadrive2030.co.za.

Before submitting an error reporting form please ensure:

- *you have checked the relevant survey forms (including skip logics);*
- *you have read the accompanying documentation (this Summary, the sampling strategy and codebooks); and*
- *unless it is a critical potential data error, to submit potential errors in batches rather than individually to facilitate the extended data cleaning.*

Table 1. Overview of the 2024 Thrive by Five Index survey forms (enrolled children)

	Variable prefix	Survey form	Respondent	Main purpose	Planned sample size	Realised sample size	Response rate
1	<i>Not applicable as this is not a public use dataset.</i>	Child sampling (face-to-face)	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 ELOM4&5 and social and emotional (SEF) rating. In ELPs with fewer than 4 eligible children, all eligible children were assessed, and the shortfall was made up in another randomly selected ELP, where the maximum number of assessed children per ELP is 6.	1296	1388	107%
2	elom_	ELOM 4&5 (face-to-face) <ul style="list-style-type: none"> gross motor development emergent literacy and language emergent numeracy and mathematics fine motor coordination & visual motor integration cognition and executive functioning height measurement using 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%
3	sef_	Child social and emotional functioning, SEF (face-to-face)	Practitioner who knows the assessed child	Capture child social and emotional functioning levels	5189	5001	96%
4	fa_	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%
5	lp_	Learning Programme Quality Assessment, LPQA (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%
6	pri_	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%
7	pra_	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	Max. 1296	1223	94%
8	cg_	Primary caregiver interview, PCG (telephonic)	Primary caregiver of assessed child	Capture child socioeconomic status, 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%