South Africa 2024 Thrive by Five Index

Sampling Strategy

Cally Ardington

Summary

Three key principles guided the 2024 Thrive by Five Index sampling strategy:

- Nationally and provincially statistically representative sample.
- Forward comparability of key indicators.
- Replicability in terms of data required for sampling for future survey waves.

A stratified multistage sampling design was used. Within each stratum, a three-stage sampling was implemented where the sampling units were defined as follows.

1. The primary sampling units (PSU) are wards, except for in a very small number of cases where two or more wards were combined to ensure a feasible enumeration area.
2. The secondary sampling units are early learning programmes (ELPs) in selected PSUs, where an ELP is eligible if it has at least 1 child aged 50-59 months attending, is open for at least 8 hours per week and is operational.
3. The third sampling unit are children attending selected ELPs. A child is eligible if they are 50-59 months old at the time of the data collection, the primary caregiver has not refused for the child to participate in the study, the child does not have difficulties seeing, hearing or walking or moving their arms and/or legs and the child does not have significant difficulty communicating with other people and the child is present on the day of the visit.

Step 1: Selection of PSUs

1. Wards are the smallest geographic unit for sampling. All wards are included to make sure that every ELP has a non-zero probability of selection. It was decided not to remove large sparsely populated areas (e.g. national parks) prior to sampling as this would require in-depth analysis for each area to be confident that no ELPs would be missed. The preference was therefore to select the PSUs and then, where necessary, study specific PSUs to determine whether sub-areas should be removed (see below). Another advantage of this approach is that areas that are currently 'empty' (no ELPs) will not be excluded from future waves of the survey.
2. PSUs consist of wards that have more than 60 Grade 3 learners, or, when necessary, of a combination of two or more wards that together have more than 60 grade 3 learners. There is a total of 4,047 PSUs, of which 3,773 are wards and 274 are combined wards.
3. Across the country, 432 PSUs are selected. These are allocated to the provinces such that each province has a minimum of 35 PSUs and then the balance of the PSUs are allocated to the other provinces in proportion to the number of two-year-olds in each province as per the 2022 National Census. The allocations are shown in the final column in the following table.

Table 1. Percentage PSU allocation across wards

Province	Number of 2-year-olds in 2022 National Census	Proportional allocation	Equal allocation	2024 Thrive by Five Index allocation (minimum 35 PSUs per province)
Eastern Cape	147,438	13%	11%	**11%**
Free State	52,994	4%	11%	**8%**
Gauteng	251,747	21%	11%	**18%**
KwaZulu-Natal	242,580	21%	11%	**17%**
Limpopo	152,375	13%	11%	**12%**
Mpumalanga	106,369	9%	11%	**9%**
North West	79,547	7%	11%	**8%**
Northern Cape	26,023	2%	11%	**8%**
Western Cape	119,226	10%	11%	**9%**

4. Confidence interval widths for each province, depend both on the number of PSUs allocated to that province and also the degree of homogeneity of child outcomes within PSUs and ELPs in that province. Using the national design effect for the percentage of children on track from the 2021 Thrive by Five Index study, expected confidence interval widths were calculated for provincial estimates as shown in the following table.

Table 2. Calculated confidence interval widths for provincial estimates

Province	Number of PSUs	Number of ELPs	Number of children	CI width for % children on track
Eastern Cape	49	147	588	10.8%
Free State	35	105	420	12.7%
Gauteng	76	228	912	8.6%
KwaZulu-Natal	73	219	876	8.8%
Limpopo	50	150	600	10.6%
Mpumalanga	38	114	456	12.2%
North West	35	105	420	12.7%
Northern Cape	35	105	420	12.7%
Western Cape	41	123	492	11.8%
National	**432**	**1,296**	**5,184**	

5. PSUs were classified by the weighted school quintile, where weights were based on the number of Grade 3 learners in each school. There were 34 (0.8%) PSUs with unknown school quintile. These PSUs were assigned a quintile value of 5, the highest socioeconomic quintile as they mostly consisted of independent schools. In all provinces other than the Western Cape and Gauteng, quintiles 4 and 5 were collapsed into one as there were a limited number of PSUs in the top two quintiles.
6. Province and the quintile variable are used to create 38 strata. The number of PSUs per stratum is shown in the following table.

Table 3. Number of PSUs per strata

Province	Number of PSUs					
	Quintile 1	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	**Total**
Eastern Cape	201	167	269	44		681
Free State	65	87	82	34		268
Gauteng	49	64	117	88	161	479
KwaZulu-Natal	206	289	258	114		867
Limpopo	185	243	94	27		549
Mpumalanga	67	94	80	99		340
North West	85	114	128	29		356
Northern Cape	27	53	39	43		162
Western Cape	45	68	73	94	65	345
National	**930**	**1,179**	**1,140**	**572**	**226**	**4,047**

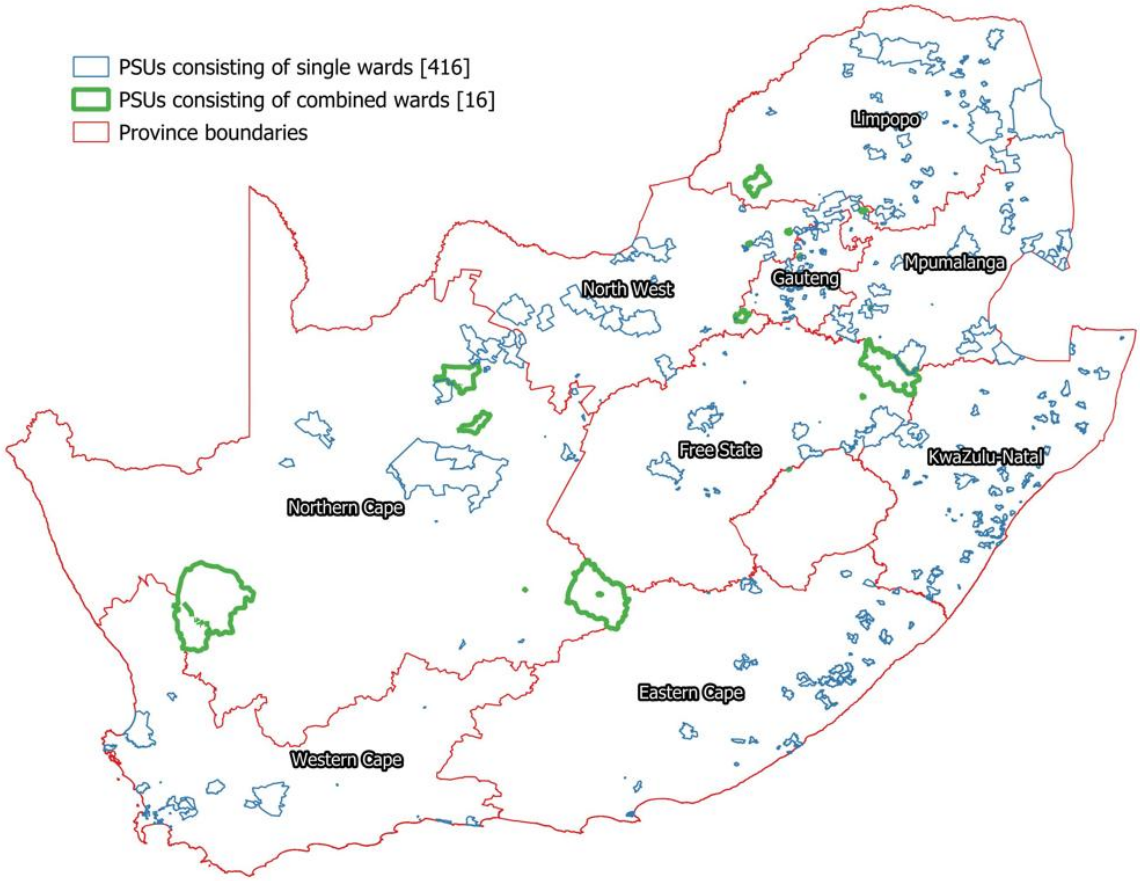
7. Where the number of PSUs allocated to a province is not divisible by four (or by five in the case of the Western Cape or Gauteng), more PSUs are allocated to the strata with a higher number of PSUs. The allocation of PSUs to be selected by province and quintile is shown in the following table.

Table 4. Number of PSUs to be selected by stratum

Province	Number of PSUs					
	Quintile 1	**Quintile 2**	**Quintile 3**	**Quintile 4**	**Quintile 5**	**Total**
Eastern Cape	12	12	13	12		49
Free State	9	9	9	8		35
Gauteng	15	15	15	15	16	76
KwaZulu-Natal	18	19	18	18		73
Limpopo	13	13	12	12		50
Mpumalanga	9	10	9	10		38
North West	9	9	9	8		35
Northern Cape	8	9	9	9		35
Western Cape	8	8	8	9	8	41
National	**101**	**104**	**102**	**101**	**24**	**432**

8. Within each stratum, PSUs are selected with probability proportional to size, where the measure of size is the number of Grade 3 learners.
9. The selected PSUs are shown in the map below, where green PSUs indicate combined wards.

Figure 1. The primary sampling units selected for the 2024 Thrive by Five Index



Step 2: Full listing of all ELPs in selected PSUs

1. The fieldwork company conducted a full listing of all ELPs in the 432 selected PSUs to create the ELP sampling frame for the 2024 Thrive by Five Index.
2. Where selected PSUs are very large and sparsely populated, the fieldwork company had the option to motivate on a case-by-case basis to remove some areas of the PSU from the listing exercise, after providing thorough documentation and agreement by DataDrive2030's Technical Lead and Sampling Specialist ahead of starting the ELP listing. But this option was never exercised.
3. The listing collected the following data from each ELP:
    a. Name, address, GPS location and contact information.
    b. Number of days open per week and number of hours open per week.
    c. Total number of learners 4-5 years old (from 4 years and one day to 4 years and 364 days or 50 – 59 months) by gender
    d. Total number of learners – Male.
    e. Total number of learners – Female.
    f. Total number of teaching staff.
4. In addition to providing the ELP sampling frame, the listing will update the status of ELPs in the 2021 ECD Census for the selected PSUs.

Step 3: Sampling of ELPs (SSU)

1. Within each stratum, the target number of ELPs was equal to three times the number of selected PSUs.
2. If all selected PSUs in a stratum have at least three ELPs with registered four-year-old children, three ELPs were randomly selected per each selected PSU.
3. For strata where some PSUs have less than three ELPs with registered four-year-old children:
    a. If there are 3 or fewer ELPs, all ELPs were selected for fieldwork (0 to 3 ELPs).
    b. The remaining sample of ELPs was allocated across the PSUs in the stratum with more than 3 ELPs by the DataDrive2030 sampling specialist.
4. This means the desired sample could be achieved without needing to replace PSUs. This is possible because the full listing (Step 2) 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 (although substitutions are common in practice, they do complicate the calculation of selection probabilities).

Step 4: Selection of children within selected ELPs (TSU)

1. Within each PSU, the aim was to assess a number of 50 to 59-month-old children equal to four times the number of selected ELPs, and equal numbers of boys and girls.
2. All present eligible (see above) children in an ELP were randomly ranked. The two highest ranking girls and two highest ranking boys were selected for assessment. If a child could not be assessed (e.g. did not assent, primary caregiver did not consent), then the next highest ranking learner of the same sex was assessed.

Replacement samples and weights

1. There will be no substitutions of PSUs.
2. All eligible ELPs within a PSU were randomly ranked. The three top ranked ELPs were selected for fieldwork. If an ELP was unavailable during fieldwork, the next highest ranking ELP was used as a replacement.
3. As child randomisation and replacement was done on site, all details were recorded in a systematic manner using a child sampling form to allow for the calculation of weights and assessment of potential bias.
4. During fieldwork, if fewer than 12 children were assessed in the first three ELPs in a selected PSU, an additional ELP was selected within the same PSU, or if this was not possible, an additional ELP was selected in another PSU within the same stratum. These were not replacements but additions to the ELP sample to reach the target of 12 children per PSU.
5. Weights will be adjusted to account for any substitutions as per points 2 and 3 above.

6. The weight is the inverse of the probability of selection. The probability of selection is the product of the following probabilities:
    - probability that a PSU is selected = (number of PSUs selected in stratum)*(Grade 3 learners in PSU)/(Grade 3 learners in stratum).
    - probability that an ELP is selected = (number of ELPs selected in that PSU)/(number of ELPs in PSU).
    - probability that a child is selected = (number of children selected in that ELP)/(number of eligible children in ELP).