

Data sources and collection

Data from multiple programmes and projects who used the ELOM tools were collated, spanning from 2019-2022. For data to be included in the meta-dataset, children had to have a complete and valid¹ ELOM 4&5 child outcomes assessment (further described below).

Since data from various studies were collated, not all studies and programmes set out to collect the same data using the same tools, consistently. It is for this reason that there are large patterns of “missingness” within the data, indicating different uses of various tools and measurements across subgroups. For example, out of the total sample of 12,265 children, 2277 children do not have height data. This is a combination of certain programmes and studies that did not intend to measure child height at all, and missing data when programmes do aim to collect information on the variable but the data is missing for unknown reasons. In order to counter this, a “data set” variable is included in the meta-data that allows the user to identify sub-sets of the data. Various tools and measurements were also used by different studies and programmes, with the common tool being the ELOM 4&5 Years assessment tool. Under each data source, we clarify the sets of tools and measurements used. First, we describe the main ELOM tools used.

Tools

The data used comes from various data collection efforts that have used the ELOM tools. The ELOM tools included in the data are 1) The ELOM 4 & 5 Years Assessment Tool, 2) The Socio-Emotional Rating Scale, 3) The Home Learning Environment (HLE) Tool, and 4) The Learning Programme Quality Tool. Additional non-ELOM tools were included in the Department of Basic Education 2021 Baseline Audit, which is described in its respective data source section.

- 1) **The ELOM 4 & 5 Years Assessment Tool** measures whether preschool children are on track for their age in key areas of development. It is a standardised tool that measures performance across five key developmental domains for children aged 50 to 59 months and 60 to 69 months. The scoring assesses a child in 23 items, across five domains, namely: gross motor development, fine motor development, emergent numeracy and mathematics, cognition and executive functioning, and emergent literacy and language. These five domains form part of the direct assessment. The Technical Manual outlines the rigorous process followed in the development of the ELOM 4&5 Assessment Tool and describes further psychometrics undertaken since the release of the tool. The tool provides a reliable and fair assessment of children regardless of their socioeconomic

¹ ELOM assessments are considered valid when the child is within the 50-69 month age criteria, the child completes the full assessment (the assessment is not exited prematurely) and when the child does not fail the WHO embedded screening questionnaire (for more information, see the ELOM 2020 technical manual).

and ethnolinguistic backgrounds. Content, construct, age, and concurrent validity (with the WPSSI-IV), as well as test-retest reliability, have been established.^[13]

- 2) **The Socio-Emotional Rating Scale** is used in conjunction with the ELOM 4 & 5 Assessment. Socio-Emotional functioning has been found to be a strong predictor of school performance. This tool is designed to be administered by someone who is familiar with the child, such as a teacher. The tool has 13 items, across three key areas: Social relations with peers and adults (i.e. the ability to cooperate without prompting and work with peers in group activities), Emotional readiness for school (i.e. the ability to communicate with adults), and self-care (independent toilet use).
- 3) **The ELOM Home Learning Environment Tool (HLE)** is a short questionnaire designed to measure home aspects associated with early language, numeracy and cognitive functioning for children aged 2-7 years old. During a 15-minute interview, the primary caregiver is asked questions about themselves and their child(ren). The ELOM HLE assesses the following: The Early Learning Resources Score (the availability of books or objects used for play), The Early Learning Activities (activities conducted with the child at home, including reading, telling stories, singing songs, going out together, playing, naming things, counting, drawing and painting), and Caregiver Time with the Child Score (the amount of time a caregiver has spent with the child in the week and during the weekend).
- 4) **The Learning Programme Quality (LPQ) Tool** has been designed to measure the quality of group learning programmes targeting children aged 3 to 5 years. Assessors spend 2 hours observing the programme, and score their observations on a three-point scale (inadequate, basic, and good) across five areas: The learning environment (learning materials and classroom set-up), Learning and teaching (session planning and progress monitoring), Relationships and interactions (practitioner interactions with children, child interactions, and discipline.), Curriculum (curriculum content, alignment with the National Curriculum Framework Early Learning and Development Areas and activity plans) and Teaching strategies (teaching techniques and actions).

Data Sources

The most prominent source of data include those collected at the end of 2021 for the Thrive by Five Index and the Grade R ELOM Data (2019). Additionally, data from an evaluation trial is included. The source of the remainder of the data comes from programme and study data which has been collected data over time. This includes the collation of small datasets (N=30) to larger datasets (N=633) of programmes which have been anonymised.

We briefly discuss each of the larger datasets in turn:

- 1) The Thrive by Five Index (2021)

The Index includes child outcomes data on 5222 children sampled from 1248 early learning programmes (ELPs) / ECD sites² nationally. Data was collected between November-December 2021 across a nationally representative sample (95% confidence interval). ELPs were randomly sampled into clusters from primary schools. In each province, 48 schools (432 nationwide) were randomly selected to be used as clusters. As many ELPs as possible were then identified as possible within a 5-10km radius around each school, or, a ward in cases where there were too few ELPs. Additionally, the school sample in each province was stratified on socio-economic quintiles (1 to 5). Child assessments (ELOM 4&5, ELOM socio-emotional rating and height for age) were conducted with an average of 4 children at each of 3 randomly selected ELPs per cluster.

ECD Baseline Audit (2021)

Within each cluster of 3 ELPs, 1 ELP was selected to "audit". This involved interviews with the principal and a practitioner (the practitioner working with the 4-year-old children), an assessment of the overall environment and infrastructure plus a 2-hour observation of the quality of the early learning programme. This data is referred to as the ECD baseline audit. . All data in the audit were collected at the facility level and merged into the child-level Thrive by Five Index data.

Instruments used to collect data were ELOM 4&5 direct assessments of young children; ELOM Social-Emotional rating scales (with the child's teacher), ELOM Learning Programme Quality assessment (classroom observation tool), CAPI Questionnaire for Principals (non-ELOM tool), CAPI Questionnaire for Practitioners (non-ELOM tool), and a Facility Observation Questionnaire (non-ELOM tool). Stadiometers were used to measure height.

Description of non-ELOM tools

The principal interviews were conducted with the ELP principals and included questions on the running of the programmes. This included questions on subjects like the ELP's registration, how many teachers were at the ELP, how the facility was funded, what services the facility offered and what its operating hours were. There were also questions on the COVID-19 pandemic incorporated into the assessments, as they were run during this time.

The practitioner interviews were conducted with the teachers at the ELPs. This section of the assessment incorporated questions on practitioners' qualifications, teaching methods, demographics and attitudes towards children learning through play.

Environment observations documented the facility's infrastructure, outdoor space/play area, teaching materials and indoor toys. It included questions on subjects like the quality of the materials and toys, the safety of the spaces and whether there were any hazards in the facility.

2) Grade R ELOM Data (2019)

² ECD sites including preschools, Grade R, playgroups, creches, day mother programmes and toy library groups.

Grade R children (N=1678) across four provinces (Western Cape, KwaZulu Natal, Eastern Cape and North West Provinces of South Africa) were assessed using the ELOM 4&5 years assessment tool and Socio-Emotional Rating scales.

For this study, a sentinel approach to sampling was used. Two education districts were selected in each of the four participating provinces. These districts cover a wide range of school quintiles.

Schools were stratified by district and quintile group and ordered randomly. The first three or four schools in each randomly-ordered district-quintile stratum list were included in the sample. Since the sample sizes per stratum were unequal it was decided to have larger sample sizes in the lower-quintile groups (e.g. four schools in Q1, three schools in Q2-3 and Q4-5, respectively). Learners were sampled on the day of the actual school visit, stratified by gender, with a target of 25 children in each school.

Ideally, data should have been collected in January/February 2019, at the start of the Grade R year and before children are exposed to the Grade R curriculum. However, the final sign-off from the sponsor was only obtained at the end of February 2019. This delayed the start of the data collection process and children who participated in this study had already been exposed to one full term of Grade R. This is likely to have influenced child outcomes.

For this dataset, facility-level characteristics were not measured, however, the data was still included in the dataset due to the availability of geo data up until the municipal level. It was intended that analysts would merge secondary data into the metadata set.

3) Additional evaluation study³

The evaluation study sample consisted of both Afrikaans and IsiXhosa schools.

Selection of Afrikaans sample: As part of an existing Randomised Control Trial (targeted at Foundation phase learners in the Western Cape). The intervention has 50 Afrikaans schools in the treatment group, compared to 50 matched, comparison schools. Within each educational district, statistical techniques were used to select the comparison schools such that they matched the treatment schools as closely as possible on performance on Grade 3 systemic assessments⁴ from 2017 to 2019.

Grade R learners were assessed in Term 1 of 2022 across 50 Afrikaans schools (half of the schools in the evaluation study). These schools are all located in the four Metro and Cape Winelands educational districts.

Selection of isiXhosa sample: In addition, children from 25 isiXhosa schools were sampled in the Western Cape. Only schools in the Cape Town Metro districts were considered since most isiXhosa schools are located in Cape Town. Schools were stratified in quintiles of their average Grade 3 Systemic Evaluation performance between 2017 to 2019 and then within each stratum, five schools were randomly selected.

³ The name has been anonymised

4) Anonymised datasets from DataDrive2030's database

The meta dataset included anonymised data from smaller studies and programmes. The total number of observations included here is 6395. Assessments were conducted digitally and stored on the DataDrive2030 server. The ELOM assessments were done on tablets, enabling quicker, higher-quality data collection. The database was merged into the metadata after being anonymised. Data stored here include a variety of measures and tools from various programmes and studies.

Additionally, within this data, the geo-locations for children across their facilities were captured. Data from the 2021 ECD Census data were merged for children, enabling an analysis of facility-level factors such as environmental observations and practitioner interviews. The 2021 ECD census used the same tools as the 2021 Baseline audit, with the exception of the learning programme quality tool.