elom

early learning measurement tools

31 MARCH 2026

EARLY LEARNING OUTCOMES MEASURE TECHNICAL MANUAL

4&5 Years Assessment Tool

5TH EDITION 2026

FIFTH EDITION 2026
Developed on behalf of DataDrive2030 by: Andrew Dawes, Linda Biersteker, Matthew Snelling, Colin Tredoux & Elizabeth Girdwood with contributions from Kate Anderson, Tiffany Henning, Jasmin Moonsamy & Megan Scott.

TO CITE THIS MANUAL:
Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L. & Tredoux, C.G. (2026). Early Learning Outcomes Measure 4&5 Years Assessment Tool Technical Manual. DataDrive2030 Westlake, Cape Town. https://DataDrive2030.co.za

Please refer to www.DataDrive2030.co.za for additional information

ISBN 978-0-620-88295-8

Fifth Edition 2026

TABLE OF CONTENTS:

ACRONYMS:

APA	American Psychological Association
AERA	American Educational Research Association
ANA	Annual National Assessment
ASQ	Ages and Stages Questionnaire
C&EF	Cognition & Executive Functioning
CAPS	Curriculum Assessment Policy Statements
CEF	Cognitive and Executive Functioning
CDAT	Cambodian Developmental Assessment Test
CFA	Confirmatory Factor Analysis
CMS	Children's Memory Scale
DBE	Department of Basic Education
DIF	Differential Item Functioning
DPME	Department of Planning, Monitoring and Evaluation
DRP	Desired Results Developmental Profile
EAP ECDS	East-Asia Pacific Early Child Development Scales
EFA	Exploratory Factor Analysis
EF	Executive Functioning
EL&L	Emergent Literacy & Language
ELDS	Early Learning Development Standards
ELOM 4&5	Early Learning Outcomes Measure 4&5 Years Assessment
ELP	Early Learning Programme
EN&M	Emergent Numeracy & Mathematics
FMC&VMI	Fine Motor Coordination & Visual Motor Integration
GMD	Gross Motor Development
IDELA	International Development and Early Learning Assessment
IE	Innovation Edge
KMO	Kaiser-Meyer-Olkin
KZN	KwaZulu-Natal
NW	North West
LOLT	Language of Learning and Teaching
LMTF	Learning Metrics Task Force
MELQO	Measuring Early Learning Quality and Outcomes
NCME	National Council on Measurement in Education
NELDS	National Early Learning Development Standards
NCF	National Curriculum Framework
SACAS	South African Child Assessment Scales
SASL	South African Sign Language
SES	Socioeconomic Status
VMI	Visual Motor Integration
WC	Western Cape
WHO	World Health Organization
ZAMCAT	Zambian Child Assessment Test

IMPORTANT CHANGES IN THIS EDITION

NAME CHANGE:
The acronym used for the Early Learning Outcomes Measure in previous editions of this Manual has been changed to ELOM 4&5 to denote the age group for which it has been designed.

The metric equivalence of the ELOM 4&5 was established for the five language groups included in the 2016 standardisation sample using Rasch analysis and DIF (Snelling et al., 2019 Dawes et al., 2020): English, Afrikaans, isiXhosa, isiZulu and Setswana.

Since then, we have been able to assess children in the remaining languages . A major addition to this edition is results from DIF analyses for all 11 official languages except South African sign language. (N=15 487).

ACKNOWLEDGEMENTS

The ELOM Team is most grateful to the leadership of Innovation Edge and its Investment Committee for commissioning the development of the ELOM 4&5 in 2015, and supporting the first edition of this Technical Manual.

DataDrive2030 has supported further psychometric research on the ELOM 4&5 as well as the production of this 5th edition.

We gratefully acknowledge the Departments of Education in the Western Cape, North West and KwaZulu-Natal for permission to access schools to assess children for the standardisation of the ELOM 4&5 in 2016.

Education stakeholders were consulted on standards and domains to include in the ELOM 4&5 prior to its finalisation in 2016, which played a key role in guiding ELOM 4&5 development. They included: Professor Nirmala Rao (The University of Hong Kong); Dr Abbie Raikes (University of Nebraska Medical Centre); Dr Jane Kvalsvig (University of KwaZulu-Natal); Dr Celia Hsaio (Save the Children International); Dr Nicholas Spaull (University of Stellenbosch and UNESCO); Dr Penny Holding (University College London); Dr Louis Benjamin (Basic Concepts Unlimited); Dr Lauren Stretch (Early Inspiration).

OTHER CONTRIBUTORS

We are grateful to all the following for their input into developing the ELOM 4&5:

- Sarah Girdwood, for her expert input into the development of the sampling framework and for programming the tablet scoring system through SurveyCTO.
- Innovation Edge support staff, particularly Caro Kellerman, Rina Mehlomakulu and Heidi Everts.
- Dr Shelley O' Carroll, Wordworks, for advice on language and literacy items and suitable equivalents of Afrikaans and isiXhosa words.
- Dr Sylvia Thompson (University of Texas) and Professor Rosemary Cromarty (North West University) discussed initial sound discrimination in different languages.
- Cally Kuhne, Schools Development Unit, UCT, for advice on the mathematics items.
- Folio Translation Consultants for providing professional translation support.
- Statistical analyses for the investigation of predictive validity were conducted by Caitlin Buenk and Jürgen Becker of AX Group (https://ax-assess.com) in consultation with Andrew Dawes.

FOR PEER REVIEW OF THE TECHNICAL MANUAL

- Professor Santiago Cueto, Grupo de Análisis para el Desarrollo (GRADE) (www.grade.org.pe/), Peru; Young Lives, Department of International Development, Oxford University.
- Dr Celia Hsaio, Save The Children South Africa.
- Dr Jane Kvalsvig, School of Nursing and Public Health, University of KwaZulu-Natal.

1.1 ELOM 4&5 COMPONENTS

The ELOM 4&5 includes the following components

1. ELOM 4&5 Direct Assessment Kit
2. ELOM 4&5 Direct Assessment Kit list
3. ELOM 4&5 Direct Assessment Manual is available in all official languages of South Africa. A version for South African Sign Language (SASL) is in development
4. ELOM 4&5 Technical Manual
5. Open-source online protocol for tablet/phone-based scoring and data capture
6. Social-Emotional Functioning Rating Scale
7. Equipment required for ELOM 4&5 administration is provided in the ELOM 4&5 Direct Assessment Kit inventory.

Apart from the Test Manual and Kit, all other components and additional resources, including research papers and briefs, are available for download from the DataDrive2030 website.

Apart from Psychologists and Occupational Therapists registered with the Health Professions Council of South Africa (HPCSA), all other users must be trained and certified as specified by DataDrive2030 (see website for applications). Untrained persons may not use the ELOM 4&5.

This Technical Manual provides information on the development of the ELOM 4&5 and the construction of the Early Learning Development Standards to which programmes for young children should aspire by the end of the year prior to Grade R.

Tablets should be used for scoring the Direct Assessment, as this procedure significantly reduces errors. All data users capture is automatically uploaded to the secure database held by DataDrive2030. ELOM 4&5 Tablet software is available on application to DataDrive2030.

The ELOM Social-Emotional Functioning Rating Scale complements the ELOM 4&5 and is intended for use by programme staff who know the child well. It measures the child's self-care, relationships with peers and adults, and emotional functioning, in areas relevant for managing the school environment.

1.2 WHAT THE AGE VALIDATED, STANDARDISED ELOM 4&5 MEASURES

Prior to the development of the ELOM 4&5 in 2016, there was no validated South African instrument for measuring programme performance against Early Learning Development Standards.

The ELOM 4&5 is an age-normed, standardised instrument. In the original 2016 standardisation (see 2020 Technical Manual), norms were provided for two age groups: 50-59 months and 60-69 months. These are retained in the current edition for use primarily by Early Learning Programmes to monitor child progress. (See Appendix 1). They have also been used in research including, the Thrive by Five Index surveys.

The ELOM 4&5 includes direct assessment of children's performance and an assessment of the child's social and emotional functioning and orientation to tasks. Socio-emotional functioning in particular, is best assessed by a person who knows the child well as evaluation of these aspects of development by a stranger in a short period is not likely reliable. For this reason, the person most familiar with the child's behaviour during the ECD programme uses the Social-Emotional Functioning Rating Scale to measure this important area and to complement the Direct Assessment.

ELOM 4&5 consists of 23 items measuring indicators of the child's early development in five domains. These are recognised as key learning and developmental areas for programmes designed to support the development of young children, and are used internationally in exercises of this type (Pisani, Borisova, & Dowd, 2015); (LMTF, 2013)[1]:

1. **Gross Motor Development**[2]
2. **Fine Motor Coordination and Visual Motor Integration**
3. **Emergent Numeracy and Mathematics**
4. **Cognition and Executive Functioning**
5. **Emergent Literacy and Language**

Figure 1: What the Age Validated Standardised ELOM 4&5 Measures

1. DIRECT ASSESSMENT	2. SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE AND DIRECT ASSESSMENT	3. SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE
• Gross Motor Development • Fine Motor Coordination and Visual Motor Integration • Emergent Numeracy and Mathematics • Cognition and Executive Functioning • Emergent Literacy and Language	• Social and Emotional Development and Awareness • Approaches to Learning (persistence, attention and concentration)	• Self-care • Social Relations (adults and peers) • Emotional Readiness for School

The child's Task Orientation (aspects of Approach to Learning in Figure 1) is rated by the assessor during direct assessment. The Social-Emotional Functioning Rating Scale measures the child's capacity for self-care (toileting), relations with peers and adults, and emotional readiness for school (see Chapter 7).

ELOM 4&5 norms are provided in Appendix 1. Domains, standards, indicators, items and sources are included in Appendix 2. Appendix 3 provides ELOM 4&5 2016 Standardisation, psychometry, and statistics.

Direct Assessment Manuals for the 2016 Standardisation were translated into Setswana, isiZulu, Afrikaans and isiXhosa using accepted procedures to ensure linguistic and metric equivalence (Chapter 5). Since 2016 the ELOM 4&5 has been standardised in 11 official South African languages, excluding SASL, which is in development (Young, et al., In Press).

[1] http://steinhardt.nyu.edu/global-ties/early_childhood/melqo
[2] Gross and Fine Motor Development are components of Physical Development but are treated separately in the ELOM 4&5.

CHAPTER 2 THE DEVELOPMENT OF THE ELOM 4&5 DIRECT ASSESSMENT

ANDREW DAWES AND LINDA BIERSTEKER

The ELOM 4&5 was developed in four phases as shown in Figure 2.

Figure 2: The Development of the ELOM 4&5

PHASE 1 2015

Development of content validated ELDS, pilot item selection, and piloting

PHASE 2 2016

Age validation study in three provinces across five school quintiles and in five languages

PHASE 3 2016

Psychometric analysis, standards, norms and finalisation of the ELOM 4&5

PHASE 4 2017-2025

Additional psychometric analyses: assessment of bias in all South African languages; assessment of test-retest reliability and concurrent validity, and development of a caregiver version of the SEF assessment.

Assessments of both concurrent and predictive validity (subtypes of criterion validity).

PHASE 1 2015

2.1 DEVELOPMENT OF CONTENT VALIDATED ELDS FOR THE ELOM 4&5

Content validation refers to the extent to which the standards and indicators are age-appropriate. In this case, selected standards and indicators must be appropriate to what children should know and be able to do on entry to Grade R in the ELOM 4&5 age range. They should also cover areas that are known to be good predictors of early school performance.

This process included three steps:

1. Selection of Early Learning and Development Standards (ELDS) and indicators for children eligible for entry to Grade R (turning 60 months prior to July in the year of admission);

2. Sourcing valid, reliable assessment tools for measuring ELDS indicators that had preferably been tested for suitability in South and Southern Africa. Apart from the need for adjustments to suit local circumstances (e.g. in language items), the process did not entail the development of new items. Rather, the ELOM 4&5 drew on existing instruments that were suitable for measuring the chosen indicators and which are applicable in a diverse cultural and developmental environment.

3. Obtaining expert comment from Grade R educators, officials and other experts (both locally and internationally) on selection of standards indicators and measures prior to drawing up the instrument for piloting. Two focus groups were conducted with experienced Grade R educators who worked with children from different language and socioeconomic groups, and interviews were held with education officials. These consultations generated stakeholder opinions on the most important capabilities children must display on entering Grade R to benefit fully. They contributed to the selection of ELOM 4&5 domains.

2.2 SELECTION OF ELDS DOMAINS

Literature and policy documents on ELDS were surveyed to determine commonly used domains and standards. Our starting point was the South African National Early Learning Development Standards (NELDS) for children from birth to four years (Department of Basic Education, 2009). This had been developed through consultation with education experts from academic institutions, a content validation with NGOs, parents/caregivers, ECD practitioners from different provinces and finally, age validated (this process was led by one of our Reference Group members, Dr Jane Kvalsvig). The South African National Curriculum Framework (NCF) from birth to four years (based on the NELDS) and other public policies and guidelines, were also primary sources. As the ELOM 4&5 is designed to assess children at the end of the year prior to Grade R, the design also considered progression to the South African Curriculum Assessment Policy Statements (CAPS) for Grade R.

Domains selected for the ELOM 4&5 do not cover all those specified in the South African ELDS. Only those deemed key capabilities for children to enter Grade R were selected (and checked with key informants in the early education sector). The primary data sources for ELDS and domain selection were therefore as follows:

- ELDS contained in the NELDS 0 - 4 and associated technical report on South African NELDS development, and the NCF, in particular the section on "towards Grade R";
- National ELDS used in other countries of the SADC region (including Swaziland, Malawi, Zambia and Lesotho);
- The United Kingdom Early Years Framework;
- The California State ELDS (as an example of models in the United States);
- South African ECD standards developed for the UNICEF Going Global with Indicators for Child Well-Being project (Dawes et al 2004 a & b).

Alignment with policy in standards development is regarded as essential (e.g. Zieky, and Perie, No Date; Kagan, Castillo, Gomez, & Gowani, 2013). The criteria for inclusion of an ELOM 4&5 standard and indicator in a domain were:

- alignment with the South African policy approach;
- included with high frequency in South African, regional and other international ELDS pertinent to the ages and grade in question;
- a good fit with one of the selected domains;
- While not in the South African ELDS, it fills a gap in what is locally available and is necessary to assess for entry to Grade R without developmental disadvantage.

Following international practice, standards have been specified for ELOM 4&5 Total scores and for each domain (See Appendix 1).

2.3 SELECTION OF ITEMS FOR MEASURING CHILDREN'S PERFORMANCE FOR ELOM 4&5 DIRECT ASSESSMENT

Once domains had been decided, a scan of available instruments and test items used in similar exercises was undertaken. The criteria for inclusion of ELOM 4&5 items were as follows:

- The item must be a valid and reliable measure of the indicator;
- While recognising that bias is inevitably introduced by limited exposure to certain test items, those that tap basic skills likely to enable learning in Grade R should be included. Examples are: handling of writing materials; visual motor integration; fine motor coordination; number concept; problem-solving exercises; task orientation;
- To reduce the risk of children being unresponsive in the test situation for any reason, and also to reduce the likelihood of cultural norms for communication with adults restricting the child's response, items requiring verbal responses from children should be kept to a minimum (while including a language domain);
- The ELOM 4&5 kit should not present significant challenges in terms of complicated equipment and high costs;
- Items should, as far as possible, be culturally fair and unbiased, or have the potential for adjustment for such a purpose to be determined prior to final field study;
- ELOM 4&5 should be suitable for administration by trained persons with knowledge of young children and early childhood care and education but who are not necessarily registered professionals such as psychologists, psychometrists, or occupational therapists;
- When all indicator measures are combined, the assessment should not take longer than approximately 45 minutes.

A particular focus of the search for suitable items was on instruments being developed internationally, and those with established validity and reliability that had been developed for use in the region, or in similar socio-cultural and socio-economic contexts. The ELOM 4&5 includes items developed for similar tools including: Save the Children's International Development and Early Learning Assessment (IDELA) tool (Pisani, Borisova and Dowd, 2015), the 2015 Direct Assessment Manual devised for the Measuring Early Learning Quality and Outcomes (MELQO) Initiative pilot, and the Zambian Child Assessment Test (Fink et al, 2012), among others. Executive Functioning (EF) (including non-verbal reasoning; problem-solving, cognitive flexibility, working memory, and behavioural inhibition) is increasingly regarded as important for assessing the functioning of children prior to school as it underlies performance in a range of domains, and is known to be associated with adjustment to the school environment, and effective learning performance in school (Best, Miller, and Naglieri, 2011; Blair, 2002; Diamond and Lee, 2011). The Cognition and EF domain of the ELOM 4&5 draws on commonly used assessments of EF. Task orientation and approach to learning were assessed during the administration of the ELOM 4&5 (using a blend of IDELA and ZAMCAT checklists). Final item selection followed a process of several rounds of consultation with experts.

2.4 ESTABLISHMENT OF CULTURAL FAIRNESS AND TRANSLATION OF DIRECT ASSESSMENT ITEMS

It is necessary for instruments used to assess children from different language and socio-economic backgrounds to take account of such variations so that children from such groups are not disadvantaged, and so that false negative scores are not obtained (the child can potentially complete the item, but the way it is presented prevents this). Assessments must be administered in the child's home language and language items must be carefully constructed. Translations from the original English must follow accepted practice. English language instructions to the assessor on ELOM 4&5 administration were not translated. Translation of pilot and (revised) age validation Direct Assessment items (instructions to the child being assessed) was undertaken by a specialist translation company. However, as professional translators commonly translate into more formal versions of the language that may not accord with what is ordinarily understood by children in the target community, we referred their translations to our experienced team of assessors, who made adjustments based on commonly used and understood terms. We also had advice from a local project specialising in early literacy, and from early language development specialists.

We followed Pena's (2007) guidance as far as possible. It is necessary to address several issues when adapting a test for use across different ethno-linguistic groups.

1. **Cultural fairness:** In developing the ELOM 4&5 items, we ensured that children from different class and cultural backgrounds were likely to be equally familiar with the tasks demanded in the items. Decisions on items were based on the views of key informants and the performance of children on pilot items.

2. **Functional equivalence:** In the ELOM 4&5 pilot, we sought to establish whether the test instructions elicited the same behaviour in children from different communities.

3. **Linguistic equivalence:** Translation and independent back translation are required when different language groups participate in the assessment. In the case of language items, we asked the question: Are the words and phrases used equivalent in meaning to those in the language of translation? The instructions for tasks must be as close as possible to local usage. Professional translations may not capture the vernacular. The objective is to establish the most commonly used word or phrase in the target community.

4. **Cultural equivalence:** Here, the challenge is to ensure that items are not likely to prejudice the performance of particular groups of children due to their lack of familiarity with such tasks. Where all groups are similarly unlikely to be familiar, this is not regarded as an issue, and there are examples of these in the ELOM 4&5. For the pilot, the team considered this carefully, in discussion with other experts. Experience during the pilot phase allowed for observations of consistently problematic items, which were then adjusted. DIF analyses as described in Chapter 5 were undertaken to address this issue.

5. **Metric equivalence:** Refers to the difficulty of the item when translated into different languages (Milfont, and Fischer, 2015). This is particularly important in the case of language assessments. Pena (2007) notes, for example, a word may be used with high frequency in the original language but less so in the language of translation. Where this is the case, children speaking the latter language would be less exposed to the word and therefore disadvantaged. Hence, simply translating the English word into another language would not likely result in an item of equivalent familiarity or difficulty.

An example is provided by the Initial Sound Discrimination Item (See the ELOM 4&5 Direct Assessment Manual), which was derived from the English language Learning Metrics Task Force MELQO. For our translations into African languages, we draw on words, using the same initial sounds (e.g. 'd' for 'duck' was presented in isiXhosa as 'd' for 'dada' and, where possible, of equivalent length). Based on experience in the Age Validation phase of the study, on psychometry, and expert advice, adjustments were made to some words in African languages.

Differential Item Functioning (DIF) analyses were undertaken on the original standardisation sample languages to check for items that discriminated unfairly between children from different backgrounds but of the same ability (see 2020 Manual). These have since been updated on a large sample for all official South African languages except SASL. Procedures are described in Chapter 5.

2.5 THE ELOM 4&5 PILOT

The pilot was designed to test the performance of Direct Assessment ELOM 4&5 items and administration procedures. Experienced preschool teachers were trained to administer the Pilot Direct Assessment ELOM 4&5 in Afrikaans, English and isiXhosa to 70 children. Forty-two records were of sufficient quality for analysis.

In some domains, and to test their performance, more items were included in the Pilot than required. The approach to item retention or exclusion was to retain those that generated a range of performance (and were neither too easy nor too difficult), were easier to administer, were better understood by children, did not take excessive time, and other things considered, were deemed important to retain.

Assessor comments and investigator observations of testing were also considered. Some items did not work as well as expected – not because they were necessarily a poor measure of the construct, but rather because the test kit needed adjusting. In other instances, the test instructions did not work well and required adjustment. Finally, some items took considerable time and/or assessors found them challenging to administer consistently and validly. Bearing in mind that the ELOM 4&5 is intended for use by trained senior ECD teachers (the same level as those used in the pilot), these observations were important.

Where indicated, changes were made to administration procedures and to instructions to improve clarity. Some instructions translated from English required adjustments in Afrikaans and isiXhosa.

Pilot experience indicated that only three children could realistically be tested by one assessor each morning. The ELOM 4&5 was then prepared for age validation.

CHAPTER 3

ELOM 4&5 STANDARDISATION SAMPLE

ELIZABETH GIRDWOOD, SARAH GIRDWOOD, MATTHEW SNELLING & COLIN TREDOUX

The **goal** of the ELOM 4&5 age validation process was to construct a sample that was likely to be as representative as possible of children eligible to enter Grade R in January 2016, drawn from across South Africa's socio-economic distribution, and including five major language groups.

A two-stage clustered sample design was employed. In the first stage, and in each district, probability proportional to Grade R population size sampling was used to randomly select schools within each of the five school quintile bands. Two schools in traditional, more rural areas in North West and KwaZulu-Natal were recruited independently of this exercise to explore the influence of more "traditional" approaches to child rearing. In the second stage, learners were selected within Grade R classes using simple random sampling.

3.1 RATIONALE FOR SAMPLE INCLUSION OF FIVE QUINTILES AND FIVE LANGUAGES

Both in South Africa and internationally, children from less deprived backgrounds outperform their more deprived counterparts. Socio-economic status (SES), and particularly the educational background and literacy levels of caregivers, are well established predictors of educational outcomes both in the developed world and the global south (Bradley & Corwyn, 2002; Engle, Fernald & Alderman, 2011; McLloyd, 1998). While South Africa lacks literature on the factors that predict the outcomes of children in Grade R and the subsequent Foundation Phase grades, they are as likely to experience similar impacts of poverty on their development as those living elsewhere. Research indicates that children's performance on the Annual National Assessments (ANAs) is strongly related to the quintile ranking of their school. ANA performance for children in the lower quintiles is significantly behind that of quintiles 4 and 5 {Department of Planning Management and Evaluation (DPME), 2014}. The studies do not permit the establishment of the causes, beyond alluding to a mix of home background and school quality variables. Finally, there is clear evidence that quality preschool and early schooling significantly affect educational outcomes for poor children, including in low-income countries (e.g. Nores and Barnett 2010; Hoadley, 2013).

Based on these considerations, the Age Validation design had to consider two highly probable influences on ELOM 4&5 performance: socio-economic status and cultural background, particularly when children are reared in more traditional ways with limited exposure to early learning opportunities relevant to schooling. These children do not lack stimulation, as is sometimes claimed. Instead, the stimulation they receive differs from that which enables readiness for the schooling system. We also had to examine possible differences in the performance of boys and girls.

Several recruitment options were considered, including:

A. a representative community sample of children prior to Grade R,
B. children in early learning programmes, and
C. children enrolled in public schools at the commencement of their Grade R year.

Given the challenges of constructing a representative community sample (Option A), the selection effects associated with recruiting children already in early learning programmes (Option B), and cost and logistical issues, Option C was selected.

To explore the influence of more "traditional" approaches to child rearing that commonly do not encourage verbal and intellectual engagement with adults (e.g. Dawes et al, 2004 a & b), the study design included school quintile 1 children from isiZulu and Setswana speaking rural, traditional backgrounds. Children from five of the country's primary languages were included: Afrikaans (spoken by 13.5% of total population), English (9.6% of total population), isiXhosa (16% of total population), isiZulu (22.7% of total population), and Setswana (8% of total population). Together, these languages represent 70% of South Africa's first language speakers.[3]

As we were not able to interview parents to obtain information on household income, the five school quintile classifications (which are based on the income levels of populations served by schools) were employed as a proxy. Quintiles 1, 2 and 3 are designated 'no fee' schools serving the poorest children. Recent evidence (DPME, 2014) indicates that only children in quintile 4 and 5 schools (highest SES) benefit from their Grade R year. This suggests two interacting determinants: a) that they are attending better functioning Grade R classes than poorer children, and b) that they are from homes that are better able to support early learning prior to and during school years (these two variables are confounded as wealthier, better educated parents choose and can afford better schools).

[3]All data from StatsSA Census 2011

3.2 POWER CALCULATIONS

It is envisioned that the ELOM 4&5 will be applied to children from a range of cultural and socio-economic settings. However, as finances did not permit a national sample, three provinces were chosen for the study. The sample for this study then aimed to be representative of public school Grade R students in the target language groups who were between the ages of 54-66 months in selected school districts in North West (Setswana speakers only), the Western Cape (English, isiXhosa and Afrikaans speakers) and KwaZulu Natal (isiZulu speakers only).

Power analysis was used to determine the required sample size for reliable statistical analysis. The study design sample stratification had to be taken into account by quintile and language. GPower[4] was used for this purpose, with power set to 0.8 for all analyses (conventional level). Alpha was set to 0.05 for all analyses except t-test, where this reduced to 0.01 as a limited manual control for Type 1 error. A scaling factor of 1.4 was used to estimate N for Tukey HSD test from N in corresponding ANOVA design[5]. This basically specified the number of children required in each language and quintile group to make meaningful conclusions. Table 1 details the target sample sizes based on power calculations.

Table 1: Power Calculations

Language	**English**	**Afrikaans**	**isiXhosa**	**Western Cape Total**	**Setswana**	**isiZulu**	**Total**
Province	**Western Cape**	**Western Cape**	**Western Cape**		**North West**	**KwaZulu Natal**	
Quintiles 4 and 5 Urban	60	60	60	180	60	60	300
Quintile 3 Urban	60	60	60	180	60	60	300
Quintile 1 and 2 Urban	N/A	75	75	150	75	75	300
Rural / traditional (likely Quintile 1)	N/A	N/A	N/A	N/A	120	120	240
TOTAL	**120**	**195**	**195**	**510**	**315**	**315**	**1140**

Using the power calculations, in each quintile, the number of schools and their language of teaching and learning were determined. For example, we determined the number of schools in English quintile 3 to reach a target of 60 quintile 3 English-speaking children. Once we had selected the number of schools needed, they were then randomly sampled within each quintile, and children were sampled within each school as described below.

3.3 CONSTRUCTION OF SAMPLING FRAMES

The defined target population was used to guide the construction of sampling frames from which the samples of schools were selected. The sampling frames were based on national lists of schools that included information about school identification numbers, enrolment for the target population of Grade R pupils, school quintile information, language and school regional location. Table 2 presents the sampling approach.

[4] http://download.cnet.com/G-Power/3000-2054_4-10647044.html
[5] http://digitalcommons.wayne.edu/cgi/viewcontent.cgi?article=1291&context=jmasm

Table 2: Sampling Approach

Desired target population	Grade R children in South African public schools aged 54-66 months.
Defined target population	All children at the Grade R level attending registered public schools in three districts of South Africa.
Excluded population	All children at the Grade R level attending schools outside the three defined districts, or at independent, community ECD centres, or special schools.
Stratification variables	Education district, school quintile, language.
Sampling stages	First stage: Schools selected within strata with simple random sampling. Second stage: Children selected within schools as per 3.2.2.
Minimum cluster size	Minimum of nine children per school.

The sampling frames were then defined according to 25 Strata, in Table 3 below. Each region was split into five according to the five quintiles and the language.

Table 3: Sampling Strata

Language	Stratum			Quintile
English	1	Stratum 1	Metro East	1
	2	Stratum 2	Metro East	2
	3	Stratum 3	Metro East	3
	4	Stratum 4	Metro East	4
	5	Stratum 5	Metro East	5
Afrikaans	6	Stratum 6	Metro East	1
	7	Stratum 7	Metro East	2
	8	Stratum 8	Metro East	3
	9	Stratum 9	Metro East	4
	10	Stratum 10	Metro East	5
isiXhosa	11	Stratum 11	Metro East	1
	12	Stratum 12	Metro East	2
	13	Stratum 13	Metro East	3
	14	Stratum 14	Metro East	4
	15	Stratum 15	Metro East	5
Setswana	16	Stratum 16	Matlosana and Tlokwe	1
	17	Stratum 17	Matlosana and Tlokwe	2
	18	Stratum 18	Matlosana and Tlokwe	3
	19	Stratum 19	Matlosana and Tlokwe	4
	20	Stratum 20	Matlosana and Tlokwe	5
isiZulu	21	Stratum 21	Umlazi	1
	22	Stratum 22	Umlazi	2
	23	Stratum 23	Umlazi	3
	24	Stratum 24	Umlazi	4
	25	Stratum 25	Umlazi	5

3.4 ASSUMPTIONS AND INPUTS FOR A TWO-STAGE CLUSTERED SAMPLE DESIGN

Sampling accuracy requirements set down by the International Association for the Evaluation of Educational Achievement (Ross, 1991) state that the standard error of sampling for pupil tests should be of a magnitude that is equal to, or smaller than, what would be achieved by employing a simple random sample of 400 pupils. Using the sampling design tables described by Ross (1987), it is possible to determine the number of schools required for this study to achieve an adequate sample of 400 pupils. According to these tables, and using the coefficient of intraclass correlation[6] of 0.4, a minimum cluster size of 9, it was required that we sample around 184 schools. However, due to several logistical and other constraints, the maximum number of schools we could feasibly assess was 175.

Using assumptions as to the number of children that could be assessed in a morning, and the number of schools that could be visited (as per the inputs above), Table 4 below provides the final outputs. A total of 173 schools were selected across the three regions. Twenty-one assessors were dedicated to assessing children from these 173 schools. The total number of children that they could have assessed was 1275 at 141 schools. This effectively allowed us a school non-response rate of 24%. Including the children and schools to be assessed in traditional rural areas brought the total schools selected to 173 and the total number of children possible to 1575. Inputs, assumptions and all constraints are listed in Table 4 below.

Table 4: Inputs, Assumptions and Constraints

25	Number of school days available to assess Grade R children – i.e. five weeks.
3	Number of provinces for which collecting data (Western Cape (Metro East), North West (two areas in Dr Kenneth Kuanda: Matolosana and Tlokwe), and KwaZulu-Natal (Umlazi).
21	Number of assessors and tablets on which the assessment is scored.
3	Number of children one assessor can feasibly assess per day (test takes around 45 min, and schools are only open in the morning).
9	Minimum of nine children will be assessed per school.
1575	This is the maximum number of children we can assess, assuming each assessor assesses three children a day for 25 days.
2	The maximum number of assessors will be sent to each school (this is related to space requirements – e.g. two quiet spots per school).
175	The maximum number of schools we can assess assuming nine children per school.
4	Assessors assigned to the North West and KwaZulu-Natal to be dedicated to assessing children in traditional rural areas (two in each).
300	Number of children to be sampled from traditional rural areas in the North West and KwaZulu-Natal (150 in each).
1200	Minimum number of children to be assessed (as per power calculations).

[6]The coefficient of intraclass correlation (rho) referred to a measure of the tendency of pupil characteristics to be more homogeneous within schools than would be the case if pupils were assigned to schools at random. In South Africa, this ranges between 0.4 and 0.6.

Table 5: Child Assessment Logistics

Child Assessment Logistics											
			Normal schools			Traditional rural schools			Total		
	Schools selected	Schools per week per assessor	Assessors	Total Children	Total schools	Assessors	Total schools	Total children	Total schools	Total children	%
WC	64	1.67	9	675	75				75	675	43%
NW	43	1.67	4	300	33	2	16	150	49	450	29%
KZN	37	1.67	4	300	33	2	16	150	49	450	29%
Total	180		17	1275	141	4	32	300	173	1575	100%

(WC = Western Cape; NW = North West; KZN = KwaZulu-Natal)

3.5 ETHICS

The study protocol was approved by the University of Cape Town (UCT) Faculty of Humanities Research Ethics Committee on 23 November 2015 with Reference Number PSY2015-048. The Provincial Education Departments gave their permission, and each school was approached directly for permission to assess children. Informed consent was also sought from parents and guardians.

Parents were informed on their consent form that if they did not return it, it would be understood that they would not object to the assessment of their child. In such cases, the child was assessed. This is known as passive consent – a procedure approved by the Ethics Committee, as there was no risk to the child. Children were only assessed if they were willing to participate following a verbal explanation of the procedure (assent to participate). Approval of all studies included in data used for construction of 6-month standards, assessment of bias, test-retest reliability and concurrent validity was obtained from the same Committee at later points.

3.6 SELECTION OF PUPILS WITHIN SCHOOLS

Selected schools were contacted, and permission to assess children was sought. Those who refused were replaced.

Assessors were trained to select a minimum of nine pupils, between the ages of 4.5 and 5.5 years, at each school. First, they obtained the list of children attending Grade R in 2016, their birthdates, and gender. The list was narrowed down by only selecting children born between the target dates, and from this, random samples were selected, firstly of girls and then of boys. The Age Validation sample is shown in Table 6.

All sampled children were screened for disabilities likely to affect performance on the ELOM 4&5 using four modified questions from the World Health Organization Ten Point Disability Screen (Durkin et al, 1995).

1. **WHO Screen:** *"Compared with other children, does the child have difficulty seeing, either in the daytime or at night?"* **ELOM 4&5 modification**: *"Did this child seem to have difficulty seeing?"*

2. **WHO Screen:** *"Does the child appear to have difficulty with hearing?"* **ELOM 4&5 modification:** *"Did this child appear to have difficulty with hearing?"*

3. **WHO Screen:** *"When you tell the child to do something, does he/she seem to understand what you are saying?"* **ELOM 4&5 modification:** *"When you told this child to do something, did he/she seem to have difficulty understanding what you were saying?"*

4. **WHO Screen:** *"Does the child have difficulty in walking or moving his/her arms, or does he/she have weakness and/or stiffness in the arms or legs?"* **ELOM 4&5 modification:** *"Did this child have difficulty in walking or moving his/her arms, or did he/she have weakness and/or stiffness in the arms or legs?"*

Children who were assessed as positive on any one of these indicators were excluded from the psychometric analyses as noted in Chapter 5.

Table 6: ELOM 4&5 Direct Assessment Age Validation Targets and Percentages of Target Reached in Final Sample

TOTALS														
Province	**Western Cape**								**North West**		**Kwa Zulu-Natal**		**All**	
Language	English	% Target	Afrikaans	% Target	Xhosa	% Target	Total	% Target	Setswana	% Target	Zulu	% Target	Total	% Target
Quintiles 4 and 5 Urban	115	192%	40	67%	50	83%	205	114%	83	138%	201	335%	489	163%
Quintile 3 Urban	35	58%	74	123%	106	177%	215	119%	60	180%	69	115%	426	142%
Quintile 1 and 2 Urban	N/A	N/A	29	39%	99	132%	128	85%	75	150%	55	73%	264	88%
Traditional/ rural (likely Quintile 1)	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	159	133%	297	124%
Total	**150**	**125%**	**143**	**73%**	**255**	**131%**	**548**	**107%**	**444**	**141%**	**484**	**154%**	**1476**	**129%**

As is evident, overall, the total number of children assessed exceeded that required in the power calculations. Only urban quintiles 1 and 2 failed to reach the target. This was because there were very few quintile 1 schools in the selected districts. Oversampling of children in rural quintile 1 schools addressed this issue.

CHAPTER 4

ASSESSOR TRAINING FOR THE AGE VALIDATION STUDY | ANDREW DAWES, LINDA BIERSTEKER & ELIZABETH GIRDWOOD

4.1 PROCESS

Three teams of assessors – one in each province - were needed for the age validation to cover the five languages and to complete the assessment within the first five weeks of the new school year. This was important so that children's exposure to Grade R learning would be minimal. Training in preparation for fieldwork took place in two steps.

1. Firstly, the field managers and senior assessors from the North West and KwaZulu-Natal provinces attended a five day training of trainers and fieldwork planning session with the Western Cape senior assessor and ELOM 4&5 team in November 2015. Training included familiarisation with the kit, item administration, tablet familiarisation and importantly, opportunities for assessing children and receiving feedback. During this time, slight adjustments were made to item administration.

2. Training of 22 assessors, in three provincial teams, took place in the first week of the January school term, prior to commencing the Age Validation study during week two. Assessors received a five-day standardised training programme and support materials, including a DVD of a child assessment. The training focused on introducing the measure, observation of expert administration (both the DVD and the senior assessors' administration), and then several opportunities to assess children while being observed by an expert.

4.2 INTER–ASSESSOR AGREEMENT

Opportunities to work towards a common understanding of the scoring requirements were built during the training. Trainee assessors worked in small groups observing each other as they practised assessing children and then comparing scores and discussing variations after each administration. Items that were variable were also picked up for discussion in general sessions, and scoring criteria were clarified. It had been agreed to exclude trainees who did not achieve a reasonable level of inter-rater agreement from the Age Validation, but this was unnecessary.

Rather than assessing large numbers of children to establish reliability, extensive training was provided to achieve reliability in assessment, as described above. A final rating of inter-assessor scoring reliability took place during the fieldwork period. This was approached slightly differently in the different provinces for practical reasons. In the Western Cape, eight assessors (including the senior assessor) watched a video of one child being assessed in English and then gave ratings. In the other provinces, it was decided to base the rating on children who spoke the local language. For the North West, seven assessors rated a single child. The senior assessor did not complete the assessment as she did not speak Setswana, so she had to be excluded. However, the six assessors were included. For KwaZulu-Natal, three isiZulu-speaking children were assessed separately by the senior assessor and one or two other assessors (out of eight) for each child. This meant that Kendall's Coefficient of Concordance could not be used, but a compound score was made out of the senior assessor and the assessors, and these indicated very high agreement between the senior assessor and the assessors.

Inter-assessor reliability was calculated using Fleiss' Kappa and Kendall's Coefficient of Concordance, as shown in Table 7 below, and revealed substantial agreement between assessors.

Table 7: Inter-assessor Reliability Scores

	Fleiss' Kappa	Kendall's W
Western Cape	K = .795	Kendall's W (.966) p < .001.
North West	K = .684	Kendall's W (.861) p < .001
KwaZulu Natal	K = .916	NA

Given these procedures, the ELOM 4&5 has been validated on a sample that is very likely to represent South African children's socio-economic backgrounds. While not representative of language, it includes languages spoken by approximately 70% of the population. Therefore, the standards and norms developed in this study are also valid for children from most linguistic backgrounds.

CHAPTER 5

PSYCHOMETRY AND STATISTICAL ANALYSES

MATTHEW SNELLING, COLIN TREDOUX, ANDREW DAWES, KATE ANDERSON, TIFFANY HENNING, JASMIN MOONSAMY, MEGAN SCOTT, CAITLIN BUENK & JÜRGEN BECKER

5.1 TABLET SCORING AND DATA COLLECTION

All data in the age validation study were collected and submitted electronically, using the SurveyCTO service (SurveyCTO, 2016). This allows forms to be digitised, stored on a central server, and then accessed by a selection of secure Android devices across a data connection. All data stored on the SurveyCTO server were encrypted, password protected, and accessed from a secure, password-protected computer. Only the data manager and project manager had access to the server and to the data.

In addition to the collection, storage and monitoring of data, SurveyCTO also allows scoring to be pre-programmed into the survey and calculated automatically. A preliminary scoring system was designed for this purpose. This scoring system translated responses into a simple numeric scale for each item. Where literature already existed to recommend scores, these were used. All scores were later transformed, using Rasch modelling, to produce an interval scoring system.

Child data were initially submitted by assessors, immediately after the assessment was concluded, or at the end of each day, by their supervisor. Data were stored on the Android device if no data connection was available. These forms were then submitted by the data manager after the data collection period. In all, 1490 records were captured.

5.2 PREPARATION OF DATA FOR ANALYSIS

All records were downloaded from the SurveyCTO server as a CSV (comma-separated value) file. The database was scanned for duplicates and incomplete records using Microsoft Excel. This was necessary due to the large number of assessments, assessors and schools. Coordination of assessments on this scale is likely to have a degree of human error, and this needed to be addressed. Records with the same first name, surname and age were deemed duplicates. All duplicates were purged, leaving 1476 records.

Due to the design of the tablet system, all records were either complete or incomplete. The latter occurred if the child failed to complete the assessment, skipped an item, or did not wish to cooperate with the assessment.

Incomplete records were accompanied by a comment from the assessor, stating why the child could not continue. The database was then scanned for incomplete records. These 18 records were purged from the database. This left 1458 records. Finally, 85 records of children with disabilities - screened at the end of each assessment - were removed (they were screened as described in section 3.5 above). These children were identified as having one or more of the four WHO disability indicators. In addition, on assessment, 98 children were identified as growth stunted (Height for Age HAZ > 2SD below the standard for their age). Analyses were conducted to determine whether viable models could be generated with their inclusion. As this was the case, these children were retained in the final sample. This left 1373 records, imported into an IBM SPSS v22 data file. Overall, the number of records analysed exceeded the number required in the power calculation, as noted above.

Data cleaning: outliers and influential values
The check for outliers took a conservative approach, seeking to discard as little data as possible. Outliers were checked descriptively using Box-and-Whisker plots and Z-Scores of the total ELOM 4&5 score, and domains. Outliers were then checked inferentially to determine whether there were more outliers, or extreme values, than was expected, 95% of the time in a random sample from a normally distributed population. Based on this criterion, no records were excluded from the dataset.

The check for influential values attempted to detect groups of scores that had an undue influence on the quintile mean scores. One possibility is that school quintiles may not be a valid indicator of the child's background.

Quintile validation
The quintile ranking of a school is based on the relative poverty of the immediate community around the school as assessed in official statistics and as prescribed in the National Norms and Standards for School Funding (SA Department of Education, 1998). However, the quintile of the school is not necessarily a valid indicator of the economic background of all of the children who attend. For example, parents from poor backgrounds with high aspirations for their children's education may choose a better school in a more advantaged area. Conversely, where only a limited number of schools are available in an area, some schools may serve children from more advantaged backgrounds than the school's quintile ranking would suggest.

To explore this issue, the Total ELOM 4&5 score means and 95% confidence intervals were computed for each sampled school. Means were then compared to identify schools within all quintiles where confidence intervals did not overlap.

We found that quintile 1 had two schools that performed better than the others in quintile 1. The effect of these schools was confirmed by removing them and observing the change in the ELOM 4&5 Total mean quintile score. The ELOM 4&5 Total mean for quintile 1 was reduced from 58 (95% CI = 55.49; 60.54) to 52 (95% CI = 50.62; 55.35). This check confirmed the influence of these schools and indicated that the children were not appropriately placed in quintile 1.

For this reason, these two schools were removed and the remaining 1331 records were included in analyses.

Quintile 4 and 5 schools serve children drawn from higher wealth areas. Quintile 3 schools and below, are fully subsidised and are known as 'No fee' schools. In 2013, 54% of South African children lived below the lower poverty line (R671 per month). In black African children, the rate is 61.7%. Although accurate data is not available, it is commonly known that many children from lower quintile areas seek better education in higher quintile schools.

However, the impoverished home context of these children is still likely to disadvantage them relative to better-off children in quintile 4 and 5 schools. The presence of these disadvantaged children could potentially lower the mean Total ELOM 4&5 and Domain scores of the combined quintile 4 and 5 group. Essentially, we needed to assess whether the child's quintile classification was valid for the age validation study. Schools were contacted to determine where children lived and whether they qualified for a fee subsidy based on their socio-economic status.

Three schools formerly thought to be in the quintile 4 and 5 group were identified as no-fee schools requiring their re- classification. Further, one child from a fourth school was identified as receiving a fee-subsidy. These children were provisionally shifted from the quintile 4 and 5 group to the quintile 2 and 3 group, to determine whether this affected the mean score of quintile 4 and 5. This approach was taken to determine the influence of the schools, rather than whether or not they were simply different. Removal of these children made no difference to either quintile group. In light of this, all children were grouped according to existing national quintile categories or no-fee status.

Finally, ELOM 4&5 total means of all schools were compared to determine whether appropriate differences existed between quintiles 1, 2, 3, 4 and 5. Two primary observations were made. Schools in quintile 4 and 5 showed significant overlap of confidence intervals. Quintile 2 and 3 schools also showed significant overlap of confidence intervals.

This evidence suggests that these quintile categories are not meaningfully different in ELOM 4&5 performance. Therefore, quintiles 4 and 5 were combined, and quintiles 2 and 3 were combined. These new groups represent the observable differences in mean ELOM 4&5 performance, based on the socioeconomic differences that each quintile is meant to represent.

The age range of the final sample was wider (50-69 months) than the original target (54-66 months). The final quintile breakdown of the sample is provided in Table 8.

Table 8: Distribution of Children by Quintile

Quintiles	N	Percentage
1 (including "Traditional" background)	114	8.56
2 and 3	756	56.80
4 and 5	461	36.64
Totals	**1331**	**100.00**

5.3 PSYCHOMETRIC METHODS

Psychometric analyses were conducted to establish the reliability, validity and fairness of the ELOM 4&5 domains and to generate ELOM 4&5 standard scores, norms and standards. These procedures followed internationally recognised standards for test development (Cicchetti, 1994; AERA, APA, and NCME,2014).

Validity is concerned with the degree to which the conceptual background of the ELOM 4&5 and the information that has been gathered allow us to develop inferences and conclusions suitable for the ELOM 4&5 and South African children. Reliability is concerned with the consistency of the ELOM 4&5 across different situations. Fairness is concerned with the bias of the ELOM 4&5 (Bond & Fox, 2015; Byrne, 2010; Field, 2013).

The most important element of reliability has already been reported previously (reliability of assessor scoring on the Direct Assessment ELOM 4&5 described in Chapter 4). Further, some content validity has already been established due to the robust way the ELOM 4&5 items and domains were selected (see Section 2.1).

Unidimensionality and internal consistency of the ELOM 4&5 domains

A key psychometric concern is whether the items can be collapsed into one measure to represent a single underlying ELOM 4&5 domain, or construct (e.g. Emergent Numeracy and Mathematics). In the literature this is referred to as Factoral Validity (Chiccetti, 1994). We used Confirmatory Factor Analysis (CFA) to investigate this (Byrne, 2010; Kline, 2011).

Further, we were concerned with the value and contribution of individual item scores to the domain. We undertook Rasch Analysis to establish a uniform interval scale of each domain (Bond & Fox, 2015). Finally, for any new assessment tool such as the ELOM 4&5, it is necessary to establish whether the Direct Assessment items work equally for children from different socio-economic backgrounds. Differential Item Functioning (DIF) techniques from Rasch Analysis were used to ensure that ELOM 4&5 items did not discriminate between children from different backgrounds, who were of the same potential ability (Bond & Fox, 2015).

ELOM 4&5 Direct assessment items were grouped, a priori, into construct domains based on relevant literature and consultations with experts. A preliminary CFA was conducted to establish the unidimensionality and internal consistency of the ELOM 4&5 domains, before the transformation of scores. A final CFA was conducted and is presented later. CFA attempts to establish whether a group of items represents a single domain by fitting a model to the observed data (Byrne, 2010). The absolute degree of model fit is represented by a likelihood ratio chi-square score (Byrne, 2010). Significance values, associated with this score, that are greater than 0.05 suggest that the model is a likely fit. Further, a number of (relative) subjective fit statistics are produced to help determine just how well the model fits the data. The recommended subjective fit statistics are the SRMR, the GFI, the CFI, and the RMSEA (Kline, 2011). This first statistic is the Standardised Root Mean Squared Residual, or SRMR. This value should be less than .05. Next is the Goodness of Fit index, or the GFI. This value should be greater than .90, but values close to 1.00 are preferred. Next is the Comparative Fit Index, or CFI. This value should also be greater than .90, and values closer to 1.00 are preferred. Finally, the Root Mean Square Error of Approximation (RMSEA) score should be lower than .05, and the upper confidence interval should be less than .08.

Preparation
Before beginning the analysis, items were checked for their real-world performance and usability. Item 1 (formerly) of the Age Validation ELOM 4&5 ("Can you tell me the name of the place where you live?") was discarded from this analysis and from the final ELOM 4&5 Standards because it did not load satisfactorily on any factor, and because it was not possible to validate the child's response to the question. This item is used in another tool to assess the child's knowledge of his/her address as a safety indicator. We found it to be an inappropriate item for the ELOM 4&5.

Preliminary Confirmatory Factor Analysis
ELOM 4&5 domains used in the preliminary CFA were:

1. Gross Motor Development,
2. Fine Motor Coordination and Visual Motor Integration,
3. Emergent Numeracy and Mathematics,
4. Cognition and Executive Functioning,
5. Social and Emotional Development and Awareness,
6. Emergent Literacy and Language.

The preliminary CFA indicated that Gross Motor Development, Fine Motor Coordination and Visual Motor Integration, Emergent Numeracy and Mathematics, and Cognition and Executive Functioning were unidimensional and internally consistent. However, Social and Emotional Development and Awareness, and Emergent Literacy and Language, were not. Items 18 and 19 (sourced from the IDELA – see Appendix 2), which represented social and emotional development and awareness, loaded on the same factor as the language items. We cannot be sure of the reason for this, but language competence likely affects performance on these items (Zillmer, Spiers, & Culbertson, 2008). These two items, in fact, strengthened the emergent literacy and language domain, and were henceforth included in that domain.

Resultant factors for further analysis were:

1. Gross Motor Development,
2. Fine Motor Coordination and Visual Motor Integration,
3. Emergent Numeracy and Mathematics,
4. Cognition and Executive Functioning,
5. Emergent Literacy and Language (the two items formerly in Social and Emotional Development and Awareness).

As noted in Chapter 1, social relations and emotional functioning in areas relevant to school are retained in the Social-Emotional Functioning Rating Scale that accompanies the ELOM 4&5.

Once it was established that groups of items represented their underlying domains, Rasch Modelling served to identify how well the individual item scores performed, and adjust them to an interval scale (Bond & Fox, 2015).

The interval status and reliability of the ELOM 4&5 domains
Rasch Modelling is concerned with the items that make up a domain (Bond & Fox, 2015). Rasch Modelling uses the ability of the child and the difficulty of the item to place each score of each item on a continuum that represents the probability of success. The probability of success increases or decreases depending on whether the item score is above or below the child's ability level on a logistic scale. Rasch Modelling then transforms the ELOM 4&5 domains from ordinal scale to an interval scale (Bond & Fox, 2015). Further, Rasch Analysis serves to strengthen evidence presented by the CFA.

A Polytomous one-parameter logistic (1PL) Partial Credit Model (PCM) was used to model responses to each ELOM item in each of the five ELOM 4&5 domains, as a function of both the child's ability and the difficulty of the item.
PCM is well-suited to items with multiple, ordered response categories that vary in number and structure across items. The model estimates item-specific step (Andrich) thresholds, which represent points on the latent trait scale where the likelihood of selecting one score category over the next becomes equal. This method allows each item to contribute uniquely to the measurement of underlying ability, while maintaining comparability on a common logit scale.The resulting logit person measures, which are interval-level estimates of ability, were then transformed into a standardised scoring system for practical use and reporting, in line with the transformation procedures described in this manual.

Several statistics are reported for this purpose (Bond & Fox, 2015). These are, the Mean Square Infit, and the Mean Square Outfit, the Point-Measure Correlation, the Variance Explained, the Person Reliability, and the Item Reliability. Acceptable values for the Mean Square Infit and Outfit should not exceed ±1.4 logits – logistic distance units – from 0 (Bond & Fox, 2015). The Point-Measure Correlation coefficients should be more than .20. The Variance Explained should account for 50% of the variance in the underlying domain, or more, and the Unexplained Variance should have an eigenvalue of less than 2.0 – indicating no more factors (Bond & Fox, 2015; Linacre, 2016). The Rasch Person Reliability and the Rasch Item Reliability scores should be greater than .50 (Bond & Fox, 2015; Linacre, 2016).

The table below summarises the results of the Rasch Modelling for each domain. All domains are adequately represented by their items and are reliable for predicting a child's position above and below a standard (Bond & Fox, 2015; Linacre, 2016).

Table 9: Rasch Model Statistics

Domain	Mean Square		Point-Measure Correlation	Rasch Reliability		Variance Explained
	Infit	Outfit		Person	Item	
Gross Motor Development	.91 - 1.2	.86 - .97	.47 – 78	.75	1.00	75.3%
Fine Motor Coordination and Visual Motor Integration	.94 - 1.08	.91 - 1.09	.55 - .83	.71	1.00	65.1%
Emergent Numeracy and Mathematics	.85 - 1.20	.74 - 1.25	.49 - .73	.63	1.00	66.8%
Cognition and Executive Functioning	.68 - 1.17	.60 - 1.19	.57 - .81	.65	1.00	60%
Emergent Literacy & Language	.91 - 1.15	.73 - 1.08	.52 - .74	.75	1.00	59.3%

Transformed Score = (proposed maximum score - proposed minimum score)/ (real maximum score – real minimum score) (Rasch Logit – real minimum score) + proposed minimum score.

This Transformed Score was constructed to make all logit scores positive, and to rescale the logit scores so that they start at zero and end at a maximum score of 20 for each domain. The maximum score for each domain was chosen to create an equal weighting between the domains.

The total contribution of each domain produces a maximum score of 100 for the ELOM 4&5. Appendix 3, Table 5 displays the untransformed difficulty logit of each item score.

Confirmation of unidimensionality and internal consistency.
After the transformation of the ELOM 4&5 scores, a final CFA was conducted to confirm the validity of the ELOM 4&5 domains under the transformed scoring system. The results of the final CFA are presented in the table below. Factor Loadings can be found in Appendix 3 Table A 3.10.

Table 10: Final Confirmatory Factor Analysis

ELOM 4&5 Domains	X^2	P	SRMR	GFI	CFI	RMSEA	10% CI RMSEA	90% CI RMSEA
Gross Motor Development	4.65	0.098	0.01	0.99	0.99	0.03	<0.01	0.07
Fine Motor Coordination and Visual Motor Integration	0.45	0.801	0.01	1.00	1.00	<0.01	<0.01	0.03
Emergent Numeracy and Mathematics	1.65	0.439	0.01	0.99	1.00	<0.01	<0.01	0.06
Cognition and Executive Functioning	4.72	0.095	0.02	0.99	0.99	0.03	<0.01	0.07
Emergent Literacy and Language	4.14	0.126	0.01	0.99	0.99	0.03	<0.01	0.07

All ELOM 4&5 domains showed satisfactory model characteristics after their transformation. Further, only Emergent Numeracy and Mathematics required the removal of records with poor model fit (person misfit ± 2.0) to produce a suitable model. This evidence suggests that some children did better on items that should have been too difficult for them, or that they failed items that were meant to be too easy for them (Bond & Fox, 2015). Some children may not have understood the instructions for some items in the Emergent Numeracy and Mathematics domain. However, as this group of children represented 12.8% of the sample, as per best practice in the field of Rasch Analysis, this domain is fit for purpose (Bond & Fox, 2015). Table A.3 in Appendix 3 details the number of records with poor model fit in each domain.

5.4 ASSESSMENT OF BIAS: DIFFERENTIAL ITEM FUNCTIONING

Establishing the fairness of the ELOM 4&5 Years Assessment tool in 11 official South African language groups

To establish the fairness of a test for use across different cultural and language groups, it must be assessed for its linguistic, cultural, functional and metric equivalence, that is, the underlying constructs being measured, must be understood in the same way and demonstrate the same psychological (factor) structure, and scale items must permit all children regardless of their background, to demonstrate their actual ability (Van der Vijver & Tanzer, 2004; Milfont & Fischer, 2010; Peña, 2007; Peña & Quinn, 1997).

Measurement invariance is the central property of metric equivalence. This was assessed for the ELOM 4&5 Years Assessment tool using analyses of Differential Item Functioning (DIF), a method used to detect bias between language groups at the item level. It assumes that children who have the same abilities should have a similar probability of responding correctly to the same item, regardless of their language group (Linacre, 2016; Magis et al., 2010). As the ELOM 4&5 Years Assessment tool was developed in English and then translated, English is considered the reference group and the other languages as focal groups for DIF analysis. That is, the performance of each language group is compared with the English home language sample.

In South Africa, children speaking African languages are disproportionately represented in the lower income groups, while English and Afrikaans children, in particular, are far more likely to live in the top two wealth quintiles. This means that language and income level are confounded – both play a role in determining children's performance on the ELOM 4&5, but household wealth is likely to be a significant influence. This is confirmed in studies of predictors of ELOM 4&5 performance (Tredoux et al., 2023; Henry & Giese, 2023).

The metric equivalence of the ELOM 4&5 was established for the five language groups included in the 2016 standardisation sample using Rasch analysis and DIF (Snelling et al., 2019 Dawes et al., 2020): English, Afrikaans, isiXhosa, isiZulu and Setswana. Since then, we have been able to assess children in the remaining languages and assemble the samples used in the analyses described here.

Sample
As reported in Table 11, the sample for DIF analyses consisted of 15 487 ELOM 4&5 assessments (7 510 males; 7 977 females) ranging in age from 49 to 70 months (mean = 58.32 months; SD =5.43).

Table 11: Sample Age in Months and Sex

N	15487	Sex	Frequency	Percent
Mean Age (months)	58.32 (SD =5.43)	Male	7510	48.5
Median Age (months)	58.00	Female	7977	51.5
Range (Min; Max)	20 (49; 70)	Total	15487	100.0

Sample size for DIF depends on the complexity of the analysis. Sireci and Rios (2013) recommend that group sizes should fall in the 200-500 range, while Belzak (2019) and Lai, et al. (2005) report that in simple models, samples of 50-100 can reliably detect DIF. Table 12 shows that for most languages, sample sizes are more than adequate. However, isiNdebele, in particular, is below 100, and findings must be treated with some caution as they may be unreliable. Xitsonga is somewhat below our target, but following Lai et al. (2005), we regard it as acceptable and not likely to generate unreliable findings.

Table 12: Sample Home Languages

Language	Frequency	Percent
Afrikaans	2329	15.0
English	1405	9.1
IsiNdebele	78	.5
IsiXhosa	2800	18.1
IsiZulu	3256	21.0
Sesotho	1106	7.1
Sesotho se leboa (Sepedi)	1190	7.7
Setswana	2620	16.9
Siswati	243	1.6
Tshivenda	289	1.9
Xitsonga	171	1.1
Total	**15487**	**100.0**

WINSTEPS® software was used to conduct the DIF analyses across all 11 languages. A partial credit 1 Parameter (1-PL) model, appropriate for items with more than two response categories, was used. Findings are summarised below. Item - Level DIFDIF heat maps for each item and language are provided for each ELOM 4&5 domain in Tables 13 through 17. Values > .50 indicate DIF and are indicated in **red**. DIF is evident for 4 of 23 items. We note DIF at the item level below.[7]

- **FMCVMI**
  - item 6 Copy Triangle: All languages. A plausible explanation is that there were few very high and few very low performing individuals on the item. Confirmation would depend on further exploration of score distributions.
- **CEF**
  - item 17 Picture puzzle completion: Siswati. The sample size is acceptable. A plausible explanation is that there were few very high and few very low performing individuals on the item.
- **ELL**
  - item 18 Expressive language: empathy: IsiXhosa, Sesotho, Tshivenda and Xitsonga. It is plausible that these are examples of benign bias occasioned by the nature of these languages.
  - item 23 Initial sound discrimination: all languages except English and Sesotho. It is plausible that these are examples of benign bias occasioned by the nature of these languages.

Table 13: Gross Motor Development

ITEM	Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sesotho	Sesotho se leboa (Sepedi)	Setswana	Siswati	Tshivenda	Xitsonga
1	0.0493	-0.113	-0.3913	0.0013	-0.113	-0.2411	-0.323	-0.2007	-0.0509	-0.3604	-0.1594
2	-0.0329	-0.0329	0.1086	-0.0329	-0.0329	0.0003	-0.063	-0.0329	-0.0329	-0.2081	0.0274
3	0.0847	0.1255	0.3813	0.0549	0.1255	0.0942	0.2597	0.1255	0.2024	0.3491	0.2299
4	-0.1049	-0.0035	-0.1035	-0.0275	0.0205	0.1519	0.1399	0.1041	-0.1133	0.2475	-0.0933

No DIF is evident for any GMD item (no Values higher than .50).

[7]A full report is available at https://datadrive2030.co.za/resources/differential-item-functioning-analyses-on-the-elom-45-across-all-11-south-african-languages/

Table 14: Fine Motor Coordination and Visual Motor Integration

ITEM	Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sesotho	Sesotho se leboa (Sepedi)	Setswana	Siswati	Tshivenda	Xitsonga
5	-0.1874	-0.1619	-0.3955	-0.3054	-0.3054	-0.4782	-0.4178	-0.3054	-0.354	-0.4172	-0.4452
6	0.7831	0.5927	0.6813	0.7831	0.7831	0.7831	0.8429	0.8354	0.8384	0.8465	0.7579
7	-0.09	-0.1877	-0.1849	-0.1773	-0.1094	-0.0182	-0.1444	-0.2303	-0.1074	-0.3252	-0.2591
8	-0.557	-0.1477	-0.0789	-0.3333	-0.3716	-0.3333	-0.2908	-0.2978	-0.4025	-0.1202	-0.0829

DIF is evident for: item 6, *Puzzle completion:* All languages.

Table 15: Emergent Numeracy and Mathematics

ITEM	Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sesotho	Sesotho se leboa (Sepedi)	Setswana	Siswati	Tshivenda	Xitsonga
9	-0.0692	-0.3161	-0.0672	-0.1258	-0.0364	0.2101	0.2129	0.1959	-0.1486	0.1578	0.1424
10	0.0508	-0.0917	-0.1094	0.2732	0.3433	0.1603	0.3306	0.3181	0.0768	0.1836	0.0824
11	0.2073	0.0709	-0.1258	0.0709	0.0916	0.0105	-0.1261	0.05	0.1698	-0.1245	0.1065
12	-0.0725	0.2785	0.3311	0.0864	-0.037	-0.0587	0.1031	-0.1223	0.1409	0.0903	0.1025
13	-0.1298	0.0612	-0.0327	-0.2949	-0.3569	-0.3284	-0.5104	-0.4325	-0.2463	-0.2949	-0.4422

No DIF is evident for any ENM item (no values higher than .50)

Table 16: Cognition and Executive Functioning

ITEM	Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sesotho	Sesotho se leboa (Sepedi)	Setswana	Siswati	Tshivenda	Xitsonga
14	0.0652	0.1802	-0.0921	0.0028	-0.0262	-0.1104	-0.1618	-0.0195	-0.0817	0.0279	-0.0601
15	-0.1132	-0.0675	0.302	-0.1546	-0.0675	0.0059	-0.0059	-0.0302	-0.144	0.1677	-0.1731
16	0.4094	0.324	-0.4372	0.0162	-0.148	-0.1504	-0.1473	-0.099	-0.2008	-0.2983	-0.0866
17	-0.2894	-0.3824	0.3349	0.1202	0.2424	0.2801	0.3622	0.149	0.5059	0.1371	0.3611

DIF is evident for: item 17 Picture puzzle completion for Siswati

Table 17: Emergent Language & Literacy

ITEM	Afrikaans	English	IsiNdebele	IsiXhosa	IsiZulu	Sesotho	Sesotho se leboa (Sepedi)	Setswana	Siswati	Tshivenda	Xitsonga
18	0.4932	0.4638	0.1984	0.8586	0.4059	0.6864	0.4277	0.5233	0.3101	0.5462	0.5462
19	-0.1757	-0.4161	-0.4743	-0.1003	-0.2461	0.0454	-0.0751	-0.0194	-0.139	-0.17	-0.0522
20	-0.0261	-0.0437	0.0369	-0.2721	-0.3644	-0.2958	-0.3452	-0.4191	-0.3262	-0.3447	-0.3492
21	-0.2393	-0.0946	-0.098	-0.4034	-0.16	-0.2963	-0.2963	-0.4898	-0.198	-0.3636	-0.381
22	-0.3415	-0.2714	-0.6003	-0.5364	-0.4243	-0.4039	-0.2787	-0.1118	-0.5429	-0.2493	-0.2676
23	0.2832	0.383	0.8542	0.4451	0.7743	0.254	0.5456	0.5026	0.9079	0.5854	0.5026

DIF is evident for:
- Item 18 Expressive language: empathy: IsiXhosa, Sesotho, Tshivenda and Xitsonga
- Item 23 Initial sound discrimination: all languages except English and Sesotho.

Domain-Level DIF

- Procedures for DIF analysis at the domain level (summed DIF across items in a domain) are described by Bond and Fox (2015) and the WINSTEPS® website[8].
- Domain-Level DIF is more relevant for assessing metric equivalence in this case because it is standard practice to use domains and not items as variables for comparing groups in ELOM 4&5 research and for reporting to Early Learning Programmes.
- Domain-Level DIF is reported in Table 18 for all languages. Values > .50 indicate DIF.

[8]See: https://winsteps.com/winman/difconcepts.htm

Table 18: Differential Item Functioning Results by Language

Language	Gross Motor Development (GMD)	Fine Motor Coordination and Visual Motor Integration (FCM&VMI)	Emergent Numeracy and Mathematics (ENM)	Cognition and Executive Functioning (CEF)	Emergent Language and Literacy (ELL)
Afrikaans	-0.0038	-0.0513	-0.0134	0.072	-0.0062
English	-0.0239	0.0954	0.0028	0.0543	0.021
IsiNdebele	-0.0049	0.022	-0.004	0.1076	-0.0831
IsiXhosa	-0.0042	-0.0329	0.0098	-0.0154	-0.0085
IsiZulu	0.0001	-0.0033	0.0046	0.0007	-0.0146
Sesotho	0.0053	-0.0466	-0.0062	0.0252	-0.0102
Sepedi	0.0136	-0.0101	0.0101	0.0472	-0.022
Setswana	-0.004	0.0019	0.0092	0.0003	-0.0142
Siswati	0.0053	-0.0255	-0.0074	0.0794	0.0119
Tshivenda	0.0281	-0.0161	0.0123	0.0344	0.004
Xitsonga	0.0046	-0.0293	-0.0084	0.0413	-0.0012

No DIF is evident at the domain level for the ELOM 4&5, with all **acceptable**. Metric equivalence is evident across different languages.

Overall, these domain-level findings support the use of the ELOM 4&5 as a reliable and fair tool for assessing child development outcomes across all 11 official South African languages (except SASL).

Summary
In sum, psychometric analyses indicated that the Direct Assessment ELOM 4&5 domains are unidimensional and internally consistent measures of their constructs; the items hold interval scale status and display adequate ability to discriminate reliably between more and less able children; and items do not discriminate unfairly between children of language backgrounds, when interpreted at the domain level.

It should be reiterated that the process of removing cases with poor model fit (person misfit ±2.0) was only undertaken under conditions necessary to produce accurate estimates of model performance and psychometric properties – Rasch Modelling, the final CFA (EN&M), and Rasch DIF (Bond & Fox, 2015). Details of "person misfit" can be seen in Appendix 3, Table A 3.7.

Age and quintile differences
After establishing that the ELOM 4&5 displayed acceptable psychometric properties, with adjustments, it was necessary to determine whether the ELOM 4&5 was useful for investigating age and quintile differences, as intended. Multi-level modelling was used for investigating these complex differences (Field, 2013). The model consisted of two levels; one level accounted for the random effect of the schools, the other level accounted for individual differences between the children. For the purpose of these analyses, age was split into two categories – 50 to 59 months, and 60 to 69 months. Quintile groups were collapsed into three groupings – quintile 1, quintile 2/3 and quintile 4/5. Gender was also added to the model in order to control for it. An interaction effect was added in order to account for any interaction between quintile and age.

The total ELOM 4&5 score and each domain score were modelled separately. The parameters for each analysis are presented in the ELOM 4&5 Psychometry and Statistical Appendix 3. Estimated Marginal Means are presented in the table below. These results suggest that quintile and age groups differ as expected. Further, in some cases, older children in lower quintiles perform similarly to younger children in higher quintiles (Total ELOM 4&5 column).

Table 19: Estimated Marginal Means

Quintile	Age	Total ELOM 4&5	GMD	FMC&VMI	EN&M	C&EF	EL&L
Q1 (n = 114)	<5 (n = 53)	37.15 ± 3.58	6.72 ± 1.10	11.88 ± 0.88	7.97 ± 1.10	4.77 ± 1.10	5.81 ± 1.17
	>5 (n = 61)	41.13 ± 3.33	8.21 ± 1.03	12.50 ± 0.82	8.08 ± 1.02	5.95 ± 1.02	6.40 ± 1.10
Q2/3 (n = 756)	<5 (n = 115)	41.27 ± 2.43	7.62 ± 0.75	11.11 ± 0.60	7.99 ± 0.74	5.58 ± 0.75	8.97 ± 0.80
	>5 (n = 641)	49.89 ± 1.03	9.98 ± 0.32	12.87 ± 0.25	9.03 ± 0.32	7.78 ± 0.32	10.24 ± 0.34
Q4/5 (n = 461)	<5 (n = 90)	48.80 ± 2.75	8.39 ± 0.85	11.81 ± 0.67	9.34 ± 0.84	7.93 ± 0.84	11.32 ± 0.90
	>5 (n = 371)	54.29 ± 1.35	9.14 ± 0.42	14.18 ± 0.33	9.84 ± 0.41	9.29 ± 0.42	11.83 ± 0.44

5.5 ASSESSING THE EFFECTS OF MATURATION ON ELOM 4&5 SCORES[9]

When assessing the effects of an early learning programme (ELP) to improve child outcomes, it is important to separate the contribution of children's normal maturation due to ageing from growth in their performance due to participation in the ELP.

Several regression analyses were conducted on ELOM 4&5 assessments collected between 2016 and 2023 to determine the growth per month one could expect in Total ELOM 4&5 score and in the 5 Domains. The regressions conducted on the Grade R 2016 ELOM 4&5 Standardisation sample dataset, in which only gender is controlled, were deemed most appropriate for benchmarking ELOM 4&5 Total and Domain scores gains due to maturation.

There were two reasons:

1. The data used to calculate the maturation effect should be derived from children who are not in ELPs, to exclude the possibility of improvement in performance due to the programme. If this is not the case, the findings will be a function of both ageing and programme exposure (as noted by van der Berg in his analyses) . The 2016 sample meets this criterion. The children were assessed in January / February of Grade R, when they would have had minimal exposure to the curriculum.

2. The sample should be randomised. While we do not know whether the 2016 sample was exposed to an ELP prior to their enrolment in Grade R, we have a random sample stratified by school quintile and covering the languages spoken by 70% of the population. Randomisation helps to minimise selection bias and gives an even chance of including children who would and who would not have had exposure to an ELP prior to Grade R.

All the other analyses were conducted on samples that include children with known ELP or Grade R programme exposure rendering regressions conducted on these samples less defensible than those undertaken on the 2016 dataset (the report on these analyses is available at https://datadrive2030.co.za/wp-content/uploads/2023/08/How-To-Maturation-July2023.pdf). Recommended maturation effects are provided in Table 20.

Table 20: Recommended Maturation Effects per month

Domain	Domain Score Maturation Effect per month
Gross Motor Development (GMD)	0.23 (95% CI:0.17 - 0.28)
Fine Motor Coordination and Visual Motor Integration (FMC & VMI)	0.23 (95% CI: 0.19 - 0.28)
Emergent Numeracy and Mathematics (ENM)	0.12 (95% CI: 0.06 - 0.17)
Cognition and Executive Function (CEF)	0.25 (95% CI: 0.19 - 0.31)
Emergent Language and Literacy (ELL)	0.21 (95% CI: 0.15-0.28)
Total score	1.04 (95% CI:0.85 - 1.22)
Note: all domain scores range from 0-20 standard score points. The total score ranges from 0 - 100 points	

[9]We acknowledge the contributions of Servaas van Der Berg and Junita Henry to these analyses.

5.6 TEST-RETEST RELIABILITY

Test-retest reliability of the ELOM 4&5 Direct Assessment was investigated by Henning and Moonsamy (2019) and reported in Anderson et al., (2021). This form of reliability involves administering one test to the same participants on two occasions and correlating the two scores yielding a coefficient of stability (Price, 2017a; Price 2017b). It is reliant on the assumptions of stability (i.e. scores being constant) and equal variances (i.e. the same error variances) being upheld in both trials (Furr & Bacharach, 2014). These assumptions provide the foundation for ascertaining whether a child is likely to achieve a similar score on the same test on two administrations (i.e. scores reliably correlate) over time. Purposive sampling was used to recruit 49 English and isiXhosa-speaking preschool children (aged 55-69 months) residing in areas served by quintile 3 (fee-exemption) public schools. They were tested one week apart in their home language.

Table 21: Test Re-Test Reliability of the Early Learning Outcomes Measure Direct Assessment

ELOM 4&5 Domain	Test-retest reliability
Gross Motor Development (GMD)	r = .50 (p < .001) 95% CI [0.29, 0.68]
Fine Motor Coordination & Visual Motor Integration (FMC &VMI)	r = .79 (p < .001) 95% CI [0.63, 0.89]
Emergent Numeracy & Mathematics (ENM)	r = .76 (p < .001) 95% CI [0.57, 0.89]
Cognition & Executive Functioning (CEF)	r = .64 (p < .001) 95% CI [0.44, 0.83]
Emergent Language & Literacy (ELL)	r = .74 (p < .001) 95% CI [0.58, 0.86]
ELOM 4&5 Total	r = .90 (p < .001) 95% CI [0.83, 0.95]

None of the confidence intervals cross zero and all p values are significant at p < .001 (Field, 2013) indicating their reliability. A test-retest correlation criterion => .75 was used in this research as an acceptable level. As is evident from Table 13, ELOM 4&5 Total, FMC &VMI, ENM and arguably ELL either exceed or meet the criterion with ELOM 4&5 Total having excellent test-retest reliability (r = .90, p < .001). Only GMD and CEF did not meet the criterion of .75.

5.7 CONCURRENT VALIDITY

Concurrent validity is a type of criterion-related validity and assesses the degree to which a scale can relate to an already established measure (Price, 2017b). As they used an older sample (aged 72 to 76 months) than that used in the standardisation of the ELOM 4&5, Anderson et al. (2021) first investigated whether ELOM 4&5 ceiling effects (Ho & Yu, 2014), would be evident for their sample. Ceiling and floor effects are limits of measurement where scores tend either towards the maximum or minimum respectively. This restricts the ability of items to measure constructs and discriminate between high and low performances (Ho & Yu, 2014). A sample of 116 Afrikaans and isiXhosa speaking children (Mean age = 75.82 months) was tested on the ELOM 4&5 to investigate possible ceiling effects as this age group is older than that used in the ELOM 4&5 standardisation. Analysis of frequency histograms for ELOM 4&5 items, domains, and Total Score showed that only items 1 (stand on one leg), 5 (Copy and cross and a square) and 7 (draw self) showed ceiling effects. Only the FMC & VMI domain (which includes Items 5 and 7) was negatively skewed, indicating ceiling effects. All other ELOM 4&5 items, domains and Total Score were normally distributed, or only slightly skewed.

Concurrent Validity of the ELOM 4&5 Direct Assessment and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV)

As no significant ceiling effects were evident, the concurrent validity study could proceed. The appropriate criterion comparison must be measuring the same construct, with the goal of finding a high correlation between the two administered to the same person at close intervals.

Anderson et al. (2021) compared children's performance on the ELOM 4&5 Direct Assessment and the Wechsler Preschool and Primary Scale of Intelligence (WPPSI-IV), which measures similar constructs (Canivez, 2014; Wechsler, 2012).

The WPPSI-IV is a standardised intelligence test for children aged between 30 and 91 months, that has not been standardised in South Africa and is only available in English. The WPPSI-IV is widely recognised as a gold-standard test, which measures similar, constructs to the ELOM 4&5. WPPSI-IV concurrent validity has been well established through comparisons with the WPPSI-III, the Differential Ability Scales, and the Bayley Scales of Infant and Toddler Development (Thorndike, 2014).

To establish the concurrent validity of the ELOM 4&5 and the WPPSI- IV, 62 children enrolled in the Drakenstein Child Health Study (DCHS[10]), aged 72 to 76 months (M = 75.05) were assessed on both instruments.

Five WPPSI-IV indices were thought to compare well with three ELOM 4&5 domains as illustrated in Table 22.

Table 22: ELOM 4&5 and WPPSI-IV Comparison

WPPSI Core Subtest	WPPSI Index	ELOM 4&5 Domain
Block Design	Visual spatial	Fine motor coordination & visual motor integration (FMC & VMI)
Matrix Reasoning	Fluid reasoning	Cognition & executive functioning (CEF)
Bug Search	Processing speed	Cognition & executive functioning (CEF)
Picture Memory	Working memory	Cognition & executive functioning (CEF)
Similarities	Verbal comprehension	Cognition & executive functioning (CEF)
Information	Verbal comprehension	Emergent literacy & language (ELL)

Results showed a very high correlation (Swank and Mullen, 2017) between ELOM 4&5 Total Score and WPPSI-IV Full Scale composite score ($r = .64, p < .001$), thus establishing concurrent validity of the ELOM 4&5. Each of the three ELOM 4&5 domains yielded a statistically significant, and high or very high correlation with the WPPSI-IV Full Scale composite score. The FMC & VMI domain of the ELOM 4&5 correlated with WPPSI-IV Block Design ($r = .34$, $p = .003$) and Bug Search ($r = .51$, $p < .001$). ELOM 4&5 CEF correlated with the WPPSI-IV Block Design subtest ($r = .37$, $p = .002$), Matrix Reasoning ($r = .35$, $p = .003$), Bug Search and Picture Memory ($r = .32$, $p = .005$). The ELOM 4&5 ELL domain correlated with the WPPSI-IV VCI composite score (incorporating the Similarities and Information subtests of the WPPSI-IV) ($r = .50$, $p < .001$).

In all, the concurrent validity of the ELOM 4&5 Direct Assessment has been established on this low-income older sample.

Assessing the reliability of individual ELOM 4&5 Direct Assessments
Three indicators of child engagement with assessment are used to decide whether ELOM 4&5 scores can be regarded as a reliable reflection of a child's ability. If all three indicators are present, the assessment is normally excluded from analysis:

1. A Total ELOM 4&5 score =/<15;
2. A Task Orientation score = 4;
3. An assessor comment that suggests a problem with the assessment that is not due to child ability.

Additional reasons for excluding a record include:

1. The child failed the disability screen;
2. The assessment was discontinued prior to completion;
3. The child was not assessed in their home language; and
4. The assessment was compromised due to assessor error.

It is the experience of the ELOM team that inter-scorer reliability is difficult to obtain on the ELOM 4&5 Direct Assessment Task Orientation items. It is therefore advised that they are only used for judging whether the assessment is a reliable indication of the child's ability.

[10]For further information on the DCHS see: http://www.paediatrics.uct.ac.za/scah/dclhs

5.8 PREDICTIVE VALIDITY OF THE ELOM 4&5[11]

Predictive validity is a type of criterion validity that indicates how well a test result is able to predict a child's performance on a different measure (the criterion) at a later point. The Roots and Shoots longitudinal study[12] which is being conducted in the Western Cape, has followed learner performance from the start of Grade R through the Foundation Phase grades. This has permitted investigation of the predictive validity of the ELOM 4&5 (average age 65.4 months) to early Grade 1 (average age 76.92 months) and to the end of Grade 2 (average age 94.56 months of age).

Sample: Roots and Shoots provided data from 585 children. randomly selected from 50 Afrikaans Language of Learning and Teaching (LOLT) schools, and 199 children randomly selected from 25 isiXhosa LOLT schools. Cases who were older than 69 months (above the ELOM 4&5 age range) were excluded (16) as were those who had no data beyond wave 1 (128). The final sample size was then 441. Given these considerations, the sample can no longer be considered random.

Sample sizes for regressions varied depending on how many waves of data were required. The regressions also used complete-case analysis but the number of observations that could be used depended on the specific variables that were modelled.

It is important to note that 73% of the sample attended no fee schools[13] and 16% attended low fee schools (Hofmeyr & Qvist, 2025). Ninety percent of the sample available for analysis can therefore be considered to be from backgrounds that are disadvantaged to a greater or lesser degree.

Measures: The ELOM 4&5 was administered to each child at the commencement of Grade R, and the newly standardised ELOM-R v1 Language and Mathematics tools (Kleineibst et al., 2025a and b) were administered when the children started Grade 1. The Early Grade Reading Assessment (EGRA) (Dubeck & Grove, 2015) and the Early Grade Mathematics Assessment (EGMA) (Platas et al., 2016) was administered when the children were in Grade 2. All administrations were conducted in the child's home language. The criteria for predictive validity were therefore positive statistically significant relationships between ELOM 4&5 Total standard scores and ELOM-R v1 Mathematics and Language percent correct scores at the start of Grade 1, and EGMA and EGRA scores in Grade 2.

Statistical Analysis: To investigate the relationship between ELOM 4&5 Total scores and Domain scores, ELOM-R v1 Mathematics and Language assessments and performance on the EGRA and EGMA tests, multiple linear regressions were conducted, controlling for age, sex, task orientation score and the Roots and Shoots measure of Socio-Economic Status[14]).

Key Findings:
ELOM 4&5 Total scores at the commencement of Grade R:

- Predicted **ELOM-R v1 Language** (β = **0.571**) and **Mathematics** (β = **0.592**) percent correct scores, as well as performance on the **EGRA** (β = **0.431**) and **EGMA** (β = **0.320**) when in Grade 2.

- Three **ELOM 4&5 domain scores** predicted performance on **ELOM-R v1 Language:** FMC&VMI (β = **0.222**), EN&M (β = **0.247**), and CEF (β = **0.229**).

- The same three **ELOM 4&5 domain scores** predicted performance on **ELOM-R v1 Mathematics:** FMC&VMI (β = **0.254**), EN&M (β = **0.256**), and CEF (β = **0.219**), as well as performance on both EGMA {FMC&VMI (β = **0.120**), EN&M (β = **0.235**), and CEF (β = **0.105**), and EGRA (FMC&VMI (β = **0.284**), EN&M (β = **0.238**), and CEF (β = **0.134**).

One would not expect the GMD domain which assesses large muscle movements and coordination to predict ELOM-R or EGRA and EGMA scores. The failure of the Early Language and Literacy (EL&L) domain to predict ELOM-R V1 Language and EGRA is unexpected and suggests that it be considered for possible revision.

In sum, the predictive validity of the ELOM 4&5 is established (in this sample) for both ELOM-R v1 tests and for EGMA and EGRA. In the case of both EGMA and EGRA, other influences not measured here such as developmental changes (uneven in the period assessed), the quality of schooling and changes in the child's home environment would be likely to influence the strength of the associations with ELOM 4&5 scores measured two years previously.

[11]Statistical analyses were conducted by Caitlin Buenk and Jürgen Becker of AX Group (https://ax-assess.com) in consultation with Andrew Dawes.
[12]https://www.rootsandshootsstudy.com
[13]Provincial departments classify schools as "No Fee" if they serve disadvantaged communities (School Quintiles 1–3). They may not charge fees. https://www.gov.za/faq/education/where-can-i-get-information-about-school-fees-and-no-fee-schools
[14]Based on 13 Household Assets.

CHAPTER 6

CONSTRUCTION OF ELDS AND ELOM 4&5 NORMS

MATTHEW SNELLING, ANDREW DAWES & LINDA BIERSTEKER

Performance standards describe what children should know and be able to do at particular levels (in this case the ELDS discussed in Chapters 1 and 2). ELOM 4&5 standards statements for each domain are derived from policy and are set out in Appendix 2 together with the items and assessment sources. Once standards statements are specified and performance on the items used to measure them has been established, it is necessary to decide on the cut scores that denote achievement of the standard (Ricker, 2006). There are a number of methods for setting performance standards and their associated cut scores (Kane, 2011; Zieky and Perie, no date). Whatever approach is used, it must be based on an acceptable quantitative methodology and involve judgements on cut scores by suitably qualified persons.

The logic for setting ELDS standard cut scores based on ELOM 4&5 performance proceeded as follows:

1. Internationally, the advice of experts is that the ELDS should be set at a level of performance attained by a representative sample of 50 – 60% of children assessed. In the case of the ELOM 4&5 study, that would be the score attained by at least 50% of the total sample (the median or middle score of the distribution). Children's performance on the ELOM 4&5 provided information that could be used in this way.

2. As the ELOM 4&5 is to be used to measure programme performance against a set of standards that children are expected to achieve, the sample median is regarded as too low. This is because it is depressed by the 65% of the sample from disadvantaged backgrounds (quintiles 1 to 3) attending 'No Fee' Schools. We know from studies of the public school system, that children in these quintile bands perform below the level of those in quintiles 4 and 5. This trend is also evident in the ELOM 4&5 data.

3. To set the standards, the question asked was: what is the most appropriate and realistic reference point for setting expected ELDS for early learning programmes delivered to children affected by socio-economic disadvantage?

ELOM 4&5 Standards are based on performance on the ELOM 4&5 Direct Assessment only. Those measured in the Social-Emotional Functioning Rating Scale of the child are not included, as they did not form part of the process. As is an accepted practice for standards setting, prior to setting provisional cut scores for each age group, Rasch Analysis was used to derive standard score distributions for ELOM 4&5 total scores, and for each of the five domains. The performance of each of the three school quintile groups (1, 2/3, 4/5), and the total Age Validation sample was determined. The distributions could then be compared to establish the proportions of children in each who would meet a standard if it were set at a particular level. A level of performance that could be realistically expected of early learning programmes while seeking to push toward an expected standard for children was decided following consultation with the Innovation Edge and inspection of the performance of the three quintile groups. Provisional ELOM 4&5 performance standards were benchmarked at the score achieved by the top 40% of children in the age validation sample (the 60th Percentile on the distribution) for presentation to an expert group including representatives from the Departments of Social Development, Basic Education and Planning, Monitoring and Evaluation in the Presidency. At a meeting held in September 2016 the expert group agreed on the benchmark. The process was also reviewed by international experts.

We provide ELDS and norms for: 50-59 months & 60-69 months age bands using the 2016 standardisation sample in Appendix 1.

As shown in Appendix 1, children whose scores fall in the top 40% of the distributions (above the 60th percentile) are classified as being **"On Track**" (green).

Those whose scores fall between the 32nd and 59th percentile are classified as **Falling Behind** (yellow) and with support they should be able to achieve it. Scores of children **Falling Far Behind** (red) are below the 32nd percentile (well below the expected standard) and are likely to need significant assistance to be **On Track**. These performance categories are also provided with the ELOM 4&5 tablet-based scoring.

Key informants interviewed stressed the importance of children entering Grade R being confident, able to follow instructions, regulate emotional expression and cooperate with peers. As noted above, these are not easily assessed in a testing situation with a stranger. Rating scales were therefore constructed for teacher assessments of children's behaviour to be used alongside the ELOM 4&5 Direct Assessment (see ELOM Social-Emotional Functioning Rating Scale). The tool is intended for use by teachers who are familiar with the child and have been able to observe them over the course of their attendance in an early learning programme. Social relations items in the tool are drawn from the Child Trends Teacher Rating (Child Trends, 2014) and the California Desired Results Developmental Profile (California Department of Education, 2008, 2010). Emotional functioning items relevant to coping with the early phases of school were selected from the South African Child Assessment Scales (SACAS) which is based on the Achenbach Child Behaviour Checklist and used in the Birth to Twenty Study (Barbarin and Richter, 2001; van der Merwe and Dawes, 2000).

The Social-Emotional Functioning Rating Scale could be adapted for parents' reports for use in home visiting programmes, and for playgroup leaders. Though norms are not set for this instrument, expected scores are provided below. Table 23 provides the item sources.

One item (not included in Psychometry) measures Self-Care. The assessor is asked to rate the child's independence on toileting: "Can this child use the toilet on her / his own."

Aspects of behaviour included in psychometric analyses are:

- Social Relations with Peers,
- Relationships and Social Interactions with Familiar Adults, and
- Emotional dimensions associated with readiness for school.

7.1 ELOM SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE VALIDITY AND RELIABILITY

The ELOM Social-Emotional Functioning Rating Scale can be used alongside the ELOM 4&5 in order to measure aspects of behaviour that requires longer-term knowledge of the child across situations, and that are not reliably measured in a one-off direct assessment. The instrument includes one item to rate the child's Self Care (degree of independent toileting), and two scales: Social Relations with Peers and Adults (SRS) (6 items) which measures children's relations with peers and adults, and the Emotional Readiness for School Scale (ER) (6 items) designed to assess aspects of emotional functioning associated with readiness for school.

Analysis is based on 261 assessments of children in the same age range as that used for the standardisation of the ELOM 4&5. The sample was constructed from ratings provided by several early childhood development organisations over a two-year period. The majority of the children are from disadvantaged backgrounds and belong to the lowest 3 income quintiles.

All items were analysed using Exploratory Factor Analysis (EFA). Reliability was assessed using Chronbach's Alpha. Due to the large sample size, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was set at .75 – or middling to meritorious (SRS = .755; EFS = .809), by the standard of Hutcheson and Sofroniou (1999). All KMO item statistics were set at .60 – well above the .50 that is recommended. The Determinant (a measure of multicollinearity), was required to be greater than 0.00001 (SRS = .151; EFS = .184). The reliability analysis was required to produce a Coefficient Alpha ( ) greater than .70 for both scales.

Reliability: All items met the above requirements with the Social Relations Scale (SRS) Chronbach = .78, and the Emotional Readiness for School Scale (ER) Chronbach = .80.

Factor Analysis: The above statistics were generated using two EFAs. The Items included in each EFA are presented below.

EFA 1

- Does this child work well with peers (can wait for their turn/manage impulsivity)?
- Does the child resolve problems with peers without becoming aggressive?
- Does the child cooperate with peers without prompting?
- Does the child seek out assistance or support from familiar adults?
- Does the child seek a familiar adult's ideas or explanations about events or experiences that are interesting to the child?
- Does the child take initiative in creating cooperative activities with a familiar adult?
- Does the child take initiative in creating cooperative activities with a familiar adult?

EFA 2

- Is it easy to understand what the child is saying?
- Does the child express needs and feelings appropriately?
- Is child independent, does child like to do things without help?
- Does the child adjust well to changes in the classroom or home routine?
- Does child approach new experiences confidently, without fear?
- Is child a self-starter?

Table 23: Social-Emotional Functioning Rating Scale Item Sources

Social Relations Scale	
Factor 1: Social Relations with Peers	
1. Does the child work well with peers (can wait for their turn/ manage impulsiveness)?	**Source: All drawn from the Social Competence Scale of the:** Child Trends (2014). Measuring elementary school students' social and emotional skills. Providing Educators with Tools to Measure and Monitor Social and Emotional Skills that Lead to Academic Success. Child Trends Publication #2014-37. Retrieved from: https://www.childtrends.org/publications/measuring-elementary-school-students- social-and-emotional-skills-providing-educators-with-tools-to-measure-and-monitor- social-and-emotional-skills-that-lead-to-academic-success five competencies and skills that help students excel in school over time: self- control, persistence, mastery orientation, academic self-efficacy, social competence (assessed here).
2. Does the child resolve problems with peers without becoming aggressive?	
3. Does the child cooperate with peers without prompting?	
Factor 2: Relationships and Social Interactions with Familiar Adults	
4. Does the child seek out assistance or support from familiar adults?	**Source:** Items are drawn from the Social and Emotional Development domain of the California Desired Results Profile (DRDP) which "assesses preschool children's developing abilities to understand and interact with others and to form positive relationships with nurturing adults and their peers" (p. iv). California Department of Education (2014). DRDP (2015): A Developmental Continuum from Early Infancy to Kindergarten Entry Calibration Version. Sacramento: California Department of Education. Retrieved from: https://www.cde.ca.gov/sp/cd/ci/ documents/drdp2015preschool.pdf DRDP items chosen describe development that typically occurs in the preschool years and early kindergarten (Grade R).
5. Does the child seek a familiar adult's ideas or explanations about events or experiences that are interesting to the child?	
6. Does the child take initiative in creating cooperative activities with a familiar adult?	

Emotional Readiness for School Scale*	
Source: South African Child Assessment Scales (SACAS) social and academic competence items (regarded by teachers consulted to be key to adjustment to school). Teachers consulted agreed that children should demonstrate these attributes prior to entering Grade R. Four items selected from the seven item Resilience scale, and three items from the ten item Academic Readiness Scale. Items that overlap with those chosen for the ELOM 4&5 Social Relations Scale were not selected.	
1. Is it easy to understand what the child is saying?	Academic Readiness
2. Does the child express needs and feelings appropriately?	Academic Readiness
3. Is the child independent, does the child like to do things without help?	Academic Readiness and Resilience
4. Does the child adjust well to changes in the classroom or home routine?	Resilience
5. Does the child approach new experiences confidently, without fear?	Resilience
6. Is the child a self-starter?	Resilience

* This scale was formerly known as the Emotional Functioning Scale; items have not changed.

Table 24: Social-Emotional Functioning Rating Scale Item Sources

Social Relations with Peers and Adults Scale			
Item	**Factor Loading (Factor Matrix)**	**Item KMO**	**Chronbach's α**
Factor 1: Social Relations with Peers			
Does the child work well with peers (can wait for their turn/manage impulsiveness)?	.75 / .76	.77	**α = .78**
Does the child resolve problems with peers without becoming aggressive?	.74 / .73	.79	
Does the child cooperate with peers without prompting?	.77 / .78	.78	
Factor 2: Relationships and Social Interactions with Familiar Adults			
Does the child seek out assistance or support from familiar adults?	.62 / .60	.72	**α = .78**
Does the child seek a familiar adult's ideas or explanations about events or experiences that are interesting to the child?	.92 / .89	.68	
Does the child take initiative in creating cooperative activities with a familiar adult?	.54 / .64	.79	

Emotional Readiness for School Scale			
Item	**Factor Loading (Factor Matrix)**	**Item KMO**	**Chronbach's α**
Factor 1: Emotional Readiness for School			
Is it easy to understand what the child is saying?	.56	.84	α = .80
Does the child express needs and feelings appropriately?	.52	.85	
Is the child independent, does the child like to do things without help?	.68	.81	
Does the child adjust well to changes in the classroom or home routine?	.55	.84	
Does the child approach new experiences confidently, without fear?	.70	.80	
Is the child a self-starter?	.78	.76	

7.2 GUIDELINES FOR INTERPRETATION ELOM SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE SCORES

Expected scores on the two Social-Emotional Functioning Rating scales have been empirically derived in analysis of 261 records (50-59 months N = 122; 60-69 months N = 133). Analyses were conducted separately on the two age groups.

Expected scores were then derived for each age group. These were based on two principles. The expected total score for both scales should either:

1. be aligned with the scale mean or,
2. not be less than the score obtained by 60% of the sample.

Findings and recommendations for expected scores are summarised in Tables 25 and 26 below.

Table 25: Descriptive findings and recommendations: Sample 50-59 months

N = 122*	SELF CARE (score range:1-4)	SOCIAL RELATIONS (score range: 6-24)	EMOTIONAL READINESS FOR SCHOOL** (score range: 0-12)
Mean (SD)	3,76 (0.57)	17,06 (3.95)	8,41 (2.84)
Mode	4	14	11
Median	4,00	17,00	9,00
Ranges	1 to 4	6 to 24	0 to 12
Expected Score	4	18 (61% total score = >17)	9 (59% total score = >9)

* Mean age = 54.52 months (SD = 2.93).
** For the ER scale, the expected score is set at 9 because it is close to the 60th percentile; 74% of children were scored 8 or higher, and 42% were scored 10 or higher.

Table 26: Descriptive findings: Sample 60-69 months

N = 133*	SELF CARE (score range:1-4)	SOCIAL RELATIONS (score range: 6-24)	EMOTIONAL READINESS FOR SCHOOL** (score range: 0-12)
Mean (SD)	3,96 (0.19)	18,46 (3.34)	9,41 (2.42)
Mode	4	18	12
Median	4,00	18,00	10,00
Ranges	3 to 4	9 to 24	3 to 12
Expected Score	4	18 (65% total score = >18)	9 (68% total score => 9)

* Mean age = 63.79 months (SD = 2.90).
** For the ER scale, the expected score is set at 9 because it is close to the mean even though 68% score 9 or higher; 79% were scored 8 or higher and 54% of children were scored 10 or higher).

7.3 CONCURRENT VALIDITY OF THE ELOM SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE

Henning and Moonsamy (2019) investigated the concurrent validity of the ELOM Social-Emotional Functioning Rating Scale by comparing ratings of the same children with ratings on a very similar instrument, the Teacher version of the Strengths and Difficulties Questionnaire (SDQ) 4-17 years (Goodman, 1997). The SDQ is a brief screening questionnaire that measures behaviours on five scales: prosocial behaviour, conduct problems, hyperactive inattention, peer problems and emotional symptoms. High concurrent validity has been established between the teacher-rated SDQ Total Difficulties Score (excluding the prosocial score) and the Rutter Questionnaire (Goodman, 1994), the Parenting Stress Index-Short Form (McSherry, Fargas, & Weatherall, 2018), the Child Behaviour Checklist (Goodman & Scott, 1999), the Revised Children's Manifest Anxiety Scale and the Parent version of the ADHD Questionnaire (Muris, Meesters, & van den Berg, 2003). The SDQ has been used in South African studies (Hoosen et al., 2018).

Fifty-nine preschool class teachers rated their children on both instruments. Half the teachers completed the Social-Emotional Functioning Rating Scale first, and the other fifty percent, the SDQ first. Results are presented in Table 27.

Table 27: Correlations between Strength and Difficulties Questionnaire Scores and the ELOM Social-Emotional Functioning Rating Scale Scores

SDQ Categories	ELOM Social-Emotional Functioning Rating Scale Emotional Readiness for School (ER) Score	ELOM Social-Emotional Functioning Rating Scale Social Relations (SR) Score
Emotional Problems Score	r = -.58 (p < .001) 95% CI [-0.75, -0.38]	r =-.11 (p = .418) 95% CI [-0.40, 0.19]
Conduct Problems Score	r = -.30 (p = .020) 95% CI [-0.55, -0.10]	r = -.49 (p < .001) 95% CI [-0.66, -0.32]
Hyperactivity Score	r = -.13 (p = .333) 95% CI [-0.37, 0.10]	r = -.51 (p < .001) 95% CI [-0.68, -0.27]
Peer Problems Score	r = -.26 (p = .333) 95% CI [-0.51, 0.07]	r = -.25 (p = .059) 95% CI [-0.46, -0.02]
Prosocial Score	r = .39 (p = .003) 95% CI [0.11, 0.64]	r = .47 (p < .001) 95% CI [0.25, 1.66]
SDQ Total Difficulties Score (excluding Prosocial Score)	r = -.48 (p < .001) 95% CI [-0.63, -0.31]	r = -.53 (p < .001) 95% CI [-0.68, -0.38]

Table 27 shows that none of the confidence intervals straddle 0, indicating that the results are reliable (Field, 2018). Except for the SDQ Prosocial and SR and ER correlations, all are negative, which is a function of the direction of scale scoring. For example, a high rating on the ER Scale indicates positive functioning, while the reverse is the case for the SDQ (a higher score indicates more difficulties). Scoring is in the same direction for SDQ Prosocial and SR Scale.

The SR Score has a high negative correlation with the Total SDQ Difficulties Score and high positive correlation with the SDQ Prosocial Score, which indicates that on these dimensions, acceptable concurrent validity of the Social-Emotional Functioning Rating Scale with the SDQ has been established. As would be expected, the SR Score is negatively correlated with the SDQ Hyperactivity Score and SDQ Conduct Problems Score, which are a very high and high correlation, respectively. The ER Score has a high negative correlation with the Total SDQ Score as expected, with the highest negative correlation between the ER Score and SDQ Emotional Problems Score.

Acceptable concurrent validity of the Social-Emotional Functioning Rating Scale, Social Relations with its SDQ Total Difficulties construct counterpart is evident. All correlations are in the expected direction. Additionally, the SDQ prosocial behaviours are positively correlated with the Social-Emotional Functioning Rating Scale Social Relations items.

REFERENCES

AERA, APA, and NCME (2014).
Standards for educational and psychological tests. Washington DC: American Psychological Association, American Educational Research Association, National Council on Measurement in Education.

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

Barbarin, O. and Richter, L.M. (2001).
Mandela's Children. Growing up in post Apartheid South Africa. New York: Routledge.

Belzak, W.C.M (2019):
Testing Differential Item Functioning in Small Samples, Multivariate Behavioral Research, DOI: 10.1080/00273171.2019.1671162

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011).
Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. Learning and individual differences, 21(4), 327- 336.

Blair, C. (2002).
School readiness: Integrating cognition and emotion in a neurobiological conceptualization of child functioning at school entry. American Psychologist, 57, 111–127 doi: 10.1037//0003-066X.57.2.111.

Bond, T., & Fox, C. M. (2015).
Applying the Rasch model: Fundamental measurement in the human sciences. New York, NY: Routledge.Bradley, R.H., & Corwyn, R.F., (2002). Socioeconomic status and child development. Annual Review of Psychology, 53, 37–399.

Breslau, J. , Javaras, K. N. , Blacker, D. , Murphy, J. M. & Normand, S. T. (2008).
Differential Item Functioning Between Ethnic Groups in the Epidemiological Assessment of Depression. The Journal of Nervous and Mental Disease, 196 (4), 297-306. doi: 10.1097/NMD.0b013e31816a490e.

Byrne, B. M. (2010).
Structural equation modeling with AMOS: Basic concepts, applications, and programming. New York, NY: Routledge.

California Department of Education (2008, 2010)
California Preschool Learning Foundations V 1 and 2. Retrieved from: http://www.cde.ca.gov/

Canivez, G. L. (2014).
Test review of Wechsler Preschool and Primary Scale of Intelligence - Fourth Edition. InJ. F. Carlson, K. F. Geisinger, & J. L. Jonson (Eds.), The nineteenth mental measurements yearbook (pp. 4-15). Nebraska: Buros Center for Testing.

Child Trends (2014).
Measuring Elementary School students' social and emotional skills. Providing educators with tools to measure and monitor social and emotional skills that lead to academic success. Retrieved from: http://www.childtrends.org/wp-content/uploads/2014/08/2014-37CombinedMeasuresApproachandTablepdf1.pdf.

Cicchetti, D.V. (1994).
Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychological Assessment, 6(4), 284.

Dawes, A., Snelling, M., Henning, T. & Moonsamy, J. (2020).
ELOM 4&5 Teacher Assessment. In Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L., Tredoux, C.G. et al. Early Learning Outcomes Measure. Technical Manual (pp. 40-44). Claremont, Cape Town: The Innovation Edge https://DataDrive2030.co.za/wp-content/uploads/2022/09/ELOM 4&5 -Technical-Manual_2020-1.pdf

Dawes,A,Bray,R.Kvalsvig,J.,Rama,S.& Richter,L.(2004b).
Indicators of children's psychosocial development in the early childhood. Phase 3 Report for UNICEF. Cape Town: Human Sciences Research Council. Retrieved from http://www.hsrc.ac.za/en/research-outputs/view/1783

Dawes,A,Bray,R.,Kvalsvig,J.,Rama,S.& Richter,L.(2004a).
Indicators of children's psychosocial development in the early childhood. Phase 1 & 2 Report for UNICEF. Cape Town: Human Sciences Research Council. Retrieved from http://www.hsrc.ac.za/en/research-outputs/view/1258

Department of Basic Education (2009).
National Early Learning and Development Standards for children birth to four years (NELDS). Pretoria: Department of Basic Education.

Department of Basic Education (2011).
Grade R Curriculum and Assessment Policy Statement Grades R-3. Pretoria: Department of Basic Education. http://www.education.gov.za/

Department of Basic Education. (2015).
The South African National Curriculum Framework for children from birth to four. Pretoria: Department of Basic Education. Pretoria: Department of Basic Education.

Department of Basic Education and Department of Performance, Monitoring and Evaluation of South Africa. (2014)
The impact of the introduction of Grade R on learning outcomes. Pretoria: Department of Performance, Monitoring and Evaluation.

Diamond, A., & Lee, K. (2011).
Interventions shown to aid executive function development in children 4 to 12 years old. Psychological Science, 333(6045), 959-964 doi: 10.1177/0963721412453722

REFERENCES

DRDP (2015).
A Developmental Continuum from Early Infancy to Kindergarten Entry. Calibration Version. California Department of Education, Early Education and Support Division and Special Education Division.

Dubeck, M. M., & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. International Journal of Educational Development, 40, 315-322

Durkin, M. S., Wang, W., Shrout, P. E., Zaman, S. S., Hasan, Z. M., Desai, P., & Davidson, L. L. (1995).
Evaluating a ten-question screen for childhood disability: reliability and internal structure in different cultures. Journal of clinical epidemiology, 48(5), 657-666.

Engle, P. L., Fernald, L. C. H., & Alderman, H., et al. (2011).
Strategies for reducing inequalities and improving developmental outcomes for young children in low and middle income countries. Lancet, 378, 1339–53.

Field, A. (2013).
Discovering statistics using IBM SPSS statistics. London, England: Sage.

Field, A. (2018).
Correlation. In Discovering statistics using IBM SPSS statistics (5th ed., pp. 333-368). London: SAGE Publications.

Furr, R. M., & Bacharach, V. R. (2014).
Chapter 6. Empirical Estimates of Reliability. In Psychometrics: an introduction (pp. 125–163). United States of America: SAGE Publications.

Goodman, R. (1994).
A modified version of the Rutter Parent Questionnaire including extra items on children's strengths: A research note. Journal of Child Psychology, 35 (8), 1483–1494. doi: https://doi.org/10.1111/j.1469-7610.1994.tb01289.x

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

Goodman, R., & Scott, S. (1999).
Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: Is Small Beautiful? Journal of Abnormal Child Psychology, 27(1), 17–24. doi:10.1023/a:1022658222914

Henning, T.J. & Moonsamy, J. R. (2019).
Early Learning Outcomes Measure (ELOM 4&5) Psychometry: Test- Retest Reliability of the ELOM 4&5 Direct Assessment and Concurrent Validity of the ELOM 4&5 Teacher Assessment. Unpublished Honours Project, Psychology Department, University of Cape Town. Retrieved from https://humanities.uct.ac.za/sites/default/files/content_migration/humanities_uct_ac_za/1233/files/HNNTIF001%2520Tiffany%2520Joy%2520Henning%2520and%2520MNSJAS001%2520Jasmin%2520Roxanne%2520Moonsamy%2520-%25202019%2520Research%2520Project%2520Final%2520P

Henry, J. & Giese, S. (2023).
Reviewing the Socio-Economic Gradient in Learning Outcomes for Children who Participated in the Thrive by Five Index. https://DataDrive2030.co.za/wp-content/uploads/2023/08/Data-Insights-Reviewing-the-Socio-Economic-Gradient-Aug-23.pdf

Hoadley, U. (2013).
Building strong foundations: Improving the quality of early education. In L. Berry, L, Biersteker, A. Dawes, L. Lake & C. Smith C (Eds) (2013). South African Child Gauge 2013 (pp. 72-77). Cape Town: Children's Institute, University of Cape Town.

Hofmeyr, H. & Qvist, J. (2025).
Roots and Shoots Wave 3 Report. Research from early learning to school outcomes. Available at: https://www.rootsandshootsstudy.com/_files/ugd/d4434f_aad383691fe04d21951257754f14eee3.pdf

Hoosen, N., Davids, E. L., de Vries, P. J., & Shung-King, M. (2018).
The Strength and Difficulties Questionnaire (SDQ) in Africa: A scoping review of its application and validation. Child and Adolescent Psychiatry and Mental Health,12(6), 1-39. doi: 10.1186/s13034-017-0212-1

Hutcheson, G. D., & Sofroniou, N. (1999).
The multivariate social scientist: Introductory statistics using generalized linear models. Sage.

Kagan, S. L., Castillo, E., Gomez, R. E., & Gowani, S. (2013).
Understanding and using early learning standards for young children globally. International Journal of Child Care and Education Policy, 7(2), 53-66.

Kane, M.T. (2011).
So much remains the same: Conception and status of validation in setting standards. In Cizek, G.J (Ed.). Setting performance standards: Concepts, methods, and perspectives, pp. 53-88. Taylor and Francis e-library.

Kline, R. B. (2011). Principles and practice of structural equation modeling. New York, NY: Guilford Publications.

Kleineibst, M., Becker, J. & Dawes, A. (2025a). ELOM-R (V1) Technical Manual 2:
Language Assessment. Westlake: Westlake Cape Town. DataDrive2030. For further information, please refer to https://DataDrive2030.co.za.

Kleineibst, M., Becker, J. & Dawes, A. (2025b).
ELOM-R (V1) Technical Manual 3: Mathematics Assessment. Westlake Cape Town. DataDrive2030. For further information, please refer to https://DataDrive2030.co.za

REFERENCES

Lai, J. S., Teresi, J., & Gershon, R. (2005).
Procedures for the analysis of differential item functioning (DIF) for small sample sizes. Evaluation & the health professions, 28(3), 283-294. https://doi.org/10.1177/0163278705278276

Linacre, J. M. (2002).
Transitional Categories and Usefully Disordered Thresholds. Retrieved from http://www.oerj.org/View?action=viewPDF&paper=2

Linacre, J. M. (2016).
Index. Retrieved from http://www. winsteps.com/index.htm

LMTF (Learning Metrics Task Force) (2013). Toward Universal Learning: A Global Framework for Measuring Learning. Report No. 2 of the Learning Metrics Task Force. Montreal and Washington: UNESCO Institute for Statistics and Center for Universal Education at the Brookings Institution.

Magis, D., Beland, S., Tuerlincks, F., & De Boeck, P. (2010).
difR: A general framework and an R package for the detection of dichotomous differential item functioning. (Version 5.1.0) [R package]. Retrieved from https://CRAN.R-project.org/package=difR.

Malawi Ministry of Gender, Children and Social Welfare (2013).
Malawi early learning and development standards (ELDS). Working Document Revised. Lilongwe: Ministry of Gender, Children and Social Welfare.

McLoyd, V. C. (1998).
Socioeconomic disadvantage and child development. American psychologist, 53(2), 185.

McSherry, D., Fargas, M., & Weatherall, K. (2018).
The strength and difficulties questionnaire (SDQ): A Proxy Measure of Parenting Stress. The British Journal of Social Work, 49(1), 96-115. doi: 10.1093/bjsw/bcy021

Milfont, T. L., & Fischer, R. (2015).
Testing measurement invariance across groups: Applications in cross-cultural research. International Journal of psychological research, 3(1), 111-130.

Muris, P., Meesters, C., & van den Berg, F. (2003).
The Strengths and Difficulties Questionnaire (SDQ). European Child & Adolescent Psychiatry, 12(1), 1–8. doi:10.1007/s00787-003-0298-2

Nores, M., & Barnett, W. (2010)
Benefits of early childhood interventions across the world: (Under) Investing in the very young. Economics of Education Review, 29, pp. 271-82.

Pena, E. D. (2007).
Lost in translation: Methodological considerations in cross-cultural research. Child Development, 78, 1255-1264.

Peña, E.D., and R. Quinn (1997).
Task Familiarity: Effects on the Test Performance of Puerto Rican and African American Children. Language, Speech, and Hearing Services in Schools 28.4: 323-332.

Pisani, L., Borisova, I., & Dowd, A.J. (2015).
International Development and Early Learning Assessment Technical Working Paper Q4. Save The Children. Retrieved from: http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/IDELA%20TECHNICAL%20WORKING%20PAPER_V4.PDF

Price, L. R. (2017a).
Reliability. In Psychometric Methods: Theory into Practice (pp.203-255). United States of America: The Guilford Press.

Price, L. R. (2017b).
Criterion, Content, and Construct Validity. In Psychometric Methods: Theory into Practice (pp. 59-104). United States of America: The Guilford Press.

Platas, L. M., Ketterlin-Geller, L. R., & Sitabkhan, Y. (2016).
Using an assessment of early mathematical knowledge and skills to inform policy and practice: Examples from the Early Grade Mathematics Assessment. International Journal of Education in Mathematics, Science and Technology, 4(3), 163-173.

Rasch, G. (1961).
On general laws and the meaning of measurement in psychology. In: Neyman J, (ed.). Proceedings of the fourth Berkeley symposium on mathematical statistics and probability (pp. 321–44). Berkeley: University of California Press.

Ricker, K. L. (2006).
Setting cut-scores: A critical review of the Angoff and modified Angoff methods. Alberta journal of educational research, 52(1), 53-64.

Ross, K.N. (1987).
Sample design. International Journal of Educational Research, 11(1), 57-75.

Ross, K.N. (1991).
Sampling Manual for the IEA International Study of Reading Literacy. Hamburg: IEA International Study of Reading Literacy Coordinating Centre.

SA Department of Education. (1998).
National norms and standards for school funding. (Government Gazette Vol.400 No.19347 12 October 1998). Retrieved from: http://www.education.gov.za/LinkClick.aspx?fileticket=ZYYtOiXHTeE%3D&tabid=188&mid=498

REFERENCES

Scott-Little, C., Kagan, S. L., & Frelow, V. S. (2006).
Conceptualization of readiness and the content of early learning standards: The intersection of policy and research. Early Childhood Research Quarterly, 21(2), 153-173 doi: 10.1016/j.ecresq.2006.04.003

Sireci, S.G. & Rios, J.A. (2013)
Decisions that make a difference in detecting differential item functioning, Educational Research and Evaluation, 19:2-3, 170-187, DOI:10.1080/13803611.2013.767621

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

SurveyCTO. (2016).
Retrieved from: http://www.surveycto. com/about.html

Swank, J., & Mullen, P. (2017).
Evaluating evidence for conceptually related constructs using bivariate correlations. Measurement and Evaluation in Counseling and Development, 50(4), 270–274. doi: 10.1080/07481756.2017.1339562

Tredoux, C., Dawes, A., Mattes, F. et al. (2024).
Are South African children on track for early learning? Findings from the South African Thrive By Five Index 2021 Survey. Child Indicators Research. 17, 601–636. https://doi.org/10.1007/s12187-023-10093-3

Snelling et al. (2019). Dawes et al. (2020).
Rasch analysis and DIF: English, Afrikaans, isiXhosa, isiZulu and Setswana.

Van de Vijver, F., & Tanzer, N. K. (2004).
Bias and equivalence in cross-cultural assessment: An overview. European Review of Applied Psychology, 54(2), 119-135.

Van der Merwe, A. & Dawes, A. (2000).
Prosocial and antisocial tendencies in children exposed to community violence. Southern African Journal of Child and Adolescent Mental Health, 12(1), 19-37.

Wechsler, D. (2012).
Wechsler preschool and primary scale of intelligence—fourth edition. San Antonio, TX: The Psychological Corporation. Retrieved https://criv. pw/vwo5.pdf

Young, A., Storbeck, C., Rogers, K., Swannack, R, Dawes, A. & Girdwood, E. (In Press)
Adaptation and pilot testing of the Early Learning Outcome Measure (ELOM) (4&5) Years Assessment tool in South African Sign Language. PLOS Global Public Health

Zieky, M. & Perie, M. A (no Date).
A Primer on Setting Cut Scores on Tests of Educational Achievement. ETS. Retrieved from: https://www.ets.org/Media/Research/ pdf/Cut_Scores_Primer.pdf

Zillmer, E. A., Spiers, M. V., & Culbertson, W. C. (2008).
Principles of neuropsychology. (2nd edition. Belmont, CA: Thomson Higher Education.

APPENDIX 1 NORMS FOR 50-59 AND 60-69 MONTHS (2016 STANDARDISATION SAMPLE)

Table 28: ELOM 4&5 Standards and Performance Bands for children aged 50-59 and 60-69 months (2016 standardisation sample)

ELOM 4&5	50 – 59 Months {N = 258; Mean age = 56,83 (SD=2.5)}			60 – 69 Months {N = 1073; Mean age = 64,98 (SD=2.25)}		
	Falling Far Behind	Falling Behind	On Track	Falling Far Behind	Falling Behind	On Track
Total	0 – 36.01	36.02 – 46.31	46.32 – 100	0 – 43.23	43.24 – 54.37	54.38 – 100
Gross Motor Development	0 – 5.40	5.41 – 8.59	8.60 – 20	0 – 7.21	7.22 – 10.53	10.54 – 20
Fine Motor Coordination and Visual Motor Integration	0 – 9.70	9.71 – 12.31	12.32 – 20	0 – 11.46	11.47 – 14.12	14.13 – 20
Emergent Numeracy and Mathematics	0 – 6.34	6.35 – 9.32	9.33 – 20	0 – 6.90	6.91 – 10.23	10.24 – 20
Cognition and Executive Functioning	0 – 4.07	4.08 – 7.16	7.17 – 20	0 – 5.84	5.85 – 9.26	9.27 – 20
Emergent Literacy and Language	0 – 6.53	6.54 – 10.25	10.26 – 20	0 – 7.97	7.98 – 11.64	11.65 – 20

The figures that follow present standard scores, Z-scores (normal distribution scores), and percentiles, for ELOM 4&5 Total and for each Domain. Performance bands are indicated with the colours used in Table 15. The median scores for each age group are plotted on each figure for comparison purposes. Medians are used to account for any skewness in the distributions. Use Table 28 to interpret the lines on each distribution.

Table 28: Distribution Interpretation Key

Standard	
Q4/5 (Median)	
Q2/3 (Median)	
Q1 (Median)	

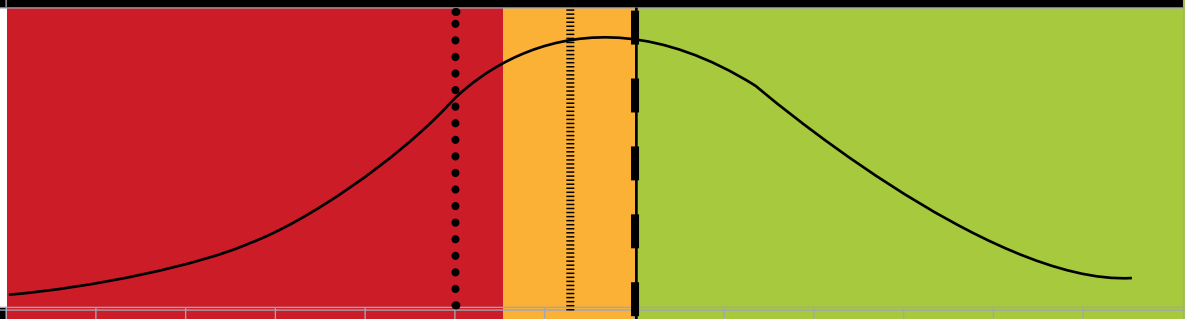
In each distribution:

- The **green** area represents the score range for children who are **On Track** (at or above the 60th percentile;
- The **yellow** area depicts those **Falling Behind** (between the 32nd and 59th percentile);
- The **red** area indicates the range of scores of children who are **Falling Far Behind** the expected standard (below the 32nd percentile).

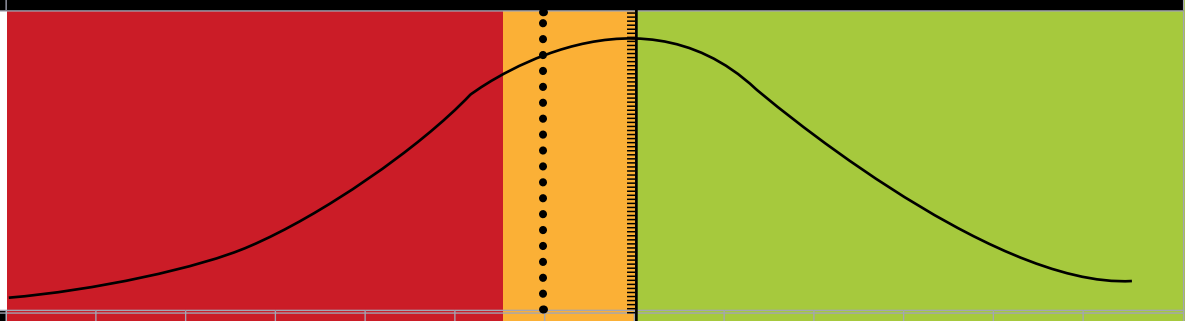
The figures also assist in visualising the positions of the school quintile groups in relation to one another on each domain and on the ELOM 4&5 Total score. They show how far programmes for children from different backgrounds would need to improve their performance if they are to reach the standard.

The standards may be revised as the ELOM 4&5 is used in the field and more data is gathered to assess the extent to which programmes are successful in assisting children to reach the expected level of performance.

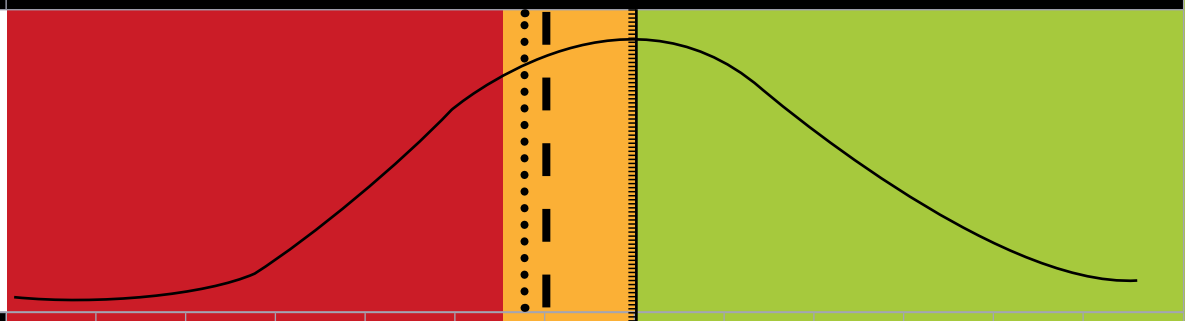
ELOM 4&5 Standard Score Distributions, for the 2016 Total Sample and for each Quintile.

60-69 Months	ELOM 4&5 TOTAL												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	9.14	16.10	23.06	30.02	36.98	43.94	**50.90**	57.86	64.82	71.78	78.74	85.70	92.66
Σ Q4/5	15.99	22.37	28.76	35.14	41.53	47.53	**54.30**	60.68	67.07	73.45	79.84	86.22	92.61
Σ Q2/3	7.15	14.27	21.40	28.52	35.65	42.77	**49.90**	57.02	64.15	71.27	78.40	85.52	92.65
Σ Q1	9.27	14.57	19.88	25.18	30.49	35.79	**41.10**	46.40	51.71	57.01	62.32	67.62	72.93.

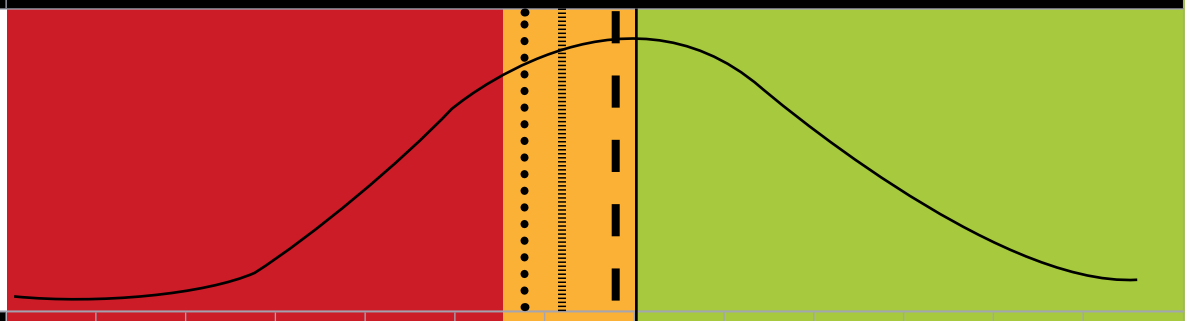
MEDIAN

60-69 Months	GROSS MOTOR DEVELOPMENT												
													
Z-Score	-1.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	1.20	3.27	5.35	7.42	**9.50**	11.57	13.65	15.72	17.80	19.87	-
Σ Q4/5	-	-	0.72	2.81	4.91	7.00	**9.10**	11.19	13.29	15.38	17.48	19.57	-
Σ Q2/3	-	-	1.62	3.69	5.76	7.83	**9.90**	11.97	14.04	16.11	18.18	-	-
Σ Q1	-	-	1.18	2.93	4.69	6.44	**8.20**	9.95	11.71	13.46	15.22	16.97	18.73

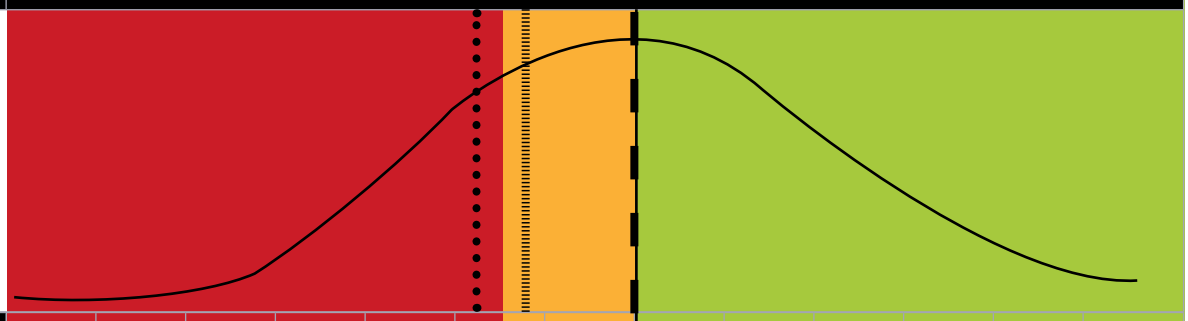
MEDIAN

60-69 Months	FINE MOTOR COORDINATION AND VISUAL MOTOR INTEGRATION												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	3.31	4.97	6.64	8.30	9.97	11.63	**13.30**	14.96	16.63	18.29	19.96	-	-
Σ Q4/5	4.59	6.17	7.76	9.34	10.93	12.51	**14.10**	15.68	17.27	18.85	-	-	.
Σ Q2/3	2.81	4.47	6.14	7.80	9.47	11.13	**12.80**	14.46	16.13	17.79	19.46	-	.
Σ Q1	2.99	4.57	6.16	7.74	9.33	10.91	**12.50**	14.08	15.67	17.25	18.84	17.39	.

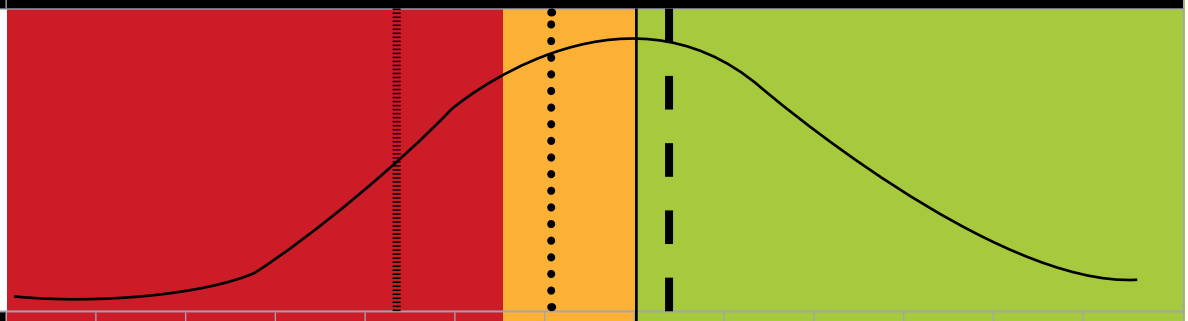
MEDIAN

60-69 Months	EMERGENT NUMERACY AND MATHEMATICS												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	0.86	2.94	5.03	7.11	**9.20**	11.28	13.37	15.45	17.54	19.62	-
Σ Q4/5	-	-	1.44	3.53	5.62	7.71	**9.80**	11.89	13.98	16.07	18.16	-	.
Σ Q2/3	-	-	0.62	2.71	4.81	6.90	**9.00**	11.09	13.19	15.28	17.38	19.47	.
Σ Q1	-	-	1.40	3.05	4.70	6.35	**8.00**	9.65	11.30	12.95	14.60	16.25	17.90

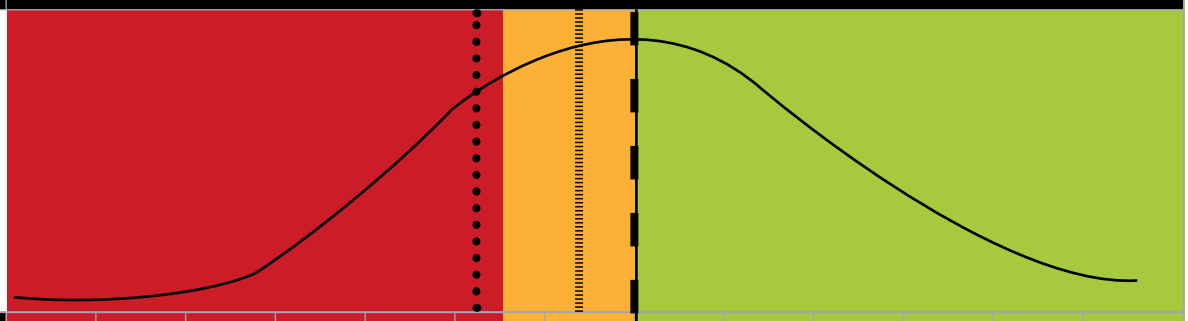
MEDIAN

60-69 Months	COGNITION AND EXECUTIVE FUNCTIONING												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	-	1.79	3.93	6.06	**8.20**	10.33	12.47	14.60	16.74	18.87	-
Σ Q4/5	-	-	0.68	2.83	4.99	7.14	**9.30**	11.45	13.61	15.76	17.92	-	.
Σ Q2/3	-	-	-	1.47	3.55	5.62	**7.70**	9.77	11.85	13.92	16.00	18.07	.
Σ Q1	-	-	-	0.42	2.25	4.07	**5.90**	7.72	9.55	11.37	13.20	15.02	16.85

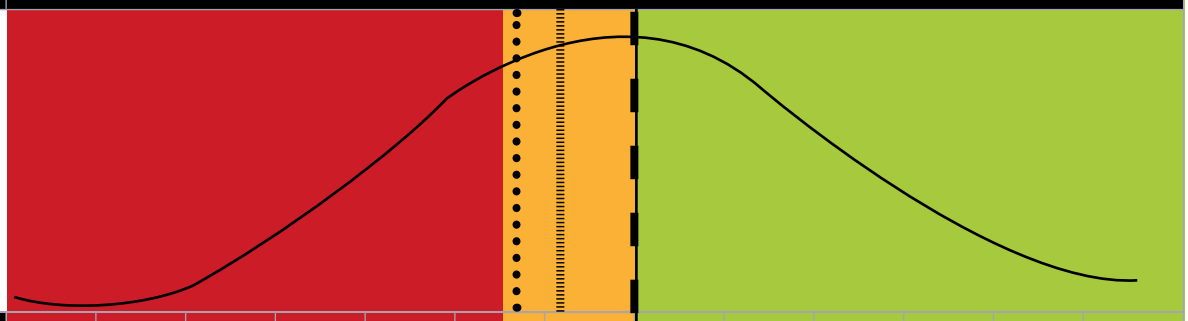
MEDIAN

60-69 Months	EMERGENT LITERACY AND LANGUAGE												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	1.32	3.61	5.91	8.20	**10.50**	12.79	15.09	17.38	19.68	-	-
Σ Q4/5	-	0.87	3.06	5.24	7.43	9.61	**11.80**	13.98	16.17	18.35	-	-	.
Σ Q2/3	-	-	1.18	3.43	5.69	7.94	**10.20**	12.45	14.71	16.96	19.22	-	.
Σ Q1	-	-	-	0.82	2.68	4.54	**6.40**	8.26	10.12	11.98	13.84	15.70	17.56

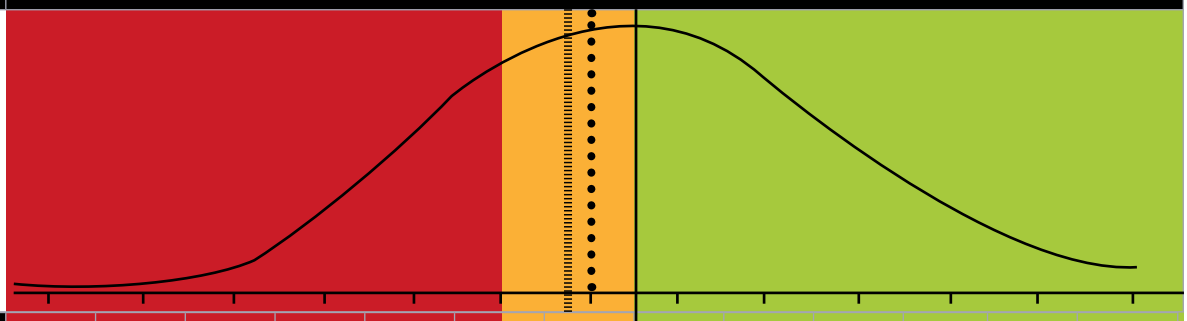
MEDIAN

50-59 Months	ELOM 4&5 TOTAL												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	4.49	10.93	17.36	23.80	30.23	36.67	**43.10**	49.54	55.97	62.41	68.84	75.28	81.71
Σ Q4/5	10.23	16.64	23.05	29.46	35.87	42.28	**48.69**	55.10	61.51	67.92	74.33	80.74	87.15
Σ Q2/3	5.48	11.48	17.47	23.47	29.46	35.46	**41.45**	47.45	53.44	59.44	65.43	71.43	77.42
Σ Q1	3.25	8.93	14.60	20.28	25.95	31.63	**37.30**	42.98	48.65	54.33	60.00	65.68	71.35

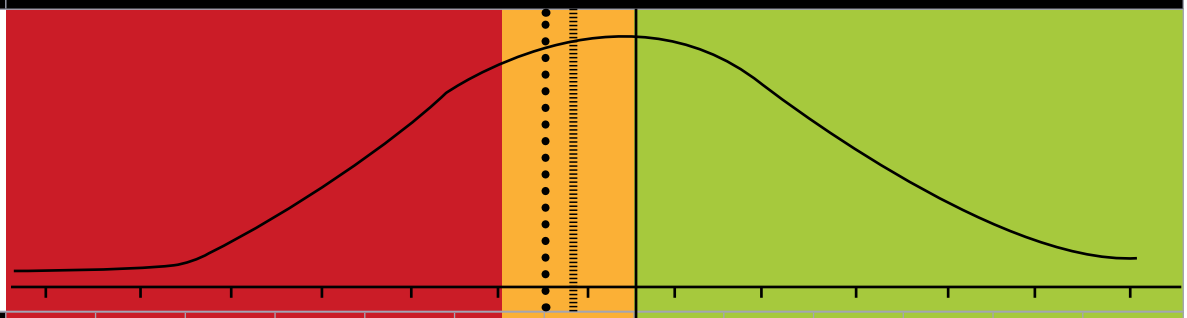
MEDIAN

50-59 Months	GROSS MOTOR DEVELOPMENT												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	-	1.62	3.61	5.60	**7.60**	9.59	11.59	13.58	15.58	17.57	19.57
Σ Q4/5	-	-	-	2.49	4.46	6.43	**8.40**	10.37	12.34	14.31	16.28	18.25	-
Σ Q2/3	-	-	-	1.41	3.44	5.47	**7.50**	9.53	11.56	13.59	15.62	17.65	19.68
Σ Q1	-	-	-	0.98	2.85	4.73	**6.60**	8.47	10.35	12.22	14.10	15.97	17.85

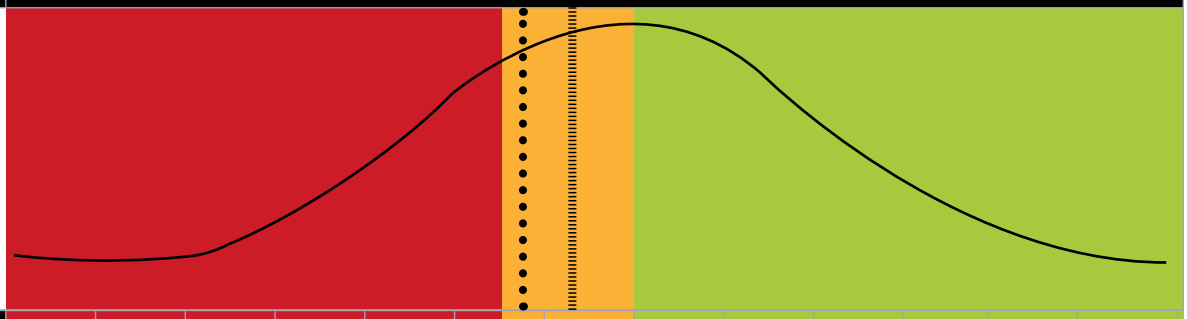
MEDIAN

50-59 Months	FINE MOTOR COORDINATION AND VISUAL MOTOR INTEGRATION												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	1.72	3.35	4.98	6.61	8.24	9.87	**11.50**	13.13	14.76	16.39	18.02	19.65	-
Σ Q4/5	1.35	3.07	4.80	6.52	8.25	9.97	**11.70**	13.42	15.15	16.87	18.60	-	-
Σ Q2/3	1.83	3.37	4.92	6.46	8.01	9.55	**11.10**	12.64	14.19	15.73	17.28	18.82	-
Σ Q1	2.12	3.75	5.38	7.01	8.64	10.27	**11.90**	13.53	15.16	16.79	18.42	-	-

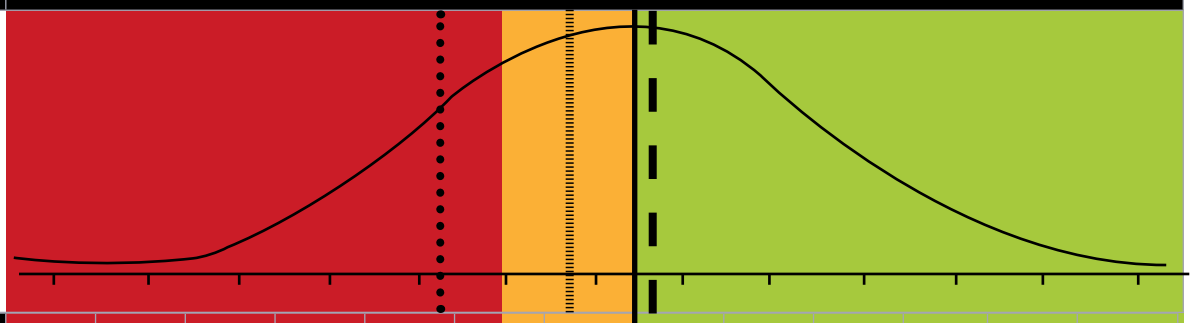
MEDIAN

50-59 Months	EMERGENT NUMERACY AND MATHEMATICS												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	0.94	2.80	4.67	6.53	**8.40**	9.59	11.59	13.58	15.58	17.57	19.57
Σ Q4/5	-	-	1.16	3.19	5.23	7.26	**9.30**	11.33	13.37	15.40	17.44	19.47	-
Σ Q2/3	-	-	0.86	2.64	4.43	6.21	**8.00**	9.78	11.57	13.35	15.14	16.92	18.71
Σ Q1	-	-	1.44	3.05	4.67	6.28	**7.90**	9.51	11.13	12.74	14.36	15.97	17.59

MEDIAN

50-59 Months	COGNITION AND EXECUTIVE FUNCTIONING												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	-	0.41	2.34	4.27	**6.20**	8.13	10.06	11.99	13.92	15.85	17.78
Σ Q4/5	-	-	-	1.91	3.91	5.90	**7.90**	9.89	11.89	13.88	15.88	17.87	19.87
Σ Q2/3	-	-	-	0.20	2.00	3.80	**5.60**	7.40	9.20	11.00	12.80	14.60	16.40
Σ Q1	-	-	-	-	1.59	3.19	**4.80**	6.40	8.01	9.61	11.22	12.82	14.43

MEDIAN

50-59 Months	EMERGENT LITERACY AND LANGUAGE												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	-	2.12	4.45	6.77	**9.10**	11.42	13.75	16.07	18.40	-	-
Σ Q4/5	-	0.14	2.37	4.59	6.82	9.04	**11.27**	13.49	15.72	17.94	-	-	-
Σ Q2/3	-	-	0.32	2.49	4.66	6.83	**9.00**	11.17	13.34	15.51	17.68	19.85	-
Σ Q1	-	-	-	0.47	2.28	4.09	**5.90**	7.71	9.52	11.33	13.14	14.95	16.76

MEDIAN

APPENDIX 2 ELOM 4&5 DOMAINS, STANDARDS, INDICATORS, ITEMS AND SOURCES

PHYSICAL DEVELOPMENT[17]
NCF[18]: Early Learning and Development Area (ELDA): Well-being.
Aim: Children should be physically strong and show ability and interest in physical activities.
Standard: SA-NELDS[19] Desired Result 6: Children begin to demonstrate physical and motor abilities and an understanding of a healthy lifestyle.

DOMAIN 1: GROSS MOTOR DEVELOPMENT (Direct Assessment)

Standard	Indicator	Assessment Items	Item Sources[20]
1.1: Children are increasingly able to use their large (gross) muscle skills.	Child shows good control and coordination in large movements.	**Direct Assessment: ELOM 4&5 Items 1, 2, 3, 4.** 1: Stand on one foot (1 point for 3-9 seconds; 2 points for 4-10 seconds). 2, 3, 4: Catch a beanbag thrown by examiner with: a) two hands against their body; b) preferred hand; c) other hand.	1: ASQ3 5**[21]. 2, 3, 4: McCarthy Scales South African adaptation**[22]. Others using same or similar methods: Lesotho[23] ELDS; CDAT[24]; EAP ECDS[25].

DOMAIN 2: FINE MOTOR DEVELOPMENT & VISUAL MOTOR INTEGRATION (VMI) (Direct Assessment)

Standard	Indicator	Assessment Items	Item Sources[20]
2.1: Children are increasingly able to use their small muscle (fine motor) skills.	Child uses small muscles competently.	**Direct Assessment ELOM Items: 5, 6, 7, 8.** 5 a) Uses sticks to copy a cross constructed with sticks; b) Uses sticks to copy a picture of a square. 6: Use a pencil to copy a triangle. 7: Use a pencil to draw self. 8: Strings beads in 40 seconds (total out of 10).	5: Beery VMI**[26] 6: Beery VMI**. Others using same or similar methods: IDELA[27]; EAP ECDS; ZAMCAT[28]. 7: This item draws on scoring of EAP ECDS and IDELA. Others using or similar methods: ASQ 3. 8: ZAMCAT**; Others using same or similar methods: ASQ 3; EAP ECDS.

[17]This document was prepared following psychometric analysis of the 2016 edition of the Age Validated ELOM 4&5.
[18]Physical Development includes both Gross and Fine Motor Coordination. Based on psychometric analysis, these are separate domains in the ELOM 4&5.
[19]Department of Basic Education. (2015). The South African National Curriculum Framework for children from birth to four. Pretoria: Department of Basic Education. Pretoria: Department of Basic Education.
[20]Department of Basic Education (2009). National Early Learning and Development Standards for children from birth to four years (NELDS). Pretoria: Department of Basic Education.
[21]"Item Sources" refers to instruments that use these or similar items; ELOM 4&5 items that are identical to the source are marked with two asterisks**.
[22]Squires, J. & Bricker, D. (2009). Ages & Stages Questionnaires: A parent-completed child monitoring system: Third Edition. Baltimore: Paul H. Brookes Publishing.
[23]Richter, L. M., Griesel, R. D., & Rose, C. B. (1994). The McCarthy scales of children's abilities: Adaptation and norms for use amongst Black South African children. The South African Journal of Occupational Therapy, 24(1), 17-30.
[24]Lesotho Ministry of Education and Training. (2014). Early learning and development standards for early childhood education and development in Lesotho, Revised Version. Maseru: Ministry of Education and Training.
[25]Rao, N. (2007). Cambodian Developmental Assessment Test. (UNICEF: Cambodia.CDAT). Rao, N., Sun, J., Ng, M., Becher, Y., Lee, D., Zhang, L. & Lau, C (2014) East-Asia Pacific Early Child Development Scales. Hong Kong: Faculty of Education, the University of Hong Kong.
[26]Beery, K. E., Buktenica, N. A. & Beery, N. A. (2010). Beery-Buktenica Developmental Test of Visual-Motor Integration. 6th Edition. Pearson.
[27]Pisani, L., Borisova, I., & Dowd, A.J. (2015). International Development and Early Learning Assessment Technical Working Paper Q4. Save The Children. Retrieved from: http://www.savethechildren.org/atf/cf/%7B9def2ebe-10ae-432c-9bd0-df91d2eba74a%7D/IDELA%20TECHNICAL%20WORKING%20PAPER_V4.PDF
[28]Fink, G., Matafwali, B., Moucheraud, C., & Zuilkowsk, S.S. (2012). The Zambian Early Childhood Development Project 2010 Assessment Final Report. Center on the Developing Child at Harvard University. Retrieved from http://www.poverty-action.org/study/zambian-early-childhood-development-project.

DOMAIN 3: EMERGENT NUMERACY AND MATHEMATICS (Direct Assessment)			
NCF: Early Learning and Development Area (ELDA): Exploring Mathematics Aim: Children show awareness of and are responsive to number and counting Children sort, classify, make comparisons and solve problems **NELDS Desired Result 5: Children are learning about mathematical concepts.**			
Standard	**Indicator**	**Assessment Items**	**Item Sources**
3.1: Children demonstrate an understanding of number concepts.	Child can count with one-to-one correspondence.	**Direct Assessment ELOM 4&5 Item 9:** Counting in Classes: Counts 3 marbles; Counts 8 buttons; Counts 15 objects from a mixed presentation of marbles, buttons and small sticks.	ELOM 4&5 item is a modification of IDELA 1:1 Correspondence item using three classes of object. Others using similar methods: MELQO[29]; EAP ECDS.
	Child can do simple calculations using addition and subtraction.	**Direct Assessment: ELOM 4&5 Item 10:** Child adds items using a picture card stimulus. Child subtracts items using a picture card stimulus.	PreGypt Battery**[30]; Others using similar methods: IDELA; EAP ECDS; CDAT.
3.2: Children begin to demonstrate an understanding of symbols, shapes, size and space.	Child can classify and match objects.	**Direct Assessment ELOM 4&5 Item 11:** Child groups stars and circles by colour and shape.	IDELA.**
	Child understands measurement terms to do with size and length, amount.	**Direct Assessment ELOM 4&5 Item 12:** From a picture: child identifies objects in a picture that are: above, under, in front of, on the side.	IDELA and MELQO.**
	Child can understand vocabulary for location.	**Direct Assessment ELOM 4&5 Item 13:** From picture stimuli: child identifies biggest, smallest, longest, shortest from pictures.	IDELA and MELQO.**

[29]Measuring Early Learning Quality and Outcomes (MELQO) http://steinhardt.nyu.edu/global-ties/early_childhood/melqo
[30]Egyptian adaptation of the Herbst Early Childhood Development Criteria Test provided by Dr Herbst.

DOMAIN 4: COGNITION & EXECUTIVE FUNCTIONING[31] (Direct Assessment)			
Early Learning and Development Area 1: Knowledge and understanding of the world. Aims: Explore and investigate their world. Explore design, make items and use technology. Explore and investigate time and place. **NELDS Desired Result 1: Children are learning how to think critically, solve problems and form concepts.**			
Standard	**Indicator**	**Assessment Items**	**Item Sources**
4.1: Children are learning how to think critically, solve problems and form concepts. 4.2: Children are developing the ability to attend to instructions, remember them and control impulses while performing a task.	Child demonstrates cognitive flexibility and working memory.	**Direct Assessment ELOM 4&5 Item: 14:** Child sorts 6 cards according to: 1) colour; 2) shape.	Dimensional Change Card Sort 1) Colour Game and 2) Shape Game DCCS** Zelatso[32]. Cards were changed from the original to a blue dog and a red car out of concern that rural children might not be familiar with boats and rabbits. Administration follows that laid down by Zelatso. Also used in EAP ECDS.
	Child demonstrates auditory discrimination, working memory and behavioural inhibition.	**Direct Assessment ELOM 4&5 Item 15:** Pencil Tapping Task: Child copies the examiner's exact sequence of taps on the table with a pencil.	ZAMCAT**; Brooker, Okello et al. 2010[33]; Others using similar methods EAP ECDS; Diamond & Taylor[34].
	Child demonstrates short-term memory.	**Direct Assessment ELOM 4&5 Item 16:** Digit Span (forward).	IDELA**. Instructions based on the Children's Memory Scale.[35] Others using same or similar methods: MELQO and CDAT.
	Child demonstrates problem-solving ability and working memory.	**Direct Assessment ELOM 4&5 Item 17:** Child assembles seven puzzles of increasing levels of difficulty.	Herbst**[36]. Others use one puzzle: IDELA; MELQO.

[31]Carlson, S. M. (2005). Developmentally sensitive measures of executive function in preschool children. Developmental Neuropsychology, 28(2), 595-616. Neuenschwander, R., Röthlisberger, M., Cimeli, P., & Roebers, C. M. (2012). How do different aspects of self-regulation predict successful adaptation to school?. Journal of Experimental Child Psychology, 113(3), 353-371.
[32]Zelazo, D.P. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. Nature. Protocols 1, 297–301.
[33]Fink et al see no 16 Zambian Child Assessment Test; Brooker, S., Okello, G., Njagi, K., et al. (2010). Improving educational achievement and anaemia of school children: design of a cluster randomised trial of school-based malaria prevention and enhanced literacy instruction in Kenya.Trials,11, 93. http://dx.doi.org/10.1186/1745-6215-11-93
[34]Diamond, A. and Taylor, C. (1996) Development of an aspect of executive control: Development of the abilities to remember what I said and to "Do as I say , not as I do". Developmental Psychobiology 29 (4): 315 – 334.
[35]Cohen, M. J. (2011). Children's Memory Scale. In Encyclopedia of Clinical Neuropsychology (pp. 556-559). Springer New York.
[36]Herbst Early Childhood Development Criteria Test. Herbst, I., & Huysamen, G. K. (2000). The construction and validation of developmental scales for environmentally disadvantaged preschool children. South African Journal of Psychology, 30(3), 19-26.

DOMAIN 5: EMERGENT LITERACY & LANGUAGE (Direct Assessment)			
National Curriculum Framework: Early Learning and Development Area: Communication. Aims:Listen to sounds and speech Listen with understanding Speak using different styles of communication Make meaning by 'reading' what they see, hear, feel, taste and touch **NELDS: Desired result 4: Children are learning to communicate effectively and use language**			
Standard	**Indicator**	**Assessment Items**	**Item Sources**
5.1: Children are able to com municate effectively and use language.	Child can speak in full sentences. Child can relate an account of events that is logical, and with correct language usage.	**Direct Assessment ELOM 4&5 Item 18:** Ability to talk about empathic response: In response to a stimulus picture of a girl crying, the child is asked to describe her feelings and actions to be taken to "help her feel better."	IDELA (with modified picture)**; Others using similar methods MELQO and EAP ECDS. Prior to age validation, items 18 and 19 were intended to measure empathic response and awareness of own feelings in the Social and Emotional Awareness domain. On factor analysis these items loaded with language items indicating strong reliance on expressive language.
		Direct Assessment ELOM 4&5 Item 19: Ability to talk about own emotions (self-awareness): The child is asked to describe a) what makes her/him feel sad and what can be done to feel better; b) what makes her/him feel happy.	
		Direct Assessment ELOM 4&5 Item 20: Child can describe what they do when they get up in the morning.	Items 20 and 21 were adapted for ELOM 4&5 following the pilot of IDELA and MELQO as these did not perform well. These items measure the same constructs and use the same format.
	Child can name common objects.	**Direct Assessment ELOM 4&5 Item 21:** Child is asked to name items to be seen inside and outside at home.	
	Child shows understanding of stories told to her/him.	**Direct Assessment ELOM 4&5 Item 22:** Assessor reads a story, after which questions are asked of the child to gauge understanding.	MELQO and IDELA.**
	Child recognises initial sounds of words.	**Direct Assessment ELOM 4&5 Item 23:** Initial Sound Discrimination task: Child is asked to state which word commences with a particular phoneme.	MELQO** IDELA; Others using or similar methods: EAP ECDS.

TASK ORIENTATION (Direct Assessment)			
NCF Early Learning and Development Area: Creativity. 1. Identify, search for and create solutions to challenges through problem solving **NELDS: Desired Result: Children are learning how to think critically, solve problems and form concepts**			
Standard	**Indicator**	**Assessment Items**	**Item Sources**
1: Children are able to persist with attention to accomplish a given task.	Child pays attention to instructions and requirements for a task. Child stays concentrated on activity and is not easily distracted. Child is motivated to complete tasks..	**Direct Assessment ELOM 4&5 Items:** Assessor observa tion of the child during assessment. 1: Did the child pay attention to the instruc tions and demonstra tions throughout the assessment? 2: Did the child stay concentrated and on task during the activities and was not easily distracted? 3: Was the child careful and diligent on tasks? Was child interested in accuracy? 4: Was the child interested and curious about the tasks throughout the assessment?	IDELA and ZAMCAT** Note that following psychometric analysis items were excluded to improve measurement of this construct.

SOCIAL-EMOTIONAL FUNCTIONING RATING SCALE			
1: SELF-CARE AND SOCIAL RELATIONS WITH PEERS AND ADULTS NCF: Early Learning and Development Area: Identity and Belonging. NELDS: Desired Results 6: Children begin to demonstrate physical and motor abilities and an understanding of a healthy lifestyle **NELDS Desired Results 3: Children demonstrate growing awareness of diversity and the need to respect and care for others**			
Standard	**Indicator**	**Assessment Items**	**Item Sources**
1.1: Children are competent in self- care.	Child can use the toilet without assistance.	**Social-Emotional Functioning Rating Scale of the child's self-care:** Can this child use the toilet on her/his own?	Constructed for ELOM 4&5
1.2: Children play co-operatively, taking turns with others.	Child cooperates and works well with peers – waits for turn, is thoughtful of others' feelings. Child resolves problems with peers without becoming aggressive, negotiates sharing.	**Social-Emotional Functioning Rating Scale of the child's relationships with peers:** 1: Works well with peers (can wait for their turn/manage impulsivity). 2: Resolves problems with peers without becoming aggressive. 3: Cooperates with peers without prompting.	Child Trends **[37] (Items presented here were finalised following Psychometry. (Relationships with peers and adults constitute a single scale).
1.3: Children are able to form relationships and interact appropriately. with adults.	Child seeks support/ assistance from familiar adults. Child solicits familiar adults' inputs about interesting experiences. Child initiates cooperative activities with familiar adults.	**Social-Emotional Functioning Rating Scale of the child's relationships with familiar adults:** 1 Child seeks assistance or support from familiar adults. 2: Child seeks a familiar adult's ideas or explanations about events or experiences that are interesting to the child. 3: Child takes initiative in creating cooperative activities with a familiar adult.	California ELDS- based Desired Results Profile** (2015)[38] Items presented here were finalised following Psychometry (Relationships with peers and adults constitute a single scale).
2: EMOTIONAL READINESS FOR SCHOOL Emotional dimensions associated with Approaches to Learning in a formal education setting NCF: Early Learning and Development Area: Identity and Belonging. NELDS: Desired Result 2: Children are becoming more aware of themselves as individuals, developing a positive self-image and learning how to manage their own behaviour.			
2.1: Children have age-appropriate emotional resources to manage formal learning environments.	Children have positive emotional functioning in areas relevant to formal learning.	**Social-Emotional Functioning Rating Scale of the child's emotional functionng in the early learning programme context.** Teacher rating of SACAS items.	South African Child Assessment Scales (SACAS)[39]: 6 items.

[37]Child Trends (2014). Measuring Elementary School students' social and emotional skills. Providing educators with tools to measure and monitor social and emotional skills that lead to academic success. http://www.childtrends.org/wp-content/uploads/2014/08/2014-37CombinedMeasuresApproachandTablepdf1.pdf.
[38]California Department of Education DRDP (2015) A Developmental Continuum from Early Infancy to Kindergarten Entry. Sacramento Calif: www.cde.ca.gov/sp/cd/ci/drdpforms.asp Accessed 03 August, 2015.
[39]Barbarin, O. (1998).The South African Child Assessment Scales (SACAS). Psychology Department, University of Michigan. Van der Merwe, A. & Dawes, A. (2000). Prosocial and antisocial tendencies in children exposed to community violence. Southern African Journal of Child and Adolescent Mental Health, 12(1), 19-37. The SACAS is based on the Child Behavior Checklist (CBCL).

APPENDIX 3 ELOM 4&5 PSYCHOMETRY AND STATISTICAL TABLES

Standard Error of Measurement (SEM) for ELOM 4&5
SEM was calculated using Cronbach's Alpha reliability for the total sample: = .80.

All children who displayed poor Rasch fit (misfit) for two or more domains (9.8% to 13.6% of sample) were excluded from this reliability estimate (see table 3). Poor fit was likely to have been due to their poor engagement in the assessment.

ELOM 4&5 standards are based on the scores obtained by children scoring in the top 40% of the distribution. To establish measurement error, standard deviations from mean standard scores were calculated for each age group. The resulting confidence intervals represent an adjustment of the standard deviation, based on Cronbach's Alpha. Confidence intervals for Total ELOM 4&5 standard scores and those for each domain are presented in Table A 4.1.

Table A 3.1: ELOM 4&5 Standards based on the Performance of the top 40% of Children with 95% Measurement Error

	50 – 59 Months			60 – 69 Months		
	Lo CI	Standard	Hi CI	Lo CI	Standard	Hi CI
ELOM 4&5 TOTAL	38.99	46.32	53.64	47.18	54.38	61.58
Gross Motor Development	6.20	8.60	10.99	8.50	10.54	12.58
Fine Motor Coordination & Visual Motor Integration	10.47	12.32	14.16	12.91	14.13	15.36
Emergent Numeracy & Mathematics	7.08	9.33	11.59	8.08	10.24	12.41
Cognition and Executive Functioning	4.93	7.17	9.39	7.09	9.27	11.44
Emergent Literacy & Language	7.92	10.26	12.61	9.80	11.65	13.50

Note: Lo CI = Lower boundary of the confidence interval; Hi CI = Upper boundary of the confidence interval. Confidence intervals should be interpreted in the context of Rasch weightings. The upper confidence interval should be viewed with caution. Due to the ability-score weighting produced by the Rasch analysis and the performance-dependent nature of measures of this type, it is implausible that children would perform better as a result of measurement error. Underperformance, as represented by the lower confidence interval, is more likely to occur in cases of poor assessment administration, and factors affecting the child during the assessment. These may include illness, fatigue and low engagement with the assessment process (Bond & Fox, 2015).

Multi-Level Modelling
For the purpose of these analyses, age was split into two categories – 50 to 59 months (<5 years), and 60 to

69 months (>5 years). As it had been established that there was no difference in the performance of children in quintiles 2 and 3, and 4 and 5 respectively, quintile groups were collapsed into three groupings – quintile 1, quintile 2/3, and quintile 4/5. Gender was added to the model to control its possible effects. To account for interaction between quintile and age, an interaction effect was added. The total ELOM 4&5 score and each domain score were modelled separately. Model parameters are presented in Tables 2a – 2f and represent the fixed effects of the multi-level model. The b parameters represent the difference between the Total Sample mean for the model, and that of the relevant group. Groups presented with a dash represent the reference category, and can be interpreted by inspecting the Total Sample b parameter. The Standard Error of the b parameter, the t statistic, the significance value (p), and 95% confidence intervals, are presented alongside the b parameter.

Table A 3.2: Multi-Level Model Parameters: ELOM 4&5 Total

	B	SE_B	T	P	95% CI	
Total sample	55.39	.77	71.75	<.001	53.88	56.91
Q1	-13.16	1.83	-7.18	<.001	-16.76	-9.56
Q2&3	-4.40	.87	-5.09	<.001	-6.10	-2.70
Q4&5	-	-	-	-	-	-
<5 Years	-5.50	1.56	-3.52	<.001	-8.56	-2.44
>5 Years	–	–	–	–	–	–
Male	-2.20	.73	-3.02	.003	-3.63	-.77
Female	–	–	–	–	–	–

Table A 3.3: Multi-Level Model Parameters: Gross Motor Development

	B	SE_B	T	P	95% CI	
Total sample	8.86	.24	37.28	<.001	8.40	9.33
Q1	-.93	.56	-1.65	.09	-2.04	.17
Q2&3	.83	.27	3.12	.002	.31	1.36
Q4&5	-	-	-	-	-	-
<5 Years	-.75	.48	-1.56	.119	-1.69	.19
>5 Years	-	-	-	-	-	-
Male	.56	.22	2.51	.012	.12	1.00
Female	–	–	–	–	–	–

Table A3.4: Multi-Level Model Parameters: Fine Motor Coordination and Visual Motor Integration

	B	SE_B	T	P	95% CI	
Total sample	14.52	.19	76.75	<.001	14.15	14.89
Q1	-1.69	.45	-3.76	<.001	-2.57	-.81
Q2&3	-1.31	.21	-6.19	<.001	-1.73	-.90
Q4&5	-	-	-	-	-	-
<5 Years	-2.37	.38	-6.20	<.001	-3.12	-1.62
>5 Years	-	-	-	-	-	-
Male	-.67	.18	-3.75	<.001	-1.02	-.32
Female	–	–	–	–	–	–

Table A3.5: Multi-Level Model Parameters: Emergent Numeracy and Mathematics

	B	SE_B	T	P	95% CI	
Total sample	10.01	.24	42.36	<.001	9.54	10.47
Q1	-1.76	.56	-3.15	.002	-2.87	-.66
Q2&3	-.82	.26	-3.09	.002	-1.34	-.30
Q4&5	–	–	–	–	–	–
<5 Years	-.51	.48	-1.06	.289	-1.44	.43
>5 Years	–	–	–	–	–	–
Male	-.32	.22	-1.45	.146	-.76	.11
Female	–	–	–	–	–	–

Table A3.6: Multi-Level Model Parameters: Cognition and Executive Functioning

	B	SE_B	T	P	95% CI	
Total sample	9.62	.23	40.56	<.001	9.16	10.09
Q1	-3.34	.56	-5.93	<.001	-4.45	-2.24
Q2&3	-1.51	.27	-5.68	<.001	-2.03	-.99
Q4&5	–	–	–	–	–	–
<5 Years	-1.36	.48	-2.84	.005	-2.30	-.42
>5 Years	–	–	–	–	–	–
Male	-.66	.22	-2.93	.003	-1.10	-.22
Female	–	–	–	–	–	–

Table A3.7: Multi-Level Model Parameters: Emergent Language and Literacy

	B	SE_B	T	P	95% CI	
Total sample	12.38	.25	48.88	<.001	11.89	12.88
Q1	-5.43	.60	-9.02	<.001	-6.61	-4.25
Q2&3	-1.59	.28	-5.60	<.001	-2.15	-1.04
Q4&5	–	–	–	–	–	–
<5 Years	-.51	.51	-1.00	.32	-1.51	.50
>5 Years	–	–	–	–	–	–
Male	-1.11	.24	-4.64	<.001	-1.58	-.64
Female	–	–	–	–	–	–

Rasch modelling: Person Misfit
Numbers and proportions of children in the sample excluded from estimates provided for each domain are presented in Table A3.8.

Table A3.8

DOMAIN	Number of Children with Poor Model Fit (%)
GROSS MOTOR DEVELOPMENT	n = 174 (13.1%)
FINE MOTOR COORDINATION & VISUAL MOTOR INTEGRATION	n = 182 (13.6%)
EMERGENT NUMERACY & MATHEMATICS	n = 170 (12.8%)
COGNITION & EXECUTIVE FUNCTIONING	n = 130 (9.8%)
EMERGENT LITERACY & LANGUAGE	n = 132 (9.9%)

Confirmation of Unidimensionality and Internal Consistency.
Factor Loadings for CFA on transformed ELOM 4&5 scores are presented in Table A4.9.

Table A3.9: Domain Factor Loadings

DOMAIN	FACTOR LOADING	ITEM NUMBER	DESCRIPTION
GROSS MOTOR DEVELOPMENT	.235	1	Standing on one leg for 10 seconds.
	.507	2	Catch bean bag both hands.
	.707	3	Catch bean bag preferred hand.
	.584	4	Catch bean bag non-preferred hand.
FINE MOTOR COORDINATION & VISUAL MOTOR INTEGRATION	.463	5	Copy cross and square.
	.450	6	Copy triangle.
	.581	7	Draw self.
	.346	8	String beads.
EMERGENT NUMERACY & MATHEMATICS	.664	9	Counting in classes.
	.638	10	Addition and subtraction.
	.327	11	Sorting and classification.
	.470	12	Spatial vocabulary.
		13	Measurement vocabulary.
COGNITION & EXECUTIVE FUNCTIONING	.345	14	Dimensional Change Card Sort.
	.735	15	Pencil tapping test.
	.333	16	Digits forward.
	.424	17	Picture puzzle completion.
EMERGENT LITERACY & LANGUAGE	.588	18	Expressive language: empathic response to distress.
		19	Expressive language: self-awareness.
	.727	20	Expressive language: describes getting up in the morning.
	.587	21	Expressive vocabulary: names familiar objects.
	.469	22	Oral comprehension: cat and mouse story.
		23	Initial sound discrimination.

Item difficulty

Table A3.11 presents the Rasch Difficulty Logit for each ELOM 4&5 item. This logit represents the probability of achieving each possible score on an item.

For example, item 1 requires children to stand on one leg for up to 10 seconds. It is more difficult to stand on one leg for longer periods of time. The -1.43 logit seen below for the item-1 score of 1, indicates that children with an average ability level have around an 85% chance of standing on one leg for 3 to 9 seconds. The .85 logit associated with a score of 2 suggests that children with an average ability level have around a 35% chance of standing on one leg for 10 seconds or more. Each domain is designed to have scores that measure a variety of difficulty levels.

A simple way of estimating the difficulty of an item based on logit values is shown in Table A4.10 on the right. For a child of average ability (around the 50th percentile), a logit of:

Table A3.10: Probabilities of test taker achieving each Logit value.

Logit:	-3.00	probability of achieving this score:	>95%
Logit:	-2.00	probability of achieving this score:	90%
Logit:	-1.00	probability of achieving this score:	75%
Logit:	0.00	probability of achieving this score:	50%
Logit:	+1.00	probability of achieving this score:	25%
Logit:	+2.00	probability of achieving this score:	10%
Logit:	+3.00	probability of achieving this score:	<5%

Table A3.11: Rasch Difficulty per Item

ITEM NUMBER	DESCRIPTION	ITEM SCORING (UN-TRANSFORMED)	ITEM SCORING (TRANSFORMED (LOGITS))
1	Standing on one leg for 10 seconds.	0	-
		1	-1.43
		2	.85
2	Catch the bean bag with both hands.	0	-
		1	-1.96
		2	-.38
		3	1.86
3	Catch the bean bag with the preferred hand.	0	-
		1	-.16
		2	2.00
		3	5.84
4	Catch the bean bag with the non-preferred hand.	0	-
		1	.42
		2	2.83
		3	7.18
5	Copy cross and square.	0	-
		1	-.42
		2	1.94
6	Copy triangle.	0	-
		1	
		2	2.75

ITEM NUMBER	DESCRIPTION	ITEM SCORING (UN-TRANSFORMED)	ITEM SCORING (TRANSFORMED)
7	Draw self.	0	-
		1	
		2	
		3	-.41
		4	
		5	1.00
		6	
		7	3.10
		8	
8	String beads.	0	-
		1	-4.99
		2	-2.77
		3	-1.21
		4	.22
		5	1.40
		6	2.60
		7	4.37
		8	6.85
		9	
		10	
9	Counting in classes.	0	-
		1	.44
		2	1.92
		3	4.29
10	Addition and subtraction.	0	-
		1	.48
		2	2.47
11	Sorting and classification.	0	-
		1	.04
		2	1.88
12	Spatial vocabulary.	0	-
		1	
		2	-1.27
		3	
		4	.43
13	Measurement vocabulary.	0	-
		1	
		2	-1.72
		3	
		4	.25
14	Dimensional Change Card Sort.	0	-
		1	-.64
		2	.59

ITEM NUMBER	DESCRIPTION	ITEM SCORING (UN-TRANSFORMED)	ITEM SCORING (TRANSFORMED)
15	Pencil tapping test.	0	-
		1	-.44
		2	
		3	
		4	.34
		5	
		6	
		7	
		8	1.33
		9	
		10	2.70
16	Digits forward.	0	-
		1	
		2	
		3	.01
		4	1.13
17	Picture puzzle completion.	0	-
		1	-.49
		2	.54
		3	
		4	
		5	2.36
		6	
		7	
18	Empathy.	0	-
		1	
		2	
		3	1.39
19	Self-awareness.	0	-
		1	
		2	.50
		3	
		4	1.73
20	Expressive language.	0	-
		1	-2.09
		2	
		3	.23
		4	
		5	
		6	1.71

ITEM NUMBER	DESCRIPTION	ITEM SCORING (UN-TRANSFORMED)	ITEM SCORING (TRANSFORMED)
21	Expressive vocabulary.	0	-
		1	
		2	-1.70
		3	
		4	
		5	.11
		6	
		7	
		8	
		9	1.33
		10	
22	Oral comprehension.	0	-
		1	
		2	-.22
		3	
		4	1.20
		5	
23	Initial sound discrimination.	0	-
		1	1.01
		2	
		3	2.17