Early Learning Outcomes Measure (ELOM) Technical Manual

Second Edition 2019
Developed on behalf of the Innovation Edge by
Andrew Dawes, Linda Biersteker, Elizabeth Girdwood,
Matthew Snelling and Colin Tedoux

This Technical Manual accompanies the ELOM Direct Assessment Manual

To Cite this Manual: Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L. & Tredoux, C.G. (2019). Early Learning Outcomes Measure. Technical Manual. Claremont, Cape Town: The Innovation Edge. http://www.innovationedge.org.za/

Please refer to **www.elom.org.za** for additional information.

Contents

APPENDICES 2
ACRONYMS 3
ACKNOWLEDGEMENTS 4

CHAPTER 1 USING THE ELOM 8

1.1 ELOM components 8
1. 2 What the age validated, standardised ELOM measures 8

CHAPTER 2 THE DEVELOPMENT OF THE ELOM 6

2.1 Development of content-validated ELDS for the ELOM 7
2.2 Selection of ELDS domains 7
2.3 Selection of items for measuring children's performance for ELOM Direct Assessment 7
2.4 Establishment of cultural fairness and translation of Direct Assessment items 8
2.5 ELOM Teacher Assessment 9
2.6 The ELOM pilot 11

CHAPTER 3 ELOM STANDARDISATION SAMPLE 12

3.1 Rationale for sample inclusion of five quintiles and five languages 12
3.2 Power calculations 13
3.3 Construction of sampling frames 13
3.4 Assumptions and inputs for a two-stage clustered sample design 15
3.5 Selection of pupils within schools 16
3.6 Ethics 16

CHAPTER 4 ASSESSOR TRAINING FOR THE AGE VALIDATION STUDY 17

4.1 Process 17
4.2 Inter – assessor agreement 18

CHAPTER 5 PSYCHOMETRY AND STATISTICAL ANALYSES 18

5.1 Tablet scoring and data collection 18
5.2 Preparation of data for analysis 19
5.3 Psychometric methods 20

CHAPTER 6 CONSTRUCTION OF ELDS AND ELOM NORMS 26

REFERENCES 40

APPENDICES

Appendix 1	ELOM domains, standards, indicators, items and sources
Appendix 2	ELOM psychometry and statistics

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
CDAT	Cambodian Developmental Assessment Test
CFA	Confirmatory Factor Analysis
CMS	Children's Memory Scale
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	Early Learning Outcomes Measure
EN&M	Emergent Numeracy & Mathematics
FMC&VMI	Fine Motor Control & 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
SES	Socioeconomic Status
VMI	Visual Motor Integration
WC	Western Cape
WHO	World Health Organisation
ZAMCAT	Zambian Child Assessment Test

ACKNOWLEDGEMENTS

The ELOM Team are most grateful to the leadership of Innovation Edge and its Investment Committee for commissioning this important process.

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

Child Assessment Teams

The Provincial Age Validation Child Assessment Teams did an excellent job in a challenging timeframe and often under difficult conditions.

Western Cape: Elizabeth Girdwood (Field Manager); Inge Sonn (Senior Assessor and Trainer); Assessors: Gene August, Ilse Ahrends, Linda Dlangamandla, Nomathemba Peter, Mphokazi Malangeni, Sivenesy Subramoney, Lumka Ndamase, Rochelle Fleishman.

North West Province: Marinel Rothman (Field Manager); Marieta Visser (Senior Assessor and Trainer); Assessors: Ntebo Dikwidi, Nontombi Duma, Matshidiso Mokgosi, Lenah Ramoleta, Kelebogile Dithejane, Grade Thakadu, Agnes Moleme, Annakie Sehenye.

KwaZulu-Natal: Pumla Mkiva (Field Manager); Dr Jane Kvalsvig (Senior Assessor and Trainer); Assessors: Mavis Buthelezi, Nondumiso Luthuli, Thamsanqa Mkhize, Ntombizonke Mncwabe, Philisiwe Ngcobo, Hlengiwe Sibisi, Sinenhlanhla Sibisi, Nozipho Sibiya, Sindi Xulu.

Participating Schools

We are particularly grateful to the many schools that agreed to participate and to the children for making the study possible. For ethical reasons they remain anonymous. Preschools used in the ELOM pilot process and for Assessor Training included: Milkwood Pre-primary, Fish Hoek; Green Curtains, Ocean View; Masakhane Educare, Masiphumelele; The Marian Institute, Bridgetown; Silindokuhle and Sicelimfundo, Mayville, Durban; Tshepo Primary School, Ikageng, Potchefstroom.

Education stakeholders were consulted on standards and domains to include in the ELOM prior to its construction, through focus groups and interviews.

Reference group members and advisors played a key role in providing guidance on ELOM 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 contribution to the Development of the ELOM:
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 and 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) on 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.

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 components

The ELOM includes the following components

1. ELOM Direct Assessment Kit
2. ELOM Direct Assessment Kit list
3. ELOM Direct Assessment Manual: Afrikaans, English; Setswana, isiXhosa and isiZulu (other languages are added when they become available)
4. ELOM Technical Manual
5. ELOM Scoring Sheet
6. Open source online protocol for tablet/phone-based scoring and data capture
7. ELOM Teacher Assessment of the Child

Equipment required for ELOM administration is provided in the ELOM Direct Assessment Kit list.

Apart from the Kit, all other components will be available for download from the Innovation Edge ELOM website from January 2017.

Users must be registered to download these components (see the website for details).

All ELOM assessors must be trained and certified as specified by The Innovation Edge (see ELOM website). Untrained persons may not use the ELOM.

This Technical Manual provides information on the development of the ELOM 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 or android mobile phones should preferably be used for scoring the Direct Assessment as this significantly reduces errors and permits scores to be uploaded automatically to a database. Software is available from Innovation Edge ELOM website. In the absence of these devices, scores may be recorded on paper using the ELOM Scoring Sheet.

The ELOM Teacher Assessment of the Child complements Direct Assessment 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 measures

Prior to the development of the ELOM, there was no validated South African instrument for measuring programme performance against Early Learning Development Standards. The ELOM is an age-normed, standardised instrument for use with children in two age groups: **50-59 months and 60-69 months.** The division into younger and older age groups addresses the need for an instrument that takes into account the different levels of development expected of the older and younger children, and enables comparison with the expected performance of age peers.

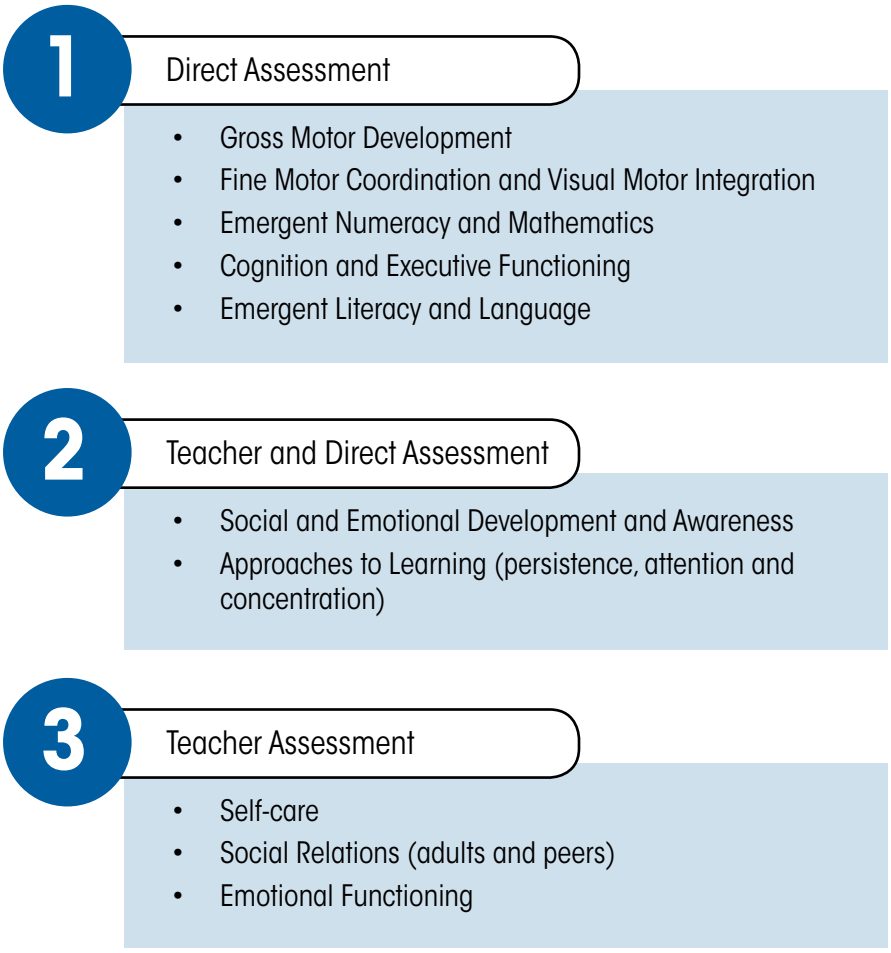
The ELOM includes both direct assessment of children's performance as well as an assessment of the child's social and emotional functioning and orientation to tasks. Socio-emotional functioning in particular, is best assessed by a person who knows the child well as evaluation of these aspects of development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

ELOM Direct Assessment 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
6. The child's Task Orientation (aspects of Approach to Learning in Figure 1) is rated by the assessor during direct assessment. The child's capacity for self-care (toileting), relations with peers and adults and emotional functioning are measured in the Teacher Assessment (see below).

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.

Figure 1: What the Age Validated Standardised ELOM Measures

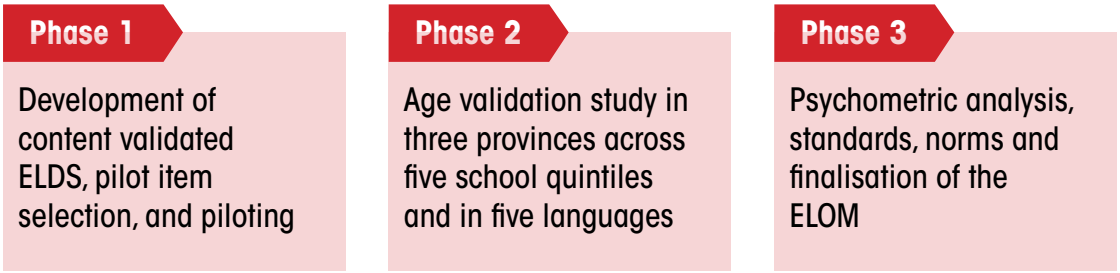


ELOM domains, standards, indicators, items and sources are included in Appendix 1. Direct Assessment Manuals for age validation were translated into Setswana, isiZulu, Afrikaans and isiXhosa using accepted procedures to ensure linguistic and metric equivalence (Chapter 5).

CHAPTER 2 THE DEVELOPMENT OF THE ELOM | Andrew Dawes and Linda Biersteker

The ELOM was developed in three phases as shown in Figure 2.

Figure 2: The Development of the ELOM



2.1 Phase 1: Development of Content Validated ELDS for the ELOM

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 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 drew on existing instruments that were suitable for measuring the chosen indicators and which are applicable in a diverse cultural and developmental affordance 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 need to display on entering Grade R in order to fully benefit. They contributed to the selection of ELOM domains.

2.2 Selection of ELDS Domains

Literature and policy documents on ELDS were surveyed in order 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) as well as other public policies and guidelines, were also primary sources. As the ELOM is designed to assess children at the end of the year prior to Grade R, the design also took into account progression to the South African Curriculum Assessment Policy Statements (CAPS) for Grade R.

Domains selected for the ELOM do not cover all those specified in South African ELDS. Only those deemed to be 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 the development of standards is regarded as essential (e.g. Zieky, and Perie, No Date; Kagan, Castillo, Gomez, & Gowani, 2013). The criteria for inclusion of an ELOM 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, several standards were specified per domain and several indicators per standard (See Appendix 1).

2.3 Selection of Items for Measuring Children's Performance for ELOM Direct Assessment

Once domains had been decided, a scan of available instruments and test items used in similar exercises was undertaken. Criteria for inclusion of ELOM 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 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 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 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 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 draws on commonly used assessments of EF. Task orientation and approach to learning was assessed during the administration of the ELOM (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 need to be carefully constructed. Translations from the original English must follow accepted practice. English language instructions to the assessor on ELOM 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 on the basis of commonly used and understood terms. We also had advice from a local project specialising in early literacy, as well as from early language development specialists.

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

1. **Cultural fairness:** In developing the ELOM items we took care to ensure 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 (Rasch and Differential Item Functioning (DIF) analysis was undertaken on the Age Validation sample).
2. **Functional equivalence:** In the ELOM 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 is 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? It is important that the instructions for tasks are 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 used 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, then this is not regarded as an issue, and there are examples of these in the ELOM. For the Pilot, this was considered carefully by the team, in discussion with other experts. Experience during the Pilot phase allowed for observations of consistently problematic items and items 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 the other 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 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 as far as 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. DIF analyses as described in Chapter 5 were undertaken to check for items that discriminated unfairly between children from different backgrounds but of the same ability.

2.5 ELOM Teacher Assessment

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 Direct Assessment (see ELOM Teacher Assessment of the Child). 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 Teacher Assessment could be adapted for parents' reports for use in home visiting programmes, and for playgroup leaders. Norms are not set for this instrument.

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 Approaches to Learning.

ELOM Teacher Assessment Validity and Reliability

The ELOM Teacher Assessment can be used alongside the ELOM Direct Assessment 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 incudes 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 Functioning Scale (EFS) (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 Direct Assessment. 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 (α) greater than .70 for both scales.

Reliability: All items met the above requirements with the Social Relations Scale (SRS) Chronbach α = .78, and the Emotional Functioning Scale (EFS) Chronbach α = .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 child is saying?
- Does child express needs and feelings appropriately?
- Is child independent, does child like to do things without help?
- Does child adjust well to changes in the classroom or home routine?
- Does child approach new experiences confidently, without fear?
- Is child a self-starter?

The first EFA (EFA 1) included all items from the SRS. The second EFA included all items from the EFS (EFA 2). As no items in either EFA violated the requirements of the EFA, all items were retained. EFA 1 revealed two valid factors, and EFA 2 revealed one valid factor. The items that contribute to these factors are presented in Table 1 with their Factor Loadings, Sampling Adequacy (KMO), and Reliability (α). The results suggest that these items represent their factors adequately and reliably.

Table 1: ELOM Teacher Assessment Item Analysis

Social Relations with Peers and Adults Scale			
Item	Factor Loading (Pattern/ Structure 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 impulsivity)?	.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 Functioning Scale			
Item	**Factor Loading (Factor Matrix)**	**Item KMO**	**Chronbach's α**
Factor 1: Emotional Functioning			
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	

2.6 The ELOM Pilot

The Pilot was designed to test the performance of Direct Assessment ELOM items as well as administration procedures. Experienced preschool teachers were trained to administer the Pilot Direct Assessment ELOM 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 or 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 taken into account. 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 in a consistent and valid manner. Bearing in mind that the ELOM 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 the English required adjustment in Afrikaans and isiXhosa.

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

The goal of the ELOM 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 each of 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 and 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 subsequent Foundation Phase, 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 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 make a significant difference to 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 take into account two highly probable influences on ELOM 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. Rather, the stimulation they receive is different 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 difficulties of constructing a representative community sample and due to the selection effects that would attend recruitment of children in early learning programmes, together with cost and logistical issues, option c) was chosen.

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 major 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 in a position 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) are benefiting 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 in fact 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 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. It had to take into account the study design sample stratification 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 2 details the target sample sizes based on power calculations.

Table 2: Power Calculations

TARGETS							
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
Quintiles 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 definition 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 3 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 3: Sampling Approach

Desired target population	Grade R children in South African public schools between the ages of 54-66 months.
Defined target population	All children at the Grade R level who are 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 as in Table 4 below. Each region was split into five according to the five quintiles and the language.

Table 4: 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 that are required for this study in order to achieve an effective 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 a number of logistical and other constraints, the maximum number of schools that 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 5 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 5 below.

Table 5: 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 assessment is scored.
3	Number of children one assessor can feasibly assess per day (test takes around 45min, 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 that 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 in the North West and in 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 6: 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).

3.6 Selection of Pupils within Schools

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

Assessors were trained to select a minimum of nine pupils, between the ages of 4.5 to 5.5 years, at each of the schools. 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 using four modified questions from the World Health Organisation 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 modification: "Did this child seem to have difficulty seeing?"
2. WHO Screen: "Does the child appear to have difficulty with hearing?" ELOM 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 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 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 7: ELOM Direct Assessment Age Validation Targets and Percentages of Target Reached in Final Sample

TOTALS														
Province	Western Cape								North West		KwaZulu-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%	142	237%	69	115%	426	142%
Quintiles 1 and 2 Urban	N/A	N/A	29	39%	99	132%	128	85%	81	108%	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	138	115%	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 | Linda Biersteker

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 Kwa-Zulu Natal provinces attended a five-day training of trainers and fieldwork planning session with the Western Cape senior assessor and ELOM 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' administrations), 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 in 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 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 included in the training, in order 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 was not Setswana speaking, so 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 8 below and revealed substantial agreement between assessors.

Table 8: 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 has been validated on a sample that is very likely to be representative of the range of socio-economic backgrounds of South African children. While not representative for language, it includes languages spoken by approximately 70% of the population. The standards and norms developed in this study are therefore also valid for children from these backgrounds.

CHAPTER 5 PSYCHOMETRY AND STATISTICAL ANALYSES | Matthew Snelling, Colin Tredoux & Andrew Dawes

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), an online service that 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. If no data connection was available, data were stored on the Android device. These forms were then submitted by the data manager at the conclusion of 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 a duplicate. All duplicates were purged. This left 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, or skipped an item.

Incomplete records could also be present if a child 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, which were 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 score, and of domains. Outliers were then checked inferentially to determine whether there were a greater number of 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, Total ELOM score means and 95% confidence intervals were computed for each school from which children were sampled. Means were then compared in order 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 Total mean quintile score. The ELOM 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 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 purposes of 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 they affected the mean score of Quintile 4 and 5. This approach was taken

7. Source: http://www.childrencount.org.za/indicator.php?id=2&indicator=14

in order 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 category or no-fee status.

Finally, ELOM Total means of all schools were compared in order 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 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 performance, based on the socioeconomic differences that each quintile is meant to represent.

The age range of the final sample was wider than the original target: 50 - 69 months. The final quintile breakdown of the sample is provided in Table 9.

Table 9: 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 domains and to generate ELOM standard scores, norms and standards. These procedures followed internationally recognised standards for test development (Ciccetti, 1994; AERA, APA, and NCME, 2014).

Validity is concerned with the degree to which the conceptual background of the ELOM, and the information that has been gathered, allow us to develop inferences and conclusions suitable for the purposes of the ELOM and South African children. Reliability is concerned with the consistency of the ELOM across different situations. Fairness is concerned with the bias of the ELOM (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 described in Chapter 4). Further, some content validity has already been established due to the robust manner in which the ELOM items and domains were selected (see Section 2.1).

Unidimensionality and internal consistency of the ELOM domains

A key psychometric concern is whether the items can be collapsed into one measure to represent a single underlying ELOM 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, it is necessary to establish whether or not 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 items did not discriminate between children from different backgrounds, who were of the same potential ability (Bond & Fox, 2015).

ELOM Direct assessment items were grouped, a priori, into construct domains on the basis of relevant literature and consultations with experts. A preliminary CFA was conducted in order to establish the unidimensionality and internal consistency of the ELOM 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^2 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 Association (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 ("Can you tell me the name of the place where you live?") was discarded from this analysis and from the final ELOM 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 her address as a safety indicator. We found it to be an inappropriate item for the ELOM.

Preliminary Confirmatory Factor Analysis

ELOM 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 1), and 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 it is likely that language competence 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 Teacher Assessment that accompanies the ELOM.

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 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 probability of success. The probability of success increases or decreases depending on whether the item score is above or below the ability level of the child on a logistic scale. Rasch Modelling then transforms the ELOM domains from an ordinal scale to an interval scale (Bond & Fox, 2015). Further, Rasch Analysis serves to strengthen evidence presented by the CFA.

A number of 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 be greater than ±1.4 logits – logistic distance units – from 0 (Bond & Fox, 2015). The Point-Measure Correlation coefficients should be in excess of .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 prediction of a child's position above and below a standard (Bond & Fox, 2015; Linacre, 2016).

Table 10: 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 0 and end at a maximum score of 20 for each domain. The maximum score for each domain was chosen in order to create an equal weighting between the domains. The total contribution of each domain produces a maximum score of 100 for the ELOM. Appendix 2, table 5 displays the untransformed difficulty logit of each item score.

Confirmation of unidimensionality and internal consistency. After the transformation of the ELOM scores, a final CFA was conducted in order to confirm the validity of the ELOM 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 2, Table 4.

Table 11: Final Confirmatory Factor Analysis

ELOM 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 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 of the 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 3 in Appendix 2, details the number of records with poor model fit in each domain.

The fairness of the ELOM domains

Differential item functioning, in Rasch Analysis, represents bias that causes some children to perform worse on some items, when compared to other children of ability equal to their own – as represented by the total domain score (Bond & Fox, 2015). The import of the bias is determined by whether it is expected for theoretical reasons, or whether the bias is particularly large. Further, some types of bias may be managed. Different groups of children may have items that counterbalance any advantage or disadvantage that they may encounter on other items. Acceptable values of Rasch Model DIF are any score less than or equal to 0.5 logits (Linacre, 2016). Further, any logit difference must be significant at the 5% level.

Table 12, presents the Differential Item Functioning for each domain and item, for gender. Gender is included to safeguard against bias caused by differential development in male and female children. No differential item functioning was found for gender.

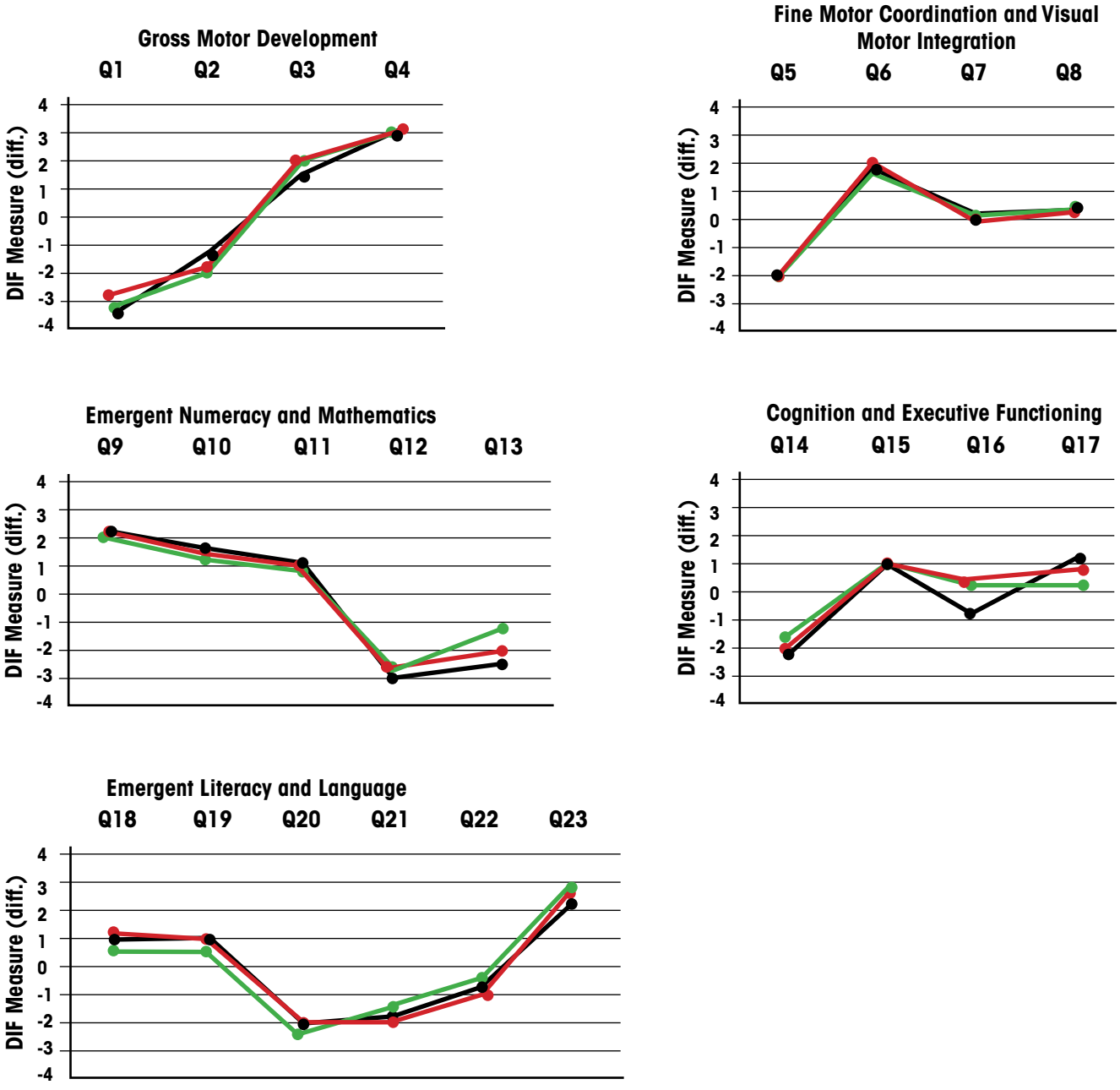
Table 12: Differential Item Functioning

Item	Domain	Differential Item Functioning Gender (Male: n = 640; Female: n = 691)	
		DIF Contrast	pRasch-Welch =
1	Gross Motor Development	.17	.25
2		-.15	.15
3		.10	.34
4		-.02	.85
5	Fine Motor Control and Visual Motor Integration	-.40	<.01
6		.00	1.00
7		.20	.06
8		.00	1.00
9	Emergent Numeracy and Mathematics	.19	.09
10		.13	.26
11		-.27	.03
12		.16	.21
13		.06	.71
14	Cognition and Executive Functioning	.13	.24
15		.12	.08
16		.00	1.0
17		-.30	<.01
18	Emergent Language and Literacy	.27	.07
19		.00	1.00
20		-.05	.67
21		.11	.30
22		-.16	.14
23		-.10	.43

Figure 3, presents the DIF for quintile group. Quintile group is included because there is some evidence to suggest that poorer children perform worse on some tasks. DIF was found for socio-economic group across a number of the figures presented below. However, as each item functions within the context of a domain, DIF is only a concern if there is no opposing DIF in other items to counterbalance the effect within the domain. After consulting the adjusted score totals associated with all noticeable cases of DIF, all DIF balanced out adequately to maintain the fairness of the domains.

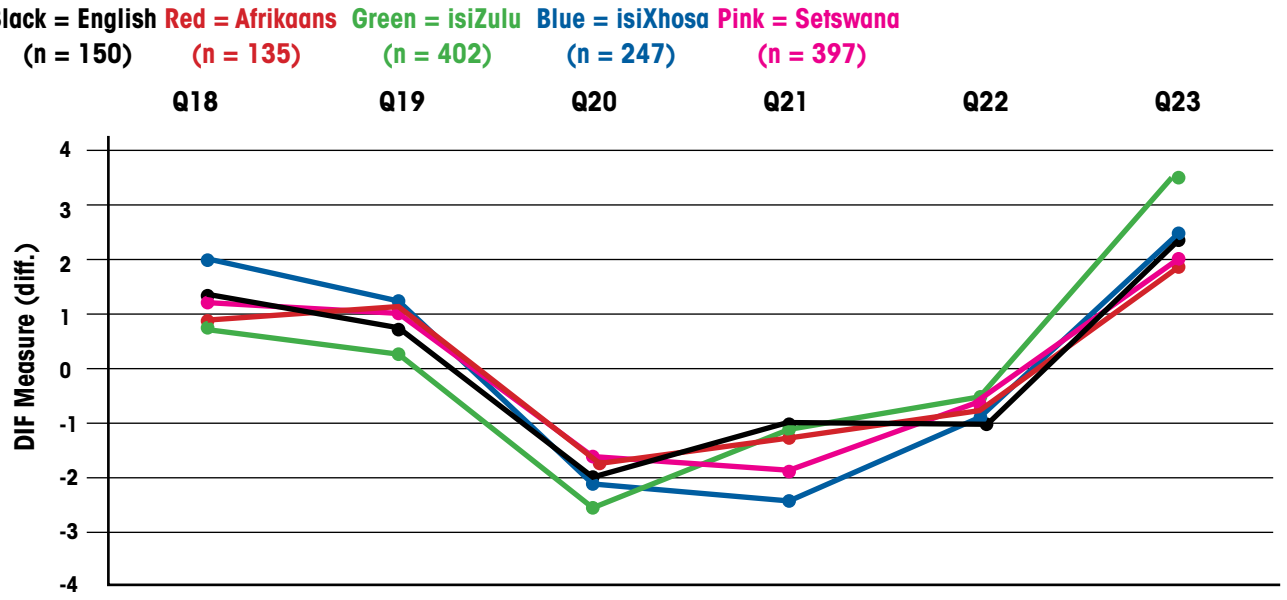
Figure 3: Differential Item Functioning by Socio-economic groups

Black = Q1 (n = 114) Red = Q2/Q3 (n = 756) Green = Q4/5 (n = 461)



In addition to the general DIF procedure presented above, a specific check was made to determine whether there were any differences between language groups on Item 23 – initial (phonetic) sound discrimination. Only isiZulu children displayed any advantage on this item. However, other items appeared to balance this advantage across the domain.

Figure 4: Differential Item Functioning by Languages



In sum, psychometric analyses indicated that the Direct Assessment ELOM 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 different socio-economic backgrounds, when interpreted at the domain level.

It should be reiterated that the process of removing records 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 2, Table 3.

Age and quintile differences

After establishing that the ELOM displayed acceptable psychometric properties, with adjustments, it was necessary to determine whether the ELOM 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 score and each domain score were modelled separately. The parameters for each analysis are presented in the ELOM Psychometry and Statistical Appendix 2. 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 column).

Table 13: Estimated Marginal Means

Quintile	Age	Total ELOM	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

5.1 Tablet Scoring and Data Collection

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 standards statements for each domain are derived from policy and are set out in Appendix 1 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 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 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 provided information that could be used in this way.

2. As the ELOM 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 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?

ELDS based on ELOM performance

The ELOM Standards are based on Age Validation sample performance on the ELOM Direct Assessment. Those measured in the Teacher Assessment 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 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 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.

Table 14 presents standard scores for the ELOM Total and for each ELOM Domain, for children in the top 40% of the age validation sample distribution (above the 60th percentile). Children at risk (red) fall well below the standard and need significant assistance to come up to the standard, while those falling behind are closer to the standard (yellow) and with support they should be able to achieve it. These performance categories are also provided with the ELOM paper-based Scoring Sheet and tablet-based scoring.

Note that the age range of the final sample (50-69 months), was wider than the original age target (54 - 66 months), which enabled standards to be set for the two age groups in Table 14.

Table 14: ELOM Standards and Performance Bands: ELOM Total and Domains

	50 – 59 Months			60 – 69 Months		
	At Risk	**Falling Behind**	**Achieving the Standard**	**At Risk**	**Falling Behind**	**Achieving the Standard**
ELOM 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 Total and for each Domain. Performance bands are indicated with the colours used in Table 14. The median scores for each age group are plotted on each figure for comparison purposes. Medians are used to account for any skewedness in the distributions. Use Table 15 to interpret the lines on each distribution.

Table 15: Distribution Interpretation Key

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

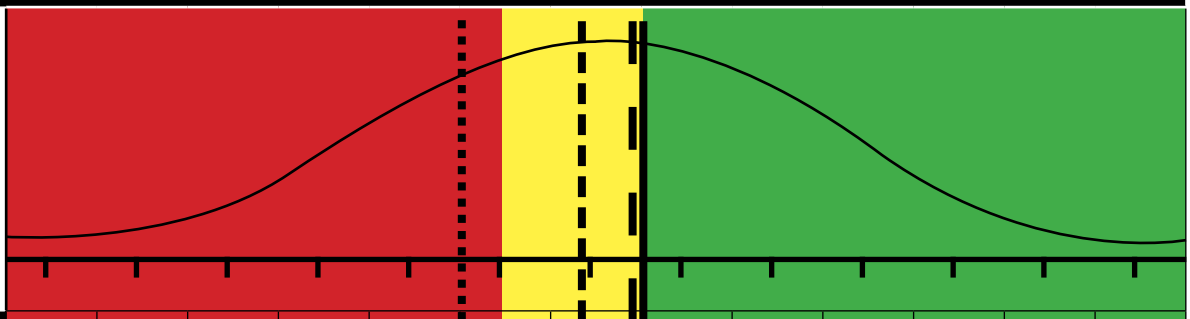
In each distribution:

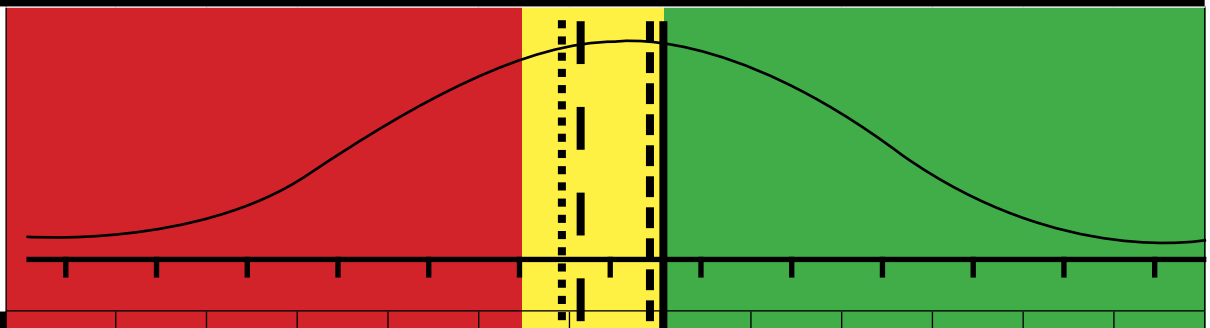
- **The green area** represents the score range for children achieving the standard (at or above the 60th percentile);
- **The yellow area** depicts those falling behind (between the 32nd and 59th percentile);
- **The red area** indicates the range of scores of children at risk (below the 32nd percentile).

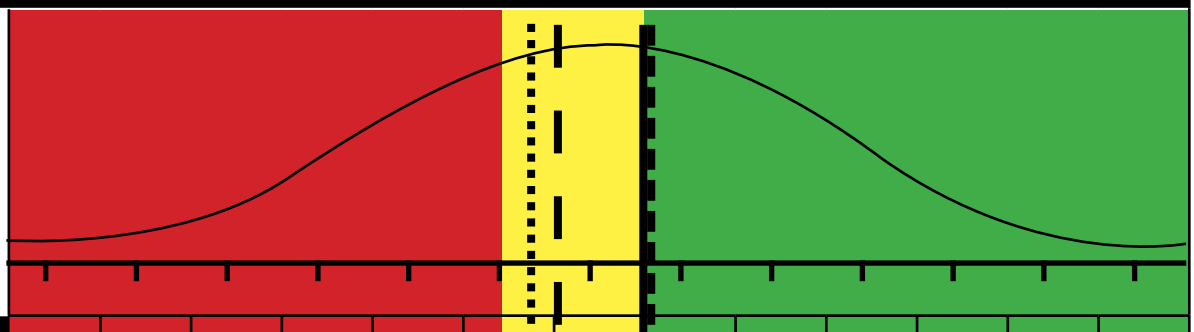
The figures also assist in visualising the positions of the school quintile groups in relation to one another on each domain and on the ELOM Total score. They show how far programmes for children from different backgrounds would need to improve their performance if they are to reach the standard.

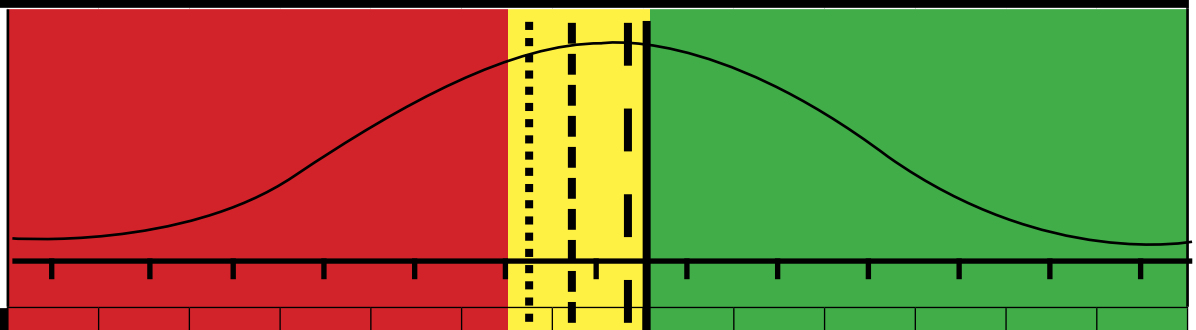
The standards may be revised as the ELOM is used in the field and more data is gathered to assess the extent to which programmes are successful in assisting children to reach the expected level of performance.

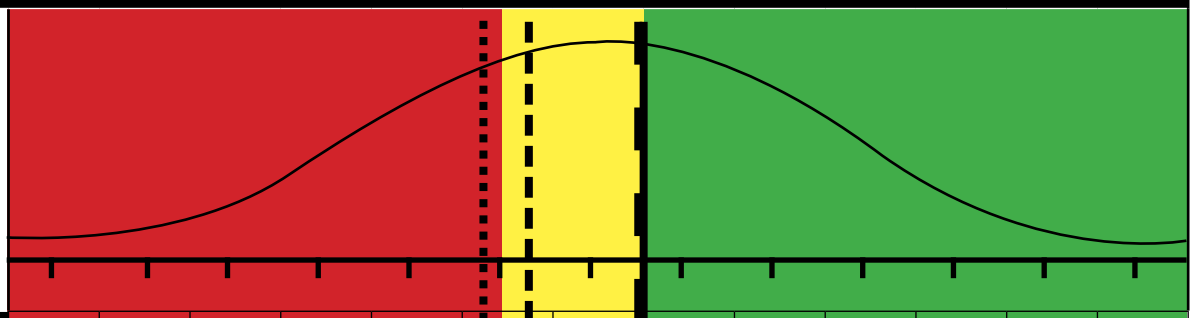
ELOM Standard Score Distributions, for the Total Sample and for each Quintile.

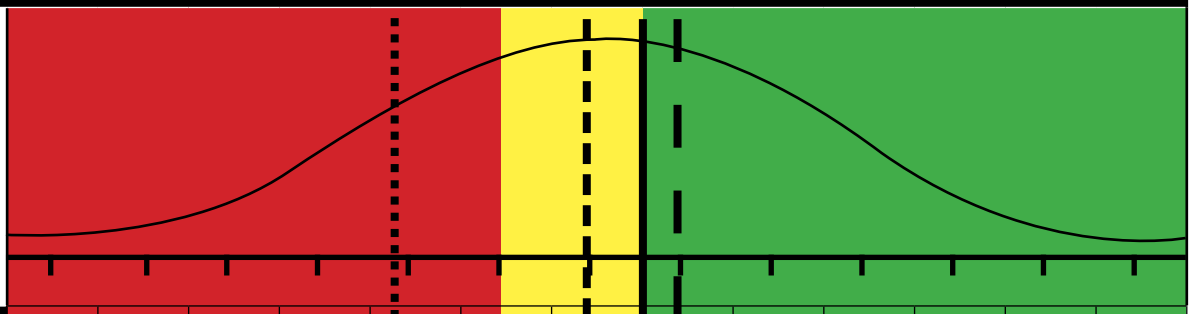
60-69 Months	ELOM Total												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	9.14	16.10	23.06	30.02	36.98	43.94	**50.90**	57.86	64.82	71.78	78.74	85.70	92.66
ΣQ4/5	15.99	22.37	28.76	35.14	41.53	47.91	**54.30**	60.68	67.07	73.45	79.84	86.22	92.61
ΣQ2/3	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
ΣQ1	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

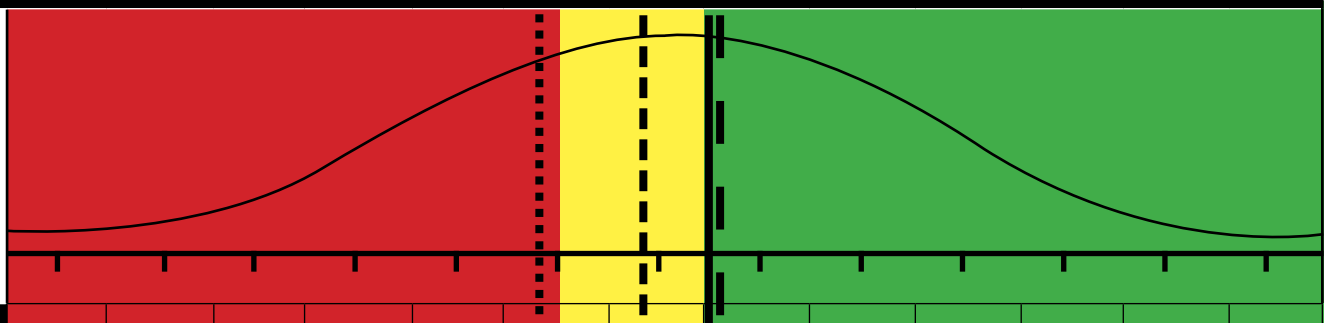
60-69 Months	Gross Motor Development												
													
Z-Score	-1.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	1.20	3.27	5.35	7.42	**9.50**	11.57	13.65	15.72	17.80	19.87	-
ΣQ4/5	-	-	0.72	2.81	4.91	7.00	**9.10**	11.19	13.29	15.38	17.48	19.57	-
ΣQ2/3	-	-	1.62	3.69	5.76	7.83	**9.90**	11.97	14.04	16.11	18.18	-	-
ΣQ1	-	-	1.18	2.93	4.69	6.44	**8.20**	9.95	11.71	13.46	15.22	16.97	18.73

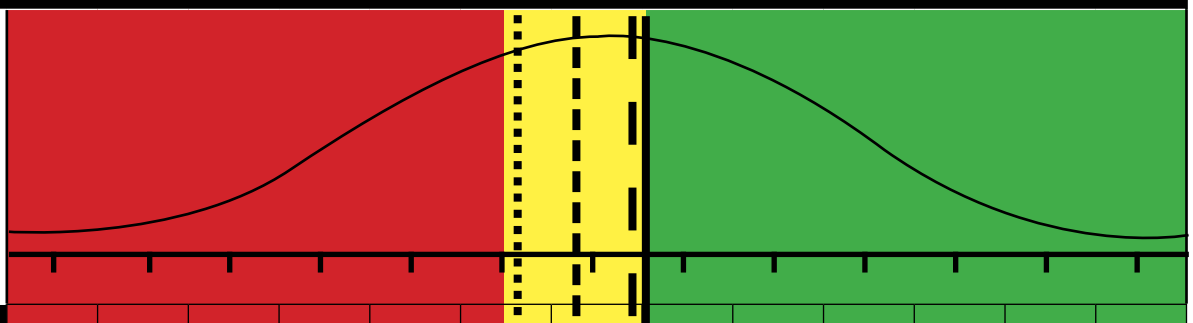
60-69 Months	Fine Motor Coordination and Visual Motor Integration												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	3.31	4.97	6.64	8.30	9.97	11.63	**13.30**	14.96	16.63	18.29	19.96	-	-
ΣQ4/5	4.59	6.17	7.76	9.34	10.93	12.51	**14.10**	15.68	17.27	18.85	-	-	-
ΣQ2/3	2.81	4.47	6.14	7.80	9.47	11.13	**12.80**	14.46	16.13	17.79	19.46	-	-
ΣQ1	2.99	4.57	6.16	7.74	9.33	10.91	**12.50**	14.08	15.67	17.25	18.84	-	-

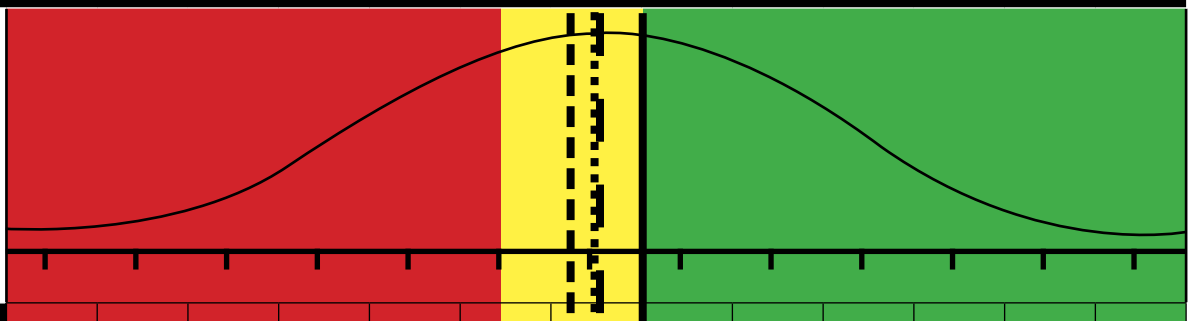
60-69 Months	Emergent Numeracy and Mathematics												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	0.86	2.94	5.03	7.11	**9.20**	11.28	13.37	15.45	17.54	19.62	-
ΣQ4/5	-	-	1.44	3.53	5.62	7.71	**9.80**	11.89	13.98	16.07	18.16	-	-
ΣQ2/3	-	-	0.62	2.71	4.81	6.90	**9.00**	11.09	13.19	15.28	17.38	19.47	-
ΣQ1	-	-	1.40	3.05	4.70	6.35	**8.00**	9.65	11.30	12.95	14.60	16.25	17.90

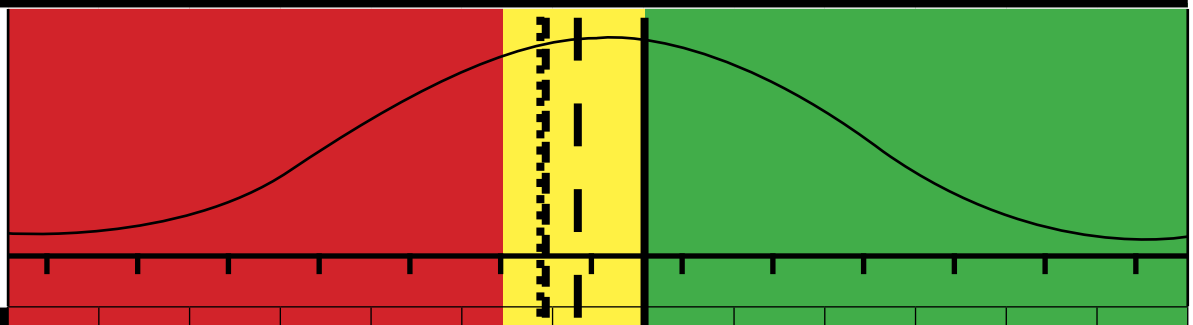
60-69 Months	Cognition and Executive Functioning												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	-	1.79	3.93	6.06	**8.20**	10.33	12.47	14.60	16.74	18.87	-
ΣQ4/5	-	-	0.68	2.83	4.99	7.14	**9.30**	11.45	13.61	15.76	17.92	-	-
ΣQ2/3	-	-	-	1.47	3.55	5.62	**7.70**	9.77	11.85	13.92	16.00	18.07	-
ΣQ1	-	-	-	0.42	2.25	4.07	**5.90**	7.72	9.55	11.37	13.20	15.02	16.85

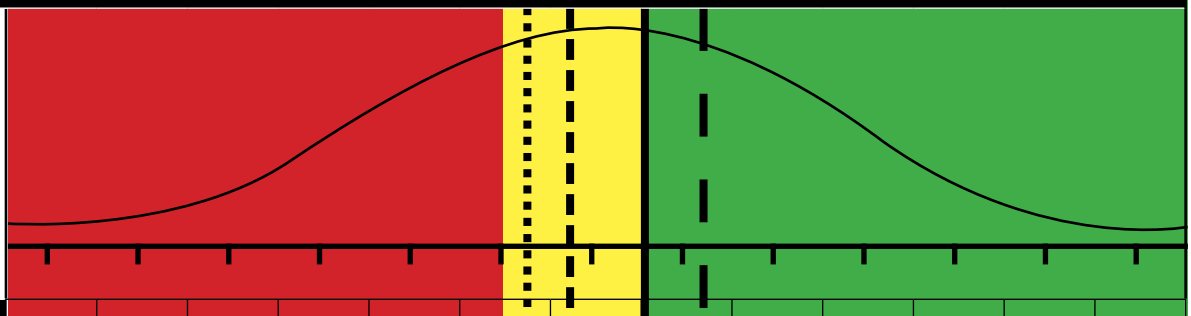
60-69 Months	Emergent Literacy and Language												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	-	-	1.32	3.61	5.91	8.20	**10.50**	12.79	15.09	17.38	19.68	-	-
ΣQ4/5	-	0.87	3.06	5.24	7.43	9.61	**11.80**	13.98	16.17	18.35	-	-	-
ΣQ2/3	-	-	1.18	3.43	5.69	7.94	**10.20**	12.45	14.71	16.96	19.22	-	-
ΣQ1	-	-	-	0.82	2.68	4.54	**6.40**	8.26	10.12	11.98	13.84	15.70	17.56

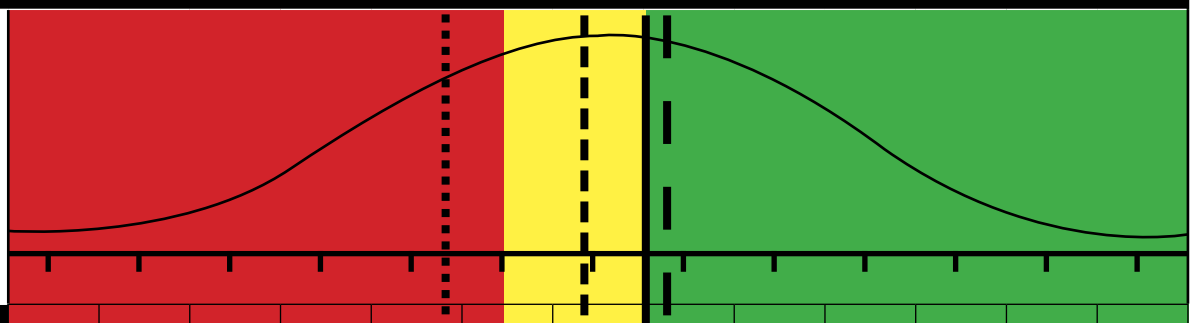
50-59 Months	ELOM Total												
													
Z-Score	-3.0	-2.5	-2.0	-1.5	-1.0	-0.5	**0**	0.5	1.0	1.5	2.0	2.5	3.0
Percentile	0.1	1	2	7	16	32	**50**	68	84	93	98	99	99.9
Standard Scores	4.49	10.93	17.36	23.80	30.23	36.67	**43.10**	49.54	55.97	62.41	68.84	75.28	81.71
ΣQ4/5	10.23	16.64	23.05	29.46	35.87	42.28	**48.69**	55.10	61.51	67.92	74.33	80.74	87.15
ΣQ2/3	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
ΣQ1	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

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
ΣQ4/5	-	-	-	2.49	4.46	6.43	**8.40**	10.37	12.34	14.31	16.28	18.25	-
ΣQ2/3	-	-	-	1.41	3.44	5.47	**7.50**	9.53	11.56	13.59	15.62	17.65	19.68
ΣQ1	-	-	-	0.98	2.85	4.73	**6.60**	8.47	10.35	12.22	14.10	15.97	17.85

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	-
ΣQ4/5	1.35	3.07	4.80	6.52	8.25	9.97	**11.70**	13.42	15.15	16.87	18.60	-	-
ΣQ2/3	1.83	3.37	4.92	6.46	8.01	9.55	**11.10**	12.64	14.19	15.73	17.28	18.82	-
ΣQ1	2.12	3.75	5.38	7.01	8.64	10.27	**11.90**	13.53	15.16	16.79	18.42	-	-

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**	10.26	12.13	13.99	15.86	17.72	19.59
ΣQ4/5	-	-	1.16	3.19	5.23	7.26	**9.30**	11.33	13.37	15.40	17.44	19.47	-
ΣQ2/3	-	-	0.86	2.64	4.43	6.21	**8.00**	9.78	11.57	13.35	15.14	16.92	18.71
ΣQ1	-	-	1.44	3.05	4.67	6.28	**7.90**	9.51	11.13	12.74	14.36	15.97	17.59

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
ΣQ4/5	-	-	-	1.91	3.91	5.90	**7.90**	9.89	11.89	13.88	15.88	17.87	19.87
ΣQ2/3	-	-	-	0.20	2.00	3.80	**5.60**	7.40	9.20	11.00	12.80	14.60	16.40
ΣQ1	-	-	-	-	1.59	3.19	**4.80**	6.40	8.01	9.61	11.22	12.82	14.43

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	-	-
ΣQ4/5	-	0.14	2.37	4.59	6.82	9.04	**11.27**	13.49	15.72	17.94	-	-	-
ΣQ2/3	-	-	0.32	2.49	4.66	6.83	**9.00**	11.17	13.34	15.51	17.68	19.85	-
ΣQ1	-	-	-	0.47	2.28	4.09	**5.90**	7.71	9.52	11.33	13.14	14.95	16.76

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.

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

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.

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/

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, 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

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.

Durkin, M. S., Wang, W., Shrout, P. E., Zaman, S. S., Hasan, Z. M., Desai, P., & Davidson, L. L. (1995). Evaluating a ten questions 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 middleincome countries. Lancet, 378, 1339–53.

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

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.

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.

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.

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.

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.

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.

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

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

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

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

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.

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.(2ndedn). Belmont, CA: Thomson Higher Education.

APPENDIX 1: ELOM EARLY LEARNING DEVELOPMENT STANDARDS INDICATORS, ASSESSMENT ITEMS, AND SOURCES[8]

PHYSICAL DEVELOPMENT[9]

NCF[10]: Early Learning and Development Area (ELDA): Well Being.
Aim: Children should be physically strong and show abilities and interest in physical activities.
Standard: SA-NELDS[11] 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[12]
1.1: Children are increasingly able to use their large (gross) muscle skills.	Child shows good control and co-ordination in large movements.	**Direct Assessment: ELOM 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**[13]. 2, 3, 4: McCarthy Scales South African adaptation**[14]. Others using same or similar methods: Lesotho[15] ELDS; CDAT[16]; EAP ECDS[17].

8. This document was prepared following psychometric analysis of the 2016 edition of the Age Validated ELOM.
9. Physical Development includes both Gross and Fine Motor Coordination. Based on psychometric analysis, these are separate domains in the ELOM.
10. 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.
11. Department of Basic Education (2009). National Early Learning and Development Standards for children birth to four years (NELDS). Pretoria: Department of Basic Education.
12. "Item Sources" refers to instruments that use these or similar items; ELOM items that are identical to the source are marked with two asterisks**.
13. Squires, J. & Bricker, D. (2009). Ages &Stages Questionnaires: A parent-completed child monitoring system: Third Edition. Baltimore: Paul H. Brookes Publishing.
14. 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.
15. 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.
16. Rao, N. (2007). Cambodian Developmental Assessment Test. (UNICEF: Cambodia.CDAT).
17. 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.

PHYSICAL DEVELOPMENT[18]

NCF[19]: Early Learning and Development Area (ELDA): Well Being.
Aim: Children should be physically strong and show abilities and interest in physical activities.
Standard: SA-NELDS[20] Desired Result 6: Children begin to demonstrate physical and motor abilities and an understanding of a healthy lifestyle.

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

Standard	Indicator	Assessment Items	Item Sources
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**[21] 6: Beery VMI**. Others using same or similar methods: IDELA[22]; EAP ECDS; ZAMCAT[23]. 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.

18. For standards purposes Physical Development has been divided into two separate domains: Gross and Fine Motor Development.
19. As for 10 above
20. As for 11 above.
21. Beery, K. E., Buktenica, N. A. & Beery, N. A. (2010). Beery-Buktenica Developmental Test of Visual-Motor Integration. 6th Edition. Pearson.
22. 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
23. 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 is able to count with one to one correspondence.	**Direct Assessment ELOM 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 item is a modification of IDELA 1:1 Correspondence item using three classes of object. Others using similar methods: MELQO[24]; EAP ECDS.
	Child is able to do simple calculations using addition and subtraction.	**Direct Assessment: ELOM Item 10:** Child adds items using a picture card stimulus. Child subtracts items using a picture card stimulus.	PreGypt Battery**[25]; Others using similar methods: IDELA; EAP ECDS; CDAT.
3.2: Children begin to demonstrate an understanding of symbols, shapes, size and space.	Child is able to classify and match objects.	**Direct Assessment ELOM 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 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 is able to understand vocabulary for location.	**Direct Assessment ELOM Item 13:** From picture stimuli: child identifies biggest, smallest, longest, shortest from pictures.	IDELA and MELQO.**

24. Measuring Early Learning Quality and Outcomes (MELQO) http://steinhardt.nyu.edu/global-ties/early_childhood/melqo

25. Egyptian adaptation of the Herbst Early Childhood Development Criteria Test provided by Dr Herbst.

DOMAIN 4: COGNITION & EXECUTIVE FUNCTIONING[26] (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 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[27]. 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 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[28]; Others using similar methods EAP ECDS; Diamond & Taylor[29].
	Child demonstrates short term memory.	D**irect Assessment ELOM Item 16:** Digit Span (forward).	IDELA**. Instructions based on the Children's Memory Scale.[30] Others using same or similar methods: MELQO and CDAT.
	Child demonstrates problem solving ability and working memory.	**Direct Assessment ELOM Item 17:** Child assembles 7 puzzles of increasing levels of difficulty.	Herbst**[31]. Others use one puzzle: IDELA; MELQO.

26. 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.

27. Zelazo, D.P. (2006). The Dimensional Change Card Sort (DCCS): a method of assessing executive function in children. Nature. Protocols 1, 297–301.

28. 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

29. 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.

30. Cohen, M. J. (2011). Children's Memory Scale. In Encyclopedia of Clinical Neuropsychology (pp. 556-559). Springer New York.

31. 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 communicate effectively and use language.	Child is able to speak in full sentences. Child is able to relate an account of events that is logical, and with correct language usage.	**Direct Assessment ELOM 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 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 Item 20: Child is able to describe what they do when they get up in the morning.	Items 20 and 21 were adapted for ELOM following pilot of IDELA and MELQO as these did not perform well. These items measure the same constructs and used the same format.
	: Child is able to name common objects.	**Direct Assessment ELOM 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 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 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 Items: Assessor observation of the child during assessment. 1: Did the child pay attention to the instructions and demonstrations throughout the assessment? 2: Did the child stay concentrated and on task during the activities and was not easily distracted? 3: Was 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.

TEACHER ASSESSMENT

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 Item	Item Sources
1.1: Children are competent in self-care.	Child can use the toilet without assistance.	**Teacher Assessment of the child's self-care:** Can this child use the toilet on her / his own?	Constructed for ELOM
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.	**Teacher Assessment 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 **[32] (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.	**Teacher Assessment of the child's relationships with familiar adults:** 1 Child seeks out 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)[33] Items presented here were finalised following Psychometry (Relationships with peers and adults constitute a single scale).
2: EMOTIONAL FUNCTIONING 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.	**Teacher Assessment of the child's emotional functioning in the early learning programme context.** Teacher rating of SACAS items.	South African Child Assessment Scales (SACAS)[34]: 6 items.

32. 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.
33. Cailifornia Department of Education DRDP (2015) A Developmental Continuum from Early Infancy to Kindergarten Entry. Sacremento Calif: www.cde.ca.gov/sp/cd/ci/drdpforms.asp Accessed 03 August, 2015.
34. 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).

Standard Error of Measurement (SEM) for ELOM

SEM was calculated using Cronbach's Alpha reliability for the total sample: $\alpha = .80$.

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 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 standard scores and those for each domain are presented in Table 1.

Table 1: ELOM Standards based on the Performance of the top 40% of Children with 95% Measurement Error Intervals

	50 – 59 Months			60 – 69 Months		
	Lo CI	Standard	Hi CI	Lo CI	Standard	Hi CI
ELOM 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 for its possible effects. To account for interaction between quintile and age, an interaction effect was added. The total ELOM 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, significance value (p), and 95% confidence intervals, are presented alongside the b parameter.

Table 2a: Multi-Level Model Parameters: ELOM 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 2b: 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 2c: 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 2d: 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 2e: 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 2f: 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 3.

Table 3: Rasch Person Misfit

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 scores are presented in Table 4.

Table 4: 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 5 presents the Rasch Difficulty Logit for each ELOM 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 way of estimating the difficulty of an item based on logit values is shown below: For a child of average ability (around the 50th percentile), a logit of:

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 5: Rasch Difficulty per Item

ITEM NUMBER	DESCRIPTION	ITEM SCORING (UN-TRANSFORMED)	RASCH DIFFICULTY LOGIT
1	Standing on one leg for 10 seconds.	0	-
		1	-1.43
		2	.85
2	Catch bean bag both hands.	0	-
		1	-1.96
		2	-.38
		3	1.86
3	Catch bean bag preferred hand.	0	-
		1	-.16
		2	2.00
		3	5.84
4	Catch bean bag 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

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

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

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