



early learning measurement tools

**JUNE 2025**

**LEARNING PROGRAMME  
QUALITY ASSESSMENT  
(LPQA) (v2):  
TECHNICAL MANUAL**



# SECOND EDITION 2025

Developed on behalf of DataDrive2030 by: Linda Biersteker, Andrew Dawes, Caitlin Buenk, Jurgen Becker, Elizabeth Girdwood, and Caylee Cook. This second edition builds upon the first edition published in 2021 by Linda Biersteker, Colin Tredoux, Frankie Mattes and Andrew Dawes.

## **TO CITE THIS MANUAL:**

Biersteker, L., Dawes, A., Buenk, C, Becker, J., Girdwood, E., Cook, C. (2025). Early Learning Outcomes Measure LPQA Technical Manual Second Edition. DataDrive2030 Steenberg, Cape Town.  
<https://DataDrive2030.co.za>

Please refer to [www.DataDrive2030.co.za](http://www.DataDrive2030.co.za) for additional information



# TABLE OF CONTENTS:

<b>INTRODUCTION</b>	<b>4</b>
<b>RATIONALE</b>	<b>5</b>
<b>SELECTION OF ITEMS</b>	<b>6</b>
<b>DEVELOPMENT OF THE LPQA</b>	<b>7</b>
Learning Programme Quality Assessment (LPQA) (v2)	8
<b>PSYCHOMETRY AND STATISTICAL ANALYSES LPQA (V2)</b>	<b>11</b>
Descriptive Statistics	11
LPQA (v2) Items	12
LPQA (v2) Item Rating Categories	13
Deriving the LPQA (v2) Domain and Total scores	14
LPQA (v2) Domain Rating Categories	15
Reliability	17
Construct Validity	18
Materials and Equipment	19
Planning and Assessment	19
Learning Programme	19
Teaching Strategies	19
Relationships and Interactions	19
Criterion Validity	19
Descriptive Statistics	20
Part 1: Statistical Analyses	21
LPQA (v2) Total Score as a Predictor of ELOM 4&5 Total Score	21
LPQA (v2) Domain Scores as Predictors of ELOM 4&5 Scores	22
Part 2: Statistical Analyses	27
LPQA (v2) Total Score as a Predictor of ELOM 4&5 Total Score	28
LPQA (v2) Domain Scores as Predictors of ELOM 4&5 Scores	28
<b>CONCLUSION</b>	<b>35</b>
<b>FUTURE RESEARCH</b>	<b>35</b>
<b>REFERENCES</b>	<b>36</b>
Appendix A –LPQA (v2) item score distributions	38
Appendix B –Distributions of LPQA (v2) item scores by quality category	39
Appendix C – Distributions of LPQA (v2) domain and total scores by quality category	40
Appendix D – Boxplots of LPQA (v2) domain scores by ELP fee band	43
Appendix E – CFA Statistical Output	45
Appendix F – Changes Made to the LPQA	55

# INTRODUCTION

This technical manual provides an overview of the ELOM Learning Programme Quality Assessment (LPQA) (v2) tool. The manual begins by outlining the development process of the LPQA (v2), including theoretical foundations and methodological refinements. This is followed by an in-depth analysis of the tool's psychometric properties, focusing on (1) internal consistency assessed through Cronbach's Alpha and McDonald's Omega, and (2) construct validity established via confirmatory factor analysis (CFA).

The second section of the manual presents a preliminary investigation into the criterion validity of the LPQA (v2), exploring its potential to predict child outcomes as measured by the Early Learning Outcomes Measure (ELOM 4&5). Using hierarchical linear regression analyses, this section examines the predictive relationships between LPQA (v2) domain and total scores and ELOM 4&5 domain and total scores, considering the confounding effects of fee band, used as a proxy for child socioeconomic status.

The LPQA (v2) observation tool is primarily for use by organisations that wish to assess the quality of an Early Learning Programme (ELP) or researchers aiming to measure programme quality (e.g., the Thrive by Five Index). The tool is designed to rate the quality of ELPs in 5 domains aligned with the South African National Curriculum Framework for Children from Birth to Four (Department of Basic Education, 2015). It includes:

1. Materials and equipment (5 items)
2. Planning and assessment (4 items)
3. Learning programme (5 items)
4. Teaching strategies (4 items)
5. Relationships and interactions (4 items)

The tool is designed for use with group learning programmes delivered directly to children aged 3-5 years, and focuses only on the provision and implementation of the learning programme and curriculum activities. The assessment takes about three hours to administer and involves a direct observation in the classroom and review of the child's assessment and planning documents.

The tool should preferably be administered by persons familiar with ELP environments. All users must be trained and accredited by DataDrive2030.



# RATIONALE

ELPs can support child learning and prepare children for a successful transition to school (Draper et al., 2024; Duncan et al., 2007). However, the quality of these programmes is a significant factor in determining whether and by how much children will benefit from their early learning experiences (Britto et al., 2011; Rao et al., 2014; Siraj et al., 2017; Von Suchodoletz et al., 2023; Zaslow et al., 2016). As of April 2022, the DBE is responsible for overseeing ECD programmes in South Africa. This shift aims to integrate early learning into the formal education system, highlighting the need for efficient tools to assess and address gaps in programme quality or in children’s developmental outcomes..

Several ELP quality assessment instruments are available and were considered when designing the LPQA tool. While appropriate for research purposes, many widely used instruments such as the ECERS-3 (Harms et al. 2018); CLASS (Pianta et al, 2008), MELE UNESCO (2018) have limitations in low-resource contexts where assessments have to be done at scale with limited expertise. First, they have not been developed with South African policy and standards in mind and do not align with the regulatory framework. Second, they require extensive training and are preferably administered by observers well-qualified in early education. Third, they may require extended observations (over several hours) of the classroom environment. The qualification requirements of assessors and the length of observation mean these instruments are costly and not feasible to implement at scale.

The need for a short, easy-to-administer, measure of programme quality aligned with the South African curriculum framework, was the impetus for the development of the ELOM LPQA tool. This short, generic LPQA (v2) tool is for use by organisations wishing to provide an indication of the quality of an ELP in the South African context. Importantly, it was not designed for mixed-age family-style groupings, for example, day mother programmes.

## Potential uses:

- For research on the quality of pre-Grade R programmes.
- To assist ECD resource and training organisations in interpreting factors that contribute to children’s ELOM 4&5 domain scores, and to identify aspects of early learning programmes that require strengthening. ELOM 4&5 To use as an element of a quality rating and improvement system, and to track improvement over time.
- To provide a reliable measure of early learning programme quality for local, provincial or national samples that can be used to track and report on improvements over time and on the quality component of Sustainable Development Goal (SDG) 4.2.



# SELECTION OF ITEMS

A review informed the selection of items in the literature and drew on the key classroom focus areas of established observational measures to support construct validity. While definitions of high-quality early childhood care and education (ECCE) vary somewhat depending on context, and include many important factors beyond the classroom, there is general agreement that the learning environment set up and resourcing, curriculum, approaches to extending learning, teacher-child interactions, and relationships are all critical contributors working together to influence programme quality and learning outcomes for children.

A comprehensive Quality Assurance and Support System (QASS) is being developed for the Department of Basic Education to assess additional aspects of quality not included here. These include Staff Qualifications and Continuing Professional Development, Safety, Parent and Community Relationships, and Leadership and Management.

A short list of items was drafted following a review of well-known measures of quality, local experience, as well as South Africa's National Curriculum Framework for Children from Birth to Four, with particular focus on the Towards Grade R phase, and National Early Learning and Development Standards (Department of Basic Education, 2009). Items were drafted to assess each of the following areas:

1. Presence in the learning environment of a variety of activities supported by materials and books (Aboud, 2006; Montie et al., 2006; Trawick-Smith et al., 2015; UNESCO, 2017).
2. Implementation of a holistic age-appropriate curriculum. To promote school readiness, there should be a targeted focus on specific school readiness skills (early mathematics and literacy) with clear learning goals (Burchinal et al., 2016) rather than a general whole child curriculum that includes these skills. Effective learning activities should be cumulative and sequenced to align with children's developmental stages (Center on the Developing Child, 2016; Phillips et al., 2017; UNESCO, 2017).
  - a. Provision of rich language and literacy experiences that are the basis for learning and later reading (Lonigan et al., 2000; Opel et al., 2009). Language skills have been found to be promoted primarily through multi-turn conversations between adults and children, including open-ended questions which lead to further discussion from the child (Cabell et al., 2015).
3. Opportunities for child-initiated activities (individual and with peers) as well as adult-led individual and small and whole group activities (Burchinal et al., 2016; Jenkins & Duncan, 2017; Phillips et al., 2017; Montie et al., 2006; Sylva et al., 2007).
4. Provision of social and emotionally supportive relationships with teaching staff, including fostering of independence and self-regulation (Durlak et al., 2011; Shala, 2013; Spilt et al., 2014; Wolf et al., 2018). Supportive interactions that model regulating negative emotions and avoiding harsh punitive exchanges have been found to support social skills (Burchinal et al., 2021).
5. Sensitive, mediated adult-child interactions targeted to the developmental levels and needs of individual children (Aga Kahn Foundation, 2010; Sylva et al., 2007; UNESCO, 2017) and designed to address areas that need strengthening. This requires careful planning and assessment of individual abilities (Grisham-Brown et al., 2006).
6. Play that promotes learning and development. A continuum of different types of play allows for this, from that which is freely chosen by children, through adult-guided play (in which adults scaffold child-led play), to adult-structured activities where the teacher designs, sets rules, and scaffolds play with a particular learning objective (Edwards & Cutter-Mackenzie, 2013; Jensen et al., 2019; Zosh et al., 2018). Highly teacher-controlled, direct instruction methods, such as large group worksheet-based academic activities, should be avoided as they have been linked with stress and reduced motivation in preschool children (Elkind, 1986; Stipek et al., 1995).

## Selected items were required to:

- Be easily observable (direct observation or documentary records) and with differentiated levels for scoring.
- Be strongly associated with overall quality scores and child outcomes related to readiness to benefit from Grade R.
- Align with the SA National Curriculum Framework, focus on Grade R competencies and the ECD programme sections included in the Children's Act Regulations on Norms and Standards for Early Childhood Development.

# DEVELOPMENT OF THE LPQA

In March 2020, a draft LPQA was piloted in 130 ELPs before fieldwork was halted due to the COVID-19 lockdown. Data from this pilot were used to conduct exploratory factor analysis to determine whether the items showed internal validity and reliably measured the same construct of “quality”. Item revisions were made (Alexander, 2020).

Between September and November 2021, the revised LPQA v1 was administered in over 500 randomly selected ELPs across all nine provinces in South Africa as part of the Thrive by Five 2021 Index. For each ELP, child outcomes data were also collected (approximately four children per site). An initial confirmatory factor analysis (CFA) on 477 records identified five factors (subscales):

1. Learning Environment (5 items)
2. Assessment for learning and teaching (2 items)
3. Relationships and interactions (4 items)
4. Curriculum (5 items)
5. Teaching strategies (5 items)

Notably, the item ‘Gross motor materials’ did not load onto the Learning environment factor or contribute to any subscale. Given its importance for monitoring equipment that supports large muscle development, it was retained as a stand-alone item.

**Table 1: Original LPQA (v1) structure**

Subscale	Items
Learning environment	Room arrangement Indoor materials Developmentally appropriate materials Accessible materials Open-ended materials
Relationships and interactions	Child-child interactions Staff-child interactions Children’s efforts acknowledged Behaviour management
Curriculum	NCF & ELDAs Programme planning Daily schedule Numeracy & mathematics Language & literacy
Teaching strategies	Free play Staff engagement Group times Open-ended questions Opportunities for independence
Assessment for learning and teaching	Child observation Recording progress
	Gross motor materials

The LPQA was revised from **LPQA (v1)** to **LPQA (v2)** in preparation for data collection for the Thrive by Five Index 2024. In **LPQA (v1)**, assessors assigned ratings of **inadequate, basic, or good** to each item based on a set of predefined criteria. For example, **Figure 1** illustrates how assessors rated the item room arrangement in the original format. This approach relied heavily on assessors interpreting the criteria consistently, which raised concerns about reliability when applied at a national scale. To improve consistency and reliability, Version 2 broke down each item into multiple sub-questions with simple response options (e.g., yes or no). Figure 2 illustrates the revised format for room arrangement. Each sub-question is scored using a predefined numeric code (ranging from 0 to 2), and the total score determines the item's overall rating. For instance:

- **For sub-question 1a:**
  - 0–2 learning areas = 0
  - 3 learning areas = 1
  - 4 or more learning areas = 2
- **For 1b and 1c:**
  - Yes = 1
  - No = 0

In this example, a total score of 0–1 corresponds to an **inadequate** rating, 2–3 to **basic**, and 4 (the maximum possible score) to **good**. This process was followed for each of the 22 items in the LPQA.

**Figure 1: LPQA (v1) room arrangement item format**

## **ROOM ARRANGEMENT**

**During playtimes the room is divided into learning areas / interest centres.**

Learning areas/interest centres refer to spaces organised with equipment for different kinds of play. These can include:

- make-believe play (more than one make-believe activity counts as one area e.g. home and shop)
- big blocks
- sensory play (sand, water etc)
- an art area (painting, drawing, modelling, cutting and pasting, box construction etc)
- book area
- nature and science table
- educational toys and games (puzzles, small construction, sorting games, counting, threading etc)

It is important that quiet areas are separated from other, more noisy and active spaces. Quiet areas include books, educational toys and games, fine motor and art areas.

**Inadequate:** There are no or fewer than 3 organised learning areas.

**Basic:** Three learning areas set out for children to play in (alone or with other children)

**Good:** There are 4 or more learning areas arranged so children can use them, quiet and active areas separate.

Figure 2: LPQA (v2) room arrangement item format

Room Arrangement		
<p><b>1a How many organised learning areas or interest centres are there in the classroom?</b></p> <p>Learning areas/interest centres refer to spaces organised with equipment for different kinds of play. These can include make-believe play (home, shop, dolls); big blocks; sensory play (sand, water, slime etc.); an art area (painting, drawing, modelling, cutting and pasting, box construction etc.); book area; nature and science or theme table; educational toys and games (puzzles, small construction, sorting games, counting, threading etc.).</p> <p>More than 1 make-believe area (e.g., kitchen and clinic) only counts as ONE learning area. The same goes for blocks or construction areas.</p>	0-2 Learning areas	<input type="checkbox"/>
	3 Learning areas	<input type="checkbox"/>
	4+ Learning area	<input type="checkbox"/>
<p><b>1b Are quiet learning areas separated from active and/or noisier ones?</b></p> <p>Examples of quiet areas are books, art, and puzzles. Examples of active areas are make-believe, sensory play, and construction. If the ELP is too small for the areas to be separated properly, answer No.</p>	No	<input type="checkbox"/>
	Yes	<input type="checkbox"/>
<p><b>1c Are the learning areas set up so that children can carry out activities independently of the practitioner?</b></p> <p>Learning areas are set out with appropriate materials accessible so that children do not need practitioner assistance to get started.</p>	No	<input type="checkbox"/>
	Yes	<input type="checkbox"/>

Subsequent psychometric validation on LPQA (v2) was conducted on the 2024 *Thrive by Five Index* ELP sample. In collaboration with content specialists, several adjustments were made to the initial structure, resulting in a revised factor structure comprising five subscales with 22 items, as outlined in Table 2.

Table 2: Revised LPQA (v2) structure

Subscale	Items
Materials and equipment	Indoor materials Developmentally appropriate materials Accessible materials Open-ended materials Gross motor materials
Planning and assessment	NCF and ELDAs Programme planning Child observation Recording progress
Learning programme	Daily schedule Numeracy and mathematics Language and literacy Free play Group times
Teaching strategies	Room arrangement Staff engagement Open-ended questions Opportunities for independence
Relationships and Interactions	Child-child interactions Staff-child interactions Children's efforts acknowledged Behaviour management

The revised LPQA (v2) structure was developed through an iterative process that combined statistical analysis and theoretical considerations.

#### Key adjustments:

- In the original structure, the item room arrangement was included in the learning environment subscale. However, excluding it from this subscale significantly improved the model fit, reducing the Root Mean Square Error of Approximation (RMSEA) from 0.129 (poor fit) to 0.066 (acceptable fit). Consequently, this item was reassigned to the Teaching Strategies subscale, where it both statistically and conceptually fit.
- The original 'Assessment for Learning and Teaching' subscale had only two items: Child Observation and Recording Progress. Rather than expanding this subscale with additional items, these items were tested in alternative subscales – 'Curriculum' and 'Teaching Strategies' – to explore their fit within broader constructs. Placement in the Curriculum resulted in poorer model fit, while inclusion in teaching strategies yielded very poor fit statistics. Consequently, these items were combined with 'NCF & ELDAs' and 'programme planning' to form a new subscale – planning and assessment.
- The gross motor materials item was included in the materials and equipment subscale.

All these changes align with theoretical frameworks and were corroborated by subject matter experts and early learning specialists. LPQA (v2)

The LPQA (v2) *Tool* is designed to measure the quality of group learning programmes for children aged 3–5 years across **five domains**:

1

**Materials and equipment:** Availability, accessibility, and developmental appropriateness of learning materials, ensuring they support open-ended play and gross motor skills.

4

**Teaching strategies:** Teaching techniques and actions.

2

**Planning and Assessment:** Programme alignment with the National Curriculum Framework Early Learning Development Areas (ELDAs), structured programme planning, and monitoring of children's progress.

5

**Relationships and interactions:** The quality of relationships and interactions between practitioners and children, and among children themselves.

3

**Learning programme:** The daily schedule and the quality of numeracy, literacy, large group and free play sessions.

# PSYCHOMETRY AND STATISTICAL ANALYSES LPQA (V2)

The data used for these analyses were drawn from observations made in 1433 ELPs from across all nine provinces in South Africa. While there was a minor reduction in the sample size for the final dataset, this is unlikely to affect the reliability or validity of the results presented here.

## DESCRIPTIVE STATISTICS

The characteristics of the ELP sample are presented in Table 3. Fee bands were derived from the fee amount charged by ELPs as reported by principals. In cases where an ELP charged differential fees depending on child age, the amount charged for children aged four to six was used, as this age group includes the children of interest for the study. In this sample, the majority of ELPs were classified as Fee-band 2 (31.8%), while only 5.7% were Fee-band 5: with Fee-band 1 representing the lowest fee-paying ELPs and Fee-band 5 the highest. ELPs were most commonly located in Gauteng (19.2%) and KwaZulu-Natal (16.2%).

Table 3: ELP sample characteristics

	Counts (N)	% of Total
<b>ELP Fee-band</b>		
Fee band 1	369	25.8%
Fee band 2	456	31.8%
Fee band 3	359	25.1%
Fee band 4	168	11.7%
Fee band 5	81	5.7%
<b>Province</b>		
Eastern Cape	164	11.4%
Free State	111	7.7%
Gauteng	275	19.2%
KwaZulu-Natal	232	16.2%
Limpopo	169	11.8%
Mpumalanga	121	8.4%
North West	112	7.8%
Northern Cape	104	7.3%
Western Cape	145	10.1%

## LPQA (v2) items

Each item (n = 22) is composed of one or more sub-questions. A 3-point scale was selected to provide some range of scores, but also take into account that more nuanced and extended scales require experienced and well-qualified ECD assessors. While a more extended scale (e.g. 7 points) would have been a better practice for this sort of tool (as, for example, in the ECERS-3), finer gradations are often challenging for observers who are not experienced ECD practitioners or trainers, which renders their scores less reliable. For items with multiple sub-questions the scores of each sub-question. The overall scale for each item thus depends on the number of sub-questions. For example, the score for the item Child–child interactions are calculated by summing the responses to sub-questions 6a, 6b, and 6c. Each sub-question is scored as 1 (yes) or 0 (no), resulting in a possible total score ranging from 0–3 (see Figure 3).

**Figure 3: Child–child interactions sub-questions**

Child-child interactions		
<p><b>6a) Does the practitioner promote and encourage positive interactions among children?</b></p> <p>Examples of promoting positive child-child interaction include encouraging playing and working cooperatively or helping children who find it hard to join a group.</p>	Yes (1)	No (0)
<p><b>6b) Does the practitioner intervene to stop any negative interactions among children?</b></p>	Yes (1)	No (0)
<p><b>6c) Does the practitioner provide activities that encourage children to work together (e.g. group collage, experiment, or group building activity)?</b></p>	Yes (1)	No (0)

Item-level descriptive statistics are presented in Table 4, and histograms depicting the distribution of item scores can be seen in Appendix A.

**Table 4: LPQA item descriptive statistics**

	N	Missing	Mean	Median	SD	Min.	Max.
1. Indoor materials	1433	0	1.177	2.00	0.959	0	2
2. Developmentally appropriate materials	1433	0	1.325	2.00	0.800	0	2
3. Accessible materials	1433	0	1.100	1.00	0.749	0	2
4. Open-ended materials	1433	0	0.777	1.00	0.710	0	2
5. Gross motor materials	1433	0	2.149	3.00	1.068	0	3
6. NCF & ELDAs	1433	0	2.573	3.00	1.708	0	4
7. Programme planning	1433	0	2.021	3.00	1.230	0	3
8. Child observation	1433	0	1.254	1.00	1.358	0	3
9. Recording progress	1433	0	1.576	2.00	1.392	0	3
10. Daily schedule	1433	0	2.176	3.00	1.125	0	3
11. Numeracy and mathematics	1433	0	1.807	2.00	1.055	0	3

	N	Missing	Mean	Median	SD	Min.	Max.
12. Language and literacy	1433	0	1.863	2.00	1.025	0	3
13. Free play	1433	0	1.062	1.00	0.829	0	2
14. Group times	1433	0	1.026	1.00	0.716	0	2
15. Room arrangement	1433	0	1.916	2.00	1.472	0	4
16. Staff engagement	1433	0	1.060	1.00	0.704	0	2
17. Open-ended questions	1433	0	1.192	1.00	1.260	0	3
18. Opportunities for independence	1433	0	1.621	2.00	1.120	0	3
19. Child-child interactions	1433	0	2.595	3.00	0.701	0	3
20. Staff-child interactions	1433	0	1.890	2.00	0.979	0	3
21. Children's efforts acknowledged	1433	0	1.320	2.00	0.802	0	2
22. Behaviour management	1433	0	2.040	2.00	0.785	0	3

There are three items (items 9, 15 and 17) where assessors have the option to select *not applicable* (NA). After testing various approaches (e.g., treating NA as missing, applying a standard penalty), it became clear that each item required individual consideration as no single method proved suitable for all three.

If NA was selected for item 15 (Free Play) and item 17 (Group Times), a penalty was applied (i.e., the response was coded as 0 and not as missing) because one would usually and ideally see large group times and free play during a two-hour session.

If NA was chosen for items 9b ('Did practitioners practice positive discipline?') and/or 9c ('Are children supported to solve conflicts for themselves?'), a score of 0.33 was assigned to avoid penalising cases where behaviour management or conflict resolution was not observed within the limited observation period. Notably the overall score for item 9 (Behaviour Management) is based solely on the first three sub-questions (9a, 9b, and 9c) due to an error in the skip logic. This pro-rating reflects the equal weighting of the three sub-questions ( $1 \div 3 = 0.33$ ). Significantly, the LPQA (v2) has been adjusted for future use, and the only items where assessors will have the option to select NA will be 9c and 9d. In the future, when item 9 includes all four sub-questions (9a – 9d), NA responses for 9c and/or 9d will be pro-rated as 0.25 ( $1 \div 4 = 0.25$ ) to account for the corrected structure.

### Deriving the LPQA (v2) item rating categories

Raw item scores are rated as inadequate, basic, or good quality according to pre-determined criteria specific to each item. Using open-ended materials as an example: *Inadequate* would be assigned if the ELP had no open-ended materials, *basic* would be assigned if there were open-ended materials. However, there were fewer than four different types, and *good* would be assigned if there were four or more different types of open-ended materials found in the classroom.

Table 5 displays the percentage of ELPs (n = 1433) in each quality category on all items (See histograms in Appendix B). For example, 38.80% of ELPs were rated as inadequate on 'open-ended materials'.

Table 5: Percentage of ELPs classified as inadequate, basic, and good across items

Items	Inadequate	Basic	Good
1. Indoor materials	38.73%	4.88%	56.39%
2. Developmentally appropriate materials	21.00%	25.47%	53.52%
3. Accessible materials	23.52%	42.99%	33.50%
4. Open-ended materials	38.80%	44.66%	16.54%
5. Gross motor materials	12.14%	34.47%	53.38%
6. NCF and ELDA's	28.96%	23.31%	47.73%
7. Programme planning	29.59%	16.26%	54.15%
8. Child observation	57.36%	9.91%	32.73%
9. Recording progress	44.17%	12.63%	43.20%
10. Daily schedule	33.29%	4.47%	62.25%
11. Numeracy and mathematics	14.93%	52.13%	32.94%
12. Language and literacy	11.72%	53.45%	34.82%
13. Free play	31.40%	30.98%	37.61%
14. Group times	24.35%	48.71%	26.94%
15. Room arrangement	41.73%	39.22%	19.05%
16. Staff engagement	21.91%	50.17%	27.91%
17. Open-ended questions	56.52%	19.82%	23.66%
18. Opportunities for independence	41.24%	31.61%	27.15%
19. Child-child interactions	1.81%	27.98%	70.20%
20. Staff-child interactions	35.31%	30.98%	33.71%
21. Children's efforts acknowledged	21.28%	25.47%	53.24%
22. Behaviour management	3.21%	51.29%	45.50%

### *Deriving the LPQA (v2) domain and total scores*

Item scores are summed to derive five domain scores. These are the most important LPQA (v2) scores as they highlight the strengths and weaknesses of the programme and can be used to guide improvement.

Domain scores are summed to derive the Total LPQA (v2) score, which is common practice in similar tools, such as the ECERS.

As each domain consists of a varying number of items, averages are commonly used for reporting to improve interpretability.

In this report, domain total scores rather than domain averages are presented and analysed, since they more directly reflect the raw data and the tool's underlying structure. Descriptive statistics for LPQA (v2) domain-level and total scores are provided in Table 6.

**Table 6: LPQA (v2) domain descriptive statistics**

	Materials and equipment	Planning and Assessment	Learning Programme	Median	SD	Min.
N	1433	1433	1433	1433	1433	1433
Missing	0	0	0	0	0	0
Mean	6.53	7.42	7.93	5.79	7.85	35.50
Median	7.00	8.00	8.00	6.00	8.33	37.00
SD	3.20	4.46	3.47	3.10	2.36	13.5
Min.	0	0	0	0	0.00	1.00
Max.	11	13	13	12	11.0	60.0
Skewness	-0.385	-0.289	-0.448	0.0308	-0.777	-0.360
Std. error skewness	0.0646	0.0646	0.0646	0.0646	0.0646	0.0646
Kurtosis	-1.01	-1.23	-0.696	-0.836	0.129	-0.686
Std. error kurtosis	0.129	0.129	0.129	0.129	0.129	0.129

### *Deriving the LPQA (v2) domain rating categories*

LPQA (v2) domain and total scores are also classified into three rating categories – inadequate, basic, and good – to provide ELPs with an indication of their performance and areas for improvement. The process for determining domain rating categories involves transforming raw item scores into a standardised percentage, as follows:

- 1. Convert item rating categories to numeric codes:** Each item is initially rated as inadequate, basic, or good based on predefined criteria. These ratings are then coded numerically as 1 (inadequate), 2 (basic), or 3 (good) to standardise the scale across items.
- 2. Sum the numeric codes within each domain:** For each of the five domains, the numeric codes of the items are added together.
- 3. Calculate a percentage score:** The domain sum is divided by the maximum possible score – 12 for domains with four items ( $4 \times 3 = 12$ ) or 15 for domains with five items ( $5 \times 3 = 15$ ) – to yield a percentage.
- 4. Assign the rating category:** The percentage score determines the domain's rating category (inadequate, basic, or good), as specified in Table 7.

For the total LPQA score, the numeric codes of all 22 items are summed and divided by the maximum possible score of 66 ( $22 \times 3 = 66$ ) to obtain a percentage. Similarly, this percentage is then mapped to the inadequate, basic, or good category.

Table 7: Domain rating categories

Rating category	Threshold	Interpretation
Inadequate	Less than 60% of domain total	Room for significant improvement
Basic	60% to <80% of domain total	Some room for improvement
Good	80% or more of domain total	Small enhancements possible

The percentage of ELPs classified as inadequate, basic, and good across the five LPQA domains, along with the total score, is presented in Table 8. For example, the Materials and Equipment domain shows 27.22% of ELPs rated inadequate, 25.33% basic, and 47.45% good. The total score distribution indicates that 26.10% of ELPs are Inadequate, 41.87% Basic, and 32.03% Good overall.

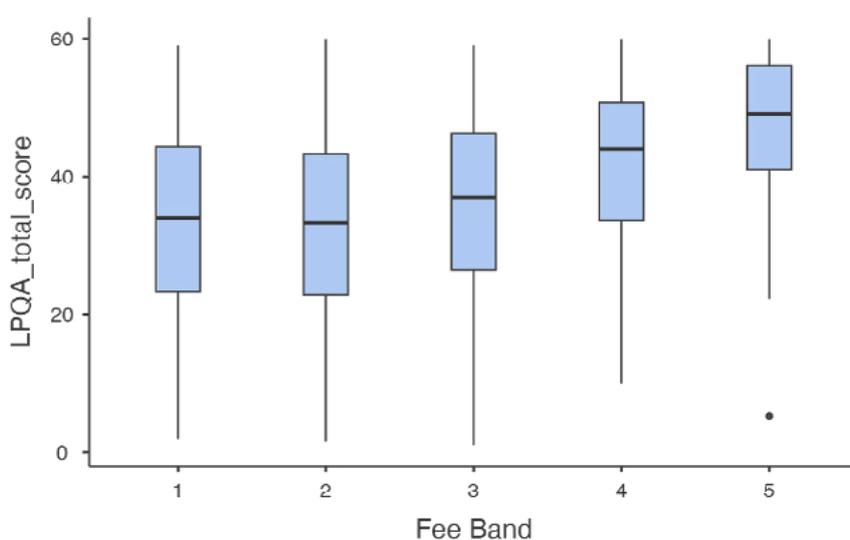
Table 8: Percentage of ELPs classified as inadequate, basic, and good across domains.

Domains	Inadequate	Basic	Good
Materials and Equipment	27.22%	25.33%	47.45%
Planning and Assessment	41.52%	19.89%	38.59%
Learning programme	21.91%	35.24%	42.85%
Teaching Strategies	55.13%	27.01%	17.86%
Relationships and interactions	18.63%	29.10%	52.27%
Total Score	26.10%	41.87%	32.03%

LPQA (v2) domain and total score distributions are depicted in Appendix C.

Figure 4, which depicts LPQA (v2) total score by ELP fee band, indicates that higher fee bands have progressively better scores<sup>1</sup>. LPQA (v2) domain scores by ELP fee band can be seen in Appendix D.

Figure 4: LPQA (v2) Total Score by ELP Fee Band.



<sup>1</sup>No fee and otherwise subsidised pre-schools fall into the first fee band and could be one of the reasons for the inflated results from this band when compared to fee band 2.

## RELIABILITY

Cronbach's alpha ( $\alpha$ ) is a measure of internal consistency and assesses whether items on a test measure the same characteristic. Cronbach's alpha value of at least 0.70 is considered acceptable reliability (Tavakol & Dennick, 2011). Reliability was also tested using McDonald's Omega ( $\omega$ ) – a more robust measure that accounts for unequal item contributions. Omega was calculated using a maximum likelihood approach to ensure stability of the factor model (Dunn et al., 2014).

The scale had a Cronbach's alpha value of 0.905 and an Omega value of 0.927, which indicates excellent reliability across all items. Alpha and Omega values at the domain level are presented in Table 9.

**Table 9: LPQA (v2) reliability**

	Cronbach's Alpha	McDonald's Omega
<b>Domain 1:</b> Materials and equipment	0.789	0.830
<b>Domain 2:</b> Planning and assessment	0.782	0.842
<b>Domain 3:</b> Learning programme	0.770	0.819
<b>Domain 4:</b> Teaching strategies	0.568	0.660
<b>Domain 5:</b> Relationships and interactions	0.685	0.719

The LPQA (v2) demonstrates strong overall reliability, suggesting that items collectively measure a cohesive construct. Domains 1, 2, and 3 exhibit good reliability, with alpha values ranging from 0.770 to 0.789 and omega values from 0.819 to 0.842. However, alpha values for Domains 4 ( $\alpha = 0.568$ ,  $\omega = 0.660$ ) and 5 ( $\alpha = 0.685$ ,  $\omega = 0.719$ ) demonstrate questionable to acceptable reliability. Omega, which is relatively higher across all domains, reinforces the scale's reliability when alpha is marginal.

**Table 10: Change in Cronbach's alpha if item dropped**

	Alpha if Item Dropped
<b>Domain 1: Materials and equipment (<math>\alpha = 0.789</math>)</b>	
Indoor materials	0.728
Developmentally appropriate materials	0.719
Accessible materials	0.766
Open-ended materials	0.747
Gross motor materials	0.789
<b>Domain 2: Planning and assessment (<math>\alpha = 0.782</math>)</b>	
NCF and ELDAs	0.771
Programme planning	0.720
Child observation	0.712
Recording progress	0.717
<b>Domain 3: Learning programme (<math>\alpha = 0.770</math>)</b>	
Daily schedule	0.730
Numeracy and mathematics	0.691
Language and literacy	0.682
Free play	0.760
Group times	0.762

Domain 4: Teaching strategies ( $\alpha = 0.568$ )	
Room arrangement	0.573
Staff engagement	0.469
Open-ended questions	0.475
Opportunities for independence	0.481
Domain 5: Relationships and interactions ( $\alpha = 0.685$ )	
Child-child interactions	0.621
Staff-child interactions	0.646
Children's efforts acknowledged	0.598
Behaviour management	0.616

Table 10 illustrates the impact of removing individual items affects Cronbach's alpha within each domain. For most domains, dropping any item reduces alpha below the original value, confirming that all items contribute positively to internal consistency and should be retained. *Room arrangement* was the only item that, if removed, would increase the alpha value of Domain 4 from 0.568 to 0.573. However, removing this item would only boost alpha modestly. It could compromise the content validity of the Teaching Strategies domain (i.e., its ability to capture the construct fully) without substantially enhancing overall reliability.

## CONSTRUCT VALIDITY

Construct validity refers to the extent to which a tool accurately measures what it's intended to. Construct validity was assessed using Confirmatory Factor Analysis (CFA) to determine whether the hypothesised model fitted the data well.

Table 11: CFA statistics and interpretation.

Statistic	Interpretation (Barrett, 2007; Hu & Bentler, 1999; Tavakol & Wetzel, 2020)
Chi-square ( $\chi^2$ )	An overall test of the fit of observed variance within and between items to a specified statistical model. Smaller values with non-significant p-values are considered indicative of model fit. However, this test is considered highly sensitive and often shows misfit for models that are typically tested in larger samples or with complex factor structures. For this reason, fit indices such as RMSEA, CFI, and TLI are usually considered more important for assessing CFA model fit.
Factor Loadings	Represents the strength and direction of the relationship between each item and its corresponding latent factor, with higher values indicating stronger relationships. Loadings should ideally be high and statistically significant.
Standardised Loadings	Factor loadings transformed to a standardised, universally comparable scale. Standardised loadings above 0.3 are typically considered acceptable.
RMSEA	An Absolute Fit Index where a value of 0 indicates a perfect model. Values closer to 0 indicate better model fit. Root Mean Square Error of Approximation (RMSEA) values below 0.08 indicate an acceptable fit. Values below 0.05 indicate a strong fit.
SRMR	Standardised Root Mean Square Residual (SRMR) assesses the difference between the observed and predicted correlations. Values below 0.08 is considered a good fit.
CFI & TLI	The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are both fit statistics that compare the fit of a factor model to a baseline model. These values range from 0 to 1; values above 0.90 are considered acceptable, and those above 0.95 are considered excellent.

### **Materials and equipment**

All factor loadings were significant ( $p < 0.001$ ), with standardised estimates ranging from 0.548 to 0.790, indicating the latent construct is well-measured. All fit indices were strong (CFI = 0.986, TLI = 0.971, RMSEA = 0.065, SRMR = 0.022).

### **Planning and assessment**

For the Planning and Assessment domain, a strong positive relationship was observed between 'child observation' and 'recording progress,' indicating shared variance beyond what the latent factor explained. This likely reflects conceptual overlap, as both involve ongoing formative assessment practices used by teachers to understand and support individual child development. We therefore allowed these items to covary in the model based on theoretical and empirical justification. Standardised factor loadings ranged from 0.588 to 0.792 (all  $p < 0.001$ ), demonstrating that the observed indicators strongly and reliably reflect the underlying construct. Model fit indices further confirmed excellent fit to the data (CFI = 0.999, TLI = 0.994, RMSEA = 0.034, SRMR = 0.006).

### **Learning programme**

Standardised loadings ranged from 0.479 to 0.806 (all  $p < 0.001$ ), indicating solid measurement of the latent factor. All fit indices were strong (CFI = 0.990, TLI = 0.981, RMSEA = 0.049, SRMR = 0.021).

### **Teaching strategies**

Standardised loadings ranged from 0.406 to 0.675 (all  $p < 0.001$ ), supporting construct validity. Fit indices indicated an excellent model fit (CFI = 0.996, TLI = 0.989, RMSEA = 0.028, SRMR = 0.012).

### **Relationships and interactions**

Standardised loadings ranged from 0.555 to 0.648 (all  $p < 0.001$ ), demonstrating strong construct validity. Fit indices suggested an almost perfect fit (CFI = 1.000, TLI = 1.006, RMSEA = 0.000, SRMR = 0.003), but these results should be interpreted with caution. Importantly, collapsing multiple items into a single indicator (i.e., summing multiple sub-questions to get an overall score for an item) reduces the number of observed variables and decreases the model's complexity. This simplification can artificially inflate factor loadings and fit indices.

All five models demonstrate a good to excellent fit, supporting the construct validity of the LPQA (v2) tool across its domains. Fit indices are summarised in Table 12 (See Appendix E for detailed statistical output). Taken together, these results suggest strong construct validity of the LPQA (v2) measure based on the factor analytical results.

**Table 12: Summary of fit indices**

<b>Domain</b>	<b>CFI</b>	<b>TLI</b>	<b>RMSEA</b>	<b>SRMR</b>
Materials and equipment	0.986	0.971	0.065	0.022
Planning and assessment	0.999	0.994	0.034	0.006
Learning programme	0.990	0.981	0.049	0.021
Teaching strategies	0.996	0.989	0.028	0.012
Relationships and interactions	1.000	1.006	0.000	0.003

## **CRITERION VALIDITY**

Criterion-related validity assesses how well a tool predicts an external criterion or outcome. Higher-quality ELPs should result in improved child outcomes, as assessed by the ELOM 4&5 (Dawes et al., 2020). However, it is very important to note that the literature typically does not report high correlations between ELP quality measures and learning outcomes (Zaslow et al., 2016). This is due to the influence of a range of other factors on performance on tests such as ELOM 4&5, including the ELOM 4&5 characteristics of the learners themselves, and household factors such as parent education, SES, and access to learning opportunities.

### Descriptive statistics

The ELOM 4&5 assesses the performance of children aged 50–69 months in five important developmental domains: gross motor development (GMD), fine motor coordination and visual-motor integration (FMC and VMI), emergent numeracy and mathematics (ENM), cognition and executive functioning (CEF), and emergent literacy and language (ELL). Descriptive statistics for the five domains and the ELOM 4&5 total score are presented in Table 13. ELOM 4&5

**Table 13: ELOM 4&5 descriptives Statistics**

	Domain 1: GMD	Domain 2: FMC and VMI	Domain 3: ENM	Domain 4: CEF	Domain 5: ELL	Total
N	5189	5189	5189	5189	5189	5189
Missing	0	0	0	0	0	0
Mean	7.92	11.2	8.40	7.27	10.9	45.6
Median	7.60	10.8	7.92	6.63	10.9	45.4
Standard deviation	3.94	3.76	4.09	4.22	4.46	14.1
Minimum	0.00	1.00	0.00	0.00	0.00	6.64
Maximum	20.0	20.0	20.0	20.0	20.0	99.1
Skewness	0.471	0.0596	0.375	0.456	-0.0701	0.181
Std. error skewness	0.0340	0.0340	0.0340	0.0340	0.0340	0.0340
Kurtosis	-0.0751	-0.495	-0.386	-0.362	-0.750	-0.273
Std. error kurtosis	0.0680	0.0680	0.0680	0.0680	0.0680	0.0680

Characteristics of the child sample are presented in Table 14. Following recent research by DataDrive2030 (Henry & Giese, 2023), the fee-band serves as a proxy for socio-economic status. In this sample, the majority of children attended a Fee-band 2 ELP (32.7%). Only 5.8% of children attended an ELP classified as Fee-band 5. Child gender distribution (n = 5189) was nearly equal, with 49.1% male and 50.9% female, reflecting a balanced sample in terms of sex.

**Table 14: Child sample characteristics**

	Counts (N)	% of Total
<b>ELP Fee-band</b>		
Fee band 1	1357	26.2%
Fee band 2	1697	32.7%
Fee band 3	1247	24.0%
Fee band 4	587	11.3%
Fee band 5	301	5.8%
<b>Child Gender</b>		
Male	2549	49.1%
Female	2640	50.9%

## Part 1: Statistical analyses

As an initial exploratory analysis, hierarchical linear regression analyses were conducted to examine whether LPQA (v2) domain and total scores predicted ELOM 4&5 domain and total scores. ELOM 4&5 data were aggregated per ELP, and all predictors were standardised using z-score standardisation (subtracting the mean and dividing by the standard deviation) to ensure comparability across variables. This transformation resulted in standardised variables with a mean of 0 and a standard deviation of 1, preserving the original distribution shape while enabling direct comparison of predictors measured on different scales. For each criterion, a two-step regression model was employed (Petrocelli, 2003). The assumptions of linearity and homoscedasticity were tested for all regression analyses.

### LPQA (v2) total score as a predictor of ELOM 4&5 total score

Regression analyses first assessed whether the LPQA (v2) total score predicted the ELOM 4&5 total score, with fee band entered as a control variable in Model 1 and standardised LPQA (v2) total score added in Model 2 (See Table 15).

Model 1 was statistically significant, with fee band accounting for 9.6% of the variance in ELOM 4&5 total scores (*adjusted*  $R^2 = 0.096$ ,  $F(4, 1428) = 38.80$ ,  $p < .001$ ). Higher fee bands (i.e., higher socio-economic status) were associated with progressively larger increases in ELOM 4&5 total scores.

In Model 2, adding LPQA (v2) total score significantly improved the model, explaining an additional 2% of the variance (*adjusted*  $R^2 = 0.115$ ,  $F(5, 1427) = 38.02$ ,  $p < .001$ ). LPQA (v2) total score was a significant predictor ( $b = 1.558$ ,  $p < .001$ ,  $\beta = 0.146$ ). Importantly though, fee band, which serves as a proxy for child SES, is related to both LPQA (v2) and ELOM 4&5 scores, acting as a confounding variable that complicates the predictive relationship between LPQA (v2) and total scores. This confounding effect of SES suggests that LPQA (v2) total scores do not independently predict ELOM 4&5 total scores.

Table 15: Regression results with ELOM 4&5 total score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band1)	43.161	0.527	[42.128, 44.195]	< 0.001***	
Fee band2	1.265	0.709	[-0.125, 2.655]	0.074	0.055
Fee band3	1.785	0.750	[0.314, 3.257]	0.017*	0.073
Fee band4	7.451	0.942	[5.604, 9.298]	< 0.001***	0.225
Fee band5	12.798	1.242	[10.362, 15.233]	< 0.001***	0.278
<b>Model 2</b>					
(Intercept)	43.401	0.523	[42.375, 44.427]	< 0.001***	
Fee band2	1.377	0.701	[0.001, 2.752]	0.05*	0.060
Fee band3	1.491	0.744	[0.032, 2.951]	0.045*	0.061
Fee band4	6.495	0.947	[4.637, 8.353]	< 0.001***	0.196
Fee band5	11.216	1.260	[8.743, 13.688]	< 0.001***	0.244
LPQA Total	1.558	0.277	[1.014, 2.102]	< 0.001***	0.146

Note:  $b$  = unstandardised coefficient,  $SE$  = standard error, 95%  $CI$  = confidence range,  $p$  = statistical significance,  $\beta$  (beta) = standardised coefficient.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Table 16: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.098	0.096	38.80	< .001
Model 2	0.118	0.115	38.02	<.001

### LPQA (v2) domain scores as predictors of ELOM 4&5 scores

A series of regression analyses examined whether LPQA (v2) domain scores predicted ELOM 4&5 total and ELOM 4&5 domain scores. Fee band was controlled for in Model 1, and standardised LPQA (v2) domain scores were added simultaneously in Model 2.

ELOM 4&5 Total Score: When total score was the criterion (Table 17), Model 1 showed fee band explaining 9.6% of the variance ( $adjusted R^2 = 0.096$ ,  $F(4, 1428) = 38.80$ ,  $p < .001$ ). In Model 2,  $adjusted R^2$  increased; however, no individual LPQA domain score reached statistical significance ( $adjusted R^2 = 0.114$ ,  $F(9, 1423) = 21.40$ ,  $p < .001$ ).

Table 17: Regression results with ELOM 4&5 total score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	43.161	0.527	[42.128, 44.195]	< 0.001***	
Fee band 2	1.265	0.709	[-0.125, 2.655]	0.074	0.055
Fee band 3	1.785	0.750	[0.314, 3.257]	0.017*	0.073
Fee band 4	7.451	0.942	[5.604, 9.298]	< 0.001***	0.225
Fee band 5	12.798	1.242	[10.362, 15.233]	< 0.001***	0.278
<b>Model 2</b>					
(Intercept)	43.387	0.525	[42.358, 44.417]	< 0.001***	
Fee band 2	1.347	0.704	[-0.033, 2.727]	0.056	0.059
Fee band 3	1.569	0.748	[0.101, 3.037]	0.036*	0.064
Fee band 4	6.557	0.951	[4.691, 8.423]	< 0.001***	0.198
Fee band 5	11.148	1.266	[8.664, 13.631]	< 0.001***	0.242
Materials and equipment	0.551	0.400	[-0.233, 1.336]	0.168	0.052
Planning and assessment	0.675	0.353	[-0.018, 1.368]	0.056	0.063
Learning programme	0.196	0.446	[-0.678, 1.07]	0.66	0.018
Teaching strategies	0.613	0.416	[-0.202, 1.429]	0.14	0.058
Relationships and interactions	-0.185	0.349	[-0.869, 0.499]	0.596	-0.017

Note:  $b$  = unstandardised coefficient,  $SE$  = standard error,  $95\% CI$  = confidence range,  $p$  = statistical significance,  $\beta$  (beta) = standardised coefficient.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Table 18. Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.098	0.096	38.80	< .001
Model 2	0.119	0.114	21.40	< .001

### ELOM 4&5 Domain 1 (gross motor development)

For gross motor development (Table 19), Model 1 indicated that fee band accounted for 1% of the variance ( $adjusted R^2 = 0.010$ ,  $F(4, 1428) = 4.636$ ,  $p = .001$ ). None of the LPQA domains were significant, and these additional predictors did not improve the model fit ( $adjusted R^2 = 0.008$ ,  $F(9, 1423) = 2.225$ ,  $p = .018$ ).

Table 19: Regression results with ELOM 4&5 domain 1 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	8.128	0.147	[7.84, 8.416]	< 0.001***	
Fee band 2	-0.209	0.197	[-0.597, 0.178]	0.289	-0.034
Fee band 3	-0.774	0.209	[-1.184, -0.364]	< 0.001***	-0.118
Fee band 4	-0.188	0.262	[-0.703, 0.327]	0.474	-0.021
Fee band 5	0.302	0.346	[-0.377, 0.98]	0.384	0.025
<b>Model 2</b>					
(Intercept)	8.127	0.148	[7.837, 8.418]	< 0.001***	
Fee band 2	-0.213	0.198	[-0.602, 0.176]	0.284	-0.035
Fee band 3	-0.775	0.211	[-1.189, -0.361]	< 0.001***	-0.119
Fee band 4	-0.181	0.268	[-0.707, 0.345]	0.499	-0.021
Fee band 5	0.314	0.357	[-0.386, 1.014]	0.379	0.026
Materials and equipment	-0.051	0.113	[-0.272, 0.17]	0.65	-0.018
Planning and assessment	-0.067	0.100	[-0.262, 0.129]	0.503	-0.024
Learning programme	0.144	0.126	[-0.102, 0.391]	0.251	0.051
Teaching strategies	0.011	0.117	[-0.218, 0.241]	0.923	0.004
Relationships and interactions	-0.056	0.098	[-0.249, 0.137]	0.567	-0.020

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\* $p < .05$ , \*\* $p < .01$ , \*\*\* $p < .001$

Table 20: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.013	0.010	4.636	0.001
Model 2	0.014	0.008	2.225	0.018

### ELOM 4&5 domain 2 (fine motor coordination and visual-motor integration)

For fine motor coordination and visual-motor integration (Table 21), Model 1 showed that fee band explained 10.5% of the variance ( $adjusted R^2 = 0.105$ ,  $F(4, 1428) = 43.19$ ,  $p < 0.001$ ). Model 2 increased this to 12.2% ( $adjusted R^2 = 0.122$ ,  $F(9, 1423) = 23.16$ ,  $p < 0.001$ ), with *materials and equipment* emerging as a significant positive predictor ( $b = 0.235$ ,  $p = 0.02$ ,  $\beta = 0.087$ ).

Table 21: Regression results with ELOM 4&5 domain 2 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	10.423	0.133	[10.161, 10.684]	< 0.001***	
Fee band 2	0.338	0.179	[-0.014, 0.689]	0.06	0.058
Fee band 3	0.678	0.190	[0.306, 1.05]	< 0.001***	0.109
Fee band 4	2.304	0.238	[1.837, 2.771]	< 0.001***	0.274
Fee band 5	3.079	0.314	[2.463, 3.694]	< 0.001***	0.263
<b>Model 2</b>					
(Intercept)	10.485	0.133	[10.225, 10.746]	< 0.001***	
Fee band 2	0.370	0.178	[0.021, 0.719]	0.038*	0.064
Fee band 3	0.598	0.189	[0.227, 0.969]	0.002**	0.096
Fee band 4	2.051	0.241	[1.578, 2.523]	< 0.001***	0.244
Fee band 5	2.668	0.320	[2.04, 3.297]	< 0.001***	0.228
Materials and equipment	0.235	0.101	[0.036, 0.433]	0.02*	0.087
Planning and assessment	0.097	0.089	[-0.078, 0.273]	0.276	0.036
Learning programme	-0.031	0.113	[-0.252, 0.19]	0.783	-0.011
Teaching strategies	0.058	0.105	[-0.148, 0.265]	0.579	0.022
Relationships and interactions	0.129	0.088	[-0.044, 0.302]	0.143	0.048

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 22: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.108	0.105	43.19	< .001
Model 2	0.128	0.122	23.16	< .001

### ***ELOM 4&5 Domain 3 (emergent numeracy and mathematics)***

For emergent numeracy and mathematics (Table 23), Model 1 showed fee band explaining 5.4% of the variance (*adjusted*  $R^2 = 0.054$ ,  $F(4, 1428) = 21.60$ ,  $p < .001$ ). Model 2 increased this to 7.1% (*adjusted*  $R^2 = 0.071$ ,  $F(9, 1423) = 13.2$ ,  $p < .001$ ), but none of the LPQA domains were significant predictors.

Table 23: Regression results with ELOM 4&5 domain 3 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	7.828	0.148	[7.538, 8.118]	< 0.001***	
Fee band 2	0.392	0.199	[0.002, 0.782]	0.049*	0.063
Fee band 3	0.433	0.210	[0.02, 0.846]	0.04*	0.064
Fee band 4	1.458	0.264	[0.94, 1.976]	< 0.001***	0.161
Fee band 5	2.855	0.348	[2.172, 3.538]	< 0.001***	0.226
<b>Model 2</b>					
(Intercept)	7.874	0.147	[7.584, 8.163]	< 0.001***	
Fee band 2	0.413	0.198	[0.026, 0.801]	0.037*	0.066
Fee band 3	0.404	0.210	[-0.009, 0.816]	0.055	0.060
Fee band 4	1.262	0.267	[0.738, 1.786]	< 0.001***	0.139
Fee band 5	2.457	0.356	[1.76, 3.155]	< 0.001***	0.194
Materials and equipment	0.186	0.112	[-0.035, 0.406]	0.098	0.064
Planning and assessment	0.177	0.099	[-0.018, 0.371]	0.075	0.060
Learning programme	0.091	0.125	[-0.155, 0.336]	0.469	0.031
Teaching strategies	0.136	0.117	[-0.093, 0.365]	0.243	0.047
Relationships and interactions	-0.156	0.098	[-0.348, 0.036]	0.111	-0.053

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 24: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.057	0.054	21.60	< .001
Model 2	0.077	0.071	13.2	< .001

#### ***ELOM 4&5 Domain 4 (cognition and executive functioning)***

For *cognition and executive functioning* (Table 25), Model 1 indicated that fee band explained 12.3% of the variance (*adjusted R*<sup>2</sup> = 0.123, *F*(4, 1428) = 51.10, *p* < .001). Model 2 increased this to 14.9% (*adjusted R*<sup>2</sup> = 0.149, *F*(9, 1423) = 28.76, *p* < .001). Teaching strategies were a significant, positive predictor (*b* = 0.245, *p* = .03,  $\beta$  = 0.083).

Table 25: Regression results with ELOM 4&5 domain 4 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	6.308	0.144	[6.026, 6.59]	< 0.001***	
Fee band 2	0.428	0.193	[0.049, 0.808]	0.027*	0.068
Fee band 3	1.333	0.205	[0.932, 1.735]	< 0.001***	0.196
Fee band 4	2.496	0.257	[1.992, 3.001]	< 0.001***	0.272
Fee band 5	3.780	0.339	[3.116, 4.445]	< 0.001***	0.296
<b>Model 2</b>					
(Intercept)	6.402	0.143	[6.122, 6.682]	< 0.001***	
Fee band 2	0.450	0.191	[0.075, 0.825]	0.019*	0.071
Fee band 3	1.219	0.203	[0.82, 1.618]	< 0.001***	0.179
Fee band 4	2.157	0.258	[1.65, 2.664]	< 0.001***	0.235
Fee band 5	3.214	0.344	[2.539, 3.888]	< 0.001***	0.252
Materials and equipment	0.171	0.109	[-0.042, 0.384]	0.115	0.058
Planning and assessment	0.137	0.096	[-0.052, 0.325]	0.154	0.046
Learning programme	-0.022	0.121	[-0.26, 0.215]	0.853	-0.008
Teaching strategies	0.245	0.113	[0.023, 0.466]	0.03*	0.083
Relationships and interactions	0.107	0.095	[-0.079, 0.292]	0.26	0.036

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 26: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.125	0.123	51.10	< .001
Model 2	0.154	0.149	28.76	< .001

**ELOM 4&5 Domain 5 (emergent literacy & language)**

For *emergent literacy and language* (Table 27), Model 1 showed fee band explained 4.3% of the variance (*adjusted R*<sup>2</sup> = 0.043, *F*(4, 1428) = 17.13, *p* < .001). Model 2 increased this to 5.3% (*adjusted R*<sup>2</sup> = 0.053, *F*(9, 1423) = 9.949, *p* < .001). The results indicated that Planning and Assessment was a significant, positive predictor (*b* = 0.331, *p* = .003,  $\beta$  = 0.101).

Table 27: Regression results with ELOM 4&5 domain 5 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band 1)	10.475	0.168	[10.147, 10.804]	< 0.001***	
Fee band 2	0.316	0.225	[-0.126, 0.758]	0.160	0.045
Fee band 3	0.115	0.239	[-0.353, 0.583]	0.629	0.015
Fee band 4	1.381	0.299	[0.794, 1.969]	< 0.001***	0.135
Fee band 5	2.782	0.395	[2.008, 3.557]	< 0.001***	0.195
<b>Model 2</b>					
(Intercept)	10.500	0.168	[10.171, 10.829]	< 0.001***	
Fee band 2	0.326	0.225	[-0.115, 0.767]	0.147	0.046
Fee band 3	0.123	0.239	[-0.346, 0.592]	0.607	0.016
Fee band 4	1.269	0.304	[0.673, 1.865]	< 0.001***	0.124
Fee band 5	2.495	0.405	[1.701, 3.288]	< 0.001***	0.175
Materials and equipment	0.010	0.128	[-0.24, 0.261]	0.936	0.003
Planning and assessment	0.331	0.113	[0.109, 0.552]	0.003**	0.101
Learning programme	0.014	0.142	[-0.265, 0.294]	0.92	0.004
Teaching strategies	0.162	0.133	[-0.098, 0.423]	0.222	0.049
Relationships and interactions	-0.208	0.111	[-0.427, 0.01]	0.061	-0.063

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 28: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.046	0.043	17.13	< .001
Model 2	0.059	0.053	9.949	< .001

### Summary of part 1 regression results

The analyses demonstrate that LPQA (v2) total score significantly predicts ELOM 4&5 total score, adding explanatory power beyond fee band alone. This suggests that the LPQA (v2) tool, as a composite measure, is a statistically significant predictor of overall early learning outcomes. However, numerous unknown factors, beyond programme quality, likely contribute to this relationship, and thus, the association between LPQA (v2) total scores and individual learning outcomes should be interpreted cautiously. For ELOM 4&5 domains, predictive relationships varied: Materials and Equipment predict significantly higher Fine Motor Coordination & Visual Motor Integration scores (ELOM 4&5 Domain 2); Teaching Strategies predict significantly higher Cognition & Executive Functioning scores (ELOM 4&5 Domain 4); and Planning and Assessment predict significantly higher Emergent Literacy & Language scores (ELOM 4&5 Domain 5). Taken together, these results provide evidence of criterion-related validity and support the tool's relevance for evaluating ELP quality.

### Part 2: Statistical analyses

A second series of hierarchical linear regression analyses was conducted using a revised fee variable. Due to a minor reduction in the sample size of the final dataset, these analyses are based on data from 1,388 ELPs across all nine provinces in South Africa.

### LPQA (v2) Total Score as a Predictor of ELOM 4&5 Total Score

Regression analyses first assessed whether the LPQA (v2) total score predicted the ELOM 4&5 total score, with fee band entered as a control variable in Model 1 and standardised LPQA (v2) total score added in Model 2 (See Table 29).

Model 1 was statistically significant, with fee band accounting for 11.2% of the variance in ELOM 4&5 total scores (*adjusted R*<sup>2</sup> = 0.112, *F*(4, 1383) = 44.58, *p* < .001).

In Model 2, adding LPQA (v2) total score significantly improved the model fit, explaining an additional 1% of the variance (*adjusted R*<sup>2</sup> = 0.122, *F*(5, 1382) = 39.47, *p* < .001). LPQA (v2) total score was a significant predictor (*b* = 1.153, *p* < .001, *β* = 0.110).

**Table 29: Regression results with ELOM 4&5 total score as criterion**

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band 1)	41.793	0.530	[40.753, 42.833]	< 0.001***	
Fee band 2	2.707	0.834	[1.072, 4.342]	0.001**	0.097
Fee band 3	2.697	0.806	[1.116, 4.279]	< 0.001***	0.101
Fee band 4	2.387	0.812	[0.793, 3.98]	0.003**	0.089
Fee band 5	10.238	0.797	[8.675, 11.801]	< 0.001***	0.390
<b>Model 2</b>					
(Intercept: Fee band 1)	42.137	0.534	[41.09, 43.184]	< 0.001***	
Fee band 2	2.721	0.829	[1.095, 4.347]	0.001**	0.097
Fee band 3	2.454	0.804	[0.877, 4.03]	0.002**	0.092
Fee band 4	2.022	0.812	[0.429, 3.616]	0.013*	0.075
Fee band 5	9.195	0.832	[7.563, 10.827]	< 0.001***	0.350
LPQA Total	1.153	0.280	[0.604, 1.702]	< 0.001***	0.110

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

**Table 30: Summary statistics**

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.114	0.112	44.58	< .001
Model 2	0.125	0.122	39.47	< .001

### LPQA (v2) domain scores as predictors of ELOM 4&5 scores

A series of regression analyses examined whether LPQA (v2) domain scores predicted ELOM 4&5 total and ELOM 4&5 domain scores. Fee band was controlled for in Model 1, and standardised LPQA (v2) domain scores were added in Model 2.

#### ELOM 4&5 total score

When the ELOM 4&5 total score was the criterion (Table 31), Model 1 showed fee band explaining 11.2% of the variance (*adjusted R*<sup>2</sup> = 0.112, *F*(4, 1383) = 44.58, *p* < 0.001). Model 2 increased this to 12.3% (*adjusted R*<sup>2</sup> = 0.123, *F*(9, 1378) = 22.60, *p* < 0.001), with teaching strategies emerging as a significant positive predictor (*b* = 0.898, *p* = 0.031, *β* = 0.085).

Table 31: Regression results with ELOM 4&5 total score as criterion

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band 1)	41.793	0.530	[40.753, 42.833]	< 0.001***	
Fee band 2	2.707	0.834	[1.072, 4.342]	0.001**	0.097
Fee band 3	2.697	0.806	[1.116, 4.279]	< 0.001***	0.101
Fee band 4	2.387	0.812	[0.793, 3.98]	0.003**	0.089
Fee band 5	10.238	0.797	[8.675, 11.801]	< 0.001***	0.390
<b>Model 2</b>					
(Intercept: Fee band1)	42.110	0.536	[41.058, 43.161]	< 0.001***	
Fee band2	2.685	0.830	[1.056, 4.314]	0.001**	0.096
Fee band3	2.490	0.811	[0.9, 4.08]	0.002**	0.093
Fee band4	2.071	0.816	[0.47, 3.672]	0.011*	0.077
Fee band5	9.267	0.836	[7.627, 10.908]	< 0.001***	0.353
Materials and equipment	0.441	0.397	[-0.338, 1.221]	0.267	0.042
Planning and assessment	0.428	0.353	[-0.265, 1.12]	0.226	0.041
Learning programme	-0.001	0.441	[-0.866, 0.864]	0.999	0.000
Teaching strategies	0.898	0.417	[0.081, 1.716]	0.031*	0.085
Relationships and interactions	-0.449	0.346	[-1.127, 0.23]	0.195	-0.043

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 32: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.114	0.112	44.58	< .001
Model 2	0.129	0.123	22.60	< .001

#### ELOM 4&5 Domain 1 (gross motor development)

For *gross motor development* (Table 33), Model 1 indicated that fee band accounted for 0.9% of the variance (*adjusted R*<sup>2</sup> = 0.009, *F*(4, 1383) = 4.165, *p* < .001). None of the LPQA domains were significant, and these additional predictors did not improve the model fit (*adjusted R*<sup>2</sup> = 0.006, *F*(9, 1378) = 1.987, *p* = .037).

Table 33: Regression results with ELOM 4&5 domain 1 score as criterion

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band1)	7.957	0.151	[7.661, 8.253]	< 0.001***	
Fee band2	0.175	0.237	[-0.291, 0.64]	0.461	0.023
Fee band3	-0.302	0.229	[-0.752, 0.148]	0.189	-0.042
Fee band4	-0.687	0.231	[-1.141, -0.234]	0.003**	-0.095
Fee band5	0.112	0.227	[-0.333, 0.557]	0.621	0.016
<b>Model 2</b>					
(Intercept: Fee band1)	7.937	0.154	[7.635, 8.239]	< 0.001***	
Fee band2	0.178	0.238	[-0.289, 0.645]	0.455	0.024
Fee band3	-0.291	0.233	[-0.747, 0.165]	0.211	-0.040
Fee band4	-0.662	0.234	[-1.121, -0.203]	0.005**	-0.091
Fee band5	0.169	0.240	[-0.301, 0.64]	0.481	0.024
Materials and equipment	-0.030	0.114	[-0.254, 0.193]	0.790	-0.011
Planning and assessment	-0.069	0.101	[-0.267, 0.13]	0.498	-0.024
Learning programme	0.100	0.126	[-0.149, 0.348]	0.431	0.035
Teaching strategies	-0.006	0.120	[-0.24, 0.229]	0.962	-0.002
Relationships and interactions	-0.068	0.099	[-0.263, 0.126]	0.490	-0.024

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 34: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.012	0.009	4.165	< .001
Model 2	0.013	0.006	1.987	0.037

#### ELOM 4&5 Domain 2 (fine motor coordination and visual motor integration)

For *fine motor coordination and visual motor integration* (Table 35), Model 1 showed fee band explaining 13% of the variance (*adjusted R*<sup>2</sup> = 0.130, *F*(4, 1383) = 52.72, *p* < .001). In Model 2, *adjusted R*<sup>2</sup> increased; however, no individual LPQA domain score reached statistical significance (*adjusted R*<sup>2</sup> = 0.139, *F*(9, 1378) = 25.87, *p* < .001).

Table 35: Regression results with ELOM 4&5 domain 2 score as criterion

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band1)	9.839	0.132	[9.581, 10.098]	< 0.001***	
Fee band2	0.954	0.207	[0.547, 1.36]	< 0.001***	0.136
Fee band3	0.947	0.200	[0.554, 1.34]	< 0.001***	0.141
Fee band4	1.227	0.202	[0.831, 1.623]	< 0.001***	0.181
Fee band5	2.844	0.198	[2.455, 3.232]	< 0.001***	0.431
<b>Model 2</b>					
(Intercept: Fee band1)	9.939	0.133	[9.678, 10.201]	< 0.001***	
Fee band2	0.941	0.207	[0.536, 1.346]	< 0.001***	0.134
Fee band3	0.857	0.202	[0.462, 1.253]	< 0.001***	0.128
Fee band4	1.111	0.203	[0.713, 1.509]	< 0.001***	0.164
Fee band5	2.564	0.208	[2.156, 2.973]	< 0.001***	0.389
Materials and equipment	0.163	0.099	[-0.031, 0.357]	0.099	0.062
Planning and assessment	0.033	0.088	[-0.139, 0.205]	0.708	0.012
Learning programme	-0.076	0.110	[-0.291, 0.139]	0.489	-0.029
Teaching strategies	0.162	0.104	[-0.042, 0.365]	0.119	0.061
Relationships and interactions	0.072	0.086	[-0.097, 0.241]	0.403	0.027

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 36: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.132	0.130	52.72	< .001
Model 2	0.145	0.139	25.87	< .001

### ELOM 4&5 domain 3 (emergent numeracy and mathematics)

For emergent numeracy and mathematics (Table 37), Model 1 showed that fee band explained 5.6% of the variance (*adjusted R*<sup>2</sup> = 0.056, *F*(4, 1428) = 21.19, *p* < 0.001).

Table 37: Regression results with ELOM 4&5 domain 3 score as criterion

Predictor	b	SE	95% CI	p	$\beta$
<b>Model 1</b>					
(Intercept: Fee band1)	7.642	0.151	[7.347, 7.938]	< 0.001***	
Fee band2	0.657	0.237	[0.192, 1.121]	0.006**	0.085
Fee band3	0.595	0.229	[0.146, 1.044]	0.009**	0.081
Fee band4	0.306	0.231	[-0.146, 0.759]	0.185	0.041
Fee band5	2.004	0.226	[1.56, 2.448]	< 0.001***	0.277
<b>Model 2</b>					
(Intercept: Fee band1)	7.714	0.152	[7.415, 8.012]	< 0.001***	
Fee band2	0.663	0.236	[0.2, 1.125]	0.005**	0.086
Fee band3	0.562	0.230	[0.111, 1.014]	0.015*	0.076
Fee band4	0.243	0.232	[-0.212, 0.697]	0.295	0.033
Fee band5	1.764	0.237	[1.298, 2.23]	< 0.001***	0.243
Materials and equipment	0.174	0.113	[-0.048, 0.395]	0.124	0.060
Planning and assessment	0.108	0.100	[-0.089, 0.305]	0.281	0.037
Learning programme	0.056	0.125	[-0.19, 0.301]	0.655	0.019
Teaching strategies	0.214	0.118	[-0.018, 0.446]	0.071	0.074
Relationships and interactions	-0.213	0.098	[-0.406, -0.021]	0.030*	-0.074

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance,  $\beta$  (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 38: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.059	0.056	21.62	< .001
Model 2	0.076	0.070	12.52	< .001

#### ELOM 4&5 Domain 4 (cognition and executive functioning)

For *cognition and executive functioning* (Table 39), Model 1 indicated that fee band explained 14.8% of the variance (*adjusted R*<sup>2</sup> = 0.148, *F*(4, 1383) = 61.37, *p* < .001). Model 2 increased this to 16.7% (*adjusted R*<sup>2</sup> = 0.167, *F*(9, 1378) = 31.78, *p* < .001). Teaching strategies were a significant, positive predictor (*l* = 0.297, *p* = .009,  $\beta$  = 0.100).

Table 39: Regression results with ELOM 4&5 domain 4 score as criterion

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band1)	5.946	0.146	[5.66, 6.233]	< 0.001***	
Fee band2	0.733	0.230	[0.283, 1.184]	0.001**	0.093
Fee band3	1.349	0.222	[0.914, 1.785]	< 0.001***	0.180
Fee band4	1.468	0.224	[1.029, 1.906]	< 0.001***	0.193
Fee band5	3.341	0.220	[2.91, 3.772]	< 0.001***	0.452
<b>Model 2</b>					
(Intercept: Fee band1)	6.088	0.147	[5.8, 6.377]	< 0.001***	
Fee band2	0.714	0.228	[0.267, 1.16]	0.002**	0.091
Fee band3	1.226	0.222	[0.79, 1.662]	< 0.001***	0.163
Fee band4	1.310	0.224	[0.871, 1.749]	< 0.001***	0.173
Fee band5	2.939	0.229	[2.489, 3.389]	< 0.001***	0.398
Materials and equipment	0.136	0.109	[-0.077, 0.35]	0.211	0.046
Planning and Assessment	0.088	0.097	[-0.102, 0.277]	0.366	0.030
Learning programme	-0.026	0.121	[-0.263, 0.211]	0.830	-0.009
Teaching Strategies	0.297	0.114	[0.073, 0.521]	0.009**	0.100
Relationships and interactions	0.026	0.095	[-0.16, 0.212]	0.785	0.009

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

Table 40: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.151	0.148	61.37	< .001
Model 2	0.172	0.167	31.78	< .001

#### Domain 5 (emergent literacy and language)

For *emergent literacy and language* (Table 41), Model 1 showed that fee band explained 4.8% of the variance (adjusted  $R^2 = 0.048$ ,  $F(4, 1383) = 18.57$ ,  $p < .001$ ). Model 2 increased this to 5.5% (adjusted  $R^2 = 0.055$ ,  $F(9, 1378) = 10.03$ ,  $p < .001$ ). The results indicated that planning and assessment were a significant, positive predictor ( $b = 0.268$ ,  $p = .020$ ,  $\beta = 0.081$ ) and relationships and interactions were a significant, negative predictor ( $b = -0.264$ ,  $p = .019$ ,  $\beta = -0.081$ ).

Table 41: Regression results with ELOM 4&5 domain 5 score as criterion

Predictor	b	SE	95% CI	p	β
<b>Model 1</b>					
(Intercept: Fee band1)	10.409	0.172	[10.072, 10.746]	< 0.001***	
Fee band2	0.189	0.271	[-0.342, 0.719]	0.486	0.022
Fee band3	0.108	0.262	[-0.406, 0.621]	0.681	0.013
Fee band4	0.073	0.264	[-0.444, 0.59]	0.781	0.009
Fee band5	1.937	0.259	[1.43, 2.444]	< 0.001***	0.235
<b>Model 2</b>					
(Intercept: Fee band1)	10.431	0.174	[10.089, 10.773]	< 0.001***	
Fee band2	0.189	0.270	[-0.34, 0.719]	0.483	0.022
Fee band3	0.135	0.264	[-0.382, 0.652]	0.609	0.016
Fee band4	0.069	0.266	[-0.452, 0.59]	0.795	0.008
Fee band5	1.831	0.272	[1.297, 2.365]	< 0.001***	0.222
Materials and equipment	-0.002	0.129	[-0.255, 0.252]	0.990	0.000
Planning and Assessment	0.268	0.115	[0.043, 0.493]	0.020*	0.081
Learning programme	-0.054	0.143	[-0.336, 0.227]	0.705	-0.017
Teaching Strategies	0.232	0.136	[-0.034, 0.498]	0.088	0.070
Relationships and interactions	-0.264	0.112	[-0.485, -0.044]	0.019*	-0.081

Note: *b* = unstandardised coefficient, *SE* = standard error, *95% CI* = confidence range, *p* = statistical significance, *β* (beta) = standardised coefficient.

\**p* < .05, \*\**p* < .01, \*\*\**p* < .001

#### 42: Summary statistics

	R-squared	Adjusted r-squared	F. Statistic	p-value
Model 1	0.051	0.048	18.57	< .001
Model 2	0.061	0.055	10.03	< .001

#### Summary of part 2 results

Consistent with the previous round of regression analyses, the LPQA (v2) total score significantly predicted the ELOM 4&5 total score, adding explanatory power beyond fee band alone. Teaching strategies was also a significant predictor of ELOM 4&5 total score. For ELOM 4&5 domains, predictive relationships varied: teaching strategies predicted significantly higher cognition and executive functioning scores (ELOM 4&5 domain 4); and planning and assessment predicted significantly higher emergent literacy and language scores (ELOM 4&5 domain 5). Unexpectedly, higher scores for relationships and interactions predicted significantly lower emergent literacy and mathematics (ELOM 4&5 domain 3) and emergent literacy and language scores (ELOM 4&5 domain 5). However, the effect sizes remain small, consistent with findings from international literature and meta-analyses (Ulferts et al., 2019; von Suchodoletz et al., 2023; Zaslow et al., 2016), suggesting that the LPQA (v2) tool's ability to explain variance in learning outcomes is limited. Modest associations between ELP quality and learning outcomes may be attributed to numerous child- and family-level factors, such as background characteristics of the child and family, which likely influence ELOM 4&5 scores but were not accounted for in the current analyses (Ulferts et al., 2019).



## CONCLUSION

There was robust evidence for the reliability of the LPQA (v2) measure, as demonstrated by high Cronbach's Alpha and McDonald's Omega values. CFA strongly supported the internal structure of the LPQA (v2) domains, though the perfect fit observed for Relationships and Interactions raises potential concerns about model overfit, warranting further scrutiny. Hierarchical linear regression analyses were used to investigate the ability of the LPQA (v2) domain and total scores to predict early school readiness, as assessed by the ELOM 4&5. Although the LPQA (v2) total score shows a statistically significant relationship with ELOM 4&5 total score after controlling for fee band, its predictive effect is minimal, explaining only a small portion of variance and underscoring the substantial influence of numerous other unknown factors on child outcomes, as is commonly seen in international literature.

## FUTURE RESEARCH

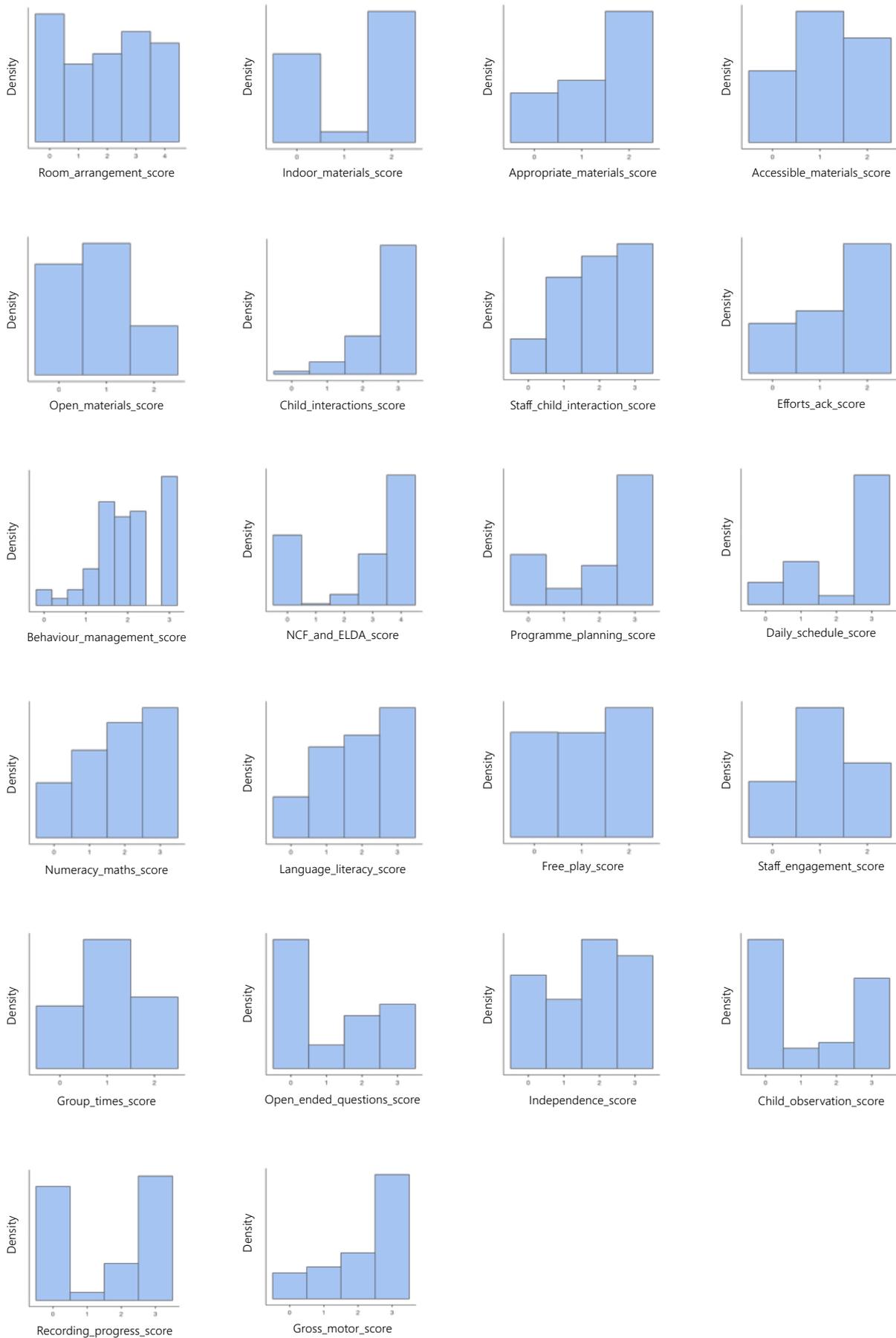
To build on the current findings, future research should prioritise ongoing validation efforts to ensure the stability and robustness of the new factor structure as well as validate recent modifications to the LPQA (v2). Additionally, investigating inter-rater bias is essential to determine whether certain raters consistently assign higher or lower scores to ELPs, which could undermine the measure's objectivity. Finally, researchers should examine whether pro-rating leads to range restriction in LPQA scores. Addressing these areas will strengthen the LPQA (v2)'s reliability, fairness and utility as a tool for assessing and improving early learning environments.

# REFERENCES

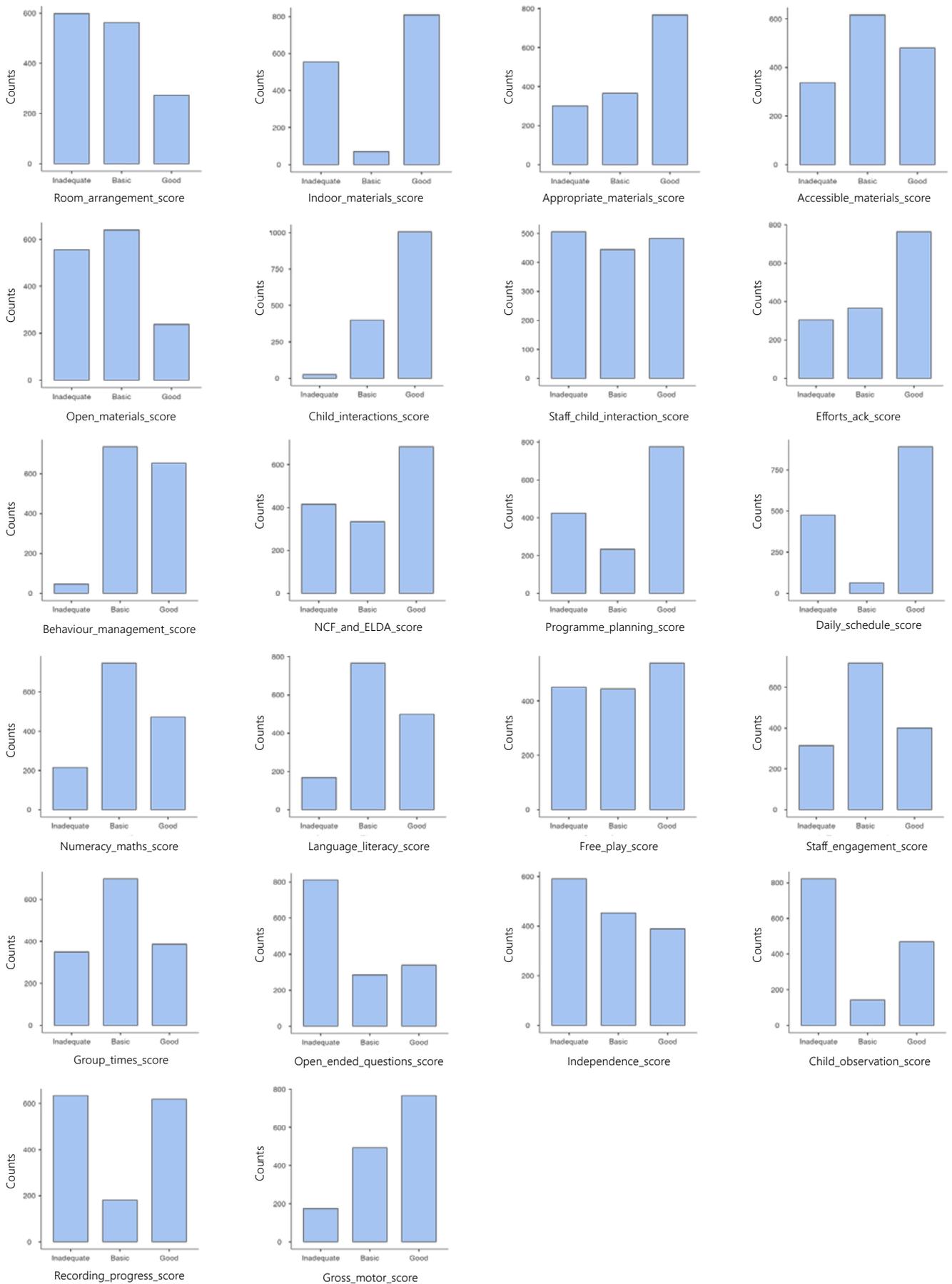
- About, F. E. (2006). Evaluation of an early childhood preschool programme in rural Bangladesh. *Early Childhood Research Quarterly*, 2, 46 – 60. <https://doi.org/10.1016/j.ecresq.2006.01.008>
- Alexander, L. (2020). Audit Quality Assessment Scale Exploratory Factor Analysis Preliminary Results Report. Cape Town: Ikapadata.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual Differences*, 42(5), 815–824. <https://doi.org/10.1016/j.paid.2006.09.018>
- Britto, P. R., Yoshikawa, H., & Boller, K. (2011). Quality of Early Childhood Development Programs in Global Contexts: Rationale for investment, conceptual framework and implications for equity. *Social Policy Report*, 25(2), 1–23. <https://doi.org/10.1002/j.2379-3988.2011.tb00067.x>
- Burchinal, M., Garber, K., Foster, T., Bratsch-Hines, M., Franco, X., & Peisner-Feinberg, E. (2021). Relating early care and education quality to preschool outcomes: The same or different models for different outcomes? *Early Childhood Research Quarterly*, 55, 35-51.
- Burchinal, M., Zaslow, M., & Tarullo, L. (2016). Quality thresholds, features and dosage in early care and education: Secondary data analyses of child outcomes. *Monographs of the Society for Research in Child Development*, 81, 1–128. <https://doi.org/10.1111/mono.12238>
- Cabell, S. Q., Justice, L. M., McGinty, A. S., DeCoster, J., & Forston, L. D. (2015). Teacher-child conversations in preschool classrooms: Contributions to children’s vocabulary development. *Early Childhood Research Quarterly*, 30, 80–92. <http://dx.doi.org/10.1016/j.ecresq.2014.09.004>
- Center on the Developing Child at Harvard University. (2016). From best practices to breakthrough impacts: A science-based approach to building a more promising future for young children and families. Retrieved from [www.developingchild.harvard.edu](http://www.developingchild.harvard.edu).
- Dawes, A., Biersteker, L., Girdwood, E., Snelling, M.J.T.L., Tredoux, C.G. et al (2020). Early Learning Outcomes Measure. Technical Manual. Claremont, Cape Town: The Innovation Edge. <http://www.innovationedge.org.za/>
- Department of Basic Education (2009) National early learning and development standards for children birth to four (NELDS) Pretoria.
- Department of Basic Education (2015) South African National Curriculum Framework for Children Birth to Four. Pretoria.
- Draper, C. E., Yousafzai, A. K., McCoy, D. C., Cuartas, J., Obradović, J., Bhopal, S., ... & Okely, A. D. (2024). The next 1000 days: building on early investments for the health and development of young children. *The Lancet*, 404(10467), 2094-2116.
- Durlak, J., Weissberg, R. P., Dymnicki, A. B., Taylor, R. D., & Schellinger, K. B. (2011). The impact of enhancing students’ social and emotional learning: A meta-analysis of school-based universal interventions. *Child Development*, 82(1), 405–432. <https://doi.org/10.1111/j.1467-8624.2010.01564.x>
- Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., Pagani, L. S., Feinstein, L., Engel, M., Brooks-Gunn, J., Sexton, H., Duckworth, K., & Japel, C. (2007). School Readiness and Later Achievement. *Developmental Psychology*, 43(6), 1428–1446. <https://doi.org/10.1037/0012-1649.43.6.1428>
- Dunn, T. J., Baguley, T., & Brunsden, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412.
- Elkind, D. (1986). Formal education and early childhood education: An essential difference. *The Phi Delta Kappan* 67, no. 9 (1986):631-636. Cited in Jenkins & Duncan (2017).
- Edwards, S. & Cutter Mackenzie, A. (2013). Pedagogical play types: What do they suggest for learning about sustainability in early childhood education? *International Journal of Early Childhood*, 45, 327 – 346.
- Grisham-Brown, J., Hallam, R. & Brookshire, R. (2006) Using authentic assessment to evidence children’s progress toward early learning standards. *Early childhood Education Journal*, 34(1), 45-47.
- Harms, T., Clifford, R. M., & Cryer, D. (2015). Early childhood environment rating scale – Third edition. Teachers College Press.
- Henry, J., & Giese, S. (2023). *Reviewing the socio-economic gradient in learning outcomes for children who participated in the Thrive by Five Index*. DataDrive2030.Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modelling*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Jenkins, J. M. & Duncan, G.R. (2017). Do pre-kindergarten curricula matter? The current state of scientific knowledge on pre-kindergarten effects. In Phillips, D., Lipsey, M., Dodge, et al. Washington DC: Brookings Institute p 37 – 44
- Jensen, H., Pyle, A., Zosh, J. M., Ebrahim, H. B., Scherman, A. Z., Reunamo, J., & Hamre, B. K. (2019). Play facilitation: The science behind the art of engaging young children (white paper). Denmark: The LEGO Foundation
- Lonigan, C. J., Burgess, S. R., & Anthony, J. L. (2000). Development of emergent literacy and early reading skills in preschool children: Evidence from a latent-variable longitudinal study. *Developmental Psychology*, 36(5), 596.
- MELQO Study Team (2018). Measure of Early Learning Environment (MELE): Manual. Paris: UNESCO.
- Montie, J. E., Xiang, Z., & Schweinhart, L. J. (2006). Preschool experience in 10 countries: Cognitive and language performance at age 7. *Early Childhood Research Quarterly*, 21(3), 313-331.

- Opel, A., Ameer, S. S., & Aboud, F. E. (2009). The effect of preschool dialogic reading on vocabulary among rural Bangladeshi children. *International Journal of Educational Research*, 48(1), 12-20.
- Petrocelli, J. V. (2003). Hierarchical multiple regression in counseling research: Common problems and possible remedies. *Measurement and Evaluation in Counseling and Development*, 36(1), 9–22. <https://doi.org/10.1080/07481756.2003.12069076>
- Phillips, D., Lipsey, M., Dodge, K., Haskings, R., Bassok, D. et al. (2017). Puzzling It Out: The current state of scientific knowledge on pre-kindergarten effects- A consensus statement. Washington DC: Brookings Institute.
- Pianta, R, La Paro, K & Hamre, B ( 2008) Classroom Assessment Scoring System Pre K. Baltimore: Paul H Brooks Publishing Co.
- Rao, N., Sun, J., Wong, J.M.S., Weekes, B., Ip, P., Shaeffer, S., Young, M., Bray, M., Chen, E. & Lee, D. (2014). Early childhood development and cognitive development in developing countries: A rigorous literature review. *Department for International Development*.
- Shala, M. (2013). The impact of preschool social-emotional development on academic success of elementary school students. *Psychology* (4).
- Siraj, I., Kingston, D., Neilsen-Hewett, C., Howard, S., Melhuish, E., & de Rosnay, M. (2017). *Fostering effective early learning: A review of the current international evidence considering quality in early childhood education and care programmes—in delivery, pedagogy and child outcomes*. University of Wollongong. Retrieved from <https://education.nsw.gov.au/early-childhood-education/whats-happening-in-the-early-childhood-education-sector/data-and-research/feel-study-literature-review-final.pdf>
- Spilt, J.L., Koomen, H.M.Y., & Harrison, L.J. (2014). Language Development in the Early School Years: The Importance of Close Relationships with Teachers. *Developmental Psychology*. DOI: 10.1037/a0038540. ©2015 by the American Psychological Association (APA).
- Stipek D., Feiler, R., Daniels, D., & Milburn, S. (1995). Effects of different instructional approaches on young children's achievement and motivation. *Child Development*, 66(1), 209–223.
- Sylva, K., Taggart, B., Siraj-Blatchford, I., Totsika, V., Ereky-Steven, K., Gilden, R., et al. (2007). Curricular quality and day-to-day learning activities in pre-school. *International Journal of Early Years Education*, 15(1), 49–65.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53–55. <https://doi.org/10.5116/ijme.4dfb.8dfd>
- Tavakol, M. & Wetzal, A. (2020). Factor analysis: A means for theory and instrument development in support of construct validity. *International Journal of Medical Education*, 11, 245–247. doi: 10.5116/ijme . 5f96.0f4a.
- Trawick-Smith, J., Wolff, J., Koschel, M. et al. (2015). Effects of toys on the play quality of preschool children: Influence of gender, ethnicity, and socioeconomic status. *Early Childhood Education Journal* 43: 249.
- Tredoux, C.G. & Mattes, F. (2022). Factor structure, internal consistency, and criterion validity of the ELOM Quality Audit Scales. Unpublished: Psychology Department, University of Cape Town.
- Ulferts, H., Wolf, K. M., & Anders, Y. (2019). Impact of process quality in early childhood education and care on academic outcomes: Longitudinal meta-analysis. *Child Development*, 90(5), 1474–1489. <https://doi.org/10.1111/cdev.13296>
- UNESCO (2017). Overview MELQO (*Measuring early learning quality and outcomes*). Paris: UNESCO.
- Von Suchodoletz, A., Lee, D. S., Henry, J., Tamang, S., Premachandra, B. & Yoshikawa, H. (2023) Early childhood education and care quality and associations with child outcomes: A meta-analysis. *Plos One*, 18(5), Article e0285985. <https://doi.org/10.1371/journal.pone.0293056>
- Wolf, S., Raza, M., Kim, S., Aber, J. L., Behrman, J., & Seidman, E. (2018). Measuring and predicting process quality in Ghanaian pre-primary classrooms using the Teacher Instructional Practices and Processes System (TIPPS). *Early Childhood Research Quarterly*, 45, 18–30. <https://doi.org/10.1016/j.ecresq.2018.05.003>
- Zaslow, M., Anderson, R., Redd, Z., Wessel, J., Daneri, P., Green, K., Cavadel, E. W., Tarullo, L., Burchinal, M., & Martinez-Beck, I. (2016). Quality thresholds, features, and dosage in early care and education: Introduction and literature review. *Monographs of the Society for Research in Child Development*, 81(2), 7–26. <https://doi.org/10.1111/mono.12236>
- Zosh, J. M.; Hirsh-Pasek, J., Hopkins, E. J., Jensen, H., Liu, C., Neale, C., Solis, L. & Whitebread, D. (2018). Accessing the inaccessible: Redefining play as a spectrum. *Frontiers in Psychology*, 9, Article 1124.

## Appendix A –LPQA (v2) item score distributions

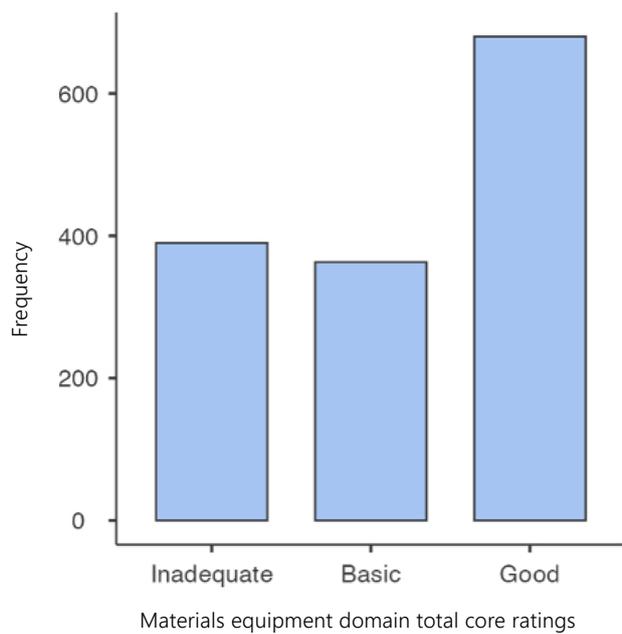


## Appendix B – Distributions of LPQA (v2) item scores by quality category

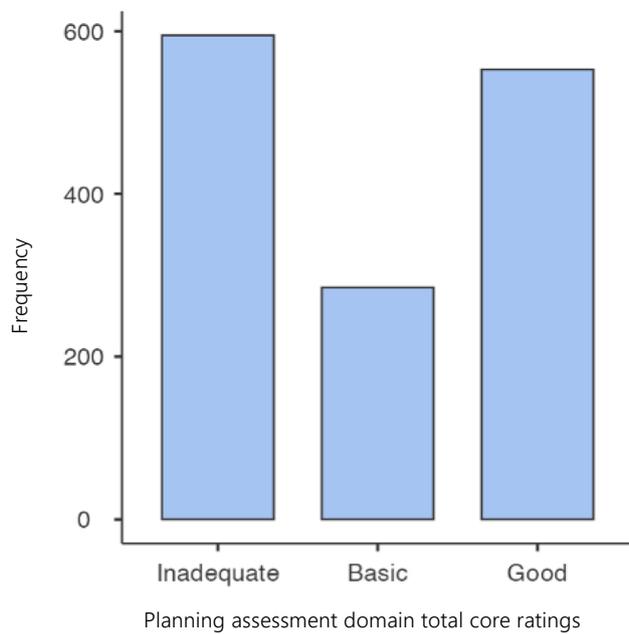


### Appendix C – Distributions of LPQA (v2) domain and total scores by quality category

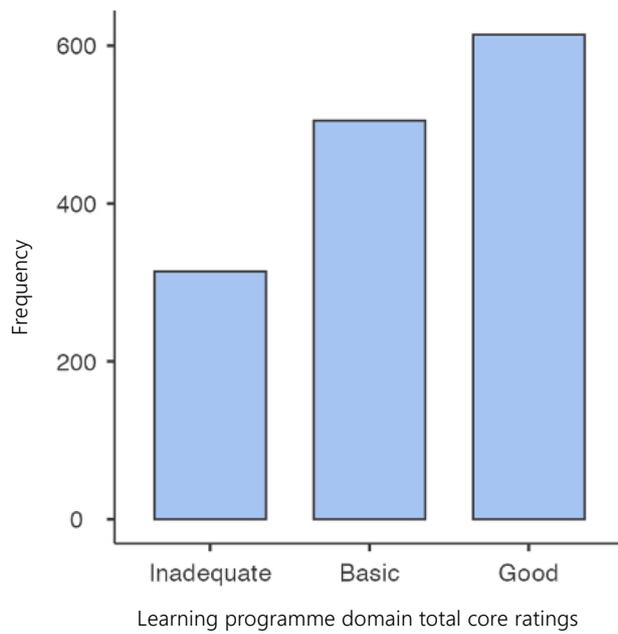
LPQA (v2) Domain 1: Materials and equipment



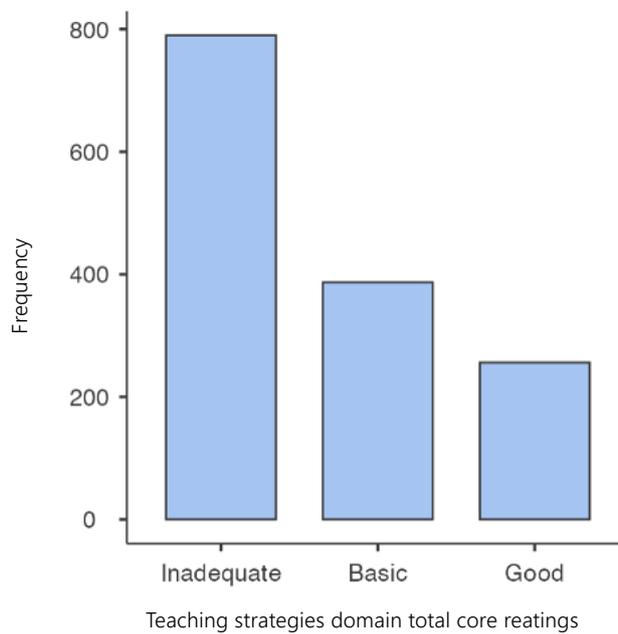
LPQA (v2) Domain 2: Planning and assessment



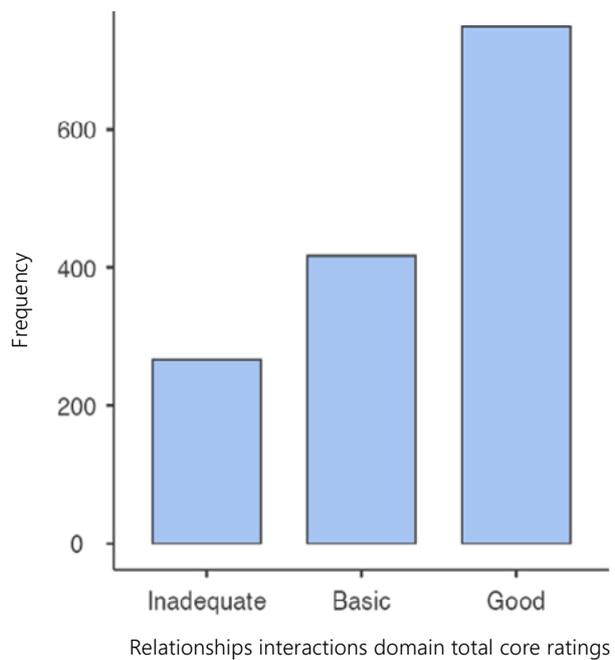
LPQA (v2) Domain 3: Learning programme



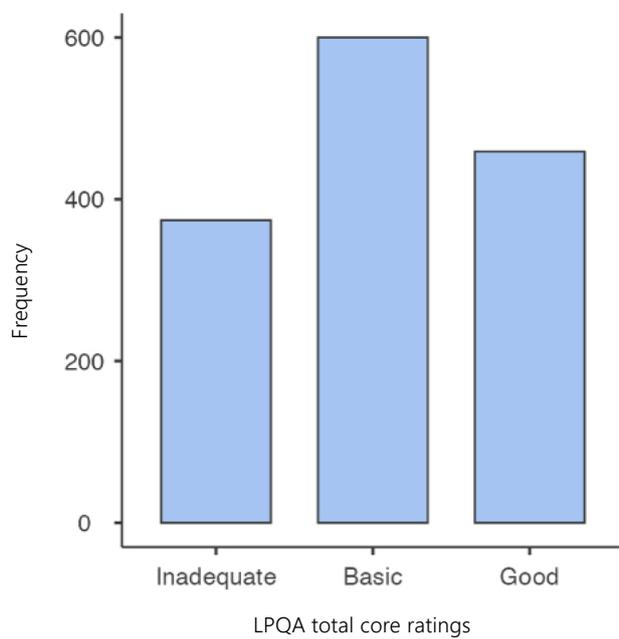
LPQA (v2) Domain 4: Teaching strategies



### LPQA (v2) Domain 5: Relationships and interactions

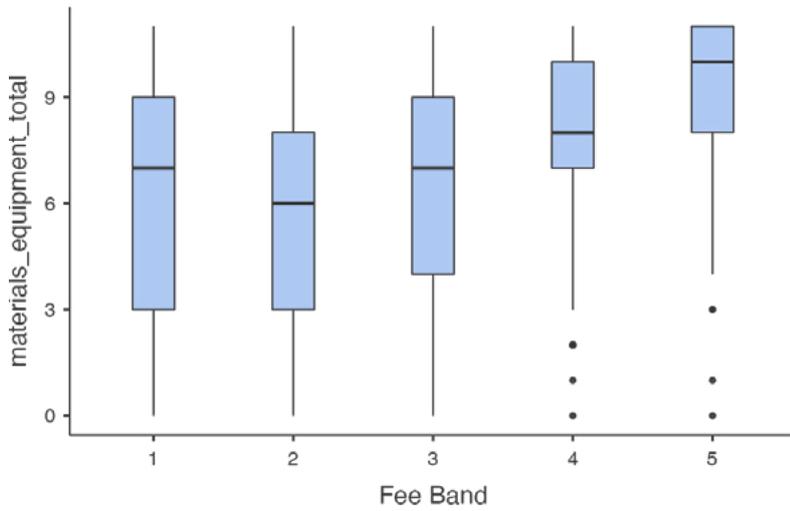


### LPQA (v2) Total score distribution

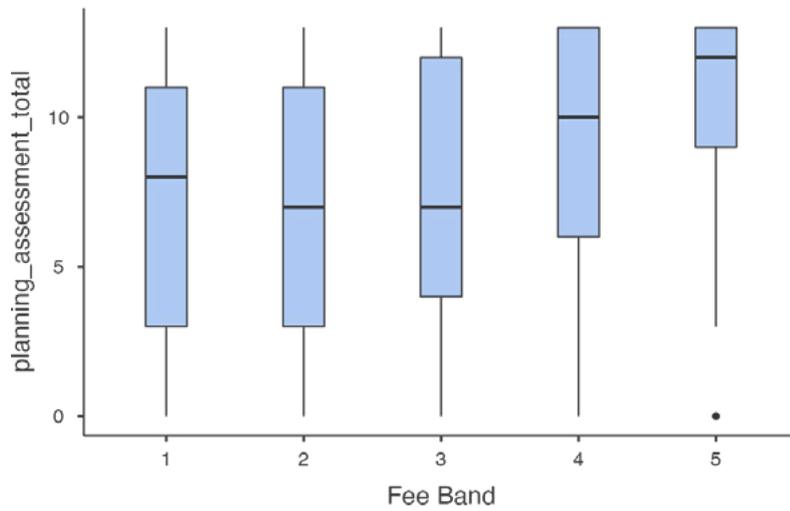


## Appendix D – Boxplots of LPQA (v2) domain scores by ELP fee band

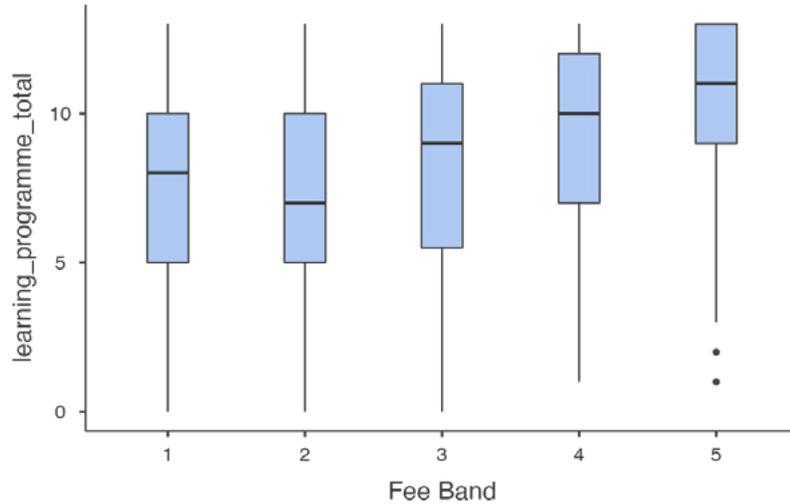
Domain 1: Materials and equipment



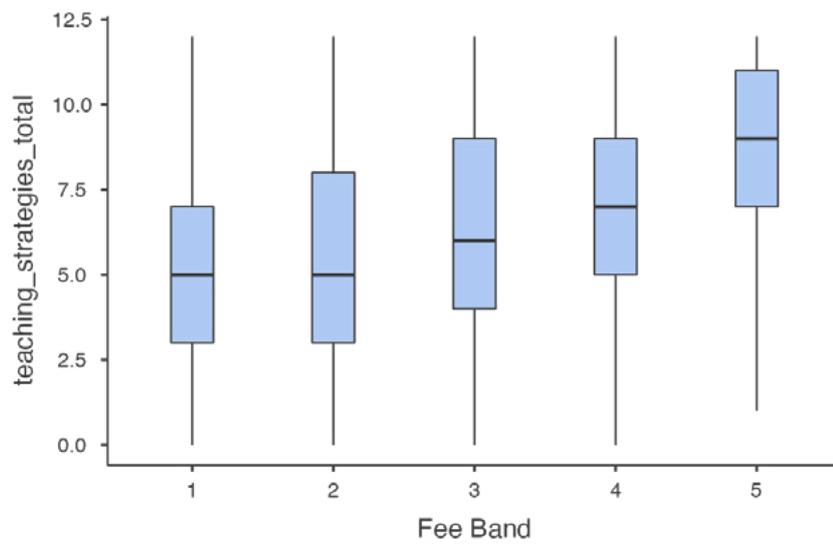
Domain 2: Planning and assessment



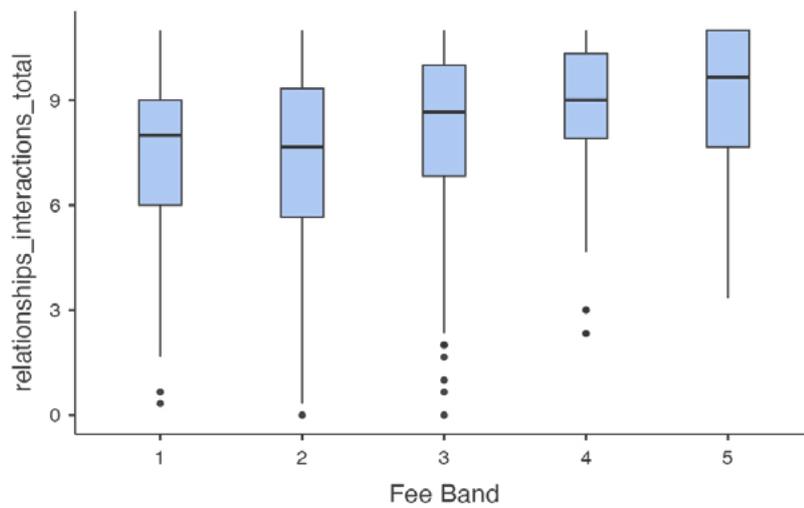
Domain 3: Learning programme



#### Domain 4: Teaching strategies



#### Domain 5: Relationships and interactions



## Appendix E – CFA statistical output

Domain 1: Materials and equipment

lavaan 0.6-19 ended normally after 19 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	10

Number of observations	1433
------------------------	------

Model Test User Model:

Test statistic	34.935
Degrees of freedom	5
P-value (Chi-square)	0.000

Model Test Baseline Model:

Test statistic	2094.276
Degrees of freedom	10
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.986
Tucker-Lewis Index (TLI)	0.971

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-7944.355
Loglikelihood unrestricted model (H1)	-7926.887
Akaike (AIC)	15908.710
Bayesian (BIC)	15961.385
Sample-size adjusted Bayesian (SABIC)	15929.618

Root Mean Square Error of Approximation:

RMSEA	0.065
90 Percent confidence interval - lower	0.045
90 Percent confidence interval - upper	0.086
P-value H <sub>0</sub> : RMSEA ≤ 0.050	0.101
P-value H <sub>0</sub> : RMSEA ≥ 0.080	0.119

Standardised Root Mean Square Residual:

SRMR	0.022
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
M_E = ~						
indr_mtrls_scr	1.000				0.724	0.755
apprprt_mtrls_	0.873	0.034	25.937	0.000	0.632	0.790
accssbl_mtrls_	0.601	0.030	19.980	0.000	0.435	0.581
open_mtrls_scr	0.654	0.029	22.816	0.000	0.474	0.667
gross_mottr_scr	0.809	0.043	18.856	0.000	0.585	0.548

Variances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.indr_mtrls_scr	0.396	0.021	18.801	0.000	0.396	0.430
.apprprt_mtrls_	0.240	0.014	16.836	0.000	0.240	0.375
.accssbl_mtrls_	0.371	0.016	23.810	0.000	0.371	0.662
.open_mtrls_scr	0.279	0.013	22.045	0.000	0.279	0.554
.gross_mottr_scr	0.797	0.033	24.289	0.000	0.797	0.699
M_E	0.524	0.034	15.209	0.000	1.000	1.000

Domain 2: Planning and assessment

lavaan 0.6-19 ended normally after 25 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	9

Number of observations	1433
------------------------	------

Model Test User Model:

Test statistic	2.705
Degrees of freedom	1
P-value (Chi-square)	0.100

Model Test Baseline Model:

Test statistic	1752.959
Degrees of freedom	6
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.999
Tucker-Lewis Index (TLI)	0.994

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-9231.918
Loglikelihood unrestricted model (H1)	-9230.566
Akaike (AIC)	18481.837
Bayesian (BIC)	18529.245
Sample-size adjusted Bayesian (SABIC)	18500.655

Root Mean Square Error of Approximation:

RMSEA	0.034
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.087
P-value H <sub>0</sub> : RMSEA ≤ 0.050	0.598
P-value H <sub>0</sub> : RMSEA ≥ 0.080	0.083

Standardised Root Mean Square Residual:

SRMR	0.006
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
Factor =~						
NCF_nd_ELDA_sc	1.000				1.147	0.672
prgrmm_plnng_	0.849	0.045	18.773	0.000	0.974	0.792
chld_bsrvtn_sc	0.727	0.041	17.876	0.000	0.834	0.615
rcrdng_prgrss_	0.714	0.041	17.240	0.000	0.819	0.588

Covariances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.chld_observation_score ~ .rcrdng_prgrss_	0.507	0.047	10.895	0.000	0.507	0.421

Variances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.NCF_nd_ELDA_sc	1.599	0.085	18.885	0.000	1.599	0.549
.prgrmm_plnng_	0.563	0.048	11.705	0.000	0.563	0.372
.chld_bsrvtn_sc	1.147	0.055	20.741	0.000	1.147	0.622
.rcrdng_prgrss_	1.266	0.059	21.429	0.000	1.266	0.654
Factor	1.316	0.109	12.069	0.000	1.000	1.000

### Domain 3: Learning programme

lavaan 0.6-19 ended normally after 20 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	10
Number of observations	1433

#### Model Test User Model:

Test statistic	22.421
Degrees of freedom	5
P-value (Chi-square)	0.000

#### Model Test Baseline Model:

Test statistic	1818.220
Degrees of freedom	10
P-value	0.000

#### User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.990
Tucker-Lewis Index (TLI)	0.981

#### Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-8799.505
Loglikelihood unrestricted model (H1)	-8788.294
Akaike (AIC)	17619.010
Bayesian (BIC)	17671.685
Sample-size adjusted Bayesian (SABIC)	17639.918

#### Root Mean Square Error of Approximation:

RMSEA	0.049
90 Percent confidence interval - lower	0.030
90 Percent confidence interval - upper	0.071
P-value H <sub>0</sub> : RMSEA ≤ 0.050	0.479
P-value H <sub>0</sub> : RMSEA ≥ 0.080	0.008

#### Standardised Root Mean Square Residual:

SRMR	0.021
------	-------

#### Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
L_P = ~						
daly_schdl_scr	1.000				0.688	0.611
nmercy_mths_scr	1.188	0.056	21.035	0.000	0.817	0.775
lngg_ltrcy_scr	1.201	0.056	21.260	0.000	0.826	0.806
free_play_scr	0.577	0.038	14.981	0.000	0.397	0.479
group_tims_scr	0.511	0.033	15.312	0.000	0.352	0.491

Variances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.daly_schdl_scr	0.793	0.034	23.129	0.000	0.793	0.626
.nmercy_mths_scr	0.445	0.026	16.921	0.000	0.445	0.400
.lngg_ltrcy_scr	0.367	0.025	14.899	0.000	0.367	0.350
.free_play_scr	0.529	0.021	24.993	0.000	0.529	0.771
.group_tims_scr	0.389	0.016	24.863	0.000	0.389	0.758
L_P	0.473	0.041	11.620	0.000	1.000	1.000

Domain 4: Teaching strategies

lavaan 0.6-19 ended normally after 32 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	8

Number of observations	1433
------------------------	------

Model Test User Model:

Test statistic	4.213
Degrees of freedom	2
P-value (Chi-square)	0.122

Model Test Baseline Model:

Test statistic	608.535
Degrees of freedom	6
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	0.996
Tucker-Lewis Index (TLI)	0.989

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-8373.575
Loglikelihood unrestricted model (H1)	-8371.468
Akaike (AIC)	16763.150
Bayesian (BIC)	16805.290
Sample-size adjusted Bayesian (SABIC)	16779.877

Root Mean Square Error of Approximation:

RMSEA	0.028
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.066
P-value H <sub>0</sub> : RMSEA ≤ 0.050	0.801
P-value H <sub>0</sub> : RMSEA ≥ 0.080	0.008

Standardised Root Mean Square Residual:

SRMR	0.012
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
T_S =~						
rm_rrngmnt_scr	1.000				0.597	0.406
stff_nggmnt_sc	0.795	0.075	10.550	0.000	0.475	0.675
opn_ndd_qstns_	1.160	0.112	10.406	0.000	0.693	0.550
independnc_scr	0.979	0.096	10.232	0.000	0.584	0.522

Variances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.rm_rrngmnt_scr	1.809	0.076	23.783	0.000	1.809	0.835
.stff_nggmnt_sc	0.269	0.020	13.765	0.000	0.269	0.544
.opn_ndd_qstns_	1.107	0.056	19.777	0.000	1.107	0.698
.independnc_scr	0.913	0.044	20.839	0.000	0.913	0.728
T_S	0.356	0.057	6.306	0.000	1.000	1.000

Domain 5: Relationships and interactions

lavaan 0.6-19 ended normally after 23 iterations

Estimator	ML
Optimization method	NLMINB
Number of model parameters	8

Number of observations	1433
------------------------	------

Model Test User Model:

Test statistic	0.339
Degrees of freedom	2
P-value (Chi-square)	0.844

Model Test Baseline Model:

Test statistic	880.668
Degrees of freedom	6
P-value	0.000

User Model versus Baseline Model:

Comparative Fit Index (CFI)	1.000
Tucker-Lewis Index (TLI)	1.006

Loglikelihood and Information Criteria:

Loglikelihood user model (H0)	-6488.958
Loglikelihood unrestricted model (H1)	-6488.789
Akaike (AIC)	12993.916
Bayesian (BIC)	13036.056
Sample-size adjusted Bayesian (SABIC)	13010.643

Root Mean Square Error of Approximation:

RMSEA	0.000
90 Percent confidence interval - lower	0.000
90 Percent confidence interval - upper	0.029
P-value H <sub>0</sub> : RMSEA ≤ 0.050	0.994
P-value H <sub>0</sub> : RMSEA ≥ 0.080	0.000

Standardised Root Mean Square Residual:

SRMR	0.003
------	-------

Parameter Estimates:

Standard errors	Standard
Information	Expected
Information saturated (h1) model	Structured

Latent Variables:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
R_1 = ~						
chld_ntrctns_s	1.000				0.420	0.600
stff_chld_ntr_	1.293	0.091	14.172	0.000	0.543	0.555
efforts_ck_scr	1.238	0.082	15.120	0.000	0.520	0.648
bhvr_mngmnt_sc	1.122	0.076	14.757	0.000	0.472	0.601

Variances:

	Estimate	Std.Err	z-value	P(> z )	Std.lv	Std.all
.chld_ntrctns_s	0.314	0.016	19.891	0.000	0.314	0.640
.stff_chld_ntr_	0.664	0.031	21.374	0.000	0.664	0.692
.efforts_ck_scr	0.373	0.021	17.864	0.000	0.373	0.580
.bhvr_mngmnt_sc	0.393	0.020	19.842	0.000	0.393	0.639
R_1	0.177	0.018	10.024	0.000	1.000	1.000

## Appendix F – Changes made to the LPQA

Following the Thrive by Five Index 2024, a small number of changes were made to the LPQA (v2). These changes were based on feedback from the LPQA trainers and focus groups that were held with some of the Thrive by Five Index assessors. These changes are outlined below:

### Indoor materials

**Change:** The skip logic between sub-question 2a and 2b was removed.

**Reason for change:** Assessors mentioned cases where a classroom had a good variety of materials but not enough materials for the amount of children in the class.

### Free play

**The original sub-questions were:** 15a) During free play times, do children have free choice about what to play with and what materials to use? Response options were Yes/No/NA.

15b) If yes, how many choices were the children able to make about where or how to carry out activities? Response options were 1-2 choices/3 or more choices.

**These were changed to:** 15a) Did you observe free play? Response options Yes/No.

15b) Which of the following choices could children make during free play? Response options: What to play with/ Who to play with/ Where to play.

**Reasons for the changes:** These changes removed the need for a 'not applicable' response option and helped better guide the assessors in terms of what kinds of choices children should have during free play, assisting with the reliability.

### Group times

**Change:** The sub-question "Did you observe large group times?" with response options Yes/No was added in front of the original question.

**Reason for change:** Removed the need for 'not applicable' response option in the main sub-question.

### Staff engagement

**Change:** Added a checklist with the techniques listed in the question hint. Assessors now used the checklist to mark off which techniques they saw rather than just counting and selecting the response options – 1-2 techniques or three or more techniques.

**Reason for change:** The checklist assisted assessors in identifying which techniques to look out for.