



DATA INSIGHTS | **TECHNICAL PAPER**

Reviewing the Socio-Economic Gradient in Learning Outcomes for Children who Participated in the Thrive by Five Index

By Junita Henry and Sonja Giese | **June 2023**¹

¹ This version uses Thrive by Five data corrected in March 2023. Further information is available [here](#).

INTRODUCTION

The Thrive by Five Index 2021 is the first (baseline) in a series of surveys that will monitor trends over time in the proportion of children enrolled in early learning programmes (ELPs) who are On Track for their age in key areas of development.

The Index provides population-level data on how well preschool children in South Africa (aged 50-59 months) are doing in three key developmental domains: early learning, physical growth, and social-emotional functioning.

Data on **learning outcomes** were collected using the Early Learning Measurement 4&5 Years tool (ELOM 4&5), a locally developed and standardised instrument that is aligned with the South African early learning curriculum. Each child was assessed in their home language, by a trained and accredited ELOM assessor. Data was collected on five important learning domains: (i) Gross Motor Development, (ii) Fine Motor Coordination and Visual Motor Integration, (iii) Emergent Numeracy and Mathematics, (iv) Emergent Literacy and Language, and (v) Cognition and Executive Functioning.

For **physical growth**, the Index looks at one key measure - the child's height for age. This is important because it tells us whether the child is at risk of stunting. Growth stunting is usually associated with chronic malnutrition and is known to compromise neurological and cognitive development with significant loss of an individual's potential.

Social Relations with Peers and Adults and Emotional Readiness for School were assessed

SAMPLE

The Thrive by Five Index contained a sample of 5,139 children. Data on the ELP that children attend were available for 4,926 of these children (96% of the originally analysed sample) distributed across 1,173 ELPs.

Table 1 shows the difference between children included in the refined analyses (N=4,926) and those excluded (N=213) due to lack of ELP fee data. Fee data was initially available for 4,911 children and additional efforts were made to collect missing fee data to avoid bias. By demographic characteristics, both groups had a similar distribution of boys and girls. They were also spread across all nine South African provinces. Children in the excluded

using the ELOM Social-Emotional rating scales, completed by the child's teacher.

In the absence of household level income data for children in the sample, school quintiles² were used as proxies for the probable **socio-economic background** of the children who were assessed. For the Index sampling frame, the assumption was made that the income level of children attending ELPs within each school cluster matched the income level of children attending the nearest school. In practice however, this is often *not* the case.

For this and other reasons, the Index team acknowledged that the quintile system is an imperfect measure of socio-economic status, and that the socio-economic gradient reported in the Index is likely to be an under-estimate of the true disparities in child outcomes between children in different income groups.

Since the launch of the Index in April 2022, additional data have become available on the ELPs attended by the participating children. This new data enabled the DataDrive2030 team to replace quintile ranking with alternative and more accurate measures of socio-economic status.

The current document details the methodology and outcomes of the refined analysis of the socio-economic gradient reported in the Index. This process has enabled us to more clearly characterise the nature and extent of the difference in outcomes between children from different socio-economic bands.

category are statistically significantly younger but are similarly 54 months old on average ($p < 0.05$) i.e. their distribution is different but the average is the same. Children in the excluded group however, differ by whether they receive a child grant ($p < 0.05$) and by child outcomes. In terms of outcomes, the groups differed by total ELOM score, Gross Motor Development, and Emergent Literacy and Language - where the excluded children performed worse on average. These factors suggest that the children included in the refined sample may represent an upward bias of socio-economic status, where poorer children are excluded. We were unable to contact the excluded children's ELPs to confirm fee levels.

² Every public school in South Africa is assigned a quintile ranking by the Provincial Departments of Basic Education. This ranking is based on the relative poverty levels of the community living within 3 kms of the school, with Quintile 1 (Q1) being the poorest and Quintile 5 (Q5) the wealthiest.

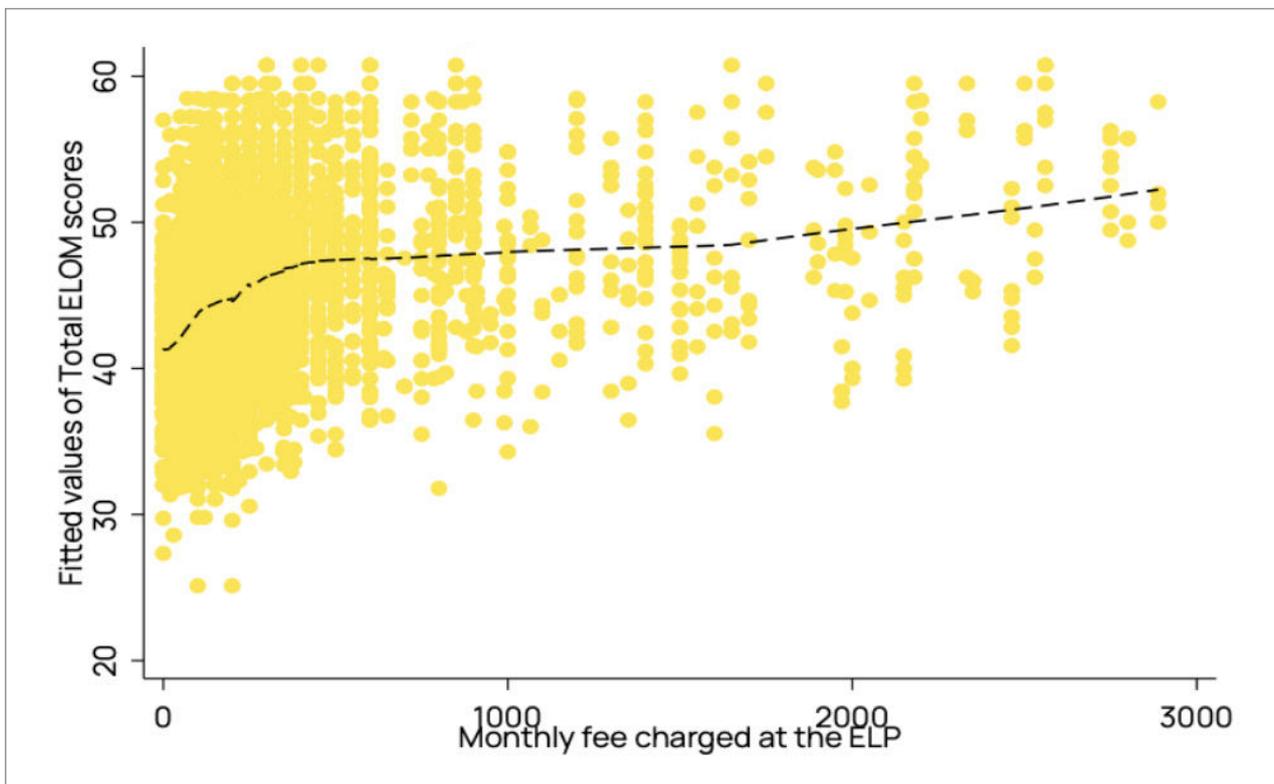
TABLE 1: AVERAGE DOMAIN SCORES BY INCLUDED AND EXCLUDED GROUP

	Included: Mean (SD) (N=4,926)	Excluded: Mean (SD) (N=213)	P-value
Child demographics			
Child gender	48% boys	49% boys	0.89
Child age in months	54.7 (2.7)	54.2 (2.6)	0.004
Child grant received	85%	92%	0.011
Child outcomes			
Total ELOM	44.48(14.44)	42.5(12.77)	0.02
Domain 1: Gross Motor Development	8.24 (3.95)	7.50 (4.01)	0.008
Domain 2: Fine Motor Coordination and Visual Motor Integration	10.90 (3.54)	10.71 (3.58)	0.43
Domain 3: Emergent Numeracy and Mathematics	8.42 (4.15)	8.03 (3.72)	0.18
Domain 4: Cognition and Executive Functioning	6.94 (4.11)	6.64 (3.67)	0.29
Domain 5: Emergent Literacy and Language	10.38 (4.53)	9.66 (4.28)	0.022

MAPPING ALTERNATIVE SOCIO-ECONOMIC VARIABLES AGAINST THE CURRENT **QUINTILE RANKING**

Figure 1 shows a lowess (Locally Weighted Scatterplot Smoothing) of the relationship between the expected total ELOM scores³ and

the fees charged at the facility. There is a positive relationship with a high amount of variation, particularly within the lower fee bands.



³ Controlling for age, gender, whether the child was stunted, whether the ELP receives the government subsidy, and provincial fixed effects clustered at the ELP level

The main proxy we used to redefine socio-economic status for each child is (1) the monthly fee charged at the ELP that the child attends, and (2) whether the ELP receives a means-tested subsidy from the Department of Social Development (DSD).⁴ Additional variables used to sense check results were: whether the ELP has access to running water, electricity for lighting, and access to a flush toilet connected to a sewage system; and whether the ELP is based in a formal or informal (in a shack) building.

Table 2 shows the average monthly fee charged at each ELP (column 2), the percentage receiving any subsidy (column 3), and the percentage of ELPs who have access to various facilities (columns 4-6), by the Department of Basic Education (DBE) quintile rating used in the report (column 1). The average monthly fee does increase with quintiles, however the range of fees and access to facilities vary substantially across quintiles.

TABLE 2: DESCRIPTIVE CHARACTERISTICS ACROSS DBE QUINTILES

(1) DBE Quintile	(2) Average monthly fee [min; max]	(3) Percent receiving a subsidy	(4) Access to running water	(5) Uses electricity for lighting	(6) Access to a flush toilet	(7) N facilities	(8) N children
1	R188 [R0;R3300]	73%	66%	81%	45%	391	1,629
2	R242 [R0;R2560]	66%	73%	85%	50%	290	1,224
3	R253 [R0;R2050]	61%	84%	88%	67%	266	1,098
4	R388 [R0;R2888]	54%	96%	96%	94%	114	465
5	R949 [R0;R5600]	29%	98%	90%	95%	116	510
Overall	R313.61	62%	78%	86%	61%	1,177	4,926

REFINED SOCIO-ECONOMIC LEVELS

Refined socio-economic levels were determined using a k-means clustering approach⁵ using monthly fees and whether the ELP receives a subsidy from the DSD. K-means clustering involves a simple unsupervised machine learning algorithm that classifies data into a number of clusters. Observations are partitioned into clusters that share similarities. The number of clusters (k) is determined beforehand. Variations of 3-6 clusters were used. Fee levels did not vary substantially when sense-checked against a model that incorporated facility access⁶ or fees only.

Table 3 displays the average characteristics of ELPs using the refined socio-economic levels. Level 1

(L1) shows the ELPs with the lowest fee level while level 5 (L5) presents those in the highest fee level. As expected, receipt of a DSD subsidy decreases while access to services increases as the ELP fee increases. The differences in service and subsidy access by fee level are much starker than by quintile, pointing to a gradient more in line with what one would expect.

A disadvantage of the updated socio-economic status (SES) levels however, is that the sample size of children in the highest level is substantially lower – decreasing from 507 children in what was classified as Quintile 5 ELPs to 149 children in ELPs charging more than R1,751 per month.

4 At the time of preparing this, the DSD was responsible for subsidy payment. Since then, this responsibility has been shifted to the Department of Basic Education.

5 The initial approach was to use a Latent Class Analysis to group categories using these variables. However, repeated models and variable compositions did not converge. This means that natural, distinct clusters based on fee and facility access variables did not exist. When a model did converge, there was only a distinction between very high fee schools (>R2,000 per month) and the remainder of the sample.

6 This model was based on N=4,482 observations and had a grouping of R0-R130; R140-R320; R340-R785; R800-R1,750; R1,888-R3,600.

TABLE 3: DESCRIPTIVE CHARACTERISTICS ACROSS ELP FEE LEVEL

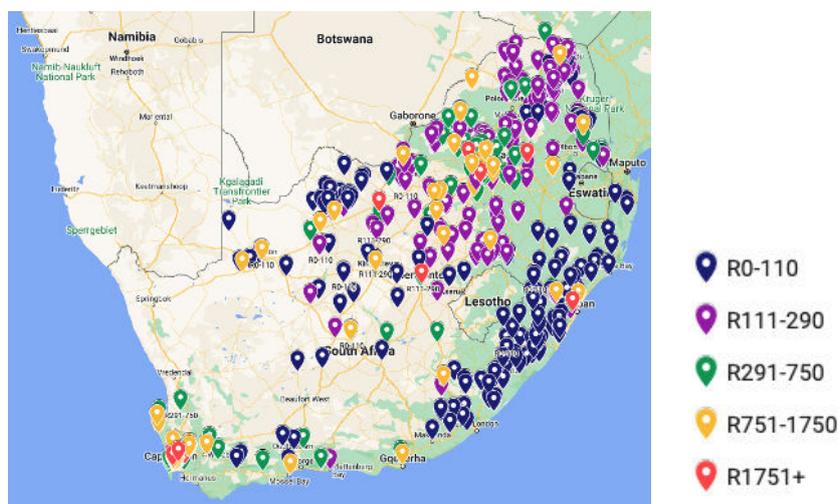
Fee level	Average monthly fees	Percent receiving subsidy	Has running water	Uses electricity for lighting	Access to a flush toilet	N Facilities	N children
L1: R0-R110	R54	83%	56%	79%	35%	398	1,662
L2: R111-R290	R185	72%	83%	85%	61%	437	1,836
L3: R291-R750	R404	35%	96%	95%	88%	238	984
L4: R751-R1,750	R1,093	7%	97%	99%	100%	72	295
L5: R1,751+	R2,623	0%	100%	97%	100%	32	149
Overall	R313.62	62%	78%	86%	61%	1,177	4,926

GEOGRAPHICAL SPREAD OF ELPs ACROSS THESE FEE LEVELS

Figure 2 shows the geographical spread of ELPs across the fee levels. Lower fee-charging schools are concentrated in the Eastern Cape and KwaZulu-Natal, as both provinces have very high proportions of young children living in poverty (82% and 73% respectively).⁷

There is more variation in ELP fee levels in the other provinces and on average, fees in the Western Cape are slightly higher. A limitation of this comparison however is that it does not account for the differences in the cost of living across provinces.

FIGURE 2: GEOGRAPHICAL DISTRIBUTION OF ELPs BY FEE LEVEL



REVISITING THE INDEX SOCIO-ECONOMIC GRADIENT USING FEE LEVELS AS A PROXY FOR SES

Replacing quintile ranking with these five fee levels as a measure of SES, we re-examined the relationship between poverty and child outcomes,

including total ELOM score, scores for each of the learning domains, socio-emotional functioning, and physical growth. All observations were weighted.

⁷ General Household Survey 2021 Children's Institute Analysis

TOTAL ELOM 4&5 SCORE

In comparison to quintile groupings, the disparities between children On Track versus Not on Track across fee groupings are far greater. Using the quintile system, 59% of children in Q5 were found to be On Track, compared to 39% in Q1. Using the

updated ELP fee levels (Figure 3) we find that 83% of children in L5 are on track (depicted by the green line), compared to only a third (34%) of learners in the poorest level. The vertical grey bars represent confidence intervals at a 5% level of significance.

FIGURE 3: TOTAL LEARNING SCORES – ON TRACK / NOT ON TRACK BY ELP FEE LEVEL

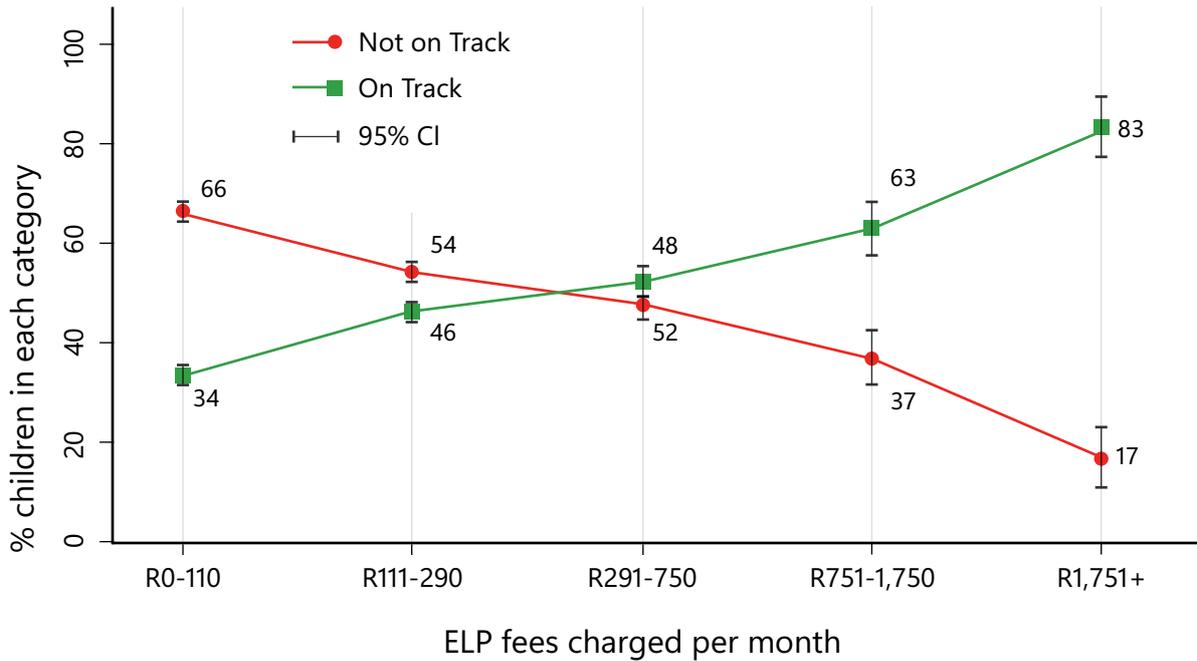
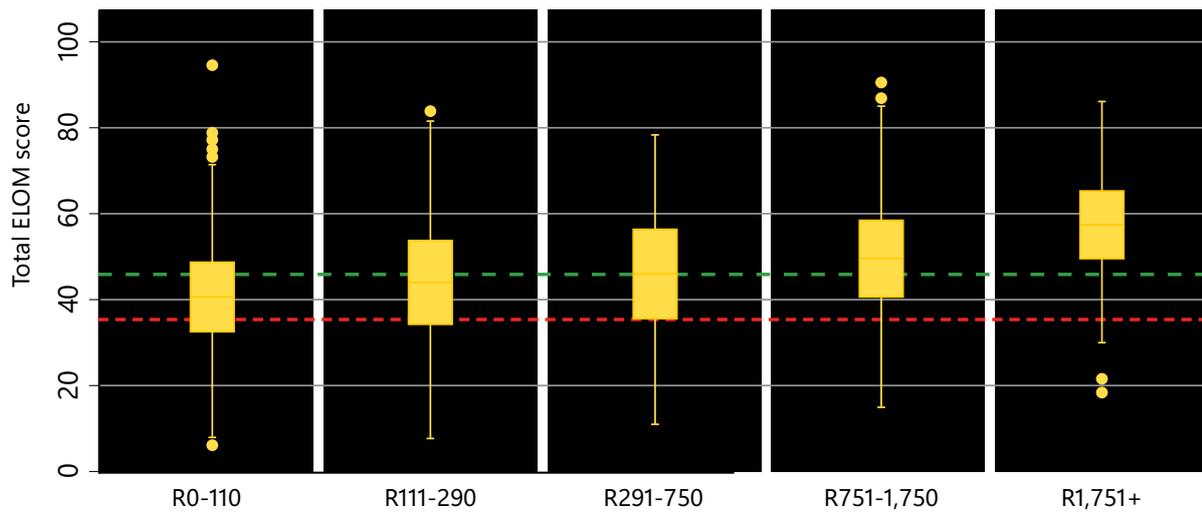


Figure 4 displays the distribution of total ELOM score across SES. The maximum total ELOM score is 100 points.⁸ The boxes contain the middle 50 percent of scores (interquartile range). The whiskers (outer lines) provide a sense of the total variation in scores. The red dashed line indicates the cut off between those Falling Far Behind (below the line) and those Falling Behind (above the line). The green dashed line indicates the cut off between those that are Falling Behind (below the line) and those On Track (above the line). For L1 (R0-R110), the

median score was 40 ELOM points and 50 percent of child scores were between 32 and 50 points. Most of these observations are below the green line. For L5, the median score was 61 ELOM points and 50 percent of scores were between 52 and 69 points – this is much higher than L1. Level 1 also has the most outlier children relative to the rest of their distribution, making the case that there are poor children who are excelling and scoring on par with L5 children despite their lower SES.

⁸ The ELOM tool consists of 5 domains, each with a maximum total of 20 points.

FIGURE 4: DISTRIBUTION OF ELOM SCORES BY ELP FEE LEVEL (INDEX DATA)

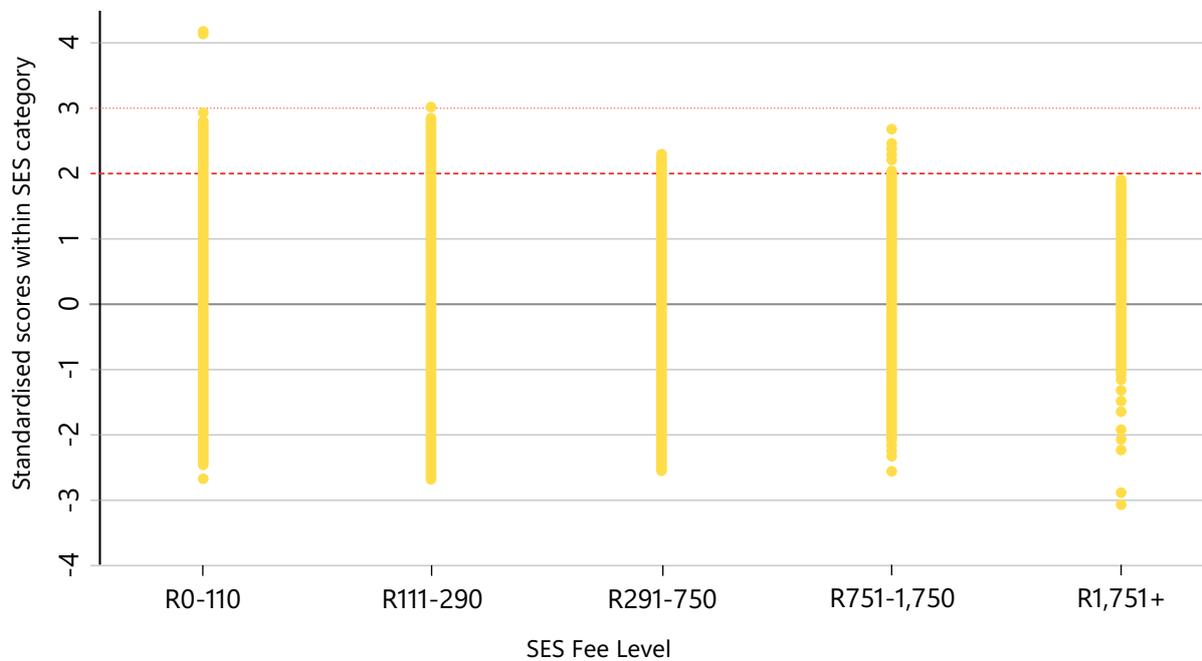


Notes: The red dashed line indicates the cut off between those Falling Far Behind and those Falling Behind. The green dashed line indicates the cut off between those that are Falling Behind and those On Track.

Figure 5 shows the standardised total ELOM scores within ELP fee levels. The image demonstrates the variability in scores within income bands, and in particular within the lower ELP fee levels. Some children in lower ELP fee levels achieve more than

2 standard deviations (SDs) and up to 4SDs above their peers, in comparison to higher ELP fee levels where variability is lower (as depicted by a shorter yellow line).

FIGURE 5: DISTRIBUTION OF STANDARDISED ELOM SCORES WITHIN ELP FEE LEVEL



MULTIVARIATE ANALYSIS: ESTIMATING THE CHANGE IN SCORES BY TOTAL ELOM AND SPECIFIC DOMAINS

Using this sample we model the change in scores by SES fee band or quintile using a multivariate regression controlling for age, gender, whether the child is stunted, whether the ELP receives the DSD subsidy, and provincial fixed effects. All regressions were clustered at the ELP level using robust standard errors. Table 4 shows the change in scores relative to either the first fee band or the first quintile. For the SES fee levels, one can see that all coefficients are significant at the 5% level except for those in Domain 1.

For example, for the total ELOM score, children in L2 score 2.2 more ELOM points than those in L1 on average. Children in L5 score 17 more points on average than those in the first fee band. When looking at quintiles, we find that there is only a distinction between the first and fifth quintile (where children in Q5 score 3 points higher on average).

TABLE 4: CHANGE IN SCORES RELATIVE TO FEE LEVEL 1 / QUINTILE 1

Difference in scores	Total ELOM score	Domain 1: Gross Motor Development	Domain 2: Fine Motor Coordination and Visual Motor Integration	Domain 3: Emergent Numeracy and Mathematics	Domain 4: Cognition and Executive Function	Domain 5: Emergent Literacy and Language
SES fee band						
Fee level 1 (R0-110)	Fee level 1 is the benchmark					
Fee level 2, R111-290	2.20***	0.11	0.39**	0.62***	0.59***	0.50*
	-0.8	-0.21	-0.17	-0.21	-0.19	-0.27
Fee level 3, R291-750	4.76***	0.31	0.92***	1.21***	1.46***	0.86***
	-1.13	-0.29	-0.25	-0.3	-0.27	-0.33
Fee level 4, R751-1,750	9.89***	0.14	1.76***	1.79***	3.46***	2.74***
	-1.37	-0.37	-0.29	-0.38	-0.38	-0.44
Fee level 5, R1,751+	17.41***	-0.12	3.08***	4.22***	5.66***	4.58***
	-1.97	-0.46	-0.35	-0.64	-0.52	-0.55
Quintile						
Quintile 1	Quintile 1 is the benchmark					
Quintile 2	0.89	-0.11	0.09	0.28	0.42**	0.21
	-0.73	-0.2	-0.16	-0.2	-0.19	-0.23
Quintile 3	0.58	0.13	-0.12	0.22	0.25	0.09
	-0.72	-0.18	-0.16	-0.21	-0.19	-0.24
Quintile 4	-0.75	-0.28	-0.05	-0.22	0.08	-0.28
	-1.01	-0.25	-0.22	-0.28	-0.26	-0.32
Quintile 5	3.09***	-0.43*	0.69***	0.58*	1.15***	1.09***
	-1.16	-0.25	-0.24	-0.34	-0.31	-0.36
Observations	4,885	4,885	4,885	4,885	4,885	4,885

Notes: Robust standard errors in parentheses. Standard errors were clustered at the ELP level. *** p<0.01, ** p<0.05, * p<0.1



ELOM DOMAIN SCORES

The table below explains the significance of each development domain.

TABLE 5: SIGNIFICANCE OF THE DEVELOPMENTAL DOMAINS

Developmental Domain	Significance
Gross Motor Development	Facilitates peer engagement through participation in games, and is associated with emotional well-being as well as with academic achievement.
Fine Motor Skills and Visual-Motor Integration	Important for coordinating the use of the hands and the eyes; and makes a specific contribution to early mathematics and early literacy.
Emergent Numeracy and Mathematics	Strongly predictive of later school success. Good math foundations are essential for a deeper understanding of more complex mathematical concepts and problem-solving.
Cognition and Executive Functioning	Helps children hold information or instructions in mind during classroom activities, focus on task-relevant stimuli during problem-solving tasks, and resist distraction.
Emergent Literacy and Language	Affects the ability to understand what is being said and read by a teacher, as well as to communicate effectively through speech and writing.

The following figures display disparities in child outcomes by ELP fee levels for each ELOM domain, and compare them to prior reports that used the DBE quintile system as a proxy for SES. The green bars indicate the percentage of children who are On Track for their development, the orange bars

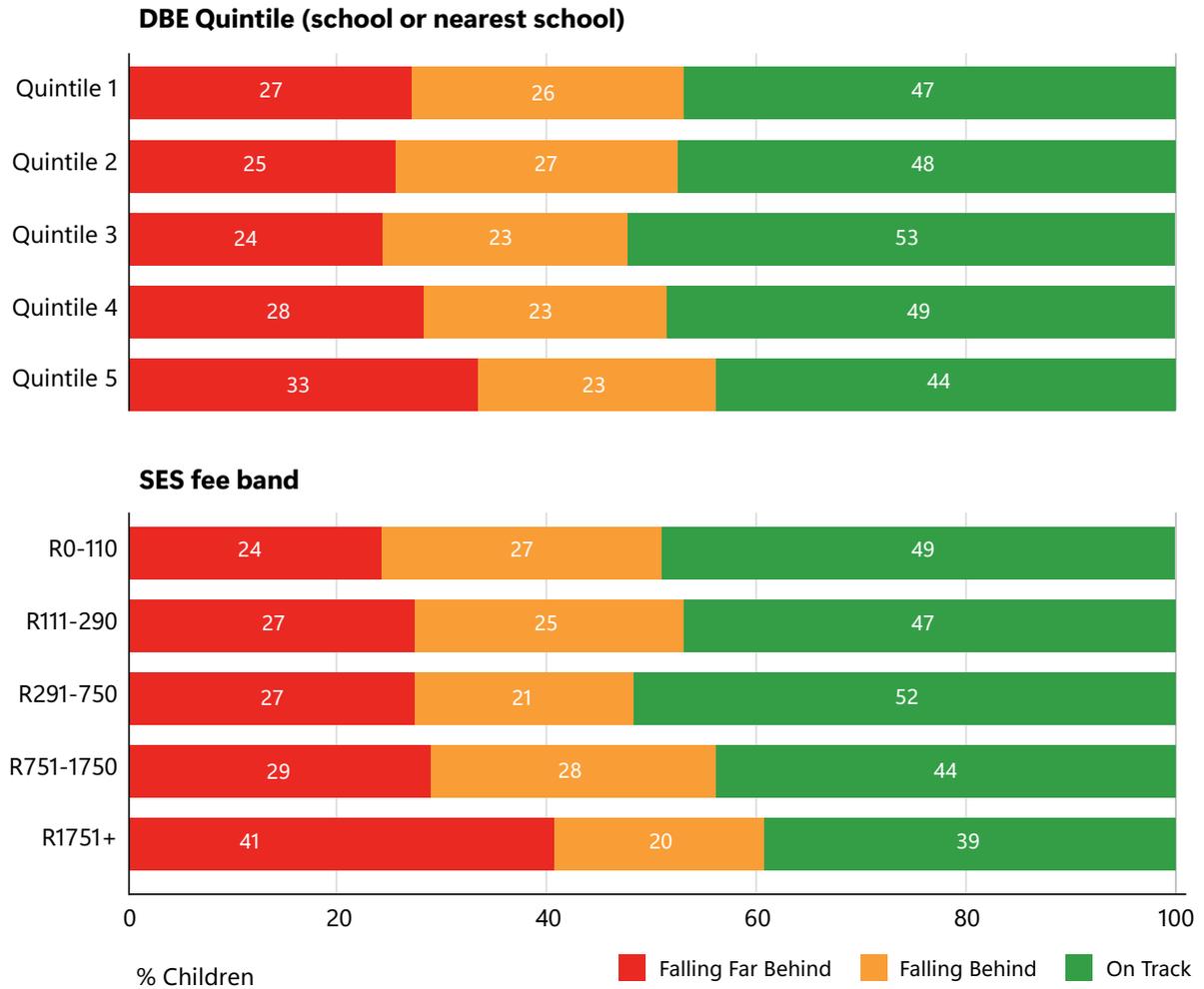
indicate the percentage of children who are Falling Behind and the red bars represent the percentage of children who are Falling Far Behind. The cut-offs for each category is outlined in the [ELOM Technical Manual](#) for children aged 50-59 months.

DOMAIN 1: GROSS MOTOR DEVELOPMENT (GMD)

In comparison to the quintile method, differences in outcomes for GMD are similar. Relative to their peers, a smaller percentage of children in the fifth quintile and highest fee paying group are On Track for their development. For example, half (49%)

of the children in L 1 are On Track, compared to 39% in ELP fee level 5. There were no statistically significant changes in scores across fee levels in the multivariate regression.

FIGURE 6: DISTRIBUTION OF GMD OUTCOMES

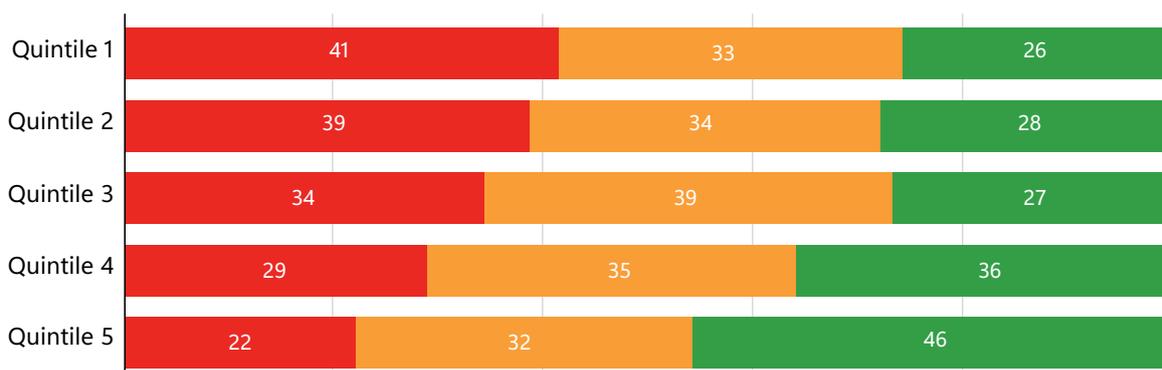


DOMAIN 2: FINE MOTOR COORDINATION AND VISUAL MOTOR INTEGRATION (FMC-VMI)

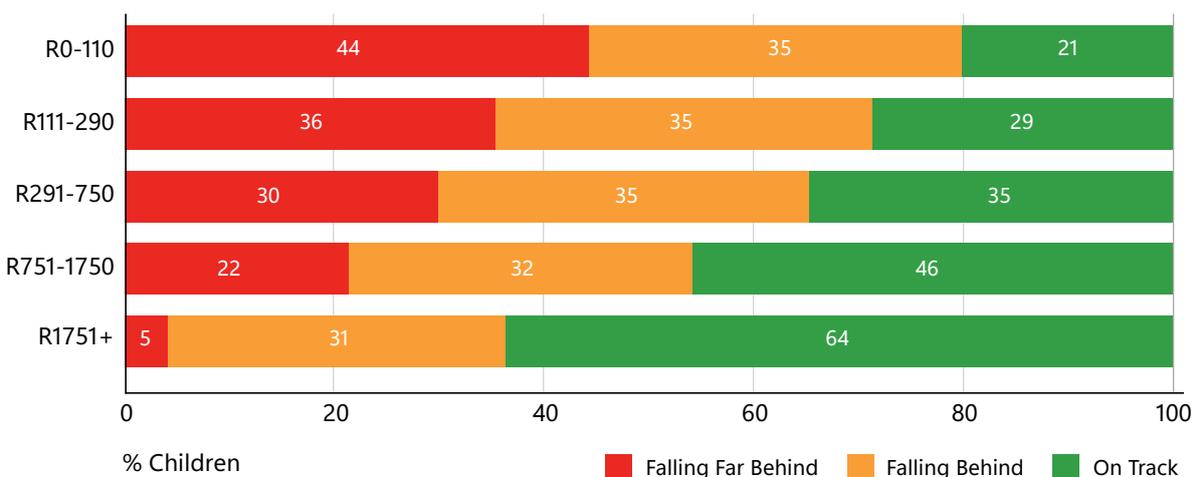
For domain FMC-VMI, a fifth of children (21%) in the lowest fee level are On Track in their development in comparison to 64% in L5. The disparities between fee levels 1 to 5 are much more pronounced than between quintiles. Multivariate regression shows that children in L5 score 3.08 points higher than those in the lowest fee level on average.

FIGURE 7: DISTRIBUTION OF FMC-VMI OUTCOMES

DBE Quintile (school or nearest school)



SES fee band

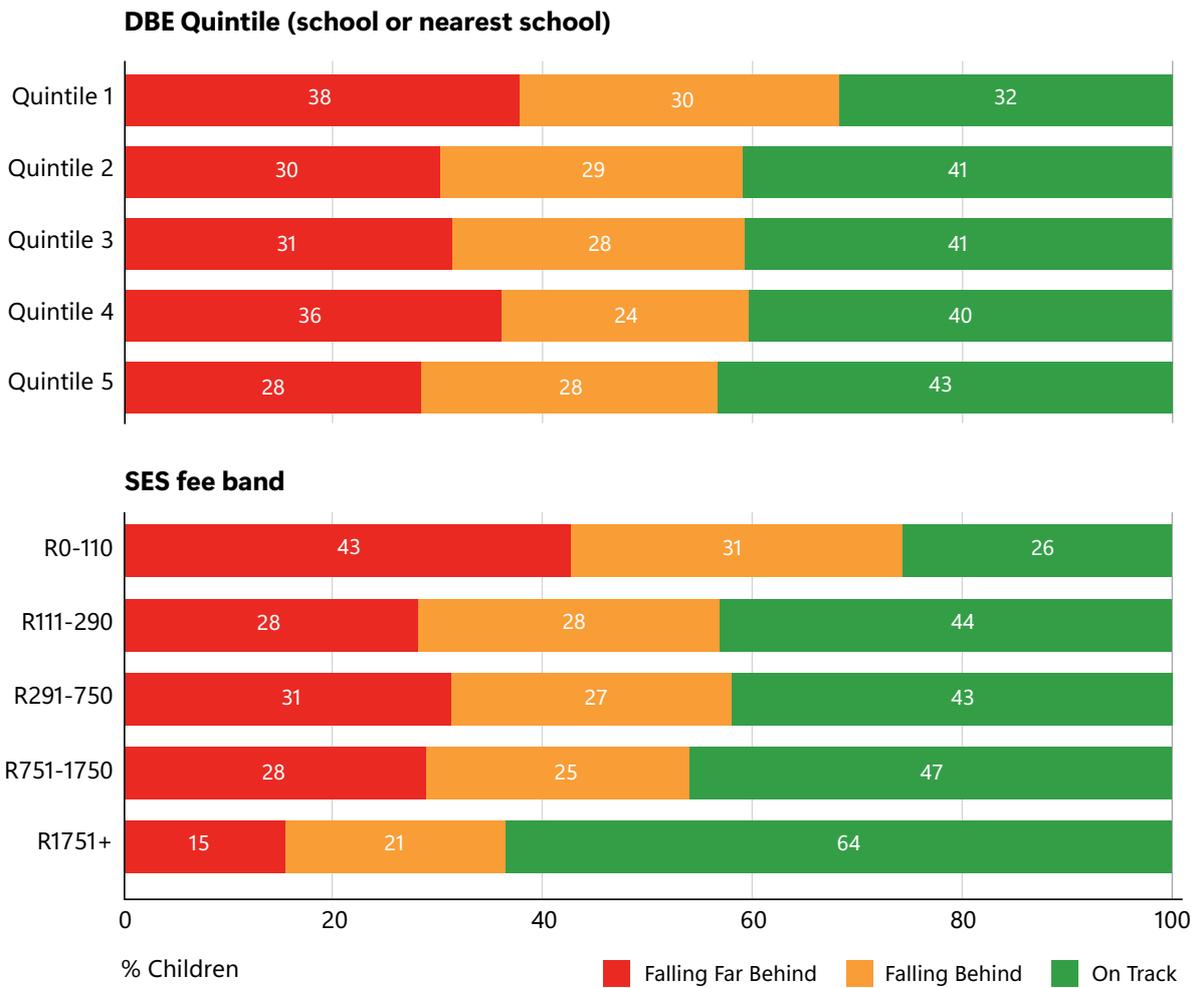


DOMAIN 3: EMERGENT NUMERACY AND MATHEMATICS (ENM)

Prior reports of differences in numeracy showed little variation across quintiles, where on average, a third of children were On Track. However, ELP fee levels show a much larger difference particularly at the lowest and highest levels. Learners in the highest level outperform other levels by almost

20% on average. Multivariate regression shows that children in L5 score 4.22 points higher than those in the lowest fee level on average. It is interesting to note little difference in ENM performance within the middle three levels.

FIGURE 8: DISTRIBUTION OF ENM OUTCOMES

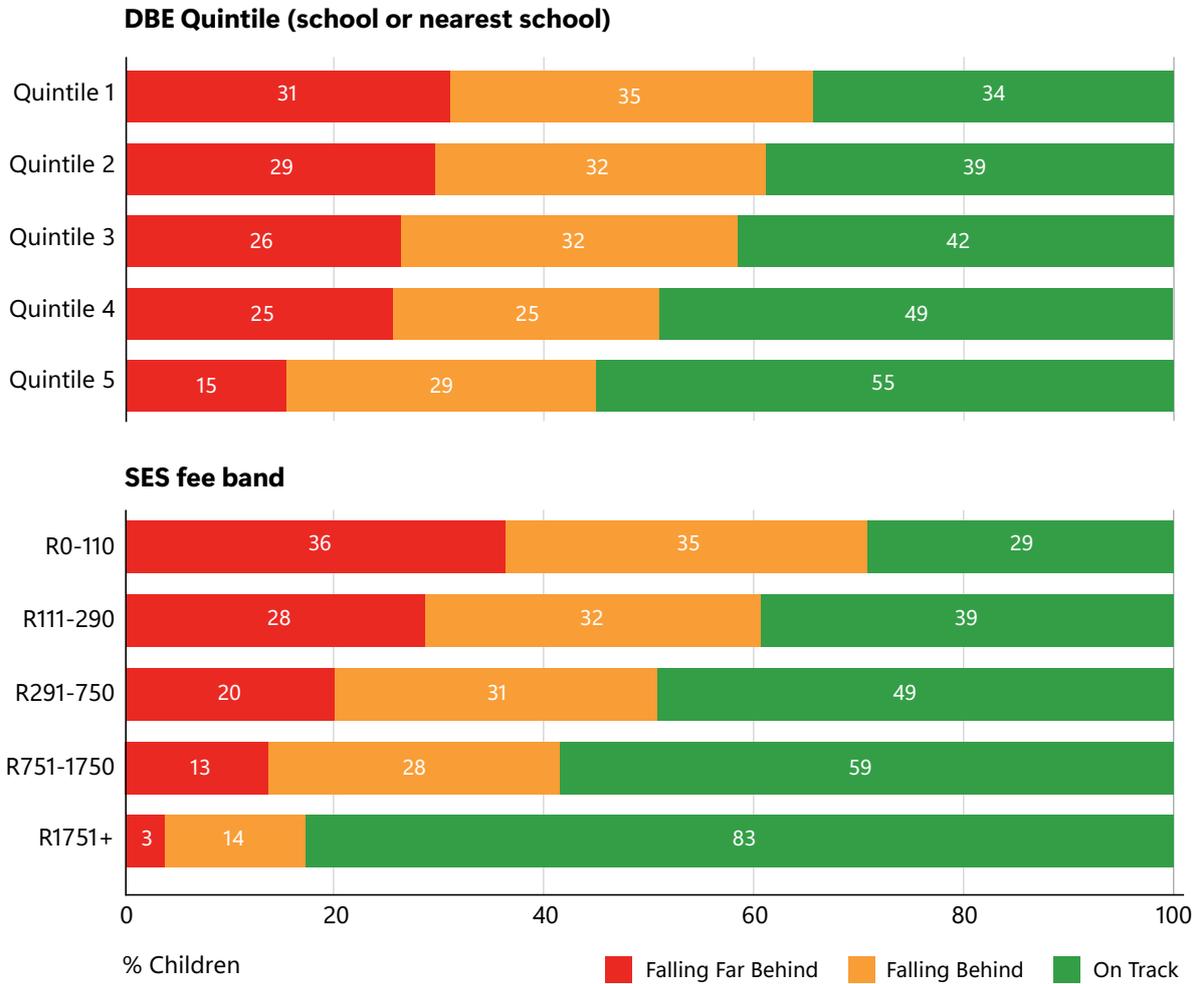


DOMAIN 4: COGNITION AND EXECUTIVE FUNCTIONING (CEF)

The largest differences across SES relate to CEF. Less than a third of children in L1 are On Track, and 36% are Falling Far Behind. In comparison, only 3% of children in the highest level are Falling Far

Behind and the vast majority (83%) are On Track. Multivariate regression shows that children in the highest fee level score 5.66 points higher than those in the lowest fee level on average.

FIGURE 9: DISTRIBUTION OF CEF OUTCOMES



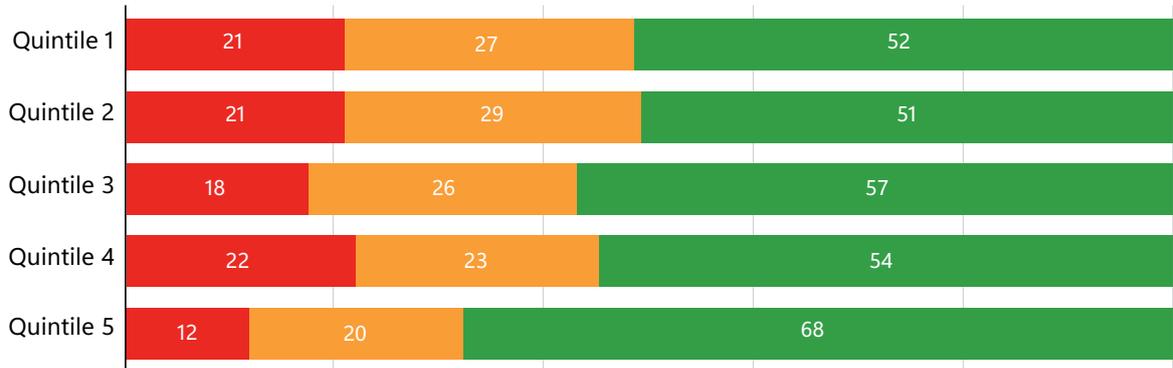
DOMAIN 5: EMERGENT LITERACY AND LANGUAGE (ELL)

Finally, differences in outcomes by ELP fee levels remain pronounced for literacy and language skills where the vast majority of children are On Track (86%) in L5 in comparison to less than half (48%) of

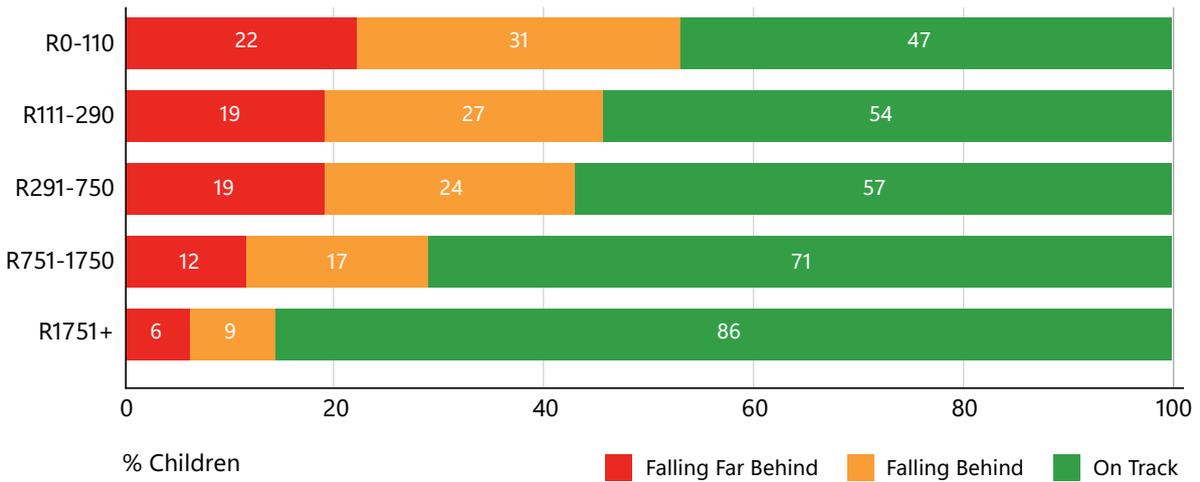
children in L1. Multivariate regression shows that children in the highest fee level score 4.58 points higher than those in the lowest fee level on average.

FIGURE 10: DISTRIBUTION OF ELL OUTCOMES

DBE Quintile (school or nearest school)



SES fee band

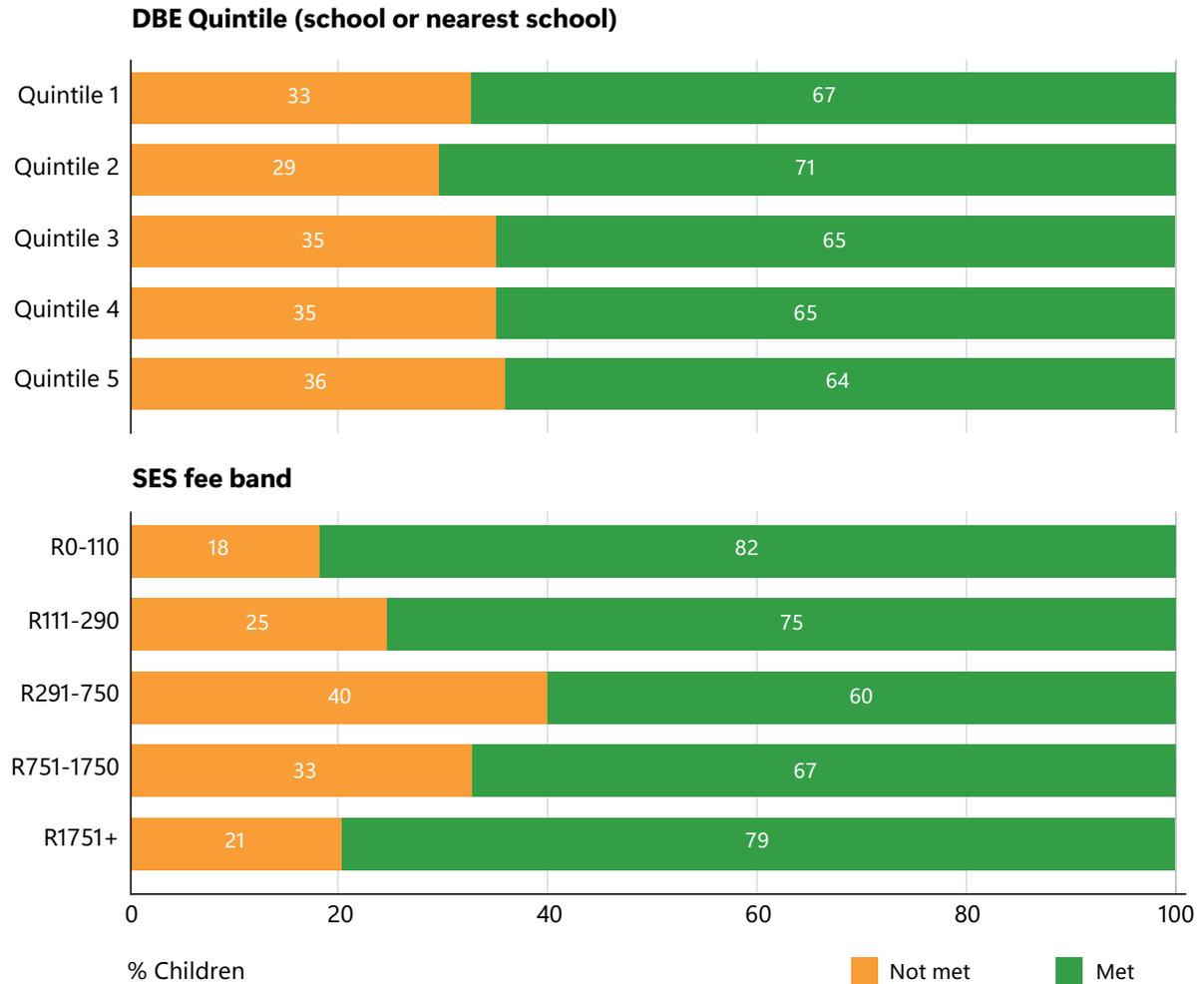


SOCIAL-EMOTIONAL FUNCTIONING: SOCIAL RELATIONS WITH PEERS AND ADULTS

The relationship between social relations and SES are less clear. However, children in the lowest ELP fee level are more likely to meet the required standards (82%). Children in mid-fee levels are the

least likely to meet standards (60% of children in the R291-750 fee group). This increases to 79% in the highest fee group.

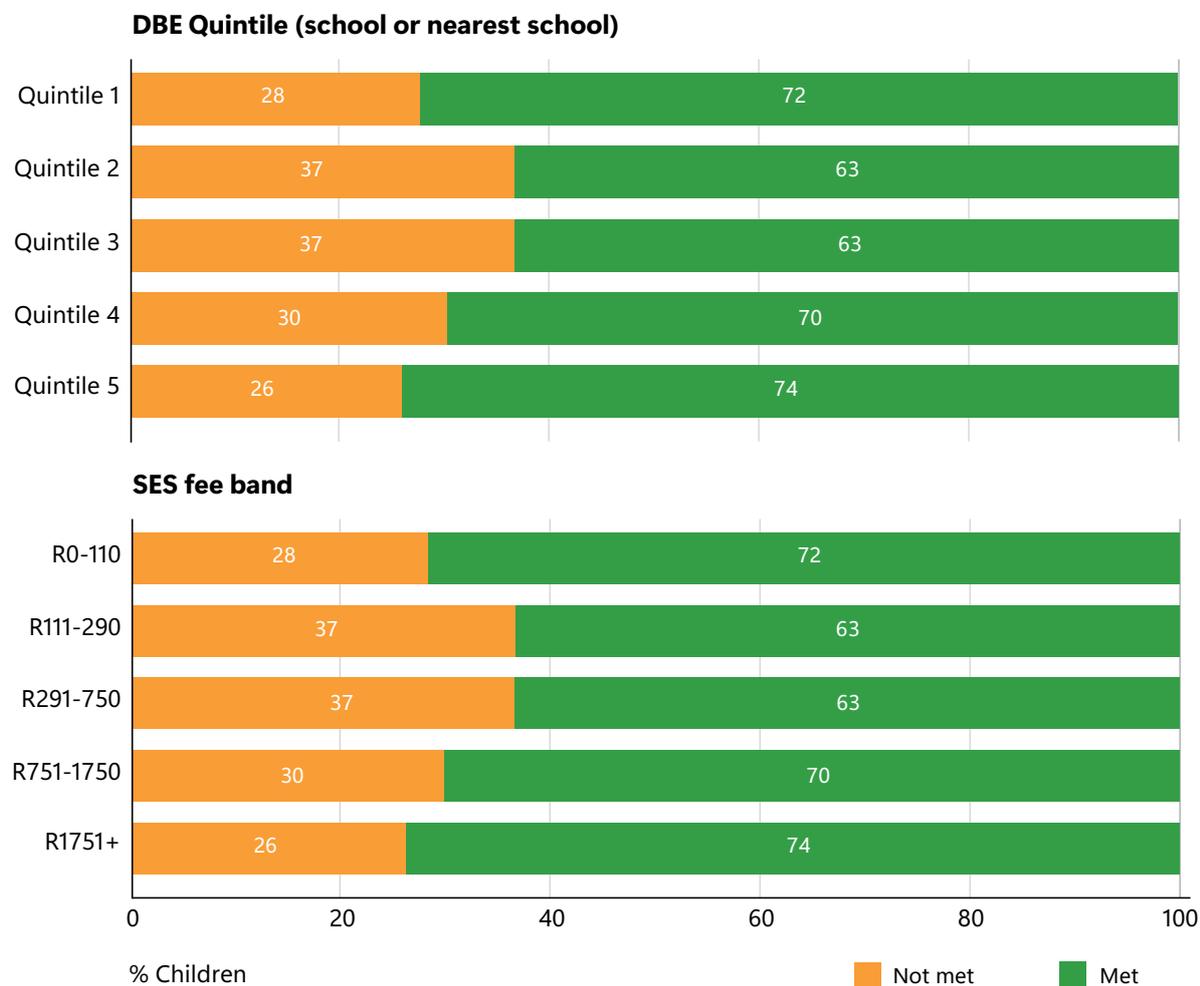
FIGURE 11: DISTRIBUTION OF SOCIAL RELATIONS OUTCOMES



SOCIAL-EMOTIONAL FUNCTIONING: EMOTIONAL READINESS FOR SCHOOL

Similarly to social relations, children in the highest and lowest fee level (74% and 72%, respectively) are more likely to meet emotional readiness standards than children in the mid-fee levels (roughly two thirds of children).

FIGURE 12: DISTRIBUTION OF EMOTIONAL READINESS OUTCOMES

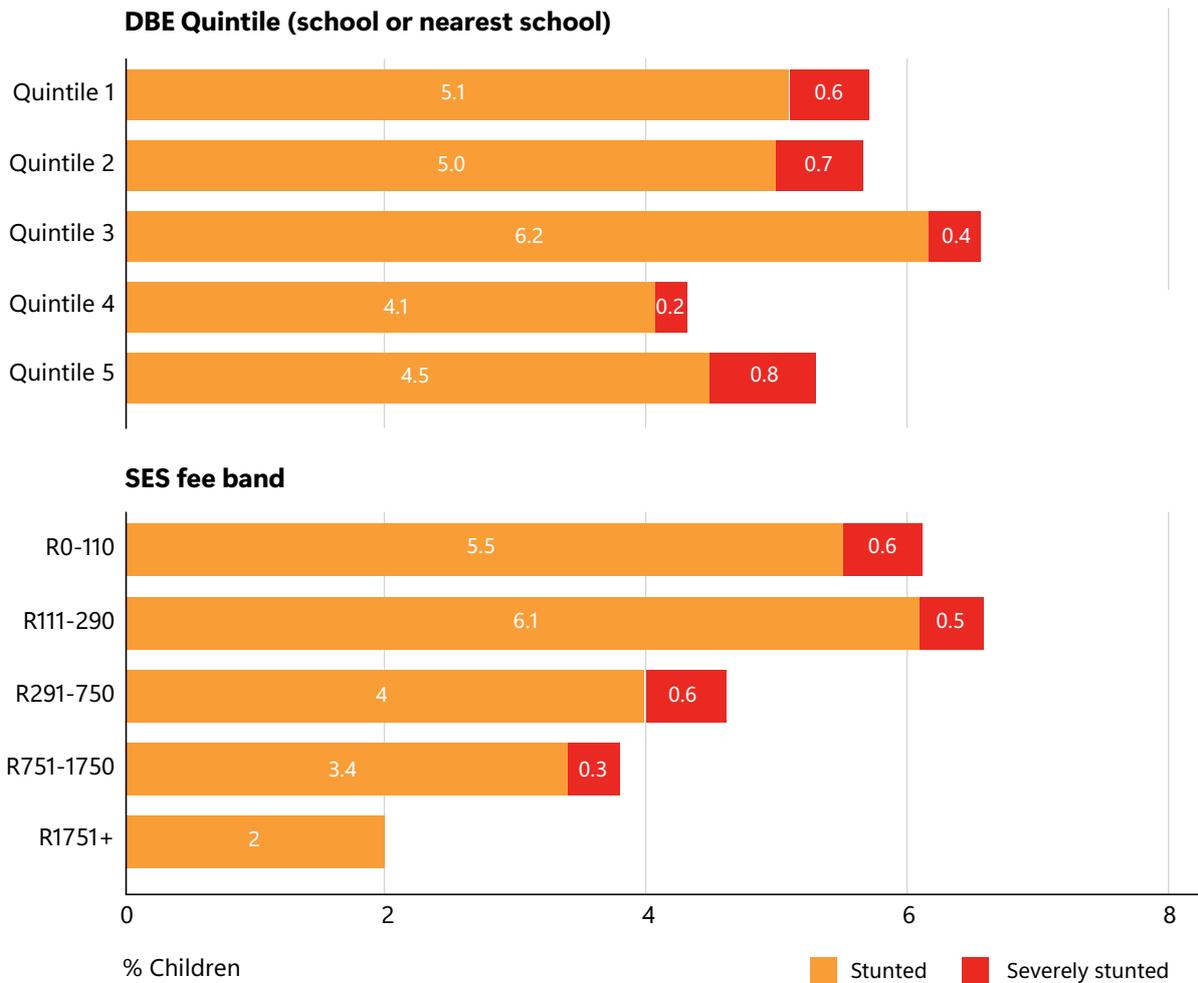


PHYSICAL GROWTH: STUNTING

There is a much starker difference in the prevalence of stunting across the fee levels than quintiles. Around 6 percent of children in the first two fee levels are moderately stunted, while 0.5% are severely stunted. These rates decrease to 2.1% in the highest fee level group for moderate stunting

and 0% for severe stunting. Children are defined as 'moderately stunted' if their height-for-age⁹ is 2 or more SDs below the World Health Organization's Child Growth Standards median and 'severely stunted' if 3 or more SDs below the median.¹⁰

FIGURE 13: DISTRIBUTION OF STUNTING PREVALENCE



9 Height-for-age scores were calculated using age in months

10 Data on child height was missing for 7 children, resulting in a total of N=5,132

LIMITATIONS

It is possible that children excluded from this analysis due to absence of fee data may be statistically different from those included. Their descriptive characteristics point to them being children from lower socio-economic backgrounds. Since their average scores are lower, it is possible that the outcome gap between higher and lower socio-economic bands estimates may be biased slightly downward and outcomes may be even more disparate with their inclusion.

At the same time, the Thrive by Five Index sample has a higher proportion of lower SES children in comparison to the ELP-enrolled child population based on the national ECD census. That is assuming that the census captures all ELPs and that there is no inherent bias in the providers that may have been missed. Additionally, these results do not account for the variation in the cost of living across provinces. Finally, it may be worth noting again that these results are reflective only of children *enrolled* in ELPs.

CONCLUSION

Using monthly fee levels at the ELP as a proxy for socio-economic status displays the clear disparities in child outcomes across income groups. These differences are much starker than the differences observed when using the DBE school quintile system as a proxy for preschool child poverty.

Moving forward, we recommend that ELOM direct child assessments be accompanied by efforts to collect information on the fee levels of the associated ELPs. These data will enrich the growing meta dataset and enable us to monitor progress in closing the opportunity gaps between children

from different socio-economic backgrounds at the point of entry into school.

Finally, the presence of outliers within lower fee levels makes the case that there are children who are doing exceptionally well within these groups – and perform similarly to their peers in higher level groups. Understanding the characteristics associated with positive outliers in lower socio-economic bands may offer insights into the kinds of interventions that could drive improved child outcomes within these contexts.

