DATA INSIGHTS

REVIEWING THE SOCIO-ECONOMIC GRADIENT IN LEARNING OUTCOMES FOR CHILDREN WHO PARTICIPATED IN THE THRIVE BY FIVE INDEX

Junita Henry and Sonja Giese
2 September 2022

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 was collected using the Early Learning Outcomes Measure (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) Early Numeracy and Mathematics, (iv) Early Literacy and Language, and (v) Cognition and Executive Function.

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 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 quintiles[1] 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, there will be many instances

[1]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.

where this is not the case. Furthermore, school quintiles within each province are assigned relative to other schools within the same province and there will therefore be differences in income levels between schools in the same quintile in different provinces.

For these 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 early learning programmes attended by participating children. This new data enabled the DD2030 team to replace quintile ranking with alternative, 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.

Sample

The Thrive by Five Index contained a sample of 5139 children. Data on the Early Learning Programme that children attend were available for 4911 of these children (97% of the originally analysed sample) distributed across 1 173 Early Learning Programmes (ELPs).

Table 1 shows the difference between children included in the refined analyses (N=4911) and those excluded (N=228) due to lack of ELP fee data. By demographic characteristics, both groups had a similar distribution of boys and girls. They are also spread across all nine South African provinces. Children in the excluded category are statistically significantly younger but similarly 54 months old on average ($p<0.05$). 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 in particular, the groups differed significantly by total ELOM score, Gross Motor Development (Domain 1) and Emergent Literacy (Domain 5), 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 socioeconomic status, where poorer children are excluded. Additional efforts are being made to collect the missing data so as to address this bias.

Table 1: Average domain scores by included and excluded group

	Included: Mean (SE) (N=4911)	Excluded: Mean (SE) (N=228)	P-value
Child demographics			
Child gender	48% boys	49% boys	0.817
Child Age	54.7 (0.04)	54.2 (0.18)	0.0019
Child grant	85%	92%	0.028
Child outcomes			
Total ELOM	44.4 (0.2)	42.5(0.85)	0.0480
Domain 1: Gross Motor Development	8.2 (0.06)	7.6 (0.27)	0.0167
Domain 2: Fine Motor Coordination and Visual Motor Integration	10.9(0.05)	10.67(0.24)	0.3467
Domain 3: Emergent Numeracy and Mathematics	7.93(0.05)	7.82(0.23)	0.6600
Domain 4: Cognition and Executive Functioning	6.93(0.06)	6.7(0.25)	0.4088
Domain 5: Emergent Literacy and Language	10.37(0.07)	9.7(0.29)	0.0298

Mapping alternative socio-economic variables against the current *quintile ranking*

The main proxy used to re-define socioeconomic status for each child is the monthly fee charged at the ELP that the child attends, and whether the ELP receives a subsidy from the Department of Social Development (DSD). Additional variables such as: whether the ELP has access to running water, electricity for lighting, access to a flush toilet connected to a sewage system, and whether the ELP has a formal or informal (in a shack) building were used to sense check results.

Table 2 shows the average monthly fee charged at each ELP (column 2), the percentage receiving any subsidy (column 3), and the percentage of facilities who have access to various facilities (columns 4-6), by the 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 of 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 of facilities	(8) N children
1	R183 [R0;R3300]	71%	66%	81%	42%	391	1627
2	R245 [R0;R2560]	65%	72%	84%	49%	287	1208
3	R252 [R0;R2050]	60%	84%	88%	65%	266	1100
4	R393 [R0;R2888]	54%	96%	96%	91%	114	469
5	R882 [R0;R3600]	25%	98%	91%	94%	115	507
Overall	R310.31	62%	78%	86%	59%	1173	4911

Refined socio-economic levels

Refined socio-economic levels were determined using a k-means clustering approach[2] 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 access[3] or fees only.

Table 3 displays the average characteristics of ELPs using the refined socio-economic levels. Level 1 (L1) shows the lowest fee level while level 5 (L5) presents those in the highest fee level. As expected, receipt of a subsidy decreases while access to services increases as the ELP fee increases. The differences by level are much starker than by quintile, pointing to a more accurate gradient. There are also larger gaps between the first and fifth quintile in access to facilities in comparison to the DBE quintile categories. In particular, access to a flush toilet (33% in level 1 versus 100% of those in level 5) while this is 42% versus 94% using quintiles.

A disadvantage of the updated SES levels however, is that the sample size of children in the highest level is substantially lower – moving from 507 to 142 children.

[2] 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. Meaning 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 (>R2000 per month) and the remainder of the sample.

[3] This model was based on N=4482 observations and had a grouping of R0-R130; R140-R320; R340-R785; R800-R1750; R1888-R3600.

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%	78%	33%	400	1670
L2: R111-R290	R185	72%	83%	85%	58%	433	1820
L3: R291-R750	R410	35%	96%	95%	87%	242	1000
L4: R751-R1750	R1123	6%	99%	99%	100%	68	279
L5: R1751+	R2623	0%	100%	97%	100%	30	142
Overall	R310	62%	78%	86%	59%	1173	4911

Revisiting the Index socio-economic gradient using fee levels as a proxy for socio-economic status, instead of quintiles

Replacing quintile ranking with these five fee levels as a measure of socio-economic status (SES), we re-examined the relationship between poverty and child outcomes, including total ELOM score, each of the learning domains, socio-emotional functioning and physical growth. All observations were weighted.

Total ELOM 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, 58% of children in Q5 were found to be On Track, compared to 38% in Q1. Using the updated SES levels (Figure 1) we find that 81% of children in level 5 are on track (depicted by the green line), compared to only a third (33%) of learners in the poorest level. The vertical grey bars represent confidence intervals at a 5% level of significance.

Figure 1: Total Learning scores – On track / Not on track by SES level

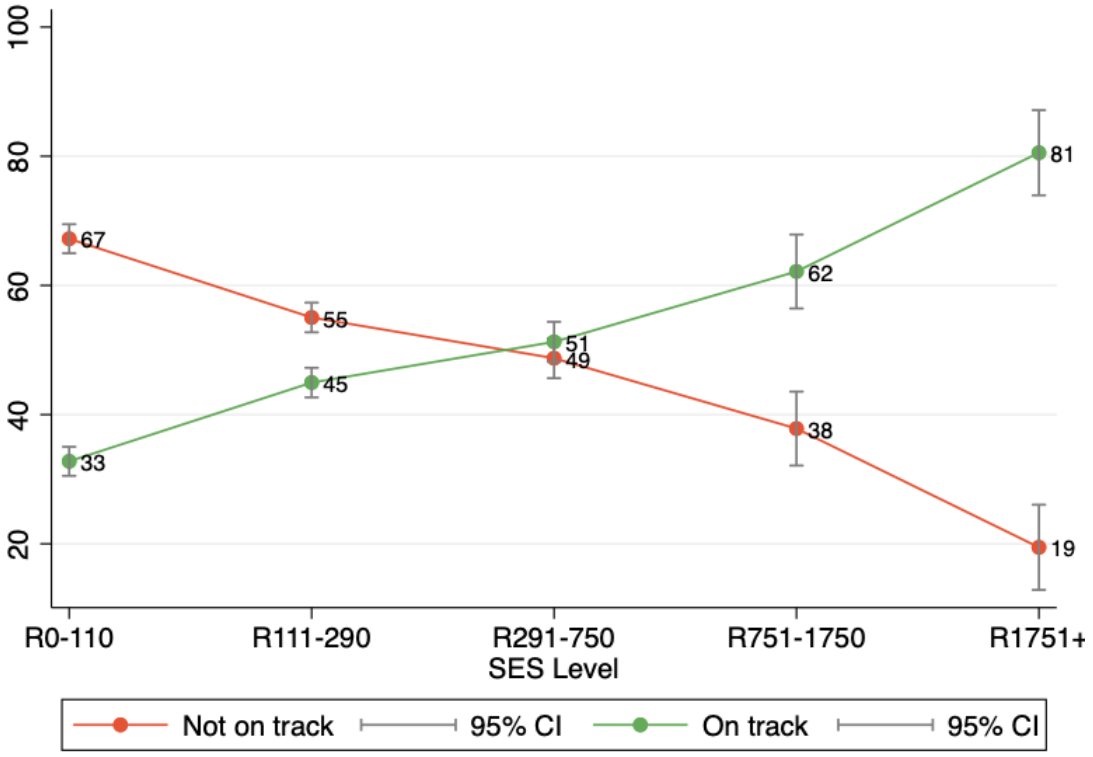
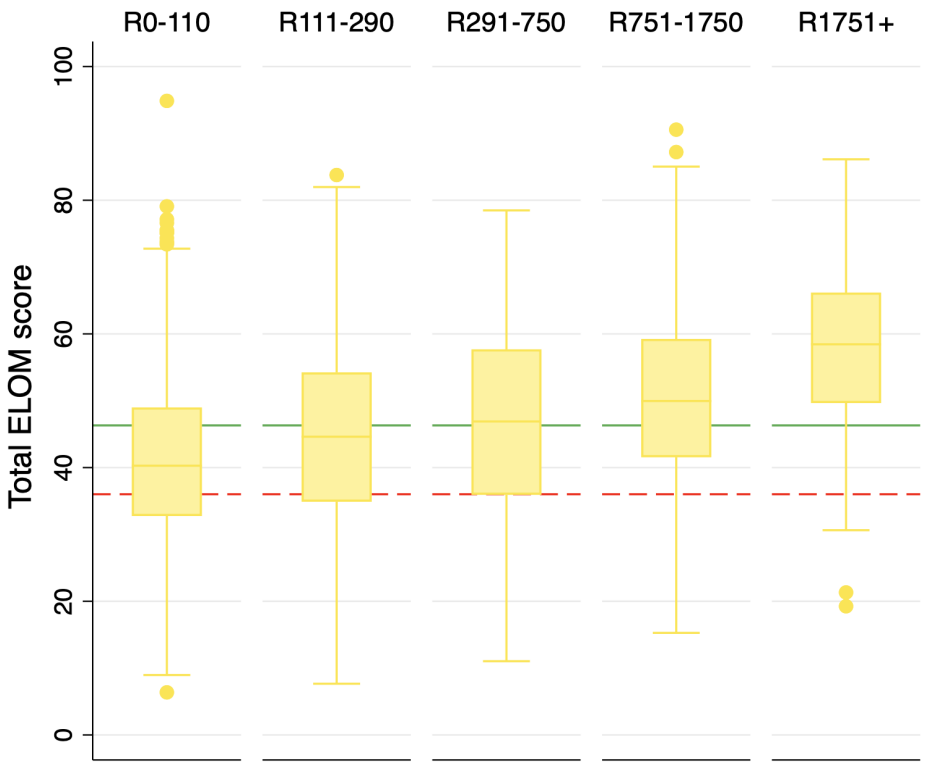


Figure 2 displays the distribution of total ELOM score across SES. The boxes contain the middle 50 percent of scores (interquartile range). The whiskers (outer lines) provide a sense of the total variation in scores.The red dashed line indicates the cut off between those falling far behind (below the line) and those falling behind (above the line). The green solid line indicates the cut off between those that are falling behind (below the line) and those on track (above the line). For the first SES level (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. In comparison to the highest SES level, the median score was 61 ELOM points and 50 percent of scores were between 52 and 69 points. This is much higher than level 1. Level 1 also has the most outlier children relative to the rest of their distribution, making the case that there are children who are excelling and on par with level 5 children despite their lower SES.

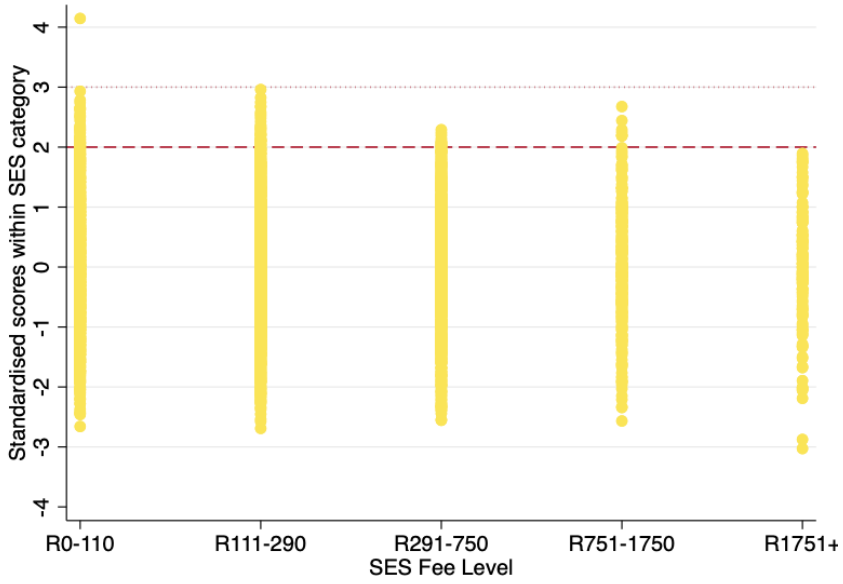
Figure 2: Distribution of ELOM scores by SES level (Index data)



Notes: The red dashed line indicates the cut off between those falling far behind and those falling behind. The green solid line indicates the cut off between those that are falling behind and those on track.

Figure 3 shows the standardised total ELOM scores within SES levels. The image demonstrates the variability in scores within income bands, and in particular within the lower SES levels. Children in lower SES levels achieve more than 2 SDs (and up to 4SDs) above their peers, in comparison to higher SES levels where variability is lower (as depicted by a shorter yellow line).

Figure 3: Distribution of standardised ELOM scores within SES level



ELOM domain scores:

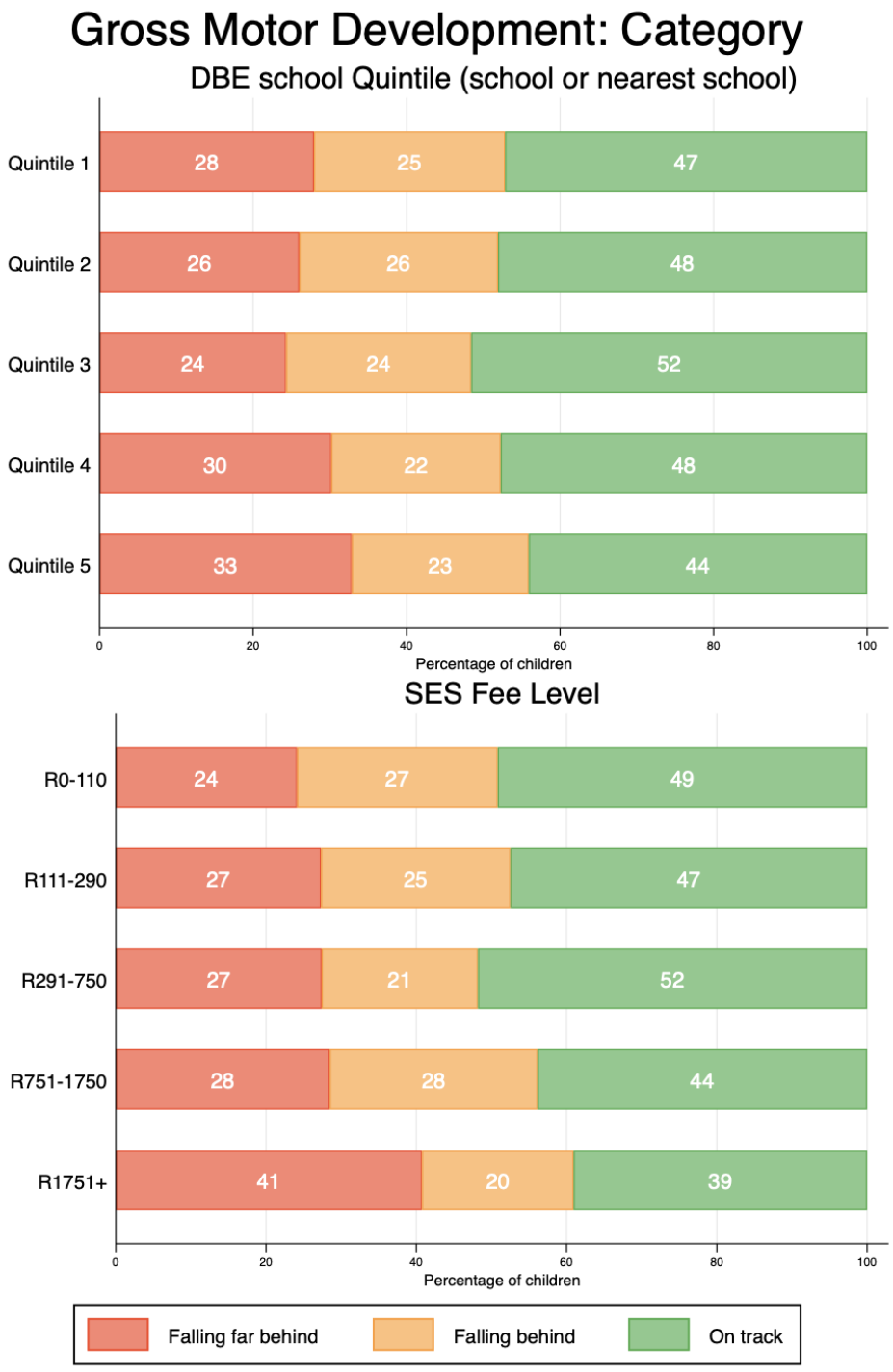
The following figures display disparities in child outcomes by SES levels for each ELOM domain, and compare them to prior reports that used the DBE quintile system as a proxy for SES. The green bars indicate the percentage of children who are "On track" for their development, the orange bars indicate 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[4] for children aged 50-59 months.

Domain 1: Gross Motor Development (Figure 4)

In comparison to the quintile method, differences in outcomes for Gross Motor Development are similar. Relative to their peers, a smaller percentage of children in the 5th quintile and highest fee paying group are on track for their development. For example, half (49%) of the children in SES level 1 are on track, whereas this is 39% in SES level 5.

[4] The ELOM 4&5 technical manual is available at www.datadrive2030.co.za/datatools

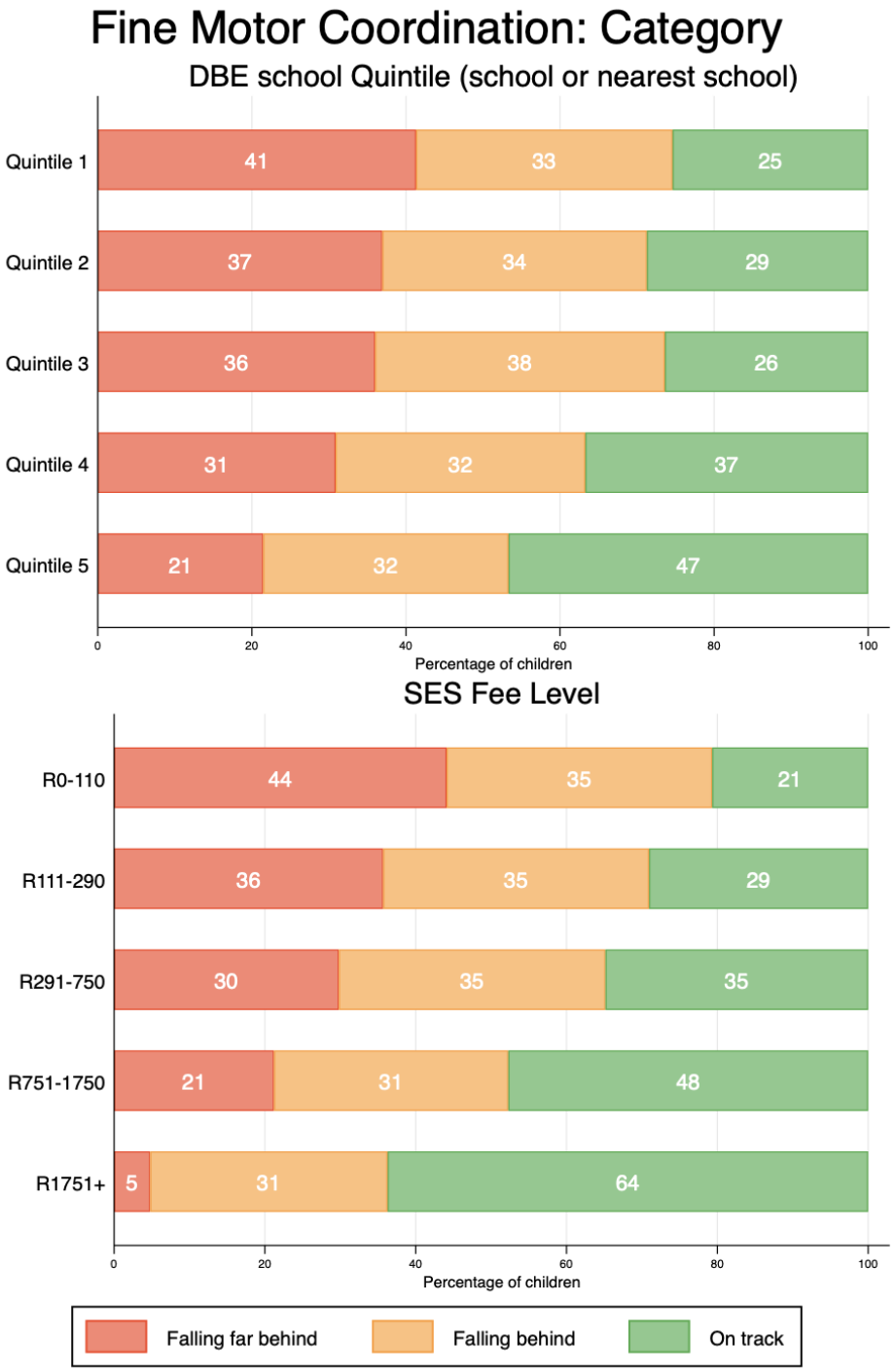
Figure 4: Distribution of Gross Motor Development outcomes by SES level



Domain 2: Fine Motor Coordination and Visual Motor Integration (Figure 5)

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 the highest fee level. The disparities between fee levels 1 to 5 are much more pronounced than between quintiles.

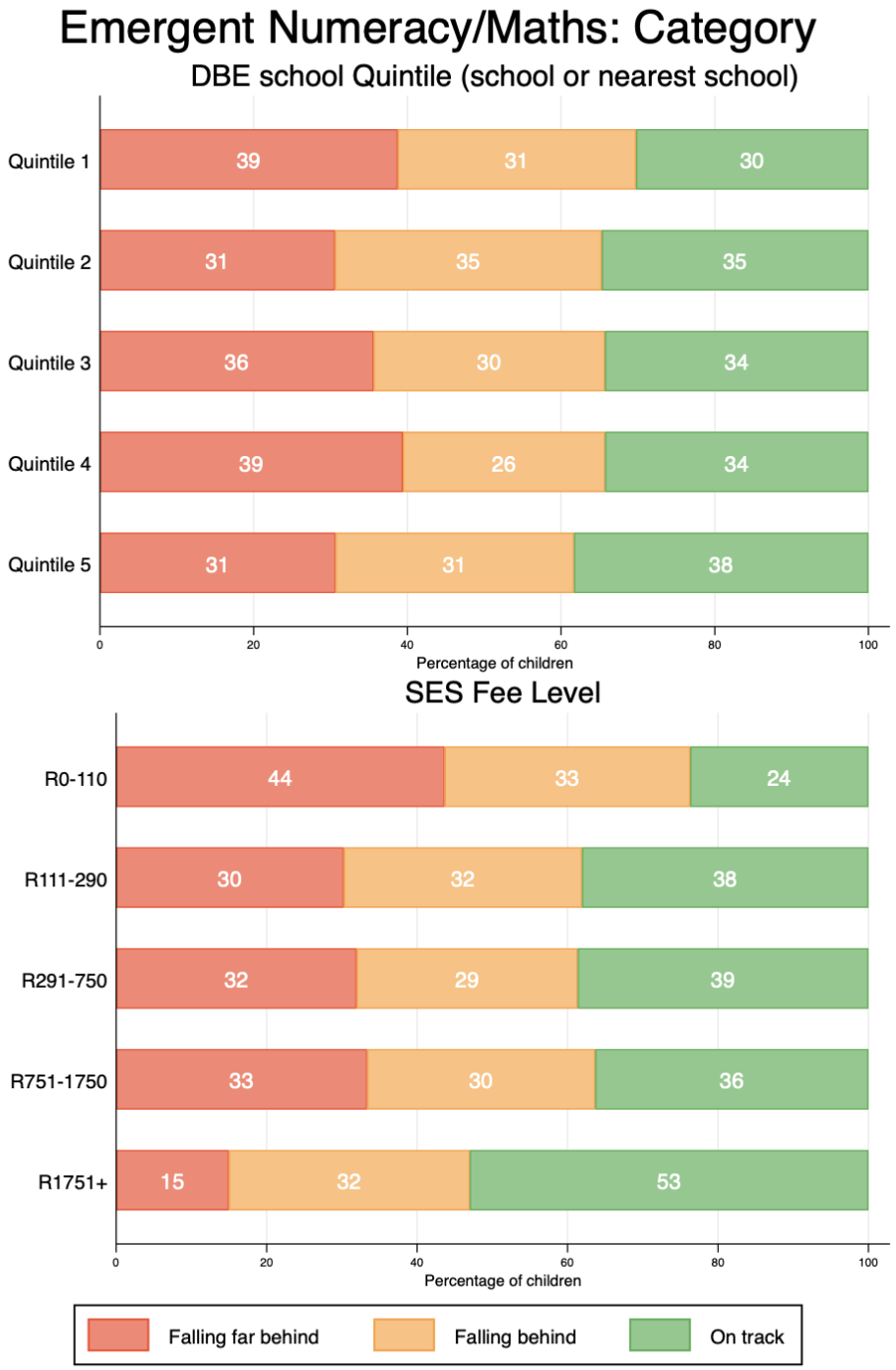
Figure 5: Distribution of Fine Motor Coordination and Visual Motor Integration outcomes by SES level



Domain 3: Emergent Numeracy and Mathematics (Figure 6)

Prior reports of differences in numeracy showed very little differences across quintiles, where on average, a third of children were on track. However, using ELP fee levels show a much larger difference particularly in the lowest and highest level. Learners in the highest level outperform other levels by almost 20% on average.

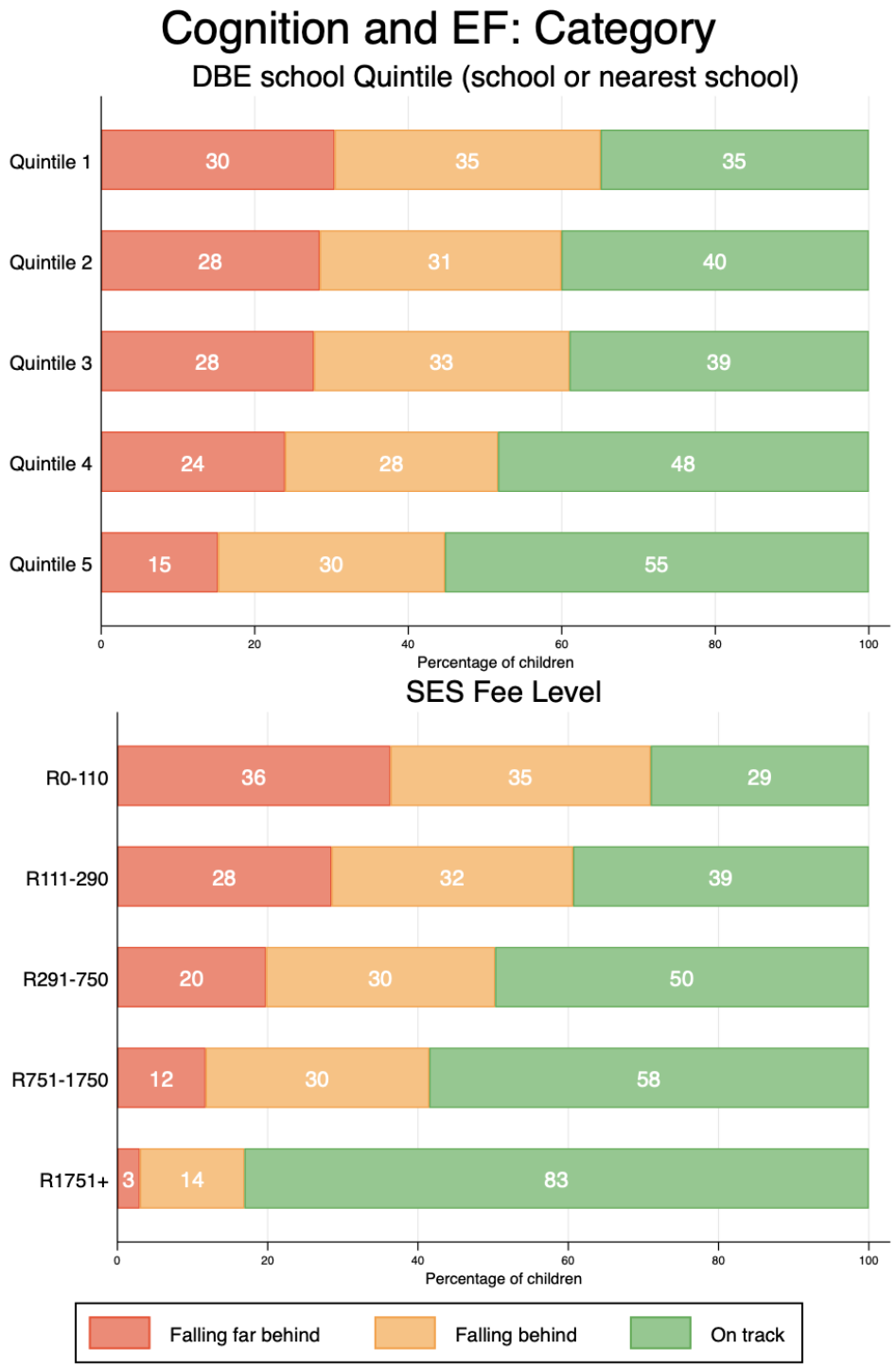
Figure 6: Distribution of Emergent Numeracy and Mathematics outcomes by SES level



Domain 4: Cognition and Executive Functioning (Figure 7)

The largest differences across SES relate to cognition and executive functioning. Less than a third of children in level 1 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.

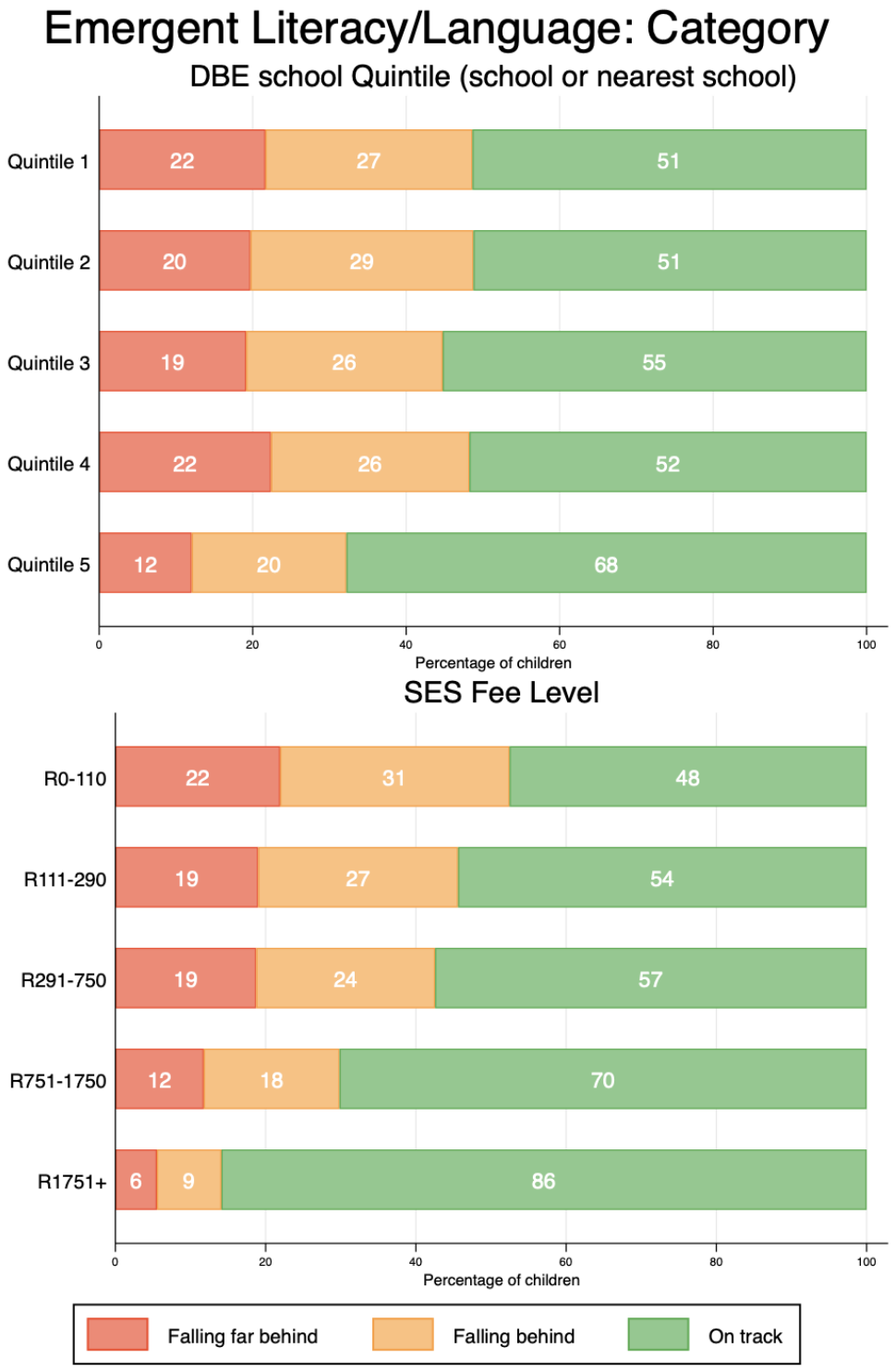
Figure 7: Distribution of Cognition and Executive Functioning outcomes by SES level



Domain 5: Emergent Literacy and Language (Figure 8)

Finally, differences in outcomes by SES levels remain pronounced for literacy and language skills where the vast majority of children are on track (86%) in level 5 in comparison to less than half (48%) of children in level 1.

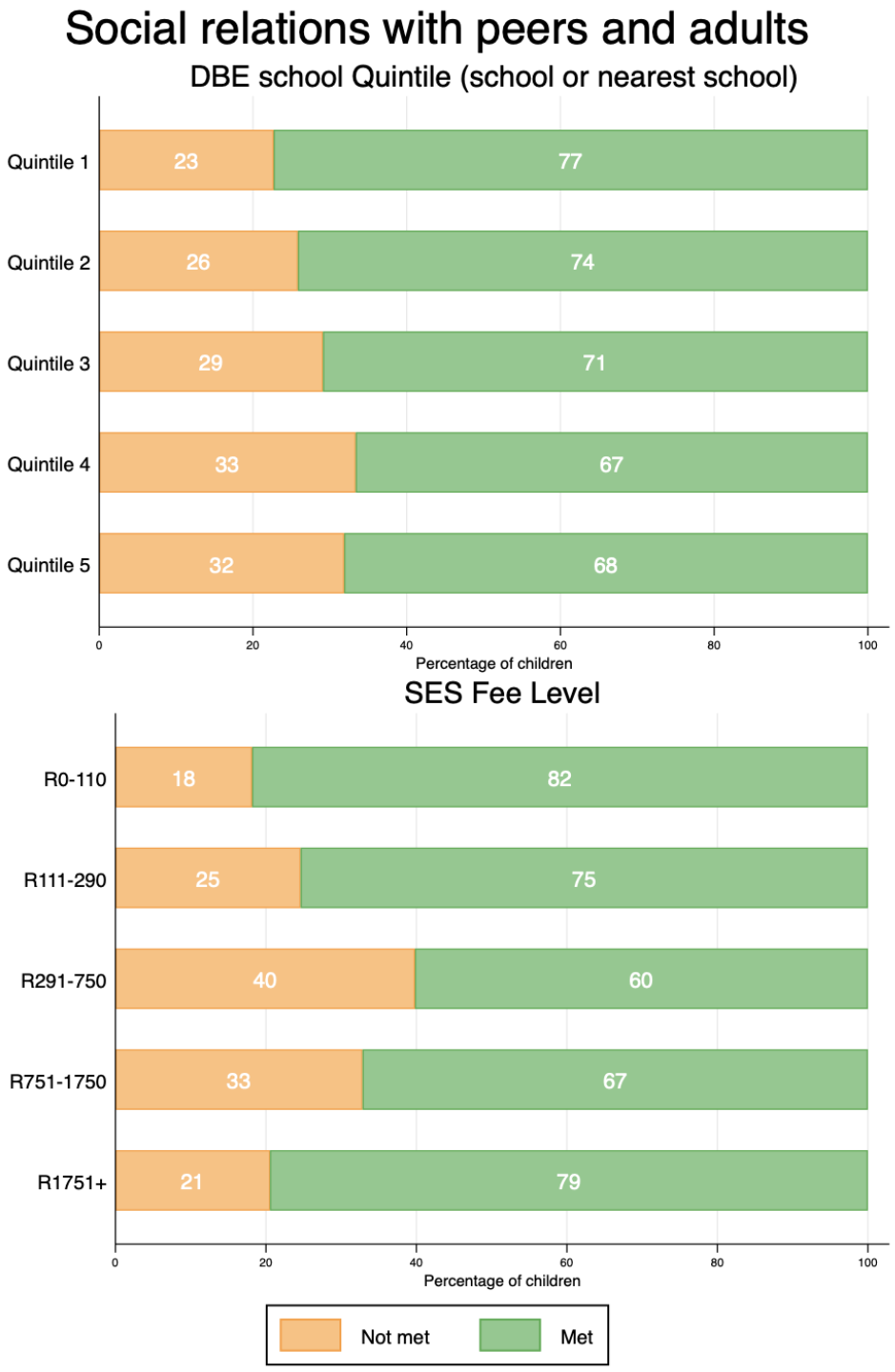
Figure 8: Distribution of Emergent Literacy and Language outcomes by SES level



Social-Emotional Functioning: Social Relations with Peers and Adults (Figure 9)

The relationship between meeting social relations standards 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.

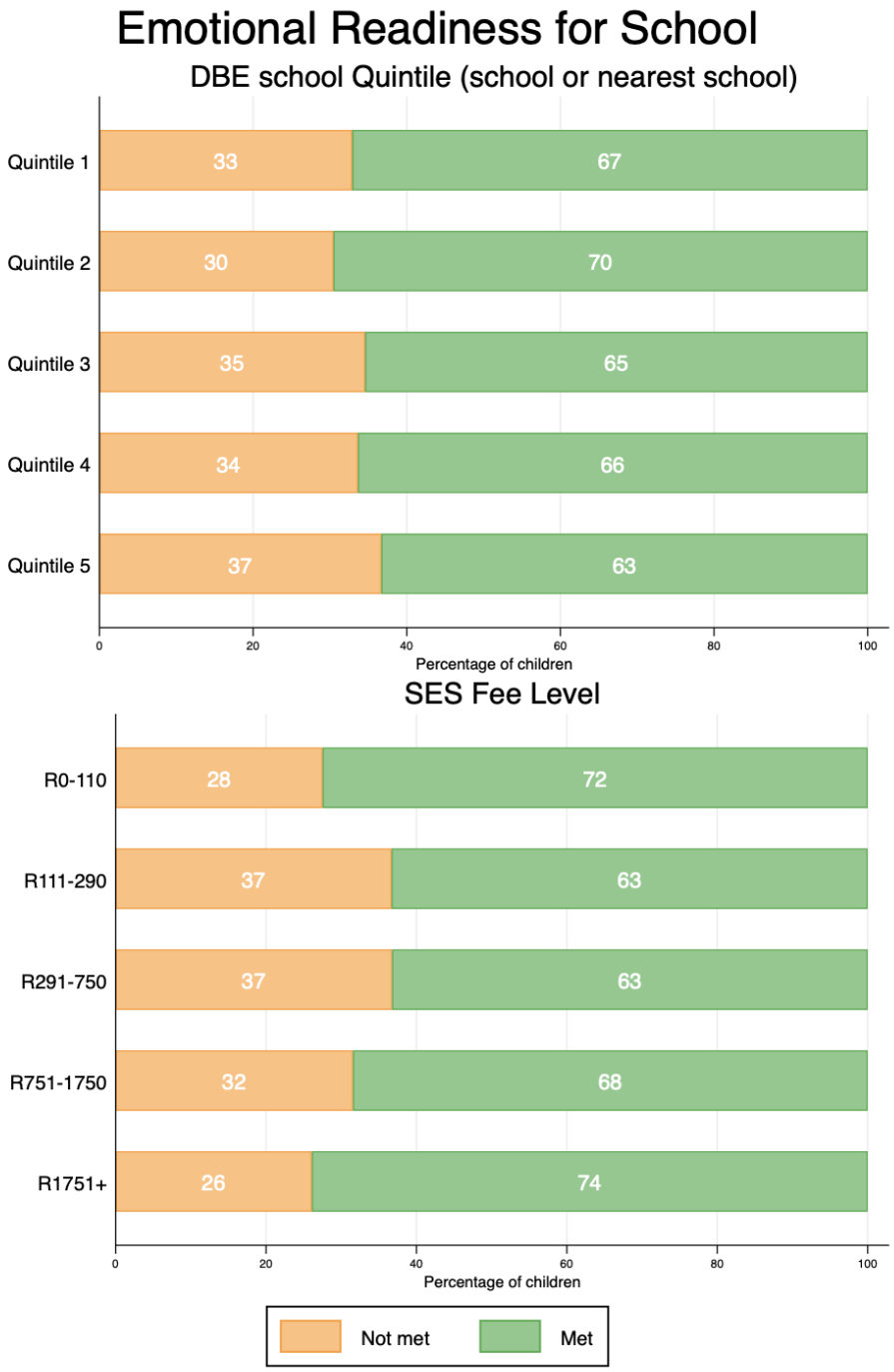
Figure 9: Distribution of Social Relations outcomes by SES level



Social-Emotional Functioning: Emotional readiness for school (Figure 10)

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).

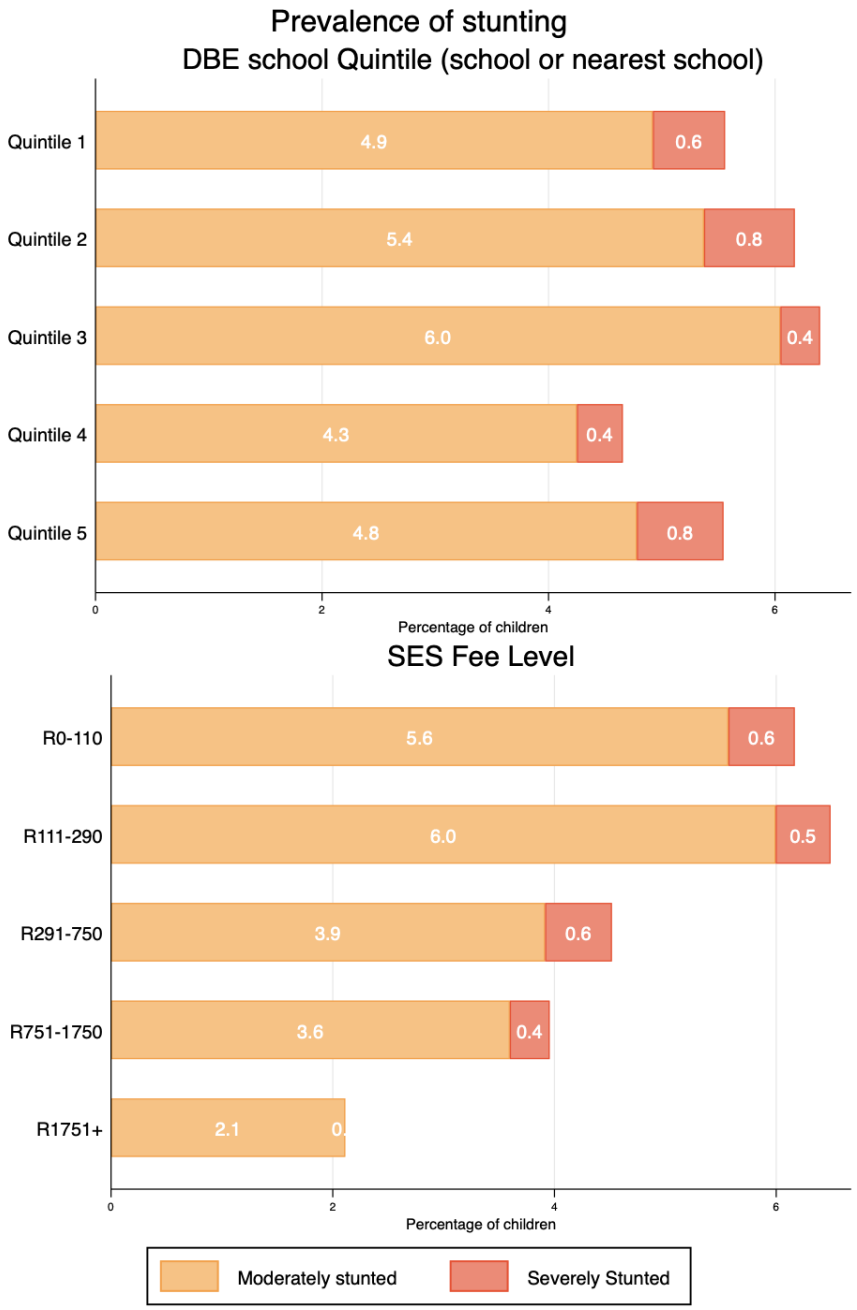
Figure 10: Distribution of Emotional Readiness outcomes by SES level



Physical growth: Stunting (Figure 11)

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, and 0.5% severely stunted. These rates decrease to 2.1% in the highest group for moderate stunting and 0% for severe stunting. Children are defined as 'Moderately stunted' if their height-for-age[5] is more than 2 SDs below the WHO Child Growth Standards median and 'Severely stunted' if below 3 SDs.[6]

Figure 11: Distribution of stunting prevalence by SES level



[5] Height-for-age scores were calculated using age in months
[6] Data on child height was missing for 7 children, resulting in a total of N=5132.

Conclusion

Using monthly fee levels at the ELP to proxy for socioeconomic status displays the clear disparities in child outcomes across income groups. These differences are much starker than using the DBE quintile system. It is possible that children excluded from this analysis may be statistically different from those included, their descriptive characteristics point to them being children from lower socioeconomic backgrounds. Since their average scores are lower, it is possible that these estimates may be slightly downward-biased and outcomes may be even more disparate with their inclusion. Additionally, the presence of outliers within lower fee levels makes the case that there are children who are doing exceedingly well within these groups – and perform similarly to those in higher level groups. As a next step, DataDrive2030 will conduct a positive deviance analysis, where we will attempt to understand the characteristics associated with positive outliers.