UNICEF Database Manager Assessment

Task 1: Maternal and Child Health Indicators Analysis

Your task is to calculate population-weighted coverage of health services (antenatal care and skilled birth attendance) for countries categorized as on-track and off-track in achieving under-5 mortality targets as of 2023.

The bar chart provides a snapshot of the relationship between maternal health indicators, specifically Skilled Birth Attendant and Antenatal Care (4+ visits) for women aged 15-49 years worldwide, based on household survey data from 2018-2022. It displays population-weighted averages for these indicators according to the SDG progress status of Under-5 Mortality Rate (U5MR), categorized as "Achieved," "On Track," and "Acceleration Needed." The findings suggest an association between maternal health indicators and U5MR status. However, these relationships are not statistically significant everywhere. Overall, the pattern indicates that countries with higher rates of skilled birth attendants and antenatal care visits tend to have lower U5MR rates. Specifically, countries that have achieved under-five mortality targets exhibit significantly higher rates of skilled birth attendance and antenatal care. While this analysis does not allow for causal claims, it does suggest that maternal health indicators play a crucial role in achieving under-five mortality targets.

The particular focus of this task is the distinction between the "Acceleration Needed" (off track) and "On Track" categories. For the "Antenatal Care 4+ visits" indicator, countries in the "On Track" category (56.5%) exhibit slightly higher coverage than those in the "Acceleration Needed" category (54.4%), but the difference is not significant, as the confidence intervals overlap.

In contrast, a different pattern emerges with the "Skilled Birth Attendant" indicator. Countries that are "On Track" have a mean coverage of 85.7%, while those requiring "Acceleration" lag behind at 63.8%. Since the confidence intervals do not overlap, these differences are statistically significant. This suggests that the "Skilled Birth Attendant" indicator is more closely related to achieving the U5MR target than the "Antenatal Care 4+ visits" indicator. However, as mentioned earlier, further research is needed to establish a causal relationship between these indicators and U5MR to make well-grounded policy recommendations.

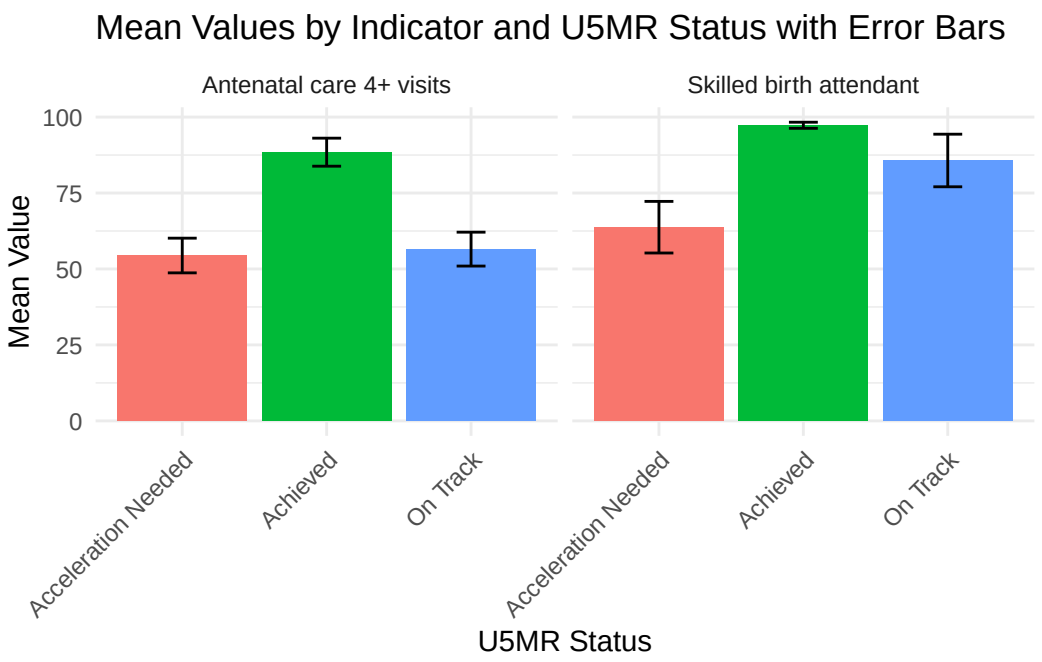

Figure 1: Mean Values by Indicator and U5MR Status with Error Bars

Task 2. Early Childhood Development Index (ECDI) Analysis

Your task is to produce a Data Perspective on the evolution of education for 4- to 5-year-old children. Particularly interesting for the Perspective is understanding how educational performance evolves month by month at these critical ages, considering both general education and specific subjects (e.g., literature and math, physical education). Methods should be sound and well-argued if necessary.

The variables outlined in the task for the Zimbabwe MICS 2019 represent the older, unrevised measure of the Early Childhood Development Index (ECDI). The ECDI is a composite index that assesses the development of children aged 36-59 months across four key domains: literacy-numeracy, physical, socio-emotional, and learning. The ECDI is calculated based on the presence or absence of specific skills in each domain. It is a binary variable, where a value of 1 indicates that a child has reached the expected level of development in all four domains, and a value of 0 indicates that they have not.

However, as per the instructions for this task (outlined in README.html), the suggested approach to aggregating the index is not based on creating binary outcomes for each domain. Instead, it involves calculating a continuous index using the arithmetic mean.

Additionally, the ECDI has recently been replaced by ECDI 2030, a new measure with a significantly different methodology. While the details of the new measure are beyond the scope

of this assessment, I will focus on producing the ECDI based on the methodology proposed in the assignment and present its main findings as outlined in the task.

How Many Children Achieved the Expected Level of Development?

Below are the findings of the ECDI and its key components for Zimbabwe MICS 2019:

Descriptive Statistics
Early Childhood Development Index and Subcomponents

Variable	N	Min	Mean	Median	SD	Max
EC6	2488	0.00	0.10	0.00	0.30	1
EC7	2490	0.00	0.11	0.00	0.31	1
EC8	2489	0.00	0.18	0.00	0.39	1
EC9	2456	0.00	0.93	1.00	0.25	1
EC10	2490	0.00	0.61	1.00	0.49	1
EC11	2490	0.00	0.87	1.00	0.34	1
EC12	2489	0.00	0.74	1.00	0.44	1
EC13	2490	0.00	0.96	1.00	0.19	1
EC14	2487	0.00	0.51	1.00	0.50	1
EC15	2486	0.00	0.63	1.00	0.48	1
Literacy and Numeracy	2490	0.00	0.13	0.00	0.25	1
Physical	2490	0.00	0.77	1.00	0.29	1
Socio-Emotional	2490	0.00	0.70	0.67	0.26	1
Learning	2490	0.00	0.80	1.00	0.32	1
ECDI	2490	0.08	0.60	0.62	0.16	1

The estimates are unweighted, as the dataset did not provide weights. It is observed that 60% of children are developmentally on track overall. (Using the official methodology, this figure is 71%, which aligns with the MICS Survey Findings Report—codes for this calculation are not presented here but can be found in the respective scripts.) While the highest percentage of children are on track in learning (80%), the lowest percentage pertains to literacy and numeracy, with only 13% on track. Interestingly, the official index showed the highest values for the physical domain, with around 90% of children in the country being on track in terms of their physical development.

Psychometric properties (reliability analysis)

We wanna measure how the final score of ECDI relates with its subcomponents by carrying out reliability analysis using Cronbach Alpha. Below is the output of the reliability analysis:

```
Reliability analysis
Call: alpha(x = cronbach.data)

  raw_alpha std.alpha G6(smc) average_r  S/N   ase mean   sd median_r
      0.31       0.3    0.25     0.099 0.44 0.022  0.6 0.16     0.11

    95% confidence boundaries
         lower alpha upper
Feldt     0.26  0.31  0.35
Duhachek  0.26  0.31  0.35

 Reliability if an item is dropped:
               raw_alpha std.alpha G6(smc) average_r  S/N alpha se   var.r
lit_num.alt         0.27      0.28    0.20     0.113 0.38    0.025 0.00014
physical.alt        0.27      0.27    0.20     0.108 0.36    0.025 0.00085
socio_emot.alt      0.24      0.23    0.17     0.093 0.31    0.026 0.00148
learn.alt           0.21      0.21    0.15     0.081 0.26    0.027 0.00108
               med.r
lit_num.alt    0.116
physical.alt   0.123
socio_emot.alt 0.100
learn.alt      0.075

 Item statistics
                  n raw.r std.r r.cor r.drop mean   sd
lit_num.alt    2490  0.51  0.55  0.24   0.14 0.13 0.25
physical.alt   2490  0.56  0.56  0.25   0.14 0.77 0.29
socio_emot.alt 2490  0.54  0.58  0.30   0.17 0.70 0.26
learn.alt      2490  0.65  0.59  0.34   0.19 0.80 0.32

Non missing response frequency for each item
                  0 0.333333333333333  0.5 0.666666666666667    1 miss
lit_num.alt    0.74              0.15 0.00              0.07 0.03    0
physical.alt   0.04              0.00 0.38              0.00 0.58    0
socio_emot.alt 0.01              0.22 0.00              0.43 0.34    0
learn.alt      0.09              0.00 0.22              0.00 0.69    0
```

The overall standardized Cronbach's alpha is 0.3, which is considered very low for an index. However, it is important to note that the approach used to produce the index in this assessment (in accordance with the instructions in README.html) does not follow the official ECDI methodology, which results in a binary measure rather than a continuous one. Nonetheless,

since the input variables are the same, these results still provide valuable insight into how the subcomponents of ECDI relate to each other.

The reliability statistics, which show the impact of dropping an item, indicate that the current ECDI is primarily driven by the literacy-numeracy and physical domains. The overall scores of the index would decrease if these items were dropped, highlighting their significance in the index.

Quality check: how many observations per month?

In order to make sure our further analysis on addressing the changes in ECDI over time by month is valid, it makes sense to check the number of observations per month.

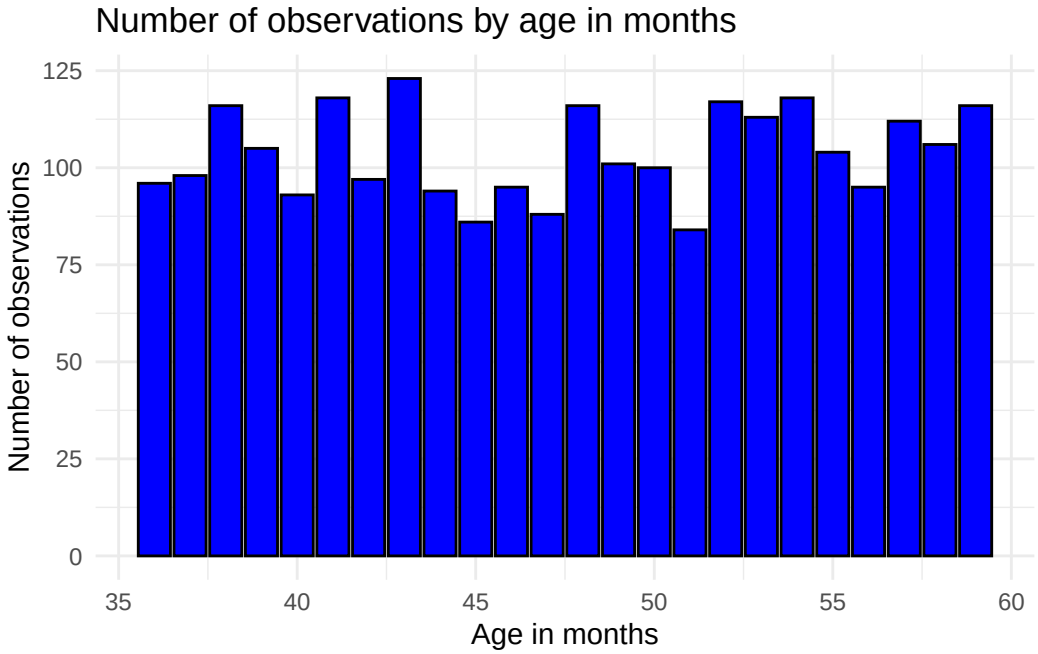

Figure 2: Number of ECDI Observations by Age in Months

The lowest number of obs is 84 and refers to 51st month, while the highest number of obs is 123 and refers to 43th month. This suggests that the data is sufficient to analyze the changes in ECDI over time by month.

Descriptive exploration: distribution of ECDI and its subcomponents by age in months

The chart below checks the distribution of ECDI score and its main components by age in months.

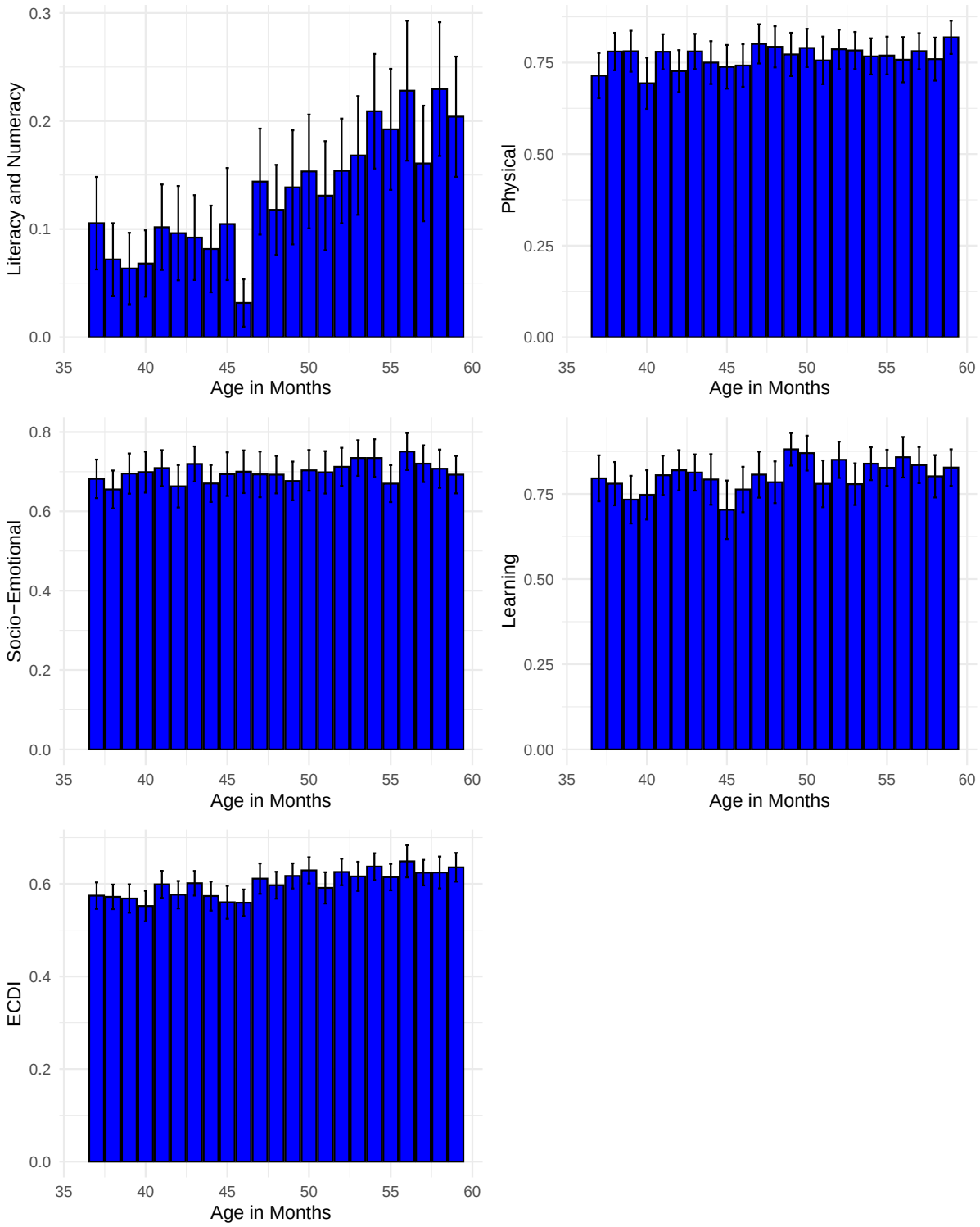

Figure 3: Distribution of ECDI and its subcomponents by Age in Months

The figure suggests that the changes in ECDI by month, as well as in most of its components, are not very substantial. Moreover, the overlapping 95% confidence intervals indicate that, in most cases, these changes are not statistically significant. This may imply that the ECDI and its components are not sensitive enough to detect small changes in child development over time. However, further research is needed to confirm this hypothesis.

Regression analysis

We ran five linear regressions, each predicting ECDI or one of its four subcomponents by age in months, to determine if changes over time in the index or any of its main components are significant. The results of the models are presented in the table below.

Regression Results
Effect of Age in Months on ECDI Subcomponents

Outcome	Beta	SE	P-Value	R-Squared	N
Literacy and Numeracy	0.007	0.001	0.00	0.037	2490
Physical	0.001	0.001	0.08	0.001	2490
Socio-Emotional	0.002	0.001	0.03	0.002	2490
Learning	0.004	0.001	0.00	0.006	2490
ECDI	0.003	0.000	0.00	0.022	2490

The table shows that age has a significant effect on ECDI and all of its subcomponents (though for the physical domain, it is significant at the 10% confidence level). However, the beta coefficients suggest that the magnitude of these changes is rather small, and the changes are not substantial in a practical sense. The highest change is observed in literacy and numeracy (beta = 0.007). This may indicate that ECDI and most of its components are not sensitive enough to detect small changes in child development over time. Given that one month is a short interval to quantify such effects, I visualized the models to better understand the patterns of child development by exploring the slopes over 24 months of children's lives.

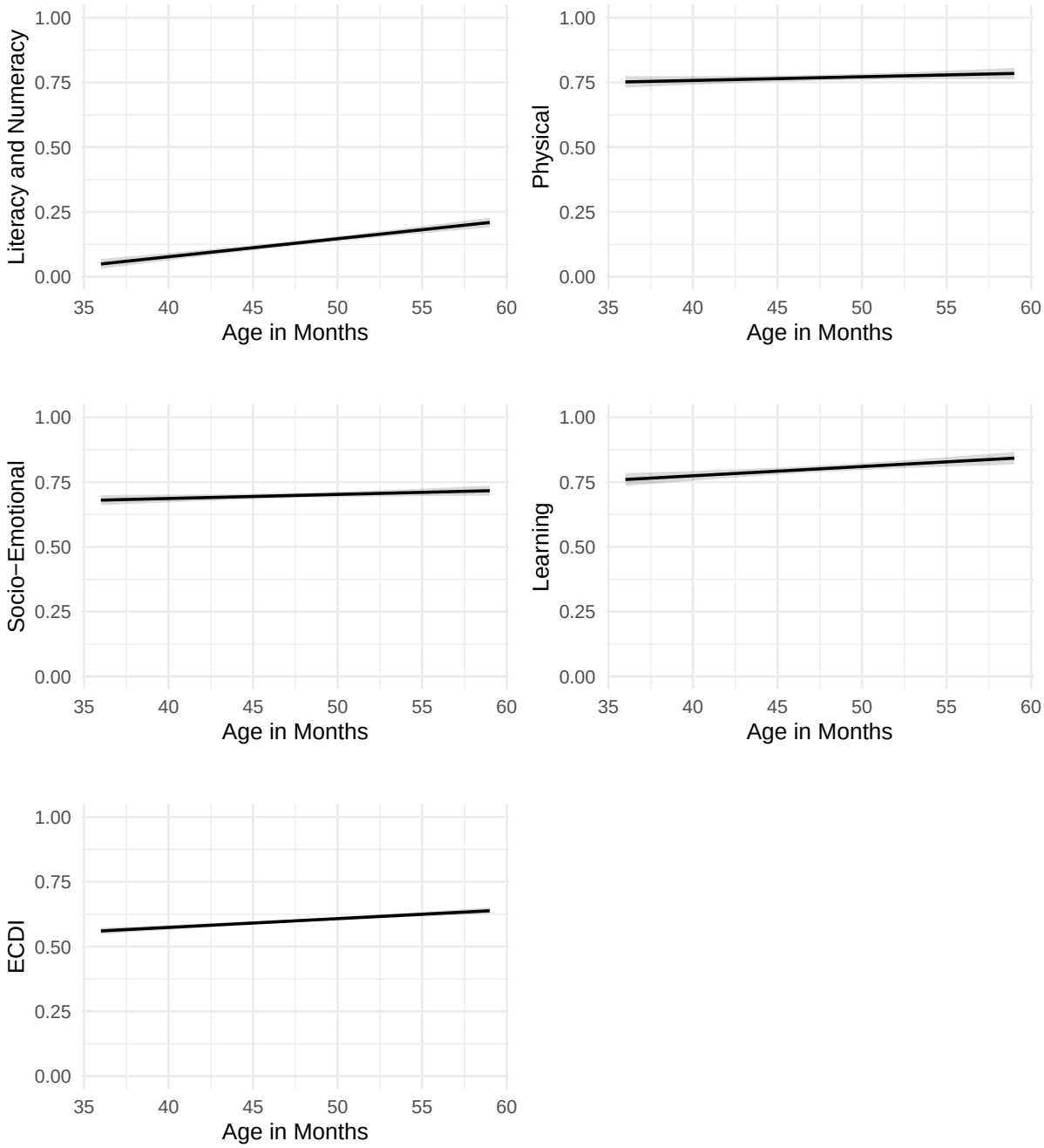

Figure 4: Effects of age in months on ECDI and its subcomponents, linear model

The charts confirm that changes over time are indeed not substantial, and the slopes for some of the componts are rather flat, with the exception of Literacy and Numeracy domain. Additionally, the low R-squared coefficients suggest that age does not explain much of the

variation in child development.

These findings are indeed somewhat counterintuitive. Better knowledge of socio-economic context is needed to interpret these findings. It needs to be noted that the effect of age in the carried out models is also confounded, as we do not have such important controls as sex, area of residence, parental education, wealth quintile, etc. But in addition to the issues related to the interpretation of the model, the findings could also potentially suggest that the proposed measures may not effectively capture changes in child development over time in terms of their reliability, validity, and other, purely psychometric, properties. This could also be a potential reason why a new measure, ECDI2030, was introduced.