ECD Odisha: Results

July 8, 2018

Contents

1 Points for discussion

Results discussion points

- What to do with reserve target children in terms of the 0-3 analysis?
- Should we include compliance in the impact evaluation? If so how?
- Decision on standardization method.
- How to explain nutrition results?
- How to explain anthropometric results?

New intervention discussion points

- What to do with reserve target children?
- What to do with children who migrated to a nearby village?

2 Survey Attrition

The sample at baseline was measured at 2170 children. The total number of responses by module of these original 2170 is given in table 1 at baseline midline and endline. In brackets for each module is the number of additional children measured in the surveys (reserve target children) who had replaced sample children who had dropped out of the treatment. For these children we have endline measures in all modules, but no data from baseline and most of midline.

Table 1: Sample by module

	(1) Baseline	(2) Midline	(3) Endline
Roster	2170	2099	2056
	(0)	(0)	(56)
HH module	2167	2089	2056
	(0)	(0)	(56)
Target child	2170	2098	2041
	(0)	(0)	(56)
Primary Care Giver	174	69	59
	(0)	(0)	(2)
Biological mother	2168	2093	2035
	(0)	(0)	(56)
Diet Tool	2163	2088	2053
	(0)	(0)	(56)
Anthro	2157	2089	2013
	(0)	(0)	(54)
ASQ	2170	2031	
	(0)	(43)	
HH environment	2170	2089	2053
	(0)	(0)	(56)
Adapted McArthur		2098	2041
		(0)	(56)
Bayley		2061	2002
		(0)	(52)
WPPSI			2002
		(0)	(52)
Raven			2012
			(54)

The attrition rate varies by whether or not we wish to include reserve target children as part of the sample. Looking at the sample who were present at baseline (excluding reserve target children) we see a rate of 5.2% for the roster, 7.7% for the bayley and 6.6% for anthropometrics. The attrition by treatment arm is given by table 2, with the implied rate of attrition in brackets.

Table 2: Attrition by treatment arm, excluding reserve target children

	Control	NE	GS & NE	IS &NE
Baseline Roster	532	537	539	556
Roster	495	514	512	536
	(0.07)	(0.04)	(0.05)	(0.04)
Bayley	484	499	499	521
	(0.09)	(0.07)	(0.07)	(0.06)
Anthro	485	499	504	526
	(0.09)	(0.07)	(0.06)	(0.05)

The reasons for attrition in the endline survey are displayed in table 3. Please note that the target sample in endline had excluded those families who had been reported permanently migrated at midline, or who for some other reason could not reasonably be expected to take part in the survey (hence why this numbers do not match with those above). The most common reasons for attrition were migration and unavailability of the respondent.

Table 3: Reasons for non-participation in the survey

Reasons for non-participation in the survey	Salepur	Bolangir	Soro	%
Migration (outside sample block)	6	6	8	32.3
Unavailability of Biological Mother	1	1	4	9.7
Death of Target Child	1	0	0	1.6
Refused	2	1	5	12.9
Respondent temporarily unavailable	10	1	11	35.5
Seasonal migration	0	0	0	0.0
House lock	0	0	1	1.6
Others	1	0	0	1.6
Duplication	3	0	0	4.8

3 Intervention Attendance

3.1 Attendance Data

The attendance data from the 0-3 intervention comes from two sources. The first and most complete is the monitoring data, which is the digitized version of the forms that home visitors and group facilitators (at least in theory) filled out at the end of each session. This contains information not only about attendance, but also items such as what activities were performed and how well each child performed. The major drawback is that forms are missing for weeks where the individual home visits were missed or when the names of children who did not attend a particular group session were omitted from the form. In order to mitigate this we have made the conservative assumption that if a form is missing for a given week, or a name missed in the group form, that that child did not attend in that week. On average there were 44 missing forms for the IS arm and 41 for the GS[1].

The second source of data is from the tracking sheet. The tracking sheet was made by the Supermentors in each district, and contains a list of target children in treatment villages by week. This was filled in with an indicator for if a child attended that week, or a short description of why not if they did not. The issue here is one of missing data. From roughly week 5 to 66 of the trial this sheet was not filled in as it was deemed too much work for the Supermentors and that the same information was (in theory) in the other monitoring data.

In this section we rely primarily on the monitoring data to get a picture of the whole duration of the trial, but for investigation of the reasons for absence we use the tracking sheets. All analysis is from the monitoring data unless otherwise specified.

3.2 Non Compliance

Home visits and group sessions display very different patterns of attendance. Non-compliance has two dimensions, "dropouts" (those who either never turned up or turned up for some but then stopped coming) and the number of sessions attended for those children who did not drop out.

[1]We have omitted NE from the analysis using monitoring forms because a small proportion of the children are listed as attending an implausible number of sessions (over 100).

3.2.1 Dropouts

The percentage of dropouts by treatment arm is shown in table 4. Here dropout is defined as attending 0 sessions or having been marked as dropped out in the tracking sheets. The average dropout rate across all treatment arms was 17.0%, but was significantly higher in the group sessions at 25.5%.

Table 4: Compliance by treatment arm (% of baseline sample)

	Assigned treatment group			
	NE	GS+NE	IS+NE	Total
Non-dropouts	88.8	74.5	85.9	83.0
Dropouts	11.2	25.5	14.1	17.0

The survival function by week of intervention (where the failure state is the last home visit or group session) is shown in figure 1.The large drop at the end of this period coincides with conclusion of the intervention, and the NE arm is currently excluded from this graph due to aforementioned data issues. From the very beginning of the trial we were less successful at getting mothers to attend the group sessions, and furthermore this gap grew over time at a monotonic rate.

Table 5: Reasons for Non Compliance (from tracking sheets)

Reasons for dropout	%
Migration	44.8
Moved in with relatives (uncle)	2.3
Refusal or not interested	18.4
Death of target child	3.4
Child never located	4.6
Distance too great	9.2
Child Over age	5.7
Mother pregnant again	3.4
Other	8.0

Figure 1: Survival Function By Treatment Arm
(failure = last session attended)

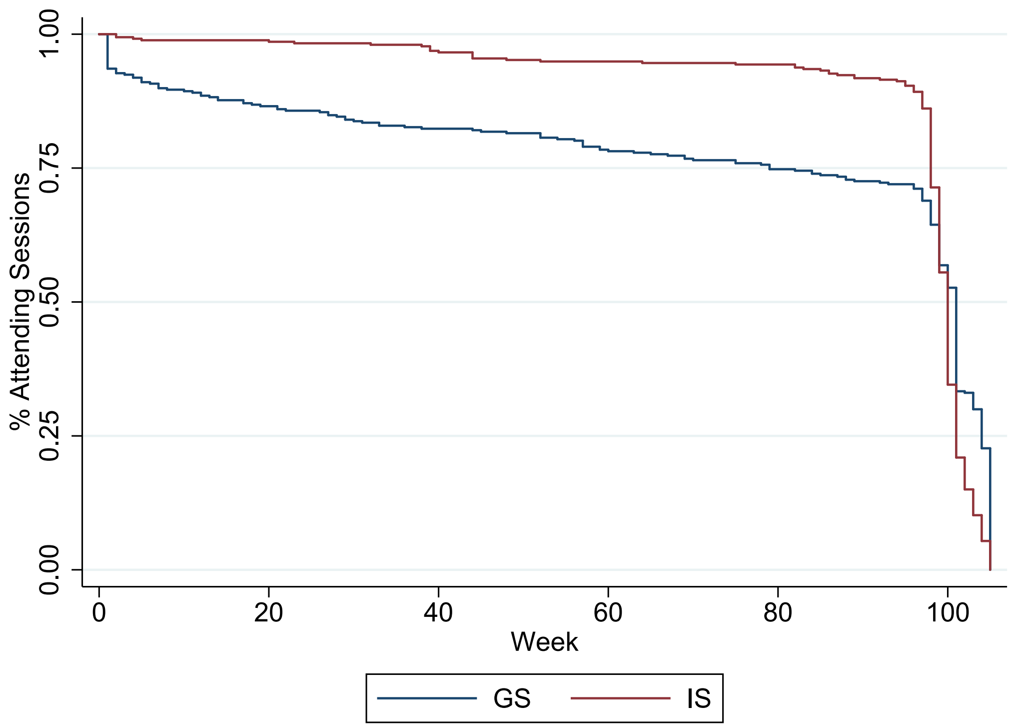


Table 5 shows information from the tracking sheets which occasionally gave a reason for dropout. This is only available for 87 children only out of total of 188. Migration is by far the largest given reason, followed by refusal.

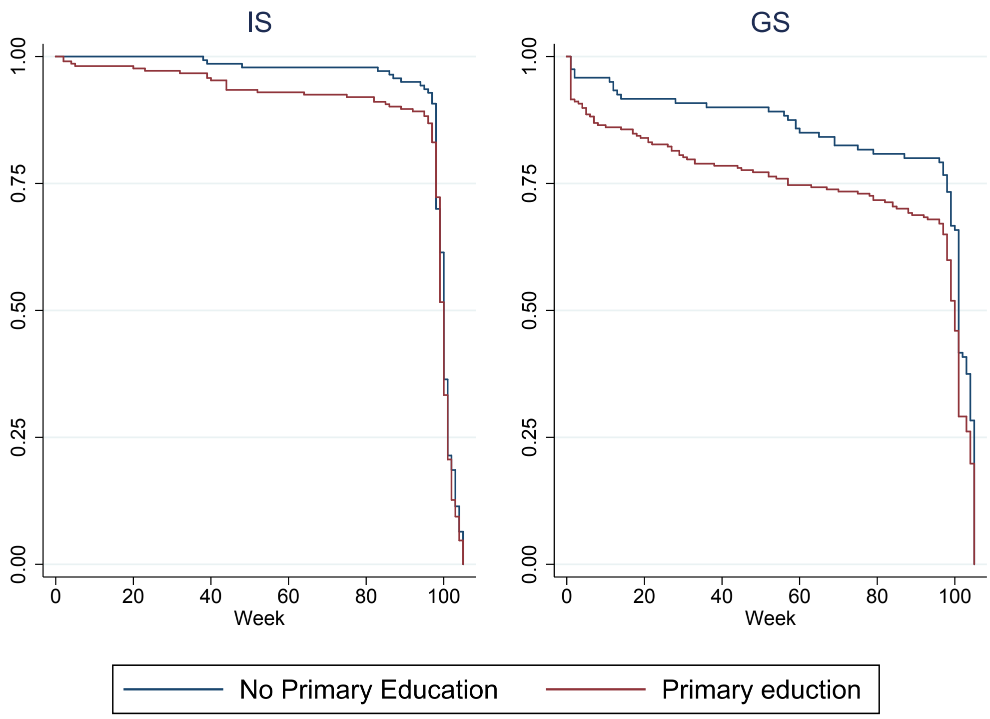
Predicting non-compliance is clearly of interest. The first way in which we have done this is to simply correlate dropout with baseline characteristics, as shown in table 6. The strongest predictors for dropout probability seem to be whether or not the biological mother of the target child is alive, a primary education dummy for the primary care giver, and a negative correlation with an indicator for if the survey respondent (a household member) to skip a meal in the last week. From this we start to get a picture of dropouts being the wealthier and better educated amongst our sample. Explanations for this could be a higher opportunity cost of the mothers time or simply that these mothers believe that a home visiting program is of less use to them or their child.

Table 6: Correlations between characteristics and dropout probability

	(1) dropout
Wealth Quintile	0.00801 (0.0107)
Biological mother is still alive?	-0.0721** (0.0333)
Mother believes intelligence difficult to change	-0.0383 (0.0260)
Someone in HH skip meal last week	-0.0429* (0.0223)
Does the household own its own toilet?	0.0173 (0.0260)
Mother completed at least prim. education	0.0600** (0.0276)
Raven 2pl IRT score	-0.0152 (0.0134)
District: Salipur	0.00896 (0.0334)
District: Soro	0.0445 (0.0348)
Constant	0.151*** (0.0511)
Observations	961
Adjusted R^2	0.020

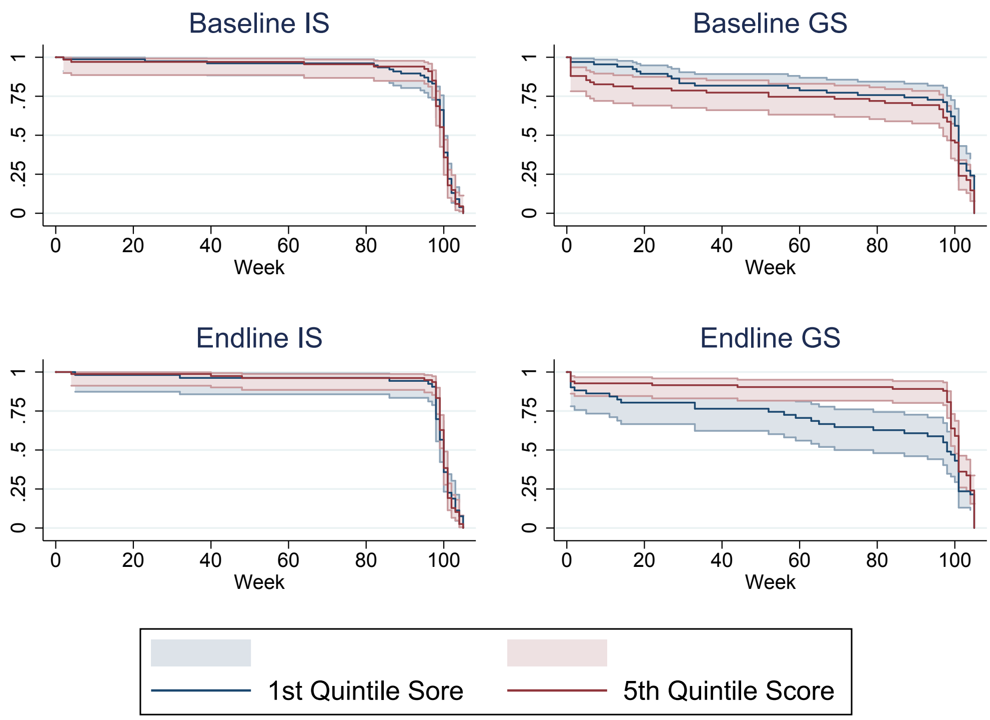
As an example, the survival function for each treatment arm broken down by whether or not the mother had at least a primary education is shown in figure 2. In the IS arm we see that a gap slowly grows over time. Contrast this with the GS, where the initial drop (whether a child has ever turned up) is 3 times as large for mothers who have a primary education, and grows as time goes on.

Figure 2: Survival Function By Treatment Arm and maternal education (failure = last session attended)



Another concern might be that non-compliance probability is affected by baseline cognitive score - for example if we only managed to treat the smarter children. A related question is what difference is there in attendance between the best performing kids at endline and the worst. Going from left to right, figure 3 shows the survival functions for the highest and lowest quintile for the ASQ problem solving subscale for the IS and then GS treatment arms. The bottom row shows the same but with quintiles calculated from the bayley cognitive subscale at endline. Here we see no discernible differences in attendance rate predicted by the ASQ score for both the IS and GS treatment arms. The same is true for the IS arm and the endline Bayley Cog quintile, but a significant gap opens up for the GS arm. This suggests that the intervention is working (people who turn up more get more out of the intervention), but is not causal evidence of the link between attendance and outcomes.

Figure 3: Cognition and Non Compliance probability: Baseline and Endline



3.2.2 Sessions attended

Beyond dropouts another dimension of program compliance is the number of sessions attended. This is simply the number of group sessions or home visits a child attended over the course of the trial, regardless of dropout status. Table 7 shows the number of sessions attended by treatment arm. This table again highlights the stark difference between the group sessions and individual sessions. 10.7% of GS children attended 80-89 sessions, the around the practical maximum. This is less than a third of the kids who managed to achieve this level of attendance in the IS arm (38.4%). Again the NE arm is displaying some odd characteristics in terms of attendance which we are still looking into.

Table 7: Sessions Attended: All target children

Sessions Attended	NE		GS+NE		IS+NE	
	Count	%	Count	%	Count	%
0	15	4.2%	37	9.9%	21	5.6%
1-9	5	1.4%	39	10.5%	9	2.4%
10-19	6	1.7%	24	6.4%	2	0.5%
20-29	6	1.7%	20	5.4%	7	1.9%
30-39	1	0.3%	18	4.8%	4	1.1%
40-49	7	2.0%	27	7.2%	7	1.9%
50-59	11	3.1%	44	11.8%	14	3.7%
60-69	19	5.3%	61	16.4%	50	13.3%
70-79	42	11.8%	63	16.9%	116	30.9%
80-89	104	29.1%	40	10.7%	144	38.4%
90-99	118	33.1%	0	0.0%	1	0.3%
100+	23	6.4%	0	0.0%	0	0.0%
Total Children	357		373		375	

The above table is not taking into account "supply" non-compliance, where a session was missed due to home vistor or group facilitator illness, training or a local holiday. An accurate count of home visitor and group session absence due to illness or training is something we have been trying to obtain for some time, but with the new intervention starting soon and Pratham being quite busy it has been quite difficult. My impression is that there should be roughly 5 missed sessions due to holidays and training in each treatment arm per year, so 47 sessions a year should have taken place (94 overall).

3.3 Spells of Attendance

As we have already seen, there is significant heterogeneity in terms of overall attendance, both across and within treatment arms. Beyond the sum of sessions attended another dimension of this is in terms of "spells" of attendance, where a spell of attendance is defined as the number of consecutive attendances without a break. The figure below illustrates how this looks in the data showing a theoretical spell of duration 4 (where 0 represents a missed and 1 an attended session).

Example Spell of Duration 4

$$\textit{Child } x \quad \cdots \quad 1 \quad 0 \quad \overline{1 \quad 1 \quad 1 \quad 1} \quad 0 \quad 1 \quad \cdots$$

There is a mechanical relationship between attendance, spells and their duration. Letting A denote attendance, s spells and d duration, with i and j be child and spell respectively. The attendance of child i is then given as

$$A_i = \sum_j s_{ij} d_{ij}$$

Why might spells matter? One argument is that there might be a difference between a child with a series of long spells and one who turns up every second week even if they have the same attendance rate. Another might be that spell duration could be a good predictor of subsequent drop out, and could highlight which children we should be worried about dropping out in the new trial. The following section looks at heterogeneity in the s and d across treatment arms and attendance quintiles.

Figure 4 plots the relationship between overall attendance and the number of separate spells a child has, along with a lowess plot. Here we see a clearly non-linear relationship between spells and attendance. With very few spells it is likely the child simply wasn't attending sessions. However those with a large number of spells have a lower overall attendance rate. Also clear from the figure is the higher number of spells in the GS arm, and the larger variance of spell number between children of the same attendance rate.

Figure 4: No. of Spells and overall attendance

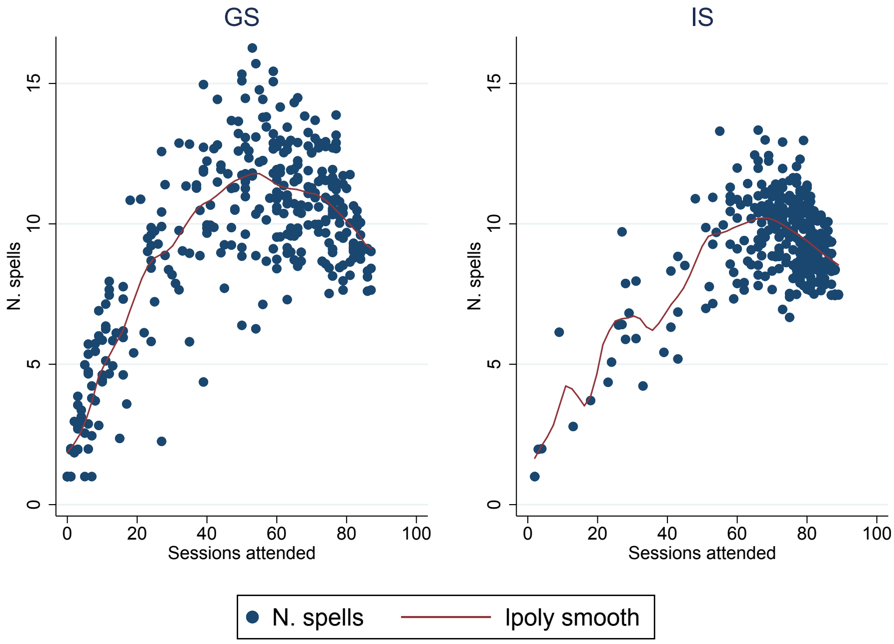
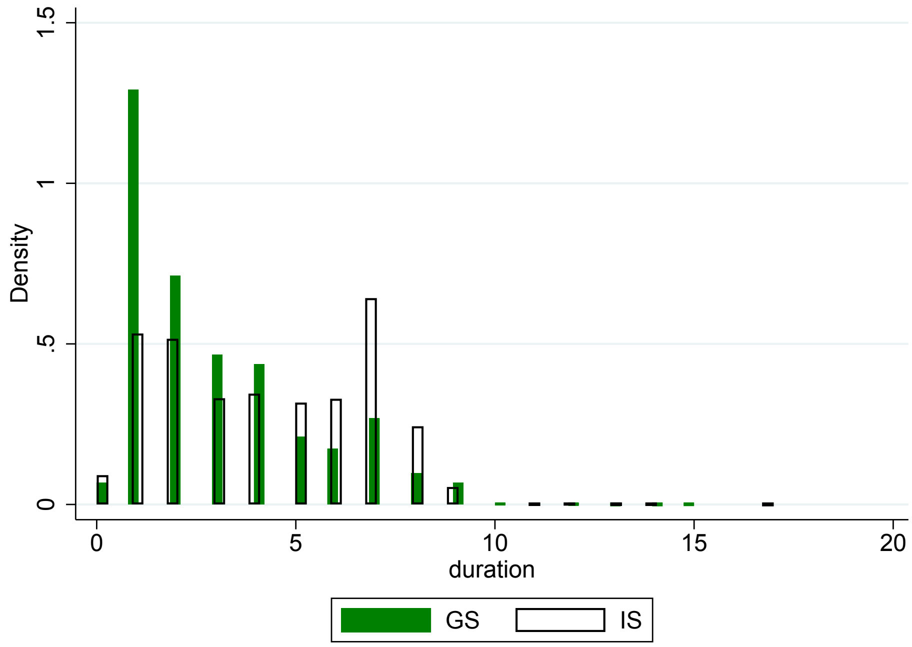


Figure 5 shows the distribution of durations of those spells. Here we can see that the distributions of spell durations for the GS arm is significantly more skewed to the left than for the IS arm, suggesting a pattern of "off and one" attendance for the GS arm not present in the IS arm.

Figure 5: Spell duration across treatment arm



3.4 Tracking Sheets: Detailed Data

In addition to the monitoring data we have more detailed tracking sheet data for part of the trial. This data is useful as it gives (in the most cases) a clear reason why a session did not occur. This data is available at the start and from roughly week 64 onwards, and as figure 6 shows, follows the same broad pattern of attendance we see in the intervention overall.

Figure 6: Box Plot of Attendance by treatment arm

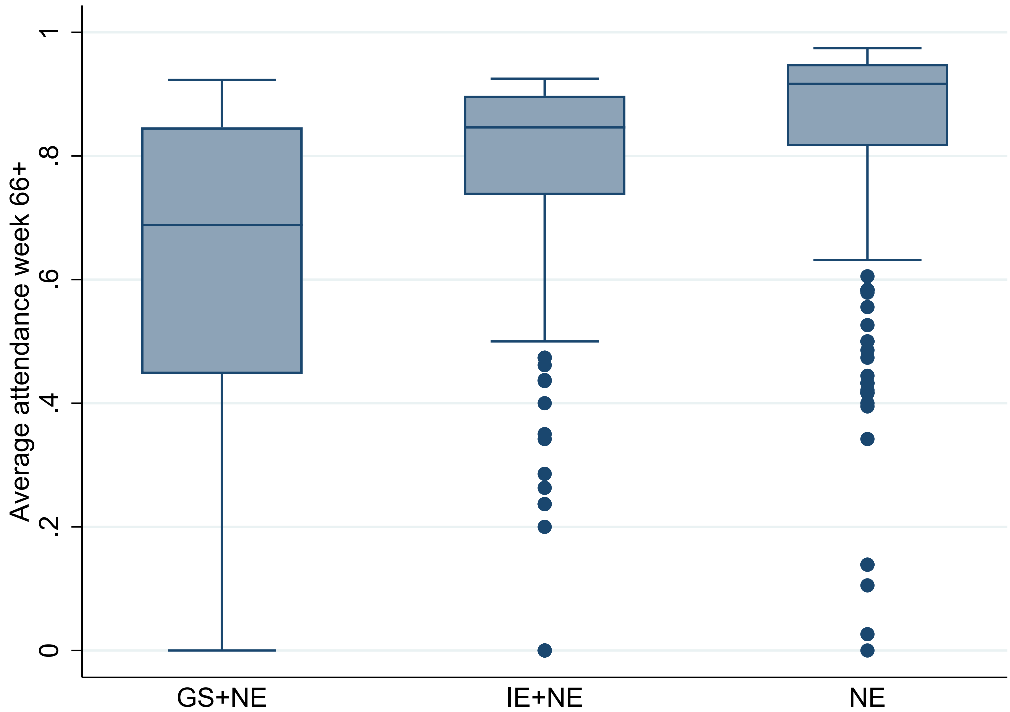


Table 8 shows the most common reasons for absence by treatment arm[2].

Table 8: Reasons for absence by treatment arm

Reasons for absence (%)	GS+NE	IS+NE	NE
Child was sleeping	1.5	0.0	0.0
Other	12.8	21.8	18.2
Migration	11.1	6.4	18.0
Home visitor ill, on leave or training	7.3	19.4	1.9
Mother Pregnant	1.2	0.7	0.5
Mother or child was sick	9.2	6.3	5.2
Mother was busy	43.5	5.6	15.9
Went to uncles house/visiting relatives	10.6	37.5	39.9
Domestic issue (unspecified)	2.8	2.2	0.4

[2]This was created by classifying common responses which were typed into the tracking sheets as the trial progressed. There is a lot of variation in terms of spellings and the quality of the responses which drives the large proportion in the "Other" category

Table 9 shows how the listed reason varies by attendance quintile. This again highlights the stark difference between the IS and GS arms - in all but the lowest attendance quintile for the GS arm, either the most or second most common reason for absence was that the mother was busy. This reason only shows up once for the IS arm. All of this suggests that (assuming that mothers in each arm were in reality equally busy) that there was a lower cost for mothers in the GS arm to not attend the weekly session, either due to the travel required or the knowledge that they were less likely to be chased up by the group facilitator. This is not true of the individual sessions, where home visitors would turn up to the household to conduct the session.

Table 9: Reasons for absence by treatment arm and attendance quintile

		Treatment Arm		
Quintile of Attendance		IS	GS	NE
1	*Most Common*	Went to uncles house	Other	Went to uncles house
	2nd Most Common	Empty Entry	Mother was busy	Migration
2	*Most Common*	Went to uncles house	Mother was busy	Went to uncles house
	2nd Most Common	Home visitor ill, on leave or training	Went to uncles house	Other
3	*Most Common*	Went to uncles house	Mother was busy	Went to uncles house
	2nd Most Common	Mother or child was sick	Other	Other
4	*Most Common*	Went to uncles house	Went to uncles house	Went to uncles house
	2nd Most Common	Mother was busy	Mother was busy	Migration
5	*Most Common*	Empty Entry	Empty Entry	Empty Entry
	2nd Most Common	n/a	Other	n/a

4 Cognitive Outcomes

The following section compares the midline and endline estimates of our treatment effects for each subscale of the Bayley using the Z score calculation method discussed in the midline report. Reserve target children have now been dropped from this analysis, and a series of "core controls" have been included to improve precision. These are a dummy for gender, mother/primary care giver having at least a primary education, their Raven Z score, and the baseline Z scores for each subscale of the ASQ. The results for this estimation are shown in table 10. In square brackets for each treatment arm is the P-value adjusted for Romano Wolf standard errors, and the stars next to the coefficient reflect that statistical significant level. Tests for differences between midline and endline for the IS and GS treatment arms are given in table 11, and tests for differences between these two arms arms at endline are given in table 12. These show that the fall in the treatment effect from midline to endline doesn't seem to be statistically significant, and that on the whole there is no difference between IS and GS treatment arms.

Table 10: Bayley Z score (outliers removed)

	Cognitive		Receptive Lan.		Expressive Lan.		Fine Motor		Gross Motor	
	Endline	Midline	Endline	Midline	Endline	Midline	Endline	Midline	Endline	Midline
NE	0.00582	0.0127	0.105	0.129	0.150	0.134	-0.0207	0.223**	0.00890	-0.0439
	(0.0926)	(0.0799)	(0.0790)	(0.0784)	(0.0794)	(0.0812)	(0.0842)	(0.0902)	(0.0777)	(0.0850)
	[1.000]	[0.984]	[0.524]	[0.302]	[0.112]	[0.250]	[0.802]	[0.014]	[1.000]	[0.984]
GS+NE	0.315***	0.417***	0.270***	0.320***	0.273***	0.310***	0.0820	0.188*	0.178*	0.0138
	(0.0887)	(0.0793)	(0.0807)	(0.0848)	(0.0860)	(0.0875)	(0.0930)	(0.0916)	(0.0819)	(0.0799)
	[0.002]	[0.000]	[0.002]	[0.006]	[0.006]	[0.004]	[0.374]	[0.076]	[0.072]	[0.868]
IS+NE	0.297***	0.450***	0.217**	0.147	0.286***	0.273**	-0.0827	0.202	0.0831	-0.0709
	(0.0888)	(0.0866)	(0.0866)	(0.0883)	(0.0893)	(0.0942)	(0.0909)	(0.0912)	(0.0789)	(0.0772)
	[0.008]	[0.000]	[0.046]	[0.250]	[0.010]	[0.018]	[0.380]	[0.110]	[0.417]	[0.364]
ASQ Communication	0.0634*	-0.00681	0.0818***	0.0380	0.107***	0.137***	0.0670*	0.0405	-0.0126	-0.0336
	(0.0376)	(0.0334)	(0.0314)	(0.0312)	(0.0334)	(0.0336)	(0.0350)	(0.0373)	(0.0311)	(0.0330)
ASQ Gross Motor	0.00203	0.0559*	-0.00355	0.00502	-0.0261	0.0173	0.0675**	0.0498	0.160***	0.209***
	(0.0328)	(0.0328)	(0.0317)	(0.0320)	(0.0312)	(0.0309)	(0.0323)	(0.0318)	(0.0313)	(0.0321)
ASQ Fine Motor	0.0200	0.0736*	-0.0112	0.0293	0.0434	0.0190	0.0384	0.0762*	0.0504	0.0198
	(0.0413)	(0.0416)	(0.0358)	(0.0360)	(0.0375)	(0.0398)	(0.0383)	(0.0400)	(0.0345)	(0.0381)
ASQ Problem Solving	0.0668**	0.0826**	0.0666**	0.0507	-0.00381	0.0197	0.136***	0.0436	0.000665	-0.0246
	(0.0334)	(0.0320)	(0.0312)	(0.0323)	(0.0351)	(0.0342)	(0.0389)	(0.0333)	(0.0326)	(0.0327)
ASQ Personal Social	0.00341	-0.0433	0.0648**	0.00282	-0.00723	0.0239	-0.0705*	-0.0352	0.0295	-0.00751
	(0.0343)	(0.0327)	(0.0309)	(0.0346)	(0.0373)	(0.0361)	(0.0358)	(0.0354)	(0.0358)	(0.0331)
PCG Raven IRT score	0.102***	0.136***	0.151***	0.169***	0.164***	0.120***	0.138***	0.125***	0.0539*	0.0581*
	(0.0372)	(0.0338)	(0.0348)	(0.0313)	(0.0328)	(0.0319)	(0.0377)	(0.0311)	(0.0298)	(0.0315)
Male	0.0126	0.0377	-0.0727	-0.181***	0.00980	-0.162***	-0.0917	-0.229***	0.0853	0.157***
	(0.0589)	(0.0566)	(0.0516)	(0.0475)	(0.0531)	(0.0564)	(0.0609)	(0.0528)	(0.0517)	(0.0537)
PCG primary ed.	0.289***	0.374***	0.295***	0.400***	0.393***	0.396***	0.238***	0.218***	0.185***	0.263***
	(0.0596)	(0.0589)	(0.0618)	(0.0559)	(0.0595)	(0.0572)	(0.0667)	(0.0643)	(0.0537)	(0.0528)
Constant	-0.186**	-0.242***	-0.128*	-0.141*	-0.244***	-0.132*	-0.0866	-0.00679	-0.152**	-0.243***
	(0.0853)	(0.0742)	(0.0772)	(0.0756)	(0.0844)	(0.0744)	(0.0806)	(0.0850)	(0.0760)	(0.0829)
Observations	1252	1258	1263	1260	1244	1260	1265	1261	1266	1253
Adjusted R^2	0.067	0.120	0.096	0.118	0.096	0.110	0.073	0.070	0.055	0.065

In square brackets is the Romano Wolf P-value for that subscale and treatment arm. Brackets are standard errors after clustering for village only. Cg stands for Cognitive subscale, RC receptive communication subscale, EC Expressive Communication, FM fine motor, GM gross motor. Z scores calculated adjusting for tester fixed effects. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 11: Z score (outliers removed): Test for differences between Endline and Midline

	Cognitive		Receptive Lan.		Expressive Lan.		Fine Motor		Gross Motor	
	GS+NE	IS+NE	GS+NE	IS+NE	GS+NE	IS+NE	GS+NE	IS+NE	GS+NE	IS+NE
Difference	-0.10	-0.15	-0.05	0.07	-0.04	0.01	-0.11	-0.28**	0.16*	0.15
P Value	0.31	0.15	0.59	0.49	0.65	0.88	0.29	0.01	0.09	0.13

Table 12: Bayley Z score: Test for differences between GS and IS (Endline)

	Cognitive	Receptive Lan.	Expressive Lan.	Fine Motor	Gross Motor
Difference	0.02	0.05	-0.01	0.16	0.09
P value	0.83	0.56	0.88	0.10	0.23

4.1 Reduced treatment effect

Putting the lack of statistical significance aside, there a few other reasons why the treatment effect at midline is larger than that at endline in some domains. A list of potential reasons are given below. The following section will go through each in turn and examine the probability of this being the case in our sample.

Potential Reasons for Fade Out

1. Standardization method (age conditional sd. of control children increased)
2. Children at top of distribution reaching end of test
3. Test less appropriate for older children (either in general or due to Basal mechanics)
4. Control improving faster than treatment

4.1.1 Alternative Standardizations

The current method of standardization is incredibly flexible, allowing for non-parametric adjustment for the effect of age on score and its standard deviation. The method is as follows. First, for each scale, tester/interviewer effects were netted out by running a regression of the raw scores on tester/interviewer dummies. Next, the residuals of these regressions were non-parametrically estimated to obtain the age-conditional mean and age-conditional standard deviation (SD) of each measure in the control group using locally weighted regression (lowess command on Stata). The Z scores of the entire sample are then calculated using by subtracting the age-conditional mean of the control group from each individual and dividing by the age-specific SD.

The following section looks at several methods of standardization and the size of the treatment effect at midline and endline to determine whether or not this method is driving our reduction in treatment effect. All of the following use the same set of “core controls” as above.

Raw Score

The simplest method would be just to not standardize at all. Table 13 looks at the effect on raw score by subscale at midline and endline. There is no adjustment for age in these regressions, but this aside the lack of a fade out in treatment effect in any of our subscales from midline to endline is very clear. Indeed for some subscales there is now a significant effect on score where at midline none existed.

Colombia Standardization

The first real alternative standardization we use is to follow the method used in the structural Columbia paper. This uses a Kernel-weighted local polynomial to adjust for age after removing tester fixed effects. The key difference between this and the original standardization method is that the age conditional means and variance is calculated using the whole sample and not just the control group. Once this Z score has been generated we then rescale it to have a mean of 0 and standard deviation of 1 for the control group to make the interpretation comparable to our main results. The results of this standardization are shown in table 14. These results paint a very similar picture to those in the main results set.

US Scaled Score

Another option is to use the age standardization given with the bayley test, and calculate the effect of the treatment on the "scaled score". The obvious flaw here is that this was designed for US kids and is generally pretty inflexible. Nevertheless, the treatment effects in this case are shown in table 15, but again these show reductions in treatment effect between midline and endline.

Latent Factor

The final processing method we use is to estimate the latent factor using a one parameter IRT[3] for each subscale at midline and endline separately and then compare the effect of our treatment. The coefficients here are not comparable to those above as it is not measured in standard deviations of the control group. The results of this method are shown in table 16. This interestingly does not show any fade out in treatment effect between midline and endline, and indeed there is an increase in the effect for some subscales.

[3] One parameter was chosen as this is the only one we could get to converge for all domains.

Table 13: Bayley Raw Score

	(1) Cg Endline	(2) Cg Midline	(3) Rc Endline	(4) Rc Midline	(5) Ec Endline	(6) Ec Midline	(7) Fm Endline	(8) Fm Midline	(9) Gm Endline	(10) Gm Midline
NE	0.125	-0.516	0.480	0.0221	0.884*	0.296	0.302	0.181	0.166	-0.537*
	(0.353)	(0.436)	(0.352)	(0.470)	(0.520)	(0.545)	(0.399)	(0.312)	(0.241)	(0.321)
GS+NE	0.944***	0.927*	0.863***	0.616	1.581***	1.334**	0.230	0.0802	0.341	-0.402
	(0.350)	(0.476)	(0.327)	(0.501)	(0.510)	(0.547)	(0.425)	(0.373)	(0.269)	(0.350)
IS+NE	1.034***	1.206***	0.634*	0.169	1.486***	0.922	-0.194	0.177	0.226	-0.634*
	(0.335)	(0.457)	(0.375)	(0.509)	(0.535)	(0.562)	(0.418)	(0.336)	(0.248)	(0.323)
ASQ Communication	0.412***	0.299*	0.580***	0.564***	0.842***	1.226***	0.495***	0.318***	0.121	0.0747
	(0.139)	(0.161)	(0.120)	(0.151)	(0.175)	(0.193)	(0.145)	(0.121)	(0.0969)	(0.112)
ASQ Gross Motor	0.0642	0.159	-0.00863	-0.0207	0.00962	0.103	0.238*	0.115	0.416***	0.605***
	(0.132)	(0.166)	(0.123)	(0.148)	(0.184)	(0.182)	(0.142)	(0.113)	(0.107)	(0.117)
ASQ Fine Motor	0.145	0.204	0.156	0.0898	0.609**	0.0776	0.352**	0.150	0.259**	0.154
	(0.172)	(0.188)	(0.161)	(0.169)	(0.237)	(0.245)	(0.145)	(0.133)	(0.112)	(0.145)
ASQ Problem Solving	0.100	0.242	-0.0173	0.00708	-0.262	-0.195	0.203	0.0697	-0.105	-0.0724
	(0.148)	(0.166)	(0.127)	(0.166)	(0.191)	(0.204)	(0.163)	(0.107)	(0.106)	(0.120)
ASQ Personal Social	0.0445	-0.241	0.232*	0.0575	0.0412	0.0295	-0.209	-0.104	0.0655	-0.0815
	(0.122)	(0.156)	(0.121)	(0.162)	(0.211)	(0.208)	(0.147)	(0.110)	(0.101)	(0.111)
PCG Raven IRT score	0.105	0.359**	0.340***	0.460***	0.573***	0.425**	0.280*	0.262**	0.0639	0.0905
	(0.125)	(0.174)	(0.112)	(0.159)	(0.150)	(0.193)	(0.149)	(0.114)	(0.0894)	(0.111)
Male	-0.123	-0.0885	-0.467**	-0.694***	-0.493*	-1.099***	-0.386	-0.671***	0.240	0.266
	(0.193)	(0.296)	(0.197)	(0.255)	(0.290)	(0.350)	(0.256)	(0.188)	(0.163)	(0.198)
PCG primary ed.	0.914***	1.463***	0.819***	1.279***	1.586***	1.928***	0.587**	0.501**	0.524***	0.942***
	(0.228)	(0.293)	(0.232)	(0.270)	(0.334)	(0.351)	(0.275)	(0.215)	(0.166)	(0.188)
Constant	63.73***	55.64***	32.05***	26.19***	34.60***	26.29***	43.80***	36.61***	60.88***	55.88***
	(0.338)	(0.369)	(0.326)	(0.394)	(0.500)	(0.439)	(0.368)	(0.307)	(0.240)	(0.310)
Observations	1275	1263	1275	1266	1275	1266	1275	1265	1275	1262
Adjusted R^2	0.063	0.058	0.082	0.059	0.089	0.092	0.047	0.037	0.051	0.057

Table 14: Columbia Z score

	(1) Cg Endline	(2) Cg Midline	(3) Rc Endline	(4) Rc Midline	(5) Ec Endline	(6) Ec Midline	(7) Fm Endline	(8) Fm Midline	(9) Gm Endline	(10) Gm Midline
NE	0.0325	0.0183	0.0641	0.129	0.109	0.126	-0.0225	0.221**	-0.00129	-0.0554
	(0.0837)	(0.0826)	(0.0725)	(0.0844)	(0.0676)	(0.0802)	(0.0766)	(0.0916)	(0.0744)	(0.0882)
GS+NE	0.296***	0.377***	0.243***	0.304***	0.278***	0.315***	0.0857	0.169*	0.144*	0.00388
	(0.0817)	(0.0913)	(0.0744)	(0.0923)	(0.0680)	(0.0878)	(0.0868)	(0.0984)	(0.0792)	(0.0823)
IS+NE	0.311***	0.418***	0.188**	0.158*	0.261***	0.252***	-0.0614	0.209**	0.0675	-0.0917
	(0.0773)	(0.0961)	(0.0866)	(0.0952)	(0.0760)	(0.0921)	(0.0859)	(0.104)	(0.0768)	(0.0858)
ASQ Communication	0.0443	-0.0154	0.0856***	0.0242	0.0911***	0.128***	0.0497	0.0303	-0.00823	-0.0269
	(0.0350)	(0.0341)	(0.0311)	(0.0315)	(0.0279)	(0.0326)	(0.0339)	(0.0370)	(0.0310)	(0.0334)
ASQ Gross Motor	0.0253	0.0515	0.0252	-0.0104	0.0162	0.0278	0.0841**	0.0352	0.156***	0.204***
	(0.0348)	(0.0317)	(0.0329)	(0.0311)	(0.0302)	(0.0317)	(0.0330)	(0.0337)	(0.0345)	(0.0306)
ASQ Fine Motor	0.0313	0.0766*	0.0200	0.0674*	0.0823**	0.0253	0.0658*	0.0891**	0.0886**	0.0748*
	(0.0458)	(0.0414)	(0.0417)	(0.0389)	(0.0379)	(0.0420)	(0.0366)	(0.0408)	(0.0389)	(0.0435)
ASQ Problem Solving	0.0894**	0.0653*	0.0691**	0.0530	0.0144	0.00524	0.139***	0.0463	0.00708	-0.0251
	(0.0376)	(0.0342)	(0.0332)	(0.0327)	(0.0303)	(0.0337)	(0.0403)	(0.0335)	(0.0346)	(0.0336)
ASQ Personal Social	0.0233	-0.0339	0.0732**	0.0175	0.00318	0.0311	-0.0517	-0.0148	0.0291	-0.0178
	(0.0330)	(0.0338)	(0.0308)	(0.0353)	(0.0335)	(0.0363)	(0.0343)	(0.0357)	(0.0346)	(0.0328)
PCG Raven IRT score	0.0513	0.117***	0.134***	0.158***	0.120***	0.116***	0.120***	0.130***	0.0472	0.0482
	(0.0326)	(0.0350)	(0.0316)	(0.0320)	(0.0255)	(0.0310)	(0.0349)	(0.0333)	(0.0292)	(0.0315)
Male	-0.00397	0.0238	-0.0860*	-0.168***	-0.0528	-0.176***	-0.0715	-0.203***	0.100*	0.138**
	(0.0529)	(0.0578)	(0.0508)	(0.0490)	(0.0470)	(0.0584)	(0.0585)	(0.0512)	(0.0524)	(0.0540)
PCG primary ed.	0.281***	0.404***	0.286***	0.415***	0.305***	0.420***	0.217***	0.215***	0.201***	0.306***
	(0.0584)	(0.0616)	(0.0598)	(0.0572)	(0.0539)	(0.0588)	(0.0633)	(0.0673)	(0.0559)	(0.0513)
Constant	-0.166**	-0.254***	-0.108	-0.155**	-0.147**	-0.144**	-0.0778	-0.00816	-0.164**	-0.248***
	(0.0809)	(0.0788)	(0.0698)	(0.0779)	(0.0712)	(0.0706)	(0.0770)	(0.0867)	(0.0735)	(0.0816)
Observations	1275	1250	1275	1253	1275	1253	1275	1252	1275	1249
Adjusted R^2	0.081	0.110	0.107	0.122	0.103	0.115	0.083	0.070	0.069	0.079

Table 15: US scaled Score

	(1) Cg Endline	(2) Cg Midline	(3) Rc Endline	(4) Rc Midline	(5) Ec Endline	(6) Ec Midline	(7) Fm Endline	(8) Fm Midline	(9) Gm Endline	(10) Gm Midline
NE	0.0563	-0.257*	0.182	0.0150	0.290*	0.128	0.153	0.0836	0.0833	-0.229
	(0.103)	(0.142)	(0.128)	(0.221)	(0.173)	(0.212)	(0.151)	(0.182)	(0.116)	(0.153)
GS+NE	0.341***	0.458***	0.396***	0.440*	0.645***	0.691***	0.213	0.160	0.219*	-0.0874
	(0.106)	(0.156)	(0.125)	(0.246)	(0.178)	(0.223)	(0.172)	(0.219)	(0.125)	(0.161)
IS+NE	0.395***	0.493***	0.272**	0.164	0.553***	0.506**	-0.00473	0.177	0.152	-0.245
	(0.0993)	(0.148)	(0.134)	(0.259)	(0.184)	(0.237)	(0.170)	(0.203)	(0.126)	(0.154)
ASQ Communication	0.0821**	-0.00423	0.160***	0.149**	0.234***	0.470***	0.155***	0.107	0.0101	-0.0685
	(0.0389)	(0.0558)	(0.0442)	(0.0708)	(0.0653)	(0.0796)	(0.0583)	(0.0713)	(0.0486)	(0.0572)
ASQ Gross Motor	0.0335	0.0885	-0.0129	0.00852	-0.00731	0.0338	0.113**	0.0930	0.221***	0.347***
	(0.0373)	(0.0544)	(0.0461)	(0.0696)	(0.0623)	(0.0725)	(0.0567)	(0.0619)	(0.0490)	(0.0534)
ASQ Fine Motor	0.0300	0.0935	0.0565	0.0323	0.229***	0.0314	0.158**	0.0785	0.147**	0.0687
	(0.0482)	(0.0628)	(0.0585)	(0.0786)	(0.0786)	(0.0977)	(0.0620)	(0.0780)	(0.0595)	(0.0697)
ASQ Problem Solving	0.0387	0.114**	0.00526	0.0739	-0.0894	-0.0605	0.117*	0.0931	-0.0312	-0.0122
	(0.0421)	(0.0524)	(0.0452)	(0.0742)	(0.0656)	(0.0801)	(0.0657)	(0.0656)	(0.0502)	(0.0574)
ASQ Personal Social	0.0366	-0.0364	0.112**	0.0512	0.0602	0.0383	-0.0487	-0.0291	0.0548	0.000767
	(0.0347)	(0.0567)	(0.0465)	(0.0756)	(0.0781)	(0.0847)	(0.0616)	(0.0648)	(0.0521)	(0.0532)
PCG Raven IRT score	0.0797**	0.211***	0.166***	0.282***	0.258***	0.229***	0.164***	0.230***	0.0459	0.0670
	(0.0347)	(0.0564)	(0.0414)	(0.0691)	(0.0596)	(0.0755)	(0.0610)	(0.0601)	(0.0442)	(0.0492)
Male	0.0110	0.0872	-0.149**	-0.270**	-0.135	-0.445***	-0.146	-0.354***	0.168**	0.213**
	(0.0591)	(0.0911)	(0.0705)	(0.110)	(0.107)	(0.134)	(0.100)	(0.0983)	(0.0788)	(0.0910)
PCG primary ed.	0.297***	0.541***	0.314***	0.614***	0.640***	0.861***	0.317***	0.312***	0.341***	0.457***
	(0.0669)	(0.0943)	(0.0859)	(0.123)	(0.125)	(0.136)	(0.111)	(0.117)	(0.0848)	(0.0941)
Constant	6.139***	6.096***	8.685***	9.073***	8.151***	7.754***	7.751***	7.470***	8.516***	8.857***
	(0.0991)	(0.134)	(0.124)	(0.196)	(0.180)	(0.183)	(0.153)	(0.180)	(0.115)	(0.156)
Observations	1275	1263	1275	1266	1275	1266	1275	1265	1275	1262
Adjusted R^2	0.078	0.101	0.085	0.071	0.098	0.108	0.064	0.050	0.065	0.072

Table 16: Latent Factor

	(1) Cg Endline	(2) Cg Midline	(3) Rc Endline	(4) Rc Midline	(5) Ec Endline	(6) Ec Midline	(7) Fm Endline	(8) Fm Midline	(9) Gm Endline	(10) Gm Midline
NE	0.0145	-0.0796	0.123	0.00558	0.133	0.0364	0.0546	0.0500	0.0541	-0.0928
	(0.0935)	(0.0734)	(0.0877)	(0.0862)	(0.0891)	(0.0782)	(0.0837)	(0.0708)	(0.0756)	(0.0766)
GS+NE	0.272***	0.173**	0.215***	0.105	0.240***	0.184**	0.0707	0.0153	0.105	-0.0604
	(0.0928)	(0.0811)	(0.0796)	(0.0906)	(0.0877)	(0.0766)	(0.0902)	(0.0846)	(0.0854)	(0.0833)
IS+NE	0.277***	0.223***	0.155*	0.0272	0.237**	0.125	-0.0185	0.0393	0.0690	-0.110
	(0.0892)	(0.0782)	(0.0925)	(0.0932)	(0.0922)	(0.0804)	(0.0869)	(0.0781)	(0.0769)	(0.0772)
ASQ Communication	0.117***	0.0487*	0.140***	0.103***	0.155***	0.167***	0.104***	0.0767***	0.0373	0.0154
	(0.0353)	(0.0272)	(0.0297)	(0.0275)	(0.0304)	(0.0274)	(0.0315)	(0.0273)	(0.0300)	(0.0261)
ASQ Gross Motor	0.0145	0.0268	0.000105	-0.00520	-0.00215	0.0131	0.0532*	0.0263	0.126***	0.146***
	(0.0328)	(0.0279)	(0.0308)	(0.0267)	(0.0321)	(0.0263)	(0.0298)	(0.0253)	(0.0324)	(0.0295)
ASQ Fine Motor	0.0378	0.0437	0.0348	0.0161	0.0991**	0.0194	0.0673**	0.0350	0.0789**	0.0441
	(0.0400)	(0.0324)	(0.0406)	(0.0311)	(0.0391)	(0.0360)	(0.0311)	(0.0309)	(0.0332)	(0.0359)
ASQ Problem Solving	0.0165	0.0386	0.00361	-0.00126	-0.0453	-0.0294	0.0480	0.0160	-0.0312	-0.0166
	(0.0353)	(0.0265)	(0.0313)	(0.0299)	(0.0329)	(0.0294)	(0.0344)	(0.0243)	(0.0323)	(0.0279)
ASQ Personal Social	0.00274	-0.0285	0.0591*	0.0139	0.000333	0.00862	-0.0422	-0.0245	0.0175	-0.00662
	(0.0311)	(0.0269)	(0.0302)	(0.0294)	(0.0362)	(0.0296)	(0.0318)	(0.0248)	(0.0307)	(0.0266)
PCG Raven IRT score	0.0389	0.0647**	0.0817***	0.0843***	0.0964***	0.0620**	0.0591*	0.0557**	0.0210	0.0260
	(0.0322)	(0.0293)	(0.0279)	(0.0284)	(0.0262)	(0.0269)	(0.0323)	(0.0254)	(0.0281)	(0.0249)
Male	-0.0186	-0.0148	-0.110**	-0.131***	-0.0683	-0.158***	-0.121**	-0.147***	0.0704	0.0767*
	(0.0507)	(0.0502)	(0.0489)	(0.0463)	(0.0496)	(0.0498)	(0.0543)	(0.0422)	(0.0500)	(0.0462)
PCG primary ed.	0.254***	0.230***	0.203***	0.234***	0.264***	0.267***	0.151**	0.115**	0.163***	0.214***
	(0.0575)	(0.0491)	(0.0574)	(0.0495)	(0.0567)	(0.0502)	(0.0595)	(0.0493)	(0.0518)	(0.0452)
Constant	-0.293***	-0.143**	-0.194**	-0.00745	-0.284***	-0.0655	-0.0646	0.0206	-0.198***	-0.0228
	(0.0888)	(0.0643)	(0.0805)	(0.0726)	(0.0851)	(0.0631)	(0.0790)	(0.0704)	(0.0749)	(0.0752)
Observations	1288	1265	1288	1265	1288	1265	1288	1265	1288	1265
Adjusted R^2	0.068	0.061	0.081	0.061	0.084	0.090	0.051	0.038	0.049	0.062

4.1.2 Reaching the end of the test

An alternative explanation is that children are now reaching the end of the test and are constrained from progressing any further. Table 17 summarizes the "ceilings" (the last question asked of the children) by subscale at midline and endline along with the final question possible, and table 18 shows the proportion of children reaching the end of the test at endline. For all subscales bar "receptive language" subscale almost all children finish the test without reaching the cap for both treatment and control.

Table 17: Summary of ceilings subscale, midline and endline

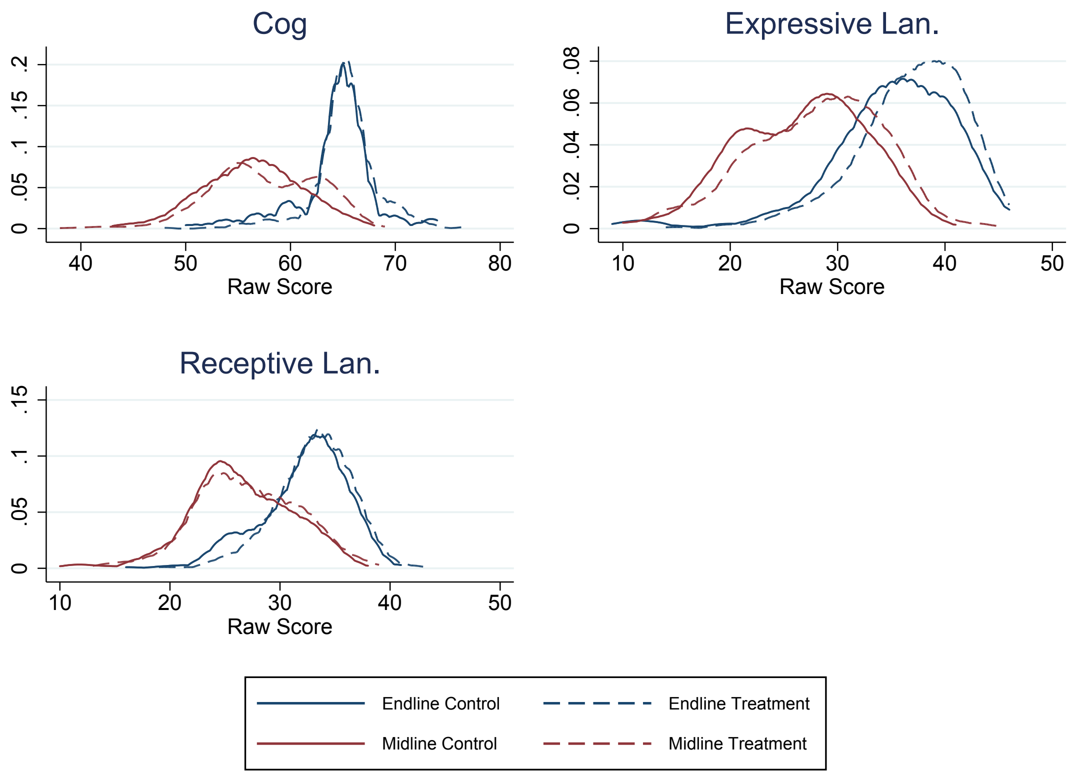
	p25	p50	p75	p90	Last Possible
Endline Cog	64	67	68	74	91
Midline Cog	64	67	69	72	91
Endline RL	33	37	40	42	49
Midline RL	30	35	39	42	49
Endline EC	36	44	46	47.5	48
Midline EC	29	36	40	42	48
Endline FM	45	47	54	57	66
Midline FM	43	44	48	51	66
Endline GM	62	63	65	67	72
Midline GM	61	62	67	68	72

Table 18: Proportion of children reaching end of test Endline

	Stimulation Treatment		*Control*	
	Proportion	Count	Proportion	Count
Cognitive	0.001	1	0.000	0
Receptive Com.	0.059	41	0.034	11
Expressive Com.	0.535	373	0.428	137
Fine Motor	0.003	2	0.003	1
Gross Motor	0.057	40	0.063	20

Another check on this would be to look at the distribution of raw scores at midline and endline to look for a bunch in the distribution. This is shown in 7, and no visible bunching is visible. Even for the expressive communication subscale, despite a large proportion of children reaching the end of the test there is no bunching at the maximum score.

Figure 7: Midline vs. Endline Raw Score Distributions



4.1.3 Control Children Catching Up

Another explanation would be that control children improved faster than those in the treatment group. Table 19 show the average gain in each subscale between midline and endline, with the standard deviation of each in brackets. Looking at this mean difference alone there does not appear to be an obvious difference between treatment and control children.

Table 19: Gains in raw score midline and endline

	(1) Control	(2) NE	(3) GS+NE	(4) IS+NE
Cog	7.85	8.41	7.70	7.55
	(4.59)	(4.46)	(4.65)	(4.45)
RC	5.69	6.22	5.91	6.12
	(4.20)	(3.69)	(4.19)	(4.24)
EC	8.40	9.08	8.57	8.99
	(5.48)	(5.36)	(5.27)	(5.08)
FM	7.34	7.62	7.50	7.16
	(3.92)	(3.70)	(3.77)	(3.98)
GM	4.89	5.45	5.51	5.73
	(3.51)	(3.14)	(2.95)	(3.29)

To examine this we plot the relationship between raw scores at midline and endline separately for treatment and control in figures 8 and 9. For all following graphs the solid red line is 45 degree line and the dashed grey line is the linear regression line. We have defined stimulation treatment as being part of either the IS or GS home visiting programs. The smaller slope of the regression line implies a "regression to the mean" type effect for each of the subscales in the bayley. Furthermore this effect seems to be more pronounced in the stimulation treatment groups than in the control, possibly because of the of the higher average score at midline.

Figure 8: Raw Score by Subscale: Midline vs. Endline, Control

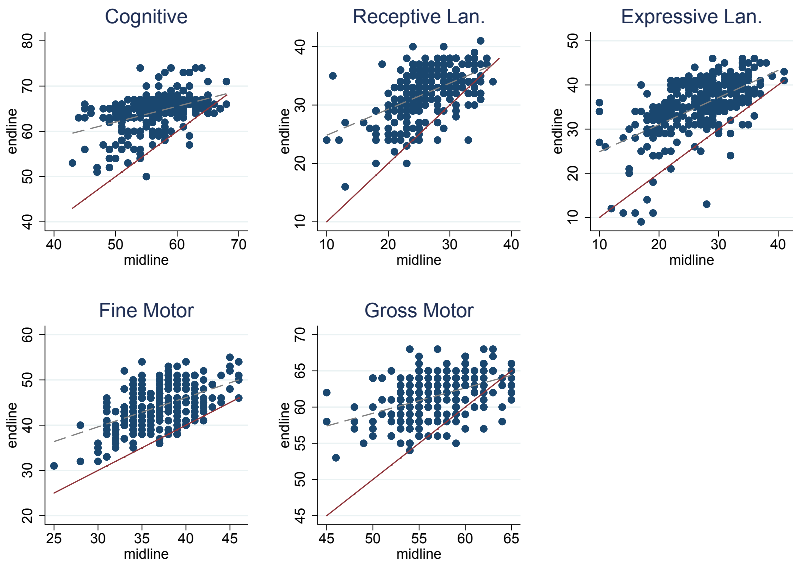
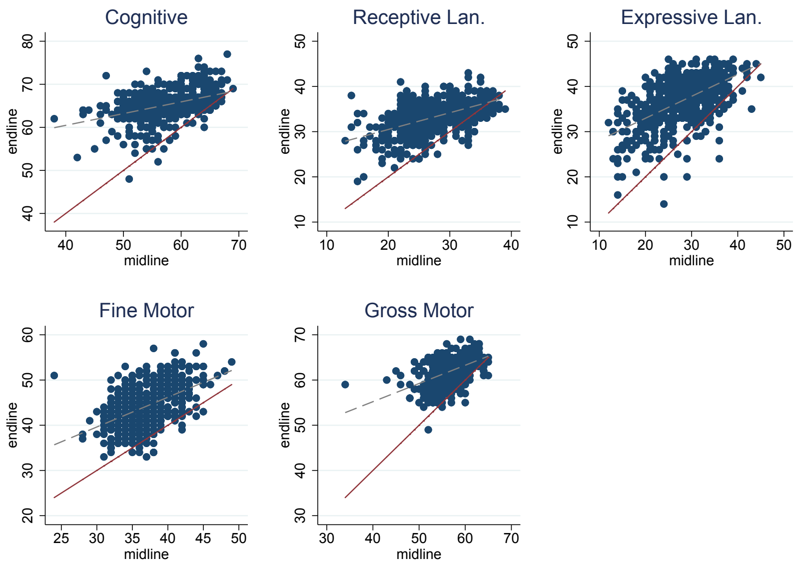


Figure 9: Raw Score by Subscale: Midline vs. Endline Stimulation Treatment



4.1.4 Changing patterns of investment

Related to this is the changing patterns of parental investment for control kids. Table 20 shows the activities part of the HOME module at midline, and table 21 shows the same for endline. The coefficients for control children are plotted in figures 10 and 11. While the treatment effects for each activity are broadly stable, it is noticeable that in a few key items (reading and drawing in particular) there is an increase in the proportion of control children who are partaking in these sorts of activities.

Table 20: Activities during the last three days (% of children), midline

		Difference from Control				
	Control Mean	NE	IS	GS	N	F-stat
Read/look at pictures	0.44	0.10**	0.19***	0.21***	1375	12.15***
	(0.50)	(0.04)	(0.04)	(0.04)		
Tell stories	0.46	-0.00	0.10**	0.08*	1375	2.74**
	(0.50)	(0.04)	(0.04)	(0.04)		
Sing	0.50	0.02	0.14***	0.09**	1375	4.70***
	(0.50)	(0.04)	(0.04)	(0.04)		
Go out	0.85	-0.03	0.02	0.02	1375	1.09
	(0.36)	(0.04)	(0.03)	(0.03)		
Play with toys	0.42	0.05	0.05	0.10**	1375	1.77
	(0.49)	(0.04)	(0.05)	(0.04)		
Draw, paint, write, play with paper/pencil	0.19	0.10***	0.14***	0.14***	1375	10.38***
	(0.39)	(0.03)	(0.03)	(0.03)		
Play to name objects, colours, count or say numbers	0.56	0.14***	0.14***	0.13***	1375	4.26***
	(0.50)	(0.04)	(0.05)	(0.04)		

Note: * p<0.1, ** p<0.05, *** p<0.01, standard errors in parentheses, clustered at the village level

Table 21: Activities during the last three days (% of children), endline

		Difference from Control				
	Control Mean	NE	IS	GS	N	F-stat
Read/look at pictures	0.59	0.14***	0.16***	0.11***	1386	5.31***
	(0.49)	(0.04)	(0.04)	(0.04)		
Tell stories	0.52	0.07	0.13***	0.08**	1386	3.70**
	(0.50)	(0.04)	(0.04)	(0.04)		
Sing	0.43	0.11***	0.16***	0.13***	1386	5.86***
	(0.50)	(0.04)	(0.04)	(0.04)		
Go out	0.84	0.04	0.00	-0.00	1386	0.89
	(0.37)	(0.03)	(0.03)	(0.03)		
Play with toys	0.27	0.05	0.16***	0.15***	1386	7.29***
	(0.44)	(0.04)	(0.04)	(0.04)		
Draw, paint, write, play with paper/pencil	0.39	0.17***	0.13***	0.15***	1386	6.82***
	(0.49)	(0.04)	(0.04)	(0.04)		
Play to name objects, colours, count or say numbers	0.64	0.15***	0.13***	0.13***	1386	5.99***
	(0.48)	(0.04)	(0.04)	(0.04)		

Note: * p<0.1, ** p<0.05, *** p<0.01, standard errors in parentheses, clustered at the village level

Figure 10: Extensive margin activities
Control midline and endline

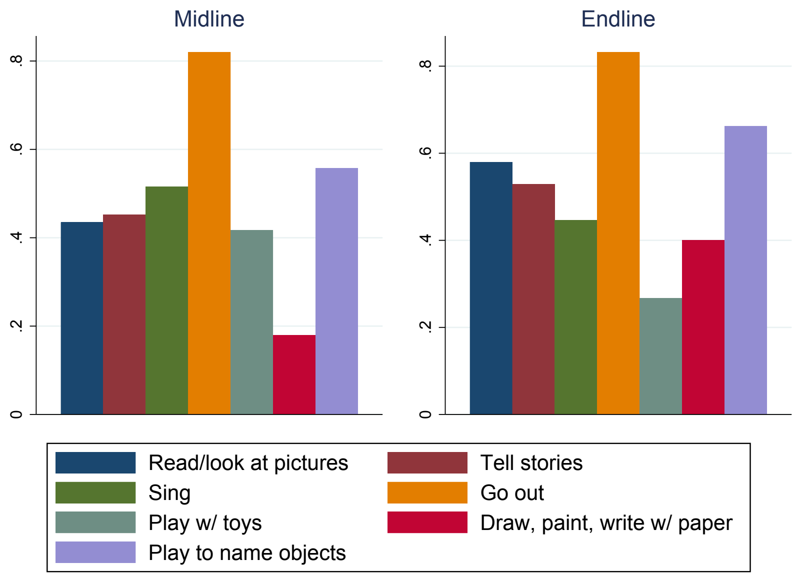
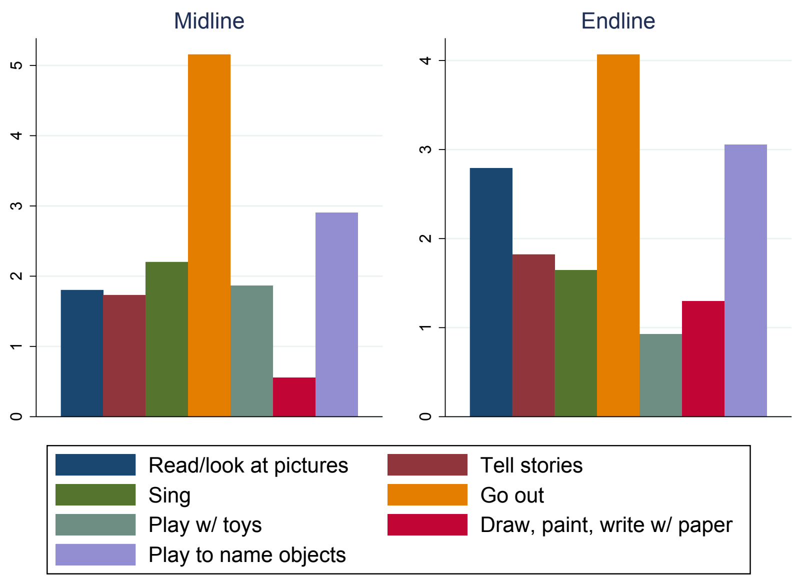


Figure 11: Intensive margin activities
Control midline and endline



4.1.5 Consistency of the Bayley

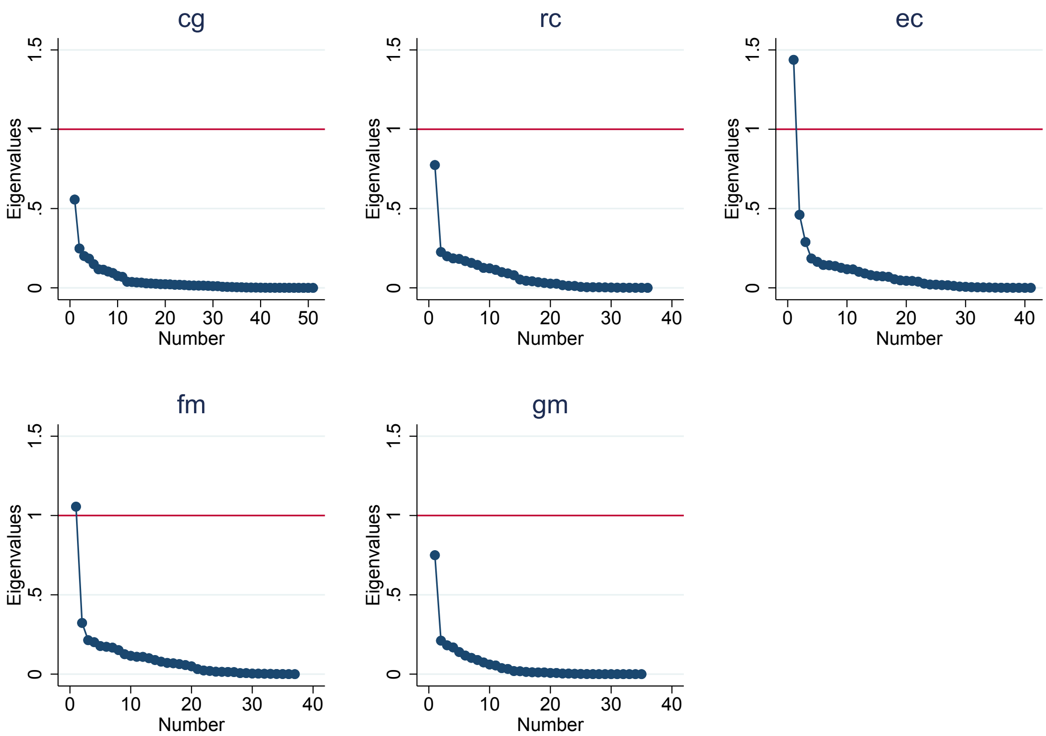
Another potential is that the bayley has stopped being as effective as a tool as the children are now older. Table 22 shows cronbachs alpha for each of the starting points of the bayley. This is broadly similar to what we saw at midline, and again indicates very good performance of the test. Figure 12 shows the scree plots from a PCA of each scale separately.[4]

Table 22: Summary of rawscore by subscale, midline and endline

	M	N	O	P	Q
Cognitive	0.9	0.7	0.8	0.8	0.8
Receptive Language	0.9	0.8	0.8	0.8	0.7
Expressive Language	1.0	0.9	0.9	0.9	0.9
Fine Motor	0.9	0.7	0.8	0.8	0.9
Gross Motor	0.9	0.7	0.8	0.8	0.8

[4] A related reason we have yet to investigate is that it may be the case that children are failing to establish the basal more as they grow older, increasing the probability of hitting the stopping rule of 5 questions in a row incorrect. This could be compounded by the closeness of starting points for older children.

Figure 12: Screeplots by subscale



Finally figures 13 to 15 show the percentile distributions of US scaled scores at midline and endline. Again if there was evidence of the test mismeasuring real ability we would see bunching at the end of the distribution. That we don't (and even for cognitive there is a falling back in scores) potentially indicates that we are finding a falling back of ability.

Figure 13: Cognitive

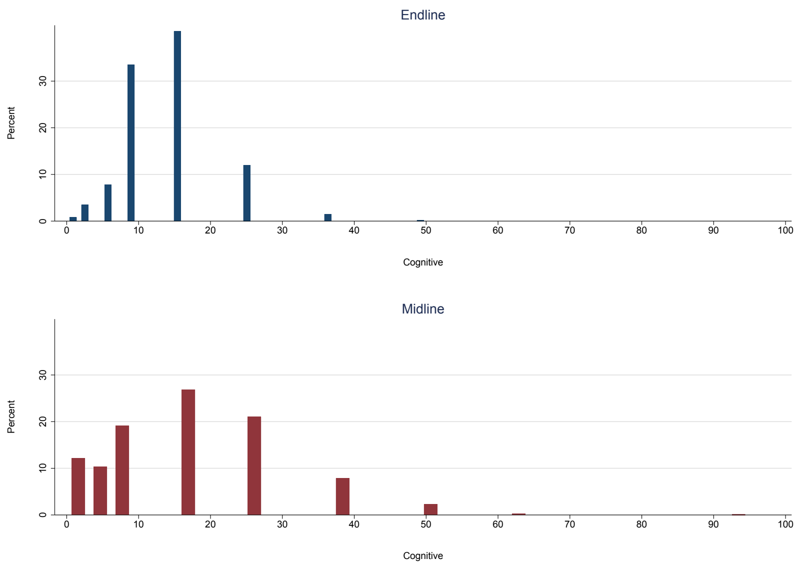


Figure 14: Language

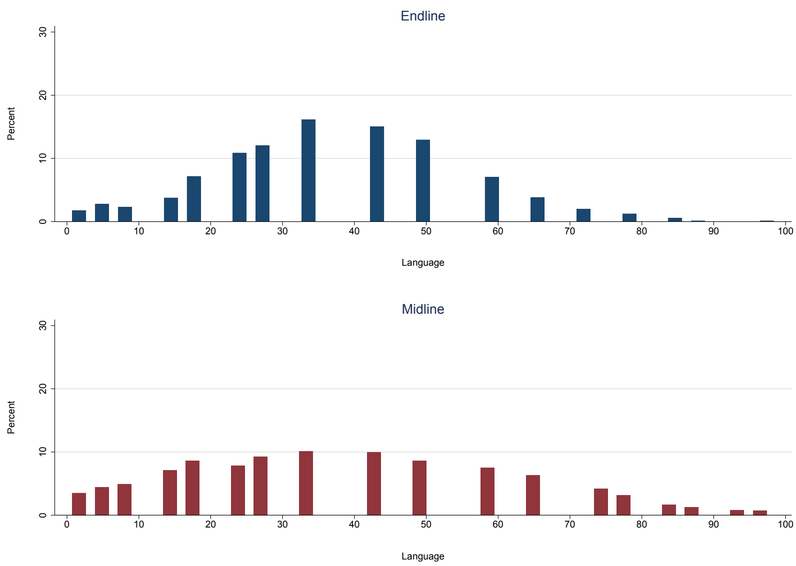
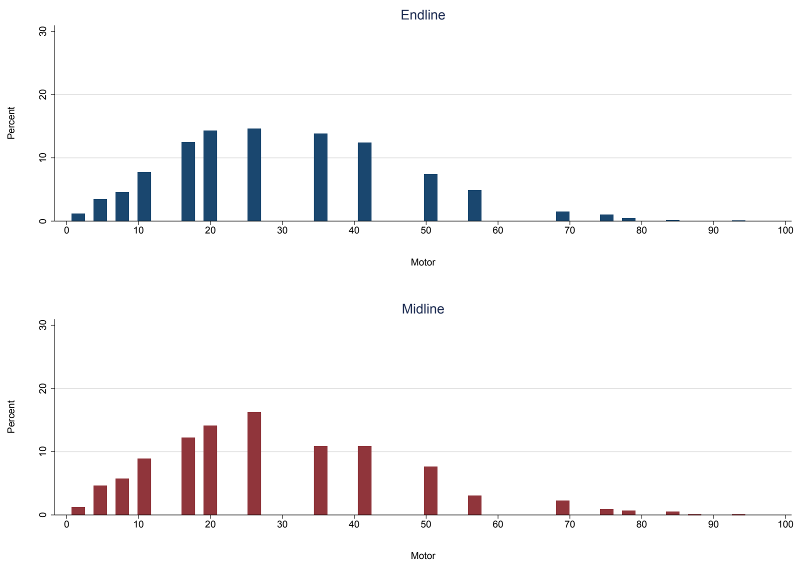


Figure 15: Motor



4.2 Mini Vocab

The Mini Vocab test is an adaption of the McArthur module, and is a mother reported measure of expressive communication. This was also conducted at midline, but updated with new words at endline to ensure that children would not hit the cap as they grew older after extensive piloting. Table 23 shows the treatment effect at midline and endline for the raw score and Z score calculated using the our original method. Here we see a fade out but only really in the GS arm, broadly consistent with the Expressive Language subscale of the bayley.

Table 23: Treatment Effect Mini Vocab Midline and Endline

	(1) MV Endline Raw Score	(2) MV Midline Raw Score	(3) MV Endline Z score	(4) MV Midline Z score
NE	1.185	2.563	0.146*	0.143*
	(1.415)	(2.356)	(0.0831)	(0.0851)
GS+NE	4.395***	7.092***	0.287***	0.359***
	(1.266)	(2.057)	(0.0767)	(0.0795)
IS+NE	4.260***	6.416***	0.301***	0.319***
	(1.331)	(2.168)	(0.0816)	(0.0841)
ASQ Communication	3.653***	5.822***	0.168***	0.179***
	(0.587)	(0.823)	(0.0338)	(0.0318)
ASQ Gross Motor	1.490**	1.324*	0.0701**	0.0667**
	(0.592)	(0.725)	(0.0338)	(0.0297)
ASQ Fine Motor	0.340	-0.0547	-0.0117	0.00869
	(0.612)	(0.924)	(0.0353)	(0.0353)
ASQ Problem Solving	0.0675	-0.0965	-0.0249	0.0160
	(0.669)	(0.829)	(0.0350)	(0.0320)
ASQ Personal Social	1.149*	1.183	0.0909***	0.0612*
	(0.603)	(0.919)	(0.0342)	(0.0337)
PCG Raven IRT score	1.190**	1.495*	0.106***	0.0829***
	(0.504)	(0.774)	(0.0280)	(0.0291)
Male	-1.106	-5.527***	0.00934	-0.189***
	(0.917)	(1.396)	(0.0489)	(0.0547)
PCG primary ed.	6.361***	5.821***	0.451***	0.247***
	(0.926)	(1.465)	(0.0584)	(0.0533)
Constant	64.79***	50.25***	-0.254***	-0.0143
	(1.366)	(2.011)	(0.0795)	(0.0772)
Observations	1287	1249	1223	1234
Adjusted R^2	0.158	0.122	0.164	0.131

Figure 16 shows the scatter plot of midline Mini Vocab scores against endline.

Figure 16: Midline vs. Endline

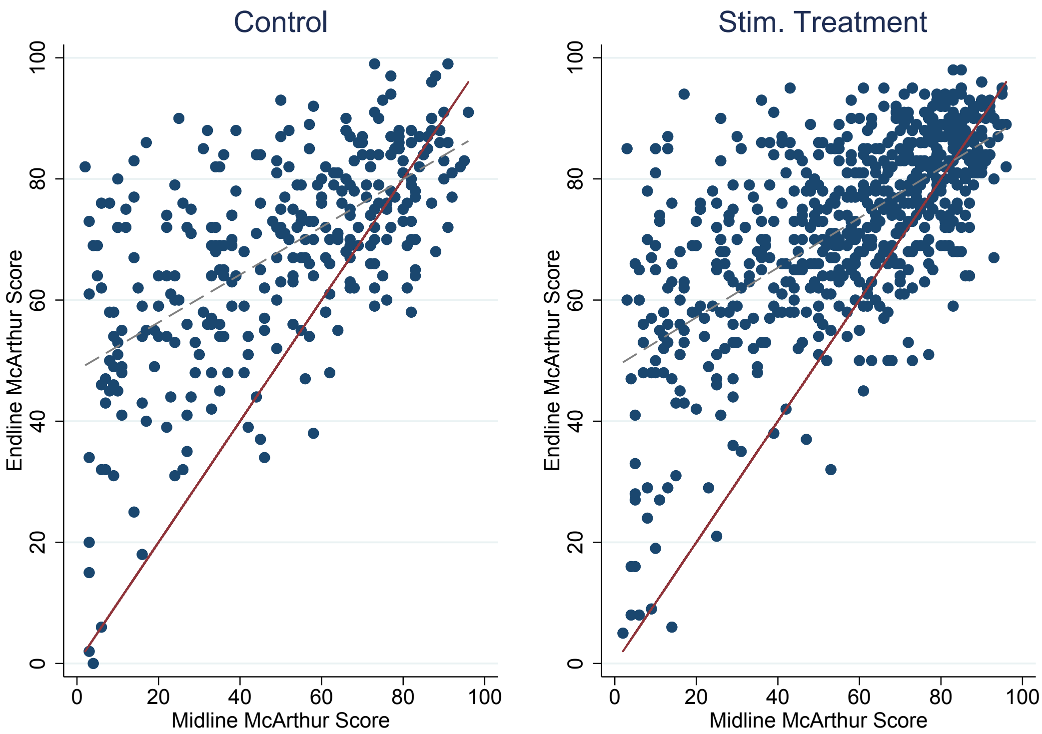


Table 24: Correlations at endline, McArthur Z score and bayley Z score

	(1) mvoc_z	(2) mvoc_z	(3) mvoc_z	(4) mvoc_z	(5) mvoc_z	(6) mvoc_z
Endline Cog. Z score	0.294***					0.0579*
	(0.0282)					(0.0305)
Endline RC. Z score		0.386***				0.161***
		(0.0254)				(0.0315)
Endline EC. Z score			0.516***			**0.424***
			(0.0257)			(0.0328)
Endline FM. Z score				0.200***		-0.0217
				(0.0268)		(0.0278)
Endline GM. Z score					0.205***	-0.0460
					(0.0278)	(0.0281)
Constant	0.133***	0.123***	0.0899***	0.178***	0.166***	0.0758***
	(0.0267)	(0.0255)	(0.0243)	(0.0268)	(0.0269)	(0.0242)
Observations	1280	1280	1280	1280	1280	1280
Adjusted R^2	0.078	0.152	0.239	0.041	0.040	0.260

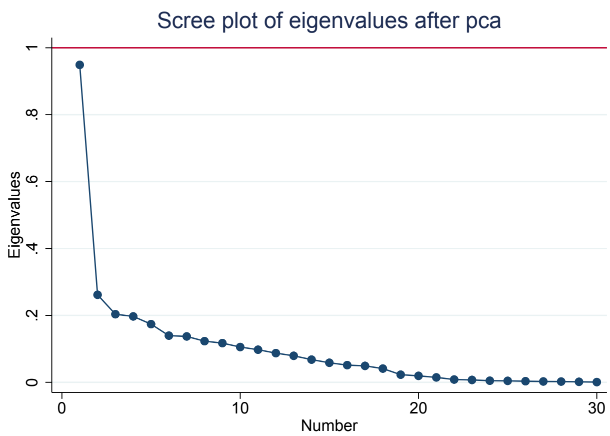
Table 25: Predicting McArthur Z score using Midline Bayley Z scores

	(1) mvoc_z	(2) mvoc_z	(3) mvoc_z	(4) mvoc_z	(5) mvoc_z	(6) mvoc_z
Midline Cog. Z score	0.234***					0.0332
	(0.0259)					(0.0283)
Midline RC. Z score		0.348***				0.0761**
		(0.0267)				(0.0331)
Midline EC. Z score			0.445***			**0.358***
			(0.0233)			(0.0286)
Midline FM. Z score				0.201***		0.0133
				(0.0269)		(0.0273)
Midline GM. Z score					0.229***	0.0793***
					(0.0266)	(0.0261)
Constant	0.137***	0.126***	0.0897***	0.152***	0.199***	0.0898***
	(0.0269)	(0.0260)	(0.0245)	(0.0271)	(0.0263)	(0.0249)
Observations	1276	1279	1279	1278	1275	1275
Adjusted R^2	0.060	0.116	0.222	0.041	0.054	0.229

4.3 WPPSI (receptive language subscale)

The WPPSI subscale for receptive language was completed after the main Bayley test at endline. We are currently aiming to use the whole WPPSI test as our main outcome in the 3-6 intervention. The minimum age range for the WPPSI was older than the youngest children we tested at endline, and so further exploration of the impact this is having on our results is required. The WPPSI shows good internal consistency, with a cronbachs alpha of 0.84. The scree plot of all the items in the WPPSI is given in figure 17, and although there is no one factor with an eigenvalue greater than one its shape suggests there is one underlying factor behind the score.

Figure 17: WPPSI Scree



Note in tables 27 and 28 the most predictive part of the bayley for the WPPSI is the receptive communication subscale - an encouraging result in terms of internal consistency.

Table 26: WPPSI: treatment effects

	(1) WPPSI Raw Score	(2) WPPSI Z score	(3) Bayley RC Z Score
NE	0.596	0.188**	0.0811
	(0.370)	(0.0945)	(0.0780)
GS+NE	1.063***	0.334***	0.254***
	(0.342)	(0.0891)	(0.0813)
IS+NE	0.608*	0.203**	0.181*
	(0.352)	(0.0865)	(0.0930)
ASQ Communication	0.389***	0.0460	0.0879***
	(0.130)	(0.0344)	(0.0328)
ASQ Gross Motor	-0.0100	0.00355	0.0244
	(0.126)	(0.0342)	(0.0342)
ASQ Fine Motor	0.0744	0.0134	0.0241
	(0.131)	(0.0370)	(0.0418)
ASQ Problem Solving	-0.0301	0.0210	0.0640*
	(0.142)	(0.0375)	(0.0342)
ASQ Personal Social	-0.0137	-0.00380	0.0781**
	(0.132)	(0.0356)	(0.0318)
PCG Raven IRT score	0.506***	0.166***	0.129***
	(0.131)	(0.0360)	(0.0325)
Male	-0.398*	-0.0918	-0.0984*
	(0.213)	(0.0573)	(0.0528)
PCG primary ed.	1.040***	0.328***	0.295***
	(0.238)	(0.0648)	(0.0623)
Constant	5.016***	-0.137	-0.106
	(0.321)	(0.0837)	(0.0754)
Observations	1275	1254	1275
Adjusted R^2	0.054	0.067	0.104

Figure 18 shows a scatter of WPPSI raw scores against each bayley subscale. Note that the WPPSI provides good variation in score for those children at the end of the receptive language subscale in the bayley.

Figure 18: Bayley vs. Wppsi

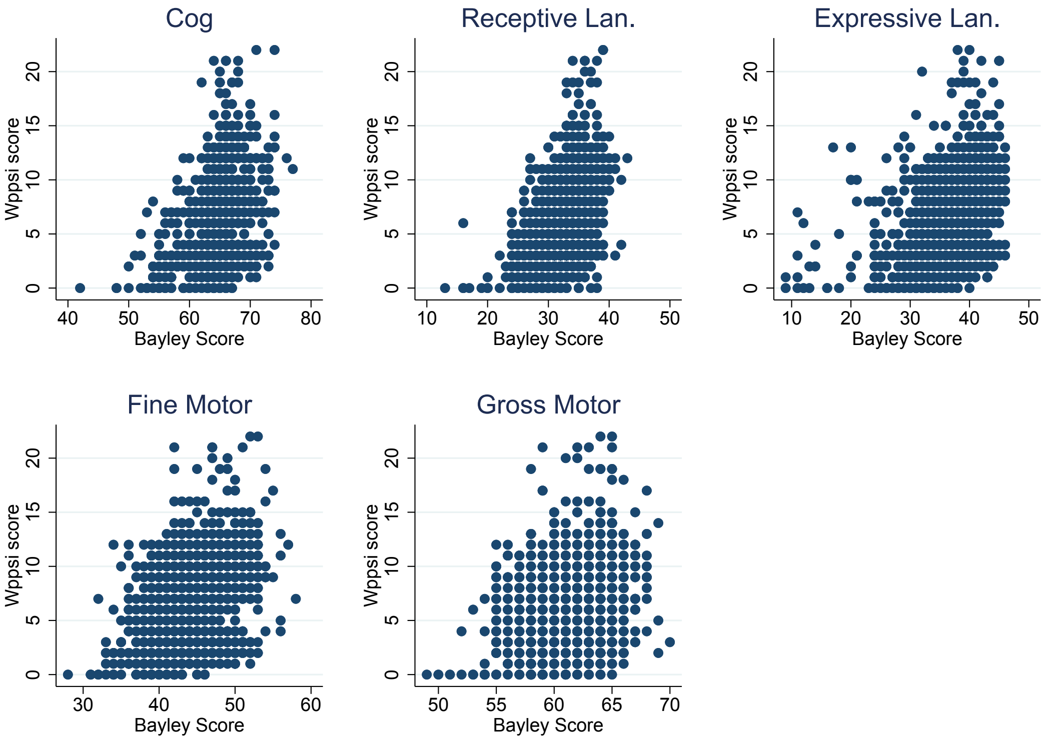


Table 27: Correlations at endline, WPPSI Z score and bayley Z score

	(1) wppsi_z	(2) wppsi_z	(3) wppsi_z	(4) wppsi_z	(5) wppsi_z	(6) wppsi_z
Endline Cog. Z score	0.332***					0.0960***
	(0.0314)					(0.0360)
Endline RC. Z score		0.412***				**0.254***
		(0.0285)				(0.0372)
Endline EC. Z score			0.342***			0.0881**
			(0.0313)			(0.0386)
Endline FM. Z score				0.322***		0.136***
				(0.0290)		(0.0327)
Endline GM. Z score					0.226***	0.0173
					(0.0310)	(0.0331)
Constant	0.144***	0.137***	0.139***	0.193***	0.182***	0.123***
	(0.0297)	(0.0285)	(0.0296)	(0.0291)	(0.0299)	(0.0286)
Observations	1280	1280	1280	1280	1280	1280
Adjusted R^2	0.080	0.140	0.085	0.087	0.039	0.169

Table 28: Predicting WPPSI Z score using Midline Bayley Z score

	(1) wppsi_z	(2) wppsi_z	(3) wppsi_z	(4) wppsi_z	(5) wppsi_z	(6) wppsi_z
Midline Cog. Z score	0.269***					0.0891***
	(0.0291)					(0.0338)
Midline RC. Z score		0.360***				**0.187***
		(0.0301)				(0.0395)
Midline EC. Z score			0.280***			0.0963***
			(0.0284)			(0.0341)
Midline FM. Z score				0.279***		0.125***
				(0.0297)		(0.0326)
Midline GM. Z score					0.189***	0.0520*
					(0.0303)	(0.0311)
Constant	0.143***	0.140***	0.142***	0.153***	0.211***	0.107***
	(0.0302)	(0.0293)	(0.0300)	(0.0299)	(0.0300)	(0.0298)
Observations	1276	1279	1279	1278	1275	1275
Adjusted R^2	0.062	0.100	0.070	0.064	0.029	0.129

4.4 Predictions

The following section looks at the predictive properties of midline measures on endline outcomes. Table 29 looks at how all the midline bayley scores together predict each endline subscale, and table 30 regresses each midline Z score against endline Z score individually.

Table 29: Predicting Endline Bayley Z Scores using Midline Bayley Z scores

	(1) Cog	(2) Receptive Lan.	(3) Expressive Lan.	(4) Fine Motor	(5) Gross Motor
Midline Cog. Z score	0.153***	-0.00215	0.0381	0.132***	0.0138
	(0.0302)	(0.0302)	(0.0280)	(0.0291)	(0.0296)
Midline RC. Z score	0.169***	0.307***	0.127***	0.185***	0.0731**
	(0.0292)	(0.0366)	(0.0313)	(0.0348)	(0.0368)
Midline EC. Z score	0.0593**	0.215***	0.381***	0.0264	0.140***
	(0.0294)	(0.0316)	(0.0282)	(0.0304)	(0.0293)
Midline FM. Z score	0.0654**	0.0963***	0.00162	0.193***	-0.00299
	(0.0299)	(0.0283)	(0.0227)	(0.0306)	(0.0292)
Midline GM. Z score	0.0963***	0.0225	0.0721***	0.0634**	0.284***
	(0.0246)	(0.0245)	(0.0234)	(0.0288)	(0.0297)
Constant	0.0874***	0.0540*	0.0771***	-0.0608*	0.0625**
	(0.0302)	(0.0312)	(0.0242)	(0.0313)	(0.0289)
Observations	1269	1269	1269	1269	1269
Adjusted R^2	0.178	0.273	0.332	0.188	0.168

Table 30: Predicting Endline Bayley Z Scores using Midline Bayley Z scores (all on each separately)

	(1) Cog	(2) Receptive Lan.	(3) Expressive Lan.	(4) Fine Motor	(5) Gross Motor
Midline Cog. Z score	0.305***	0.265***	0.265***	0.318***	0.186***
	(0.0290)	(0.0289)	(0.0279)	(0.0269)	(0.0283)
Constant	0.106***	0.113***	0.131***	-0.0310	0.0582*
	(0.0309)	(0.0338)	(0.0283)	(0.0321)	(0.0305)
Observations	1270	1270	1270	1270	1270
Adjusted R^2	0.114	0.077	0.089	0.105	0.038
Midline RC. Z score	0.354***	0.495***	0.413***	0.380***	0.265***
	(0.0268)	(0.0322)	(0.0284)	(0.0293)	(0.0341)
Constant	0.110***	0.0806***	0.114***	-0.0282	0.0481
	(0.0287)	(0.0309)	(0.0266)	(0.0310)	(0.0294)
Observations	1273	1273	1273	1273	1273
Adjusted R^2	0.135	0.237	0.189	0.133	0.069
Midline EC. Z score	0.261***	0.417***	0.493***	0.255***	0.280***
	(0.0285)	(0.0316)	(0.0254)	(0.0268)	(0.0300)
Constant	0.116***	0.0772**	0.0791***	-0.0165	0.0337
	(0.0303)	(0.0321)	(0.0247)	(0.0313)	(0.0288)
Observations	1273	1273	1273	1273	1273
Adjusted R^2	0.085	0.195	0.312	0.069	0.090
Midline FM. Z score	0.246***	0.302***	0.219***	0.351***	0.163***
	(0.0289)	(0.0287)	(0.0270)	(0.0285)	(0.0311)
Constant	0.129***	0.115***	0.148***	-0.0225	0.0675**
	(0.0307)	(0.0327)	(0.0283)	(0.0305)	(0.0290)
Observations	1272	1272	1272	1272	1272
Adjusted R^2	0.069	0.093	0.056	0.121	0.027
Midline GM. Z score	0.222***	0.195***	0.241***	0.207***	0.352***
	(0.0277)	(0.0278)	(0.0318)	(0.0264)	(0.0304)
Constant	0.185***	0.181***	0.200***	0.0510	0.112***
	(0.0303)	(0.0327)	(0.0272)	(0.0313)	(0.0278)
Observations	1269	1269	1269	1269	1269
Adjusted R^2	0.057	0.039	0.070	0.042	0.133

Table 31 looks at the impact of midline HOME module Z scores on each subscale of the bayley. These suggest that measurable parental investment matters for all subscales.

Table 31: Predicting Endline Bayley Scores using Midline HOME outcomes

	Cognitive	Cognitive	Cognitive
Midline Activity Z score	0.219***		0.202***
	(0.0246)		(0.0258)
Midline Materials Z score		0.125***	0.0615**
		(0.0274)	(0.0294)
Constant	0.0938***	0.163***	0.0992***
	(0.0304)	(0.0312)	(0.0305)
Observations	1285	1285	1285
Adjusted R^2	0.058	0.016	0.061

	Receptive Lan.	Receptive Lan.	Receptive Lan.
Midline Activity Z score	0.224***		0.185***
	(0.0285)		(0.0284)
Midline Materials Z score		0.203***	0.144***
		(0.0271)	(0.0272)
Constant	0.0885***	0.159***	0.101***
	(0.0338)	(0.0331)	(0.0333)
Observations	1285	1285	1285
Adjusted R^2	0.056	0.039	0.074

	Expressive Lan.	Expressive Lan.	Expressive Lan.
Midline Activity Z score	0.214***		0.180***
	(0.0233)		(0.0240)
Midline Materials Z score		0.181***	0.124***
		(0.0272)	(0.0279)
Constant	0.114***	0.182***	0.125***
	(0.0302)	(0.0282)	(0.0296)
Observations	1285	1285	1285
Adjusted R^2	0.060	0.036	0.075

	Fine Motor	Fine Motor	Fine Motor
Midline Activity Z score	0.144***		0.107***
	(0.0273)		(0.0286)
Midline Materials Z score		0.171***	0.137***
		(0.0307)	(0.0327)
Constant	-0.0128	0.0329	-0.000831
	(0.0331)	(0.0320)	(0.0332)
Observations	1285	1285	1285
Adjusted R^2	0.022	0.027	0.037

	Gross Motor	Gross Motor	Gross Motor
Midline Activity Z score	0.128***		0.125***
	(0.0277)		(0.0287)
Midline Materials Z score		0.0508*	0.0112
		(0.0276)	(0.0286)
Constant	0.0442	0.0847***	0.0452
	(0.0306)	(0.0295)	(0.0306)
Observations	1285	1285	1285
Adjusted R^2	0.018	0.002	0.017

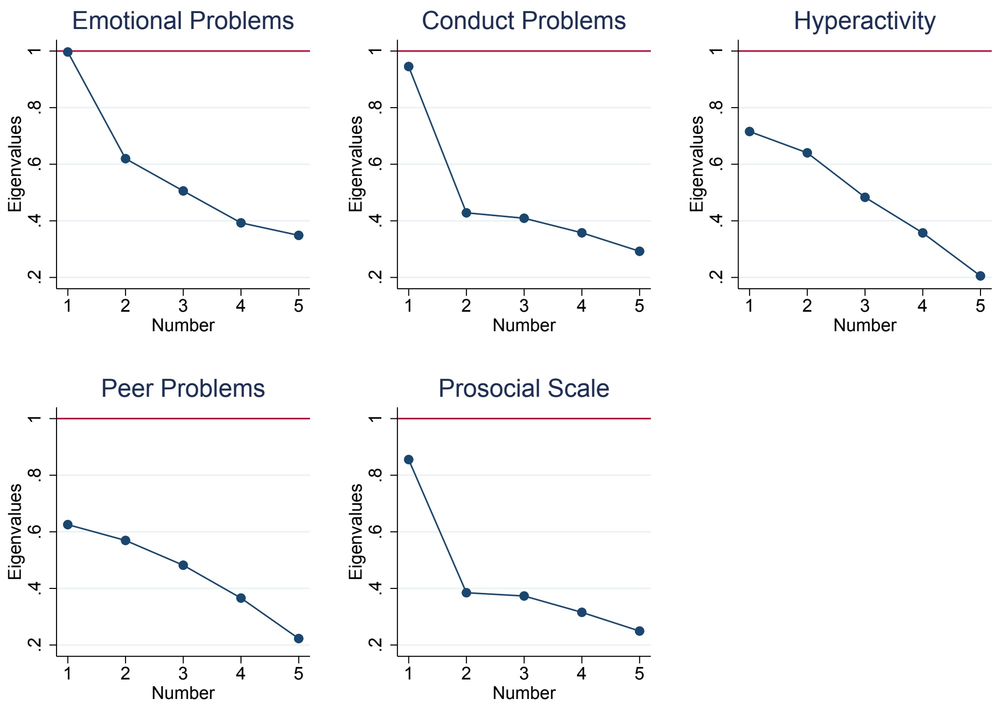
5 Social Emotional

Social emotional skills were measured using the Strengths and Difficulties questionnaire (SDQ) at endline. The SDQ is is divided into 5 subscales; Emotional Problems, Conduct Problems, Hyperactivity, Peer Problems and Pro-social. This part of the questionnaire was done for the first time at endline, and some significant adaptation and testing took place before main data collection. The cronbachs alpha for each subscale (table 32) suggests that of these 5, only three were sufficiently internally consistent - Emotional Problems, Conduct Problems and Pro-Social. As an alternate method of validation, we conduct a PCA of each subscale and examine the number of factors in each.

Table 32: Cronbachs Alpha by subscale

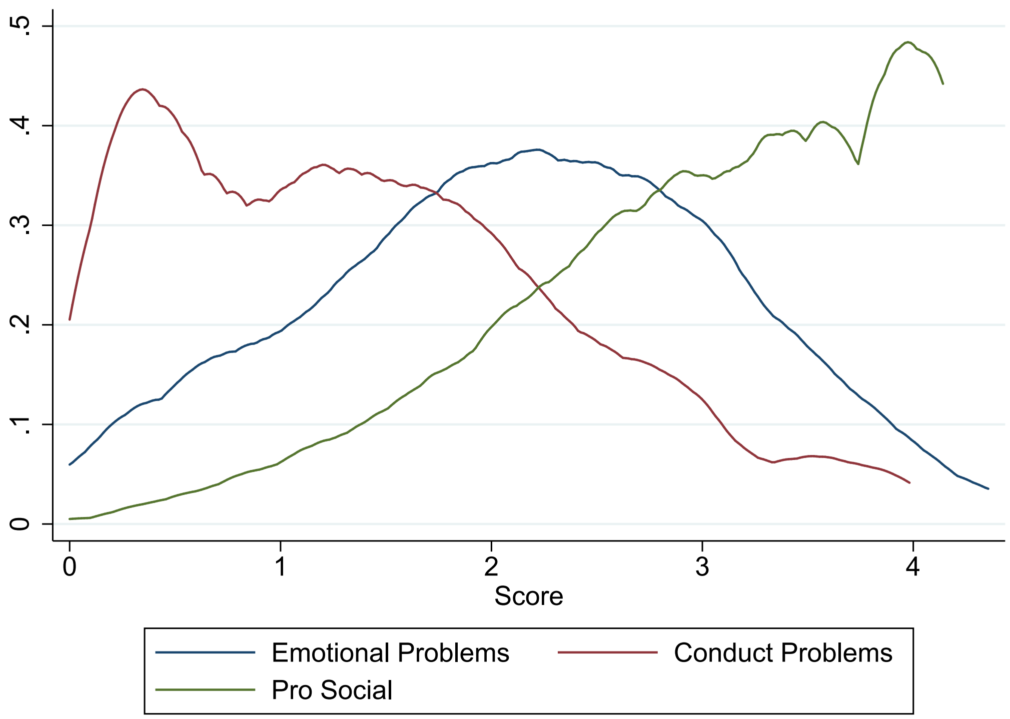
Emotional	Conduct	Hyperactivity	Peer Problems	Pro Social
0.5	0.5	0.1	0.2	0.6

Figure 19: Scree plots of PCA results, each subscale separately



For these three consistent subscales we create a score of the first component, the distributions of which are shown in figure 20. As hyperactivity and peer problems performed so badly we chose not to construct measures based on those items.

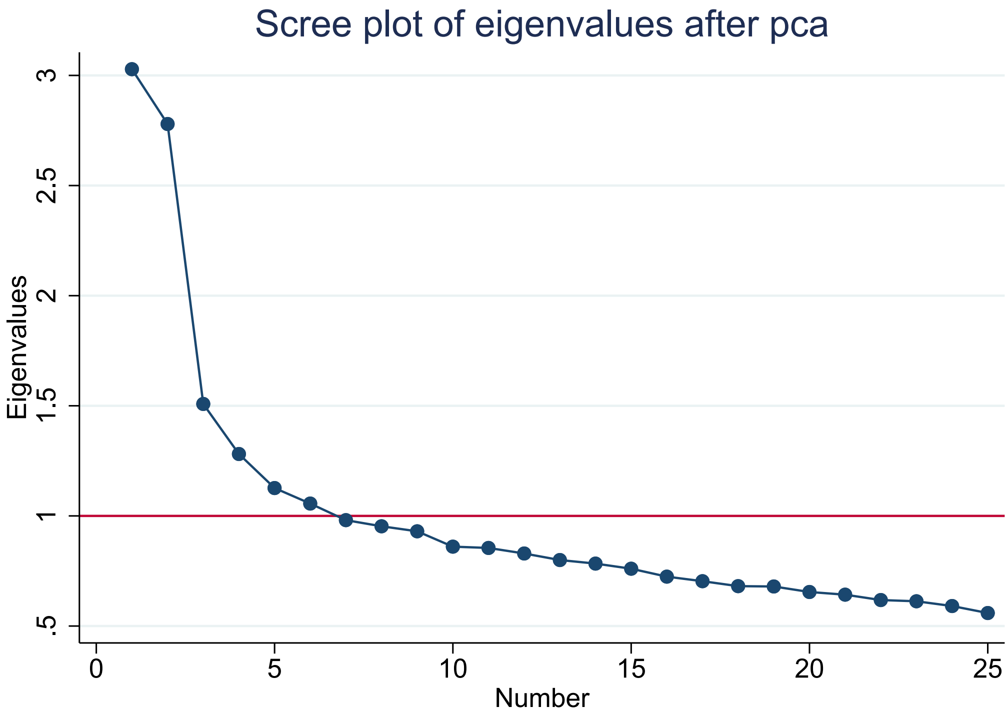
Figure 20: Distributions of Conduct, Emotional and Pro social subscales



5.1 Combined Item Analysis

Another way to process the SDQ is to ignore the division of the items in their separate subscales, and simply find the principle components of all 25 items together. This yields the following scree plot, which displays 6 factors with eigenvalues greater than 1, but 2 with eigenvalues significantly larger than the rest.

Figure 21: Scree plots of PCA results, all items

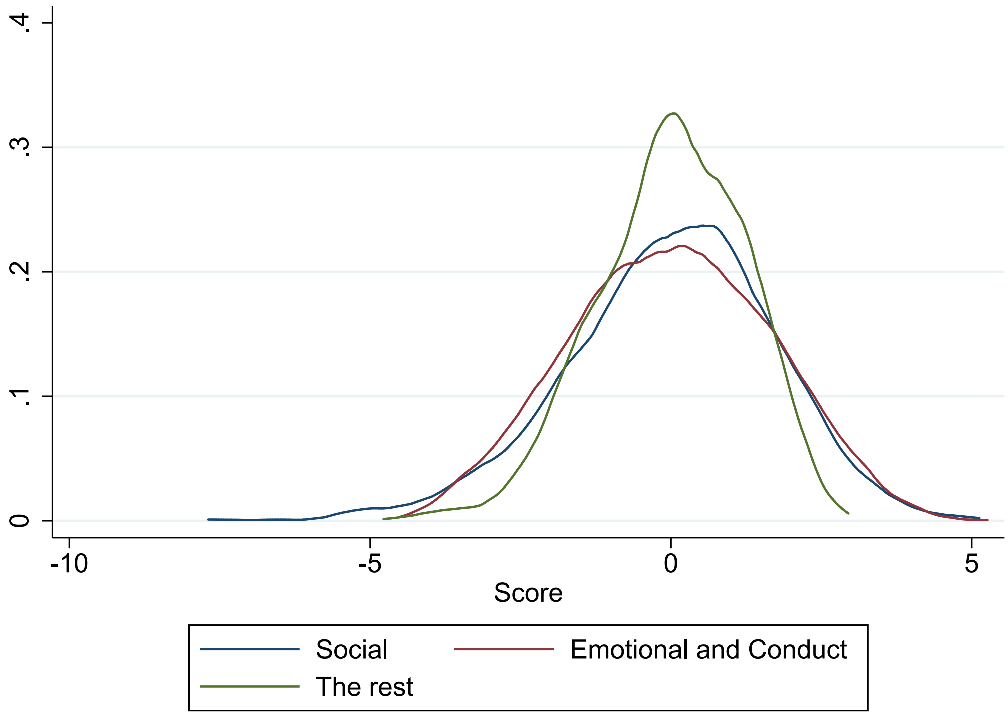


The loading of these factors onto individual items is shown in table 33. The first component loads almost entirely on items in the pro-social subscale ("Social Component"), the second on a combination of the emotional problems and conduct problems subscales ("Problem Component"). We also include a third component which seems to load on other items, but has a less clear interpretation. We generate a score for the first three components (which combined explain 29% of the variance), the densities of which are shown in figure 22.

Table 33: SDQ loadings

	"Social"	"Conduct/Emotional"	"Other "
Emotional Problems			
Often headaches	0.24	0.07	0.04
Worries often	0.20	0.18	-0.19
Often unhappy	0.15	0.33	-0.14
Nervous and clingy	0.11	0.20	-0.01
Easily scared	0.14	0.23	-0.09
Conduct Problems			
Loses temper	0.11	0.17	0.41
Well behaved	-0.30	0.20	0.10
Fights w/ others	0.14	0.34	0.12
Often argues	0.13	0.33	0.04
Spiteful to others	0.14	0.33	-0.06
Hyperactivity			
Restless/overactive	0.03	-0.07	0.46
Fidgets	0.16	0.07	0.38
Easily distracted	0.13	0.19	0.14
Can stop to think things out	-0.28	0.12	0.28
Good attention span	-0.24	0.19	0.27
Peer problems			
Prefers play alone	0.07	0.18	-0.33
Has at least one friend	-0.22	0.12	-0.23
Liked by other kids	-0.19	0.17	-0.16
Picked on/bullied by other kids	0.20	0.25	-0.02
Gets along better with adults	0.20	-0.06	0.05
Prosocial			
Considerate of others feelings	0.26	-0.16	-0.02
Readily shares	0.21	-0.11	0.15
Helpful	0.31	-0.14	-0.02
Gentle to smaller kids	0.23	-0.19	0.03
Likes to help others	0.31	-0.16	-0.03

Figure 22: Distributions of Conduct, Emotional and Pro social subscales



5.2 Treatment effects

Here we compare the treatment effects on the components estimated from each subscale individually, along with the three components returned from the PCA of all items. Table **??** shows the treatment effect for the principle component in the three subscales with acceptable Cronbachs alphas, and table 35 shows the same but for the first three components of the PCA when all items are included, labelled as by what items they weighed most heavily on.

Table 34: Treatment Effect for principle components by subscale

	(1) Emotional Problems	(2) Conduct Problems	(3) Pro Social
NE	-0.10	-0.02	0.07
	(0.08)	(0.08)	(0.07)
GS+NE	-0.06	-0.05	0.17**
	(0.07)	(0.09)	(0.07)
IS+NE	-0.11	-0.08	0.21***
	(0.07)	(0.08)	(0.07)
ASQ Communication	-0.02	0.02	0.12***
	(0.03)	(0.03)	(0.03)
ASQ Gross Motor	-0.05	0.03	-0.01
	(0.03)	(0.03)	(0.03)
ASQ Fine Motor	0.02	0.01	0.02
	(0.03)	(0.03)	(0.03)
ASQ Problem Solving	-0.01	-0.04	0.02
	(0.03)	(0.03)	(0.03)
ASQ Personal Social	0.01	-0.00	0.04
	(0.03)	(0.03)	(0.03)
PCG Raven IRT score	-0.17***	-0.10***	0.05*
	(0.03)	(0.03)	(0.03)
Male	-0.20***	0.21***	-0.15***
	(0.05)	(0.05)	(0.05)
PCG primary ed.	-0.36***	-0.45***	-0.01
	(0.06)	(0.06)	(0.06)
Constant	2.53***	1.65***	3.03***
	(0.06)	(0.07)	(0.08)
Observations	1304	1304	1304
Adjusted R^2	0.080	0.077	0.047

Table 35: Treatment Effect for first three components all items

	(1) The others	(2) Conduct and Emotional	(3) Social
NE	-0.08	-0.11	-0.01
	(0.13)	(0.14)	(0.13)
GS+NE	-0.11	-0.22*	0.12
	(0.13)	(0.13)	(0.13)
IS+NE	0.01	-0.37***	0.25**
	(0.12)	(0.13)	(0.12)
ASQ Communication	-0.06	-0.12**	0.24***
	(0.04)	(0.05)	(0.05)
ASQ Gross Motor	0.01	0.01	-0.03
	(0.04)	(0.05)	(0.05)
ASQ Fine Motor	0.02	-0.01	0.06
	(0.04)	(0.05)	(0.05)
ASQ Problem Solving	0.02	-0.08	0.05
	(0.05)	(0.05)	(0.06)
ASQ Personal Social	-0.01	-0.03	0.08
	(0.04)	(0.06)	(0.05)
PCG Raven IRT score	0.17***	-0.30***	-0.11**
	(0.04)	(0.05)	(0.05)
Male	0.12*	0.21**	-0.21**
	(0.07)	(0.08)	(0.09)
PCG primary ed.	0.23***	-0.76***	-0.47***
	(0.08)	(0.10)	(0.11)
Constant	-0.13	0.53***	0.37***
	(0.12)	(0.13)	(0.12)
Observations	1304	1304	1304
Adjusted R^2	0.027	0.121	0.065

6 Anthropometric Outcomes

For anthropometrics, as at midline, we still do not see any impact on any outcome measure. Table 36 provides a summary of our treatment effects across a variety of outcomes, figure 23 shows the distributions of Height for age Z score of reach treatment arm, and figure 24 illustrates how our sample has fallen further behind the WHO healthy reference population over time. Anthropometrics was an area which SIEF picked up on a lot after the midline report, and so explaining why the intervention has no effect on outcomes is quite important.

The following section will look at two potential mechanisms for improving child health; nutritional quality and diet, and sanitation and hygiene practices in the home.

Table 36: Proportion of children classified as underweight, stunted or wasted
Target Children

		Difference from Control				
	Control Mean	NE	IS	GS	N	F-stat
Weight-for-age	-1.22	0.02	0.06	0.02	1353	0.10
	(1.19)	(0.11)	(0.13)	(0.11)		
Height-for-age	-1.97	0.07	0.04	-0.01	1347	0.19
	(1.26)	(0.12)	(0.14)	(0.13)		
Weight-for-height	-0.16	-0.02	0.08	0.03	1344	0.38
	(1.32)	(0.10)	(0.11)	(0.10)		
Underweight (%)	24.76	-1.11	-1.91	-3.34	1353	0.30
	(43.23)	(3.71)	(3.73)	(3.65)		
Stunted (%)	49.69	-3.77	-2.11	-1.86	1347	0.22
	(50.08)	(4.60)	(4.87)	(4.74)		
Wasted (%)	6.60	-0.88	-1.75	-1.95	1344	0.41
	(24.87)	(2.11)	(2.01)	(1.93)		
Raw height (cm)	88.44	0.08	-0.00	0.21	1353	0.05
	(5.17)	(0.48)	(0.51)	(0.60)		
Raw weight (kg)	12.41	-0.05	-0.06	-0.16	1356	0.24
	(2.19)	(0.20)	(0.22)	(0.20)		

Note: * p<0.1, ** p<0.05, *** p<0.01, standard errors in parentheses, clustered at the village level

Figure 23: Height-for-Age Z-Score by Treatment
Target Children

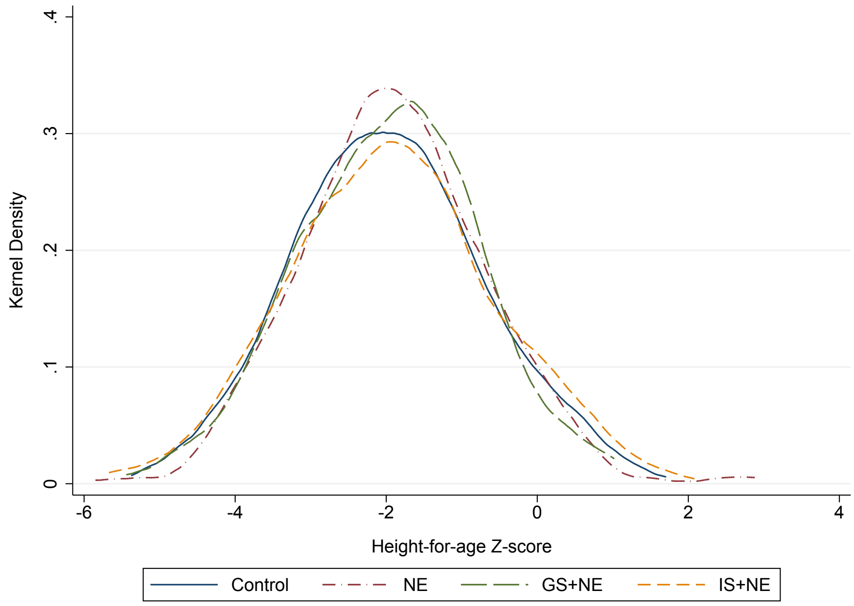
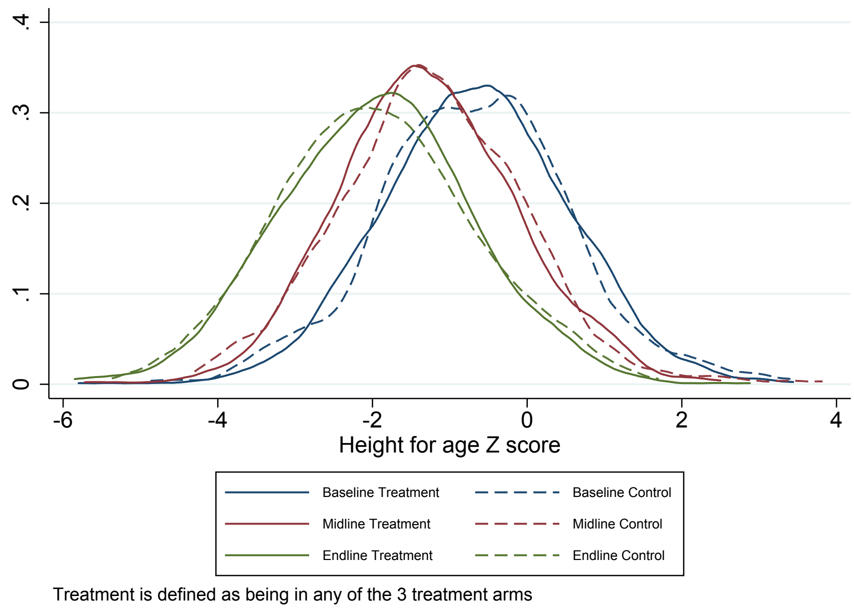


Figure 24: Height-for-Age Z-Score by Treatment Each Round
Target Children



6.1 Nutritional quality and diet

The primary outcome of the NE aspect of the intervention was to improve the diet of the children in target households. Some of this was through promoting increased dietary diversity, the other through improving the energy content of the meals. The results of these are shown in tables 37[5] and 38.

Included in table 38 is a measure of quantity. It is not entirely clear from the diet tool how to best measure quantity of food consumed by the child. Foods are listed by meals, which in turn are measured in either pieces, tablespoons, teaspoons or bowls. There is also some information on ingredients but its unclear what the nutritional value of those ingredients are.

Given 94% of all meals are measured in tablespoons, teaspoons or bowls, we've imposed the following conversion factors between the three and summed to create a rudimentary quantity measure.

A quantity variable

12 teapoons = 6 tablespoons = 1 bowl

Table 37: Dietary Diversity Measures
Target Children

		Difference from Control				
	Control Mean	NE	IS	GS	N	F-stat
Starchy Staples	100.00 (0.00)	0.00 (.)	-0.28 (0.28)	-0.56 (0.39)	1386	.
Legumes/Pulses	99.09 (9.53)	-0.55 (0.81)	-1.04 (0.86)	-2.73** (1.15)	1386	2.01
Dairy (excluding breast milk)	5.79 (23.40)	-0.24 (1.69)	2.56 (1.75)	2.33 (1.84)	1386	1.30
Meat, fish, egg	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	1386	.
Vitamin A rich fruit and vegetables	72.87 (44.53)	13.10*** (3.69)	5.13 (4.11)	4.73 (4.17)	1386	5.29***
Other fruit and vegetables	7.32 (26.08)	3.79 (2.60)	5.77** (2.92)	7.53** (3.02)	1386	2.56*
Foods made with oil, fate or butter	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	1386	.
Dietary diversity score of 0 to 2	25.91 (43.88)	-12.46*** (3.58)	-4.47 (3.79)	-5.19 (3.80)	1386	5.10***
Dietary diversity score of 3 to 4	73.78 (44.05)	12.18*** (3.62)	3.66 (3.79)	4.65 (3.84)	1386	5.06***
Dietary diversity score of 5 to 7	0.30 (5.52)	0.28 (0.51)	0.81 (0.62)	0.54 (0.56)	1386	0.70

[5] The diet tool is currently saying odd things such as that no child ate meat fish or egg. We are still doing some work cleaning this up.

Table 38: Energy Improvement Measures
Target Children

	Control Mean	Difference from Control			N	F-stat
		NE	IS	GS		
Added ghee at least one meal	0.61 (7.80)	11.67*** (2.11)	9.70*** (1.91)	9.19*** (1.92)	1386	24.18***
Added Jaggery at least one meal	0.30 (5.52)	7.59*** (2.16)	2.48*** (0.84)	2.50** (1.00)	1386	8.11***
Used arf at least one meal	0.00 (0.00)	4.39** (1.74)	2.51** (1.01)	2.24** (0.91)	1386	6.17***
Added lemon juice at least one meal	0.00 (0.00)	3.22*** (0.85)	3.62*** (1.10)	3.36*** (1.08)	1386	11.62***
Fermented food at least one meal	0.30 (5.52)	-0.30 (0.30)	-0.30 (0.30)	-0.30 (0.30)	1386	.
Used germinated grains at least one meal	0.00 (0.00)	1.75** (0.88)	0.28 (0.28)	1.12* (0.67)	1386	2.58*
Did something to improve nutriton content	0.61 (7.80)	4.07*** (1.25)	3.29*** (1.20)	3.87*** (1.40)	1386	7.06***
Did something to improve energy content	0.91 (9.53)	17.51*** (3.04)	11.90*** (2.16)	11.13*** (2.16)	1386	27.52***
Did something to improve nutrition/energy content	1.52 (12.27)	17.48*** (3.09)	14.07*** (2.42)	12.20*** (2.46)	1386	26.91***
Quantity Measure (bowls)	1.24 (0.71)	0.22*** (0.07)	0.11* (0.06)	0.11* (0.06)	1386	3.73**

Explaining Child Height

We seem to be having some effect on our mechanisms, particularly on energy improvement of foods, but no discernible effect on outcomes. Table 39 looks at the correlation between our mechanisms and child outcomes, and table 40 instruments these mechanisms using our treatment variable. Whilst measured dietary diversity seems to have no effect on child height for age Z score, the added energy content measures seem to.

Table 39: Correlations Z Height for Length Score, nutritional inputs

	(1) zlen	(2) zlen	(3) zlen	(4) zlen	(5) zlen
Child dietary diversity index score	-0.0401				-0.0907
	(0.0452)				(0.0469)
Quantity		0.245***			0.249***
		(0.0467)			(0.0479)
Did something to improve energy content			0.227*		0.135
			(0.0993)		(0.107)
Did something to improve nutriton content				0.382*	0.278
				(0.178)	(0.199)
Constant	-1.833***	-2.274***	-1.969***	-1.961***	-2.033***
	(0.143)	(0.0702)	(0.0434)	(0.0436)	(0.137)
Observations	1991	1991	1991	1991	1991
Adjusted R^2	-0.000	0.018	0.002	0.002	0.021

Table 40: Instrumenting each input with treatment assignment

	(1) zlen	(2) zlen	(3) zlen	(4) zlen
Child dietary diversity index score	0.122			
	(0.726)			
Quantity		0.220		
		(0.550)		
Did something to improve energy content			0.197	
			(0.706)	
Did something to improve nutriton content				0.474
				(2.478)
Constant	-2.310	-2.244**	-1.970***	-1.966***
	(2.144)	(0.744)	(0.0893)	(0.103)
Observations	1288	1288	1288	1288
Adjusted R^2	-0.006	0.016	-0.002	0.003
F_firststage	3.813	3.544	23.94	7.212

6.2 Sanitation Practises and illness

Some of the NE curriculum in each arm focussed on hygiene around the home. At endline as part of the household module we asked mothers a series of questions around basic sanitation practises. The table 41 shows the proportion of each treatment arm answering affirmatively to each question. Table 42 shows the impact illness's other than diarrhoea. Table 43 focuses specifically on whether or not respondents knew the 7 steps of ORS solution use (the recommended response to diarrhoea) and diarrhoea prevalence.

Table 41: Sanitation Practices

		Difference from Control				
	Control Mean	NE	IS	GS	N	F-stat
Drinking water purified	11.61	6.07*	0.71	-0.34	1401	1.24
	(32.09)	(3.66)	(3.02)	(3.19)		
Drinking water covered storage	88.39	4.66*	2.09	2.66	1401	1.24
	(32.09)	(2.47)	(2.64)	(2.49)		
HH owns toilet	65.24	-4.82	-0.92	-0.40	1331	0.23
	(47.69)	(6.25)	(5.85)	(5.68)		
A HH member openly defecates	41.08	5.88	2.34	1.99	1401	0.27
	(49.27)	(6.59)	(6.44)	(5.94)		
Washes childs hand after child defecates	17.00	9.67***	12.13***	15.66***	1401	7.85***
	(37.61)	(3.59)	(3.65)	(3.68)		

Note: * p<0.1, ** p<0.05, *** p<0.01, standard errors in parentheses, clustered at the village level

Table 42: Other Ilnesses

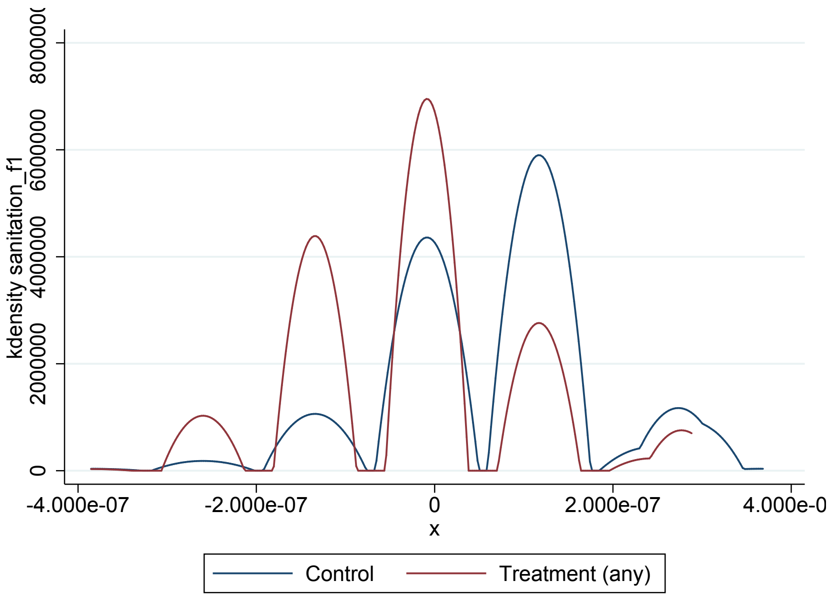
	Control Mean	Difference from Control			N	F-stat
		NE	IS	GS		
Fever	31.38	-5.47	-8.15**	-10.17***	1323	2.99**
	(46.48)	(3.72)	(4.02)	(3.50)		
Cough	46.15	-7.13*	-2.92	-8.58**	1323	1.67
	(49.93)	(4.18)	(4.37)	(4.34)		
Vomiting	13.23	-4.39*	-6.17***	-5.05**	1323	2.86**
	(33.93)	(2.29)	(2.14)	(2.28)		
Rashes	10.15	4.18	-0.74	0.15	1323	1.05
	(30.25)	(2.98)	(2.57)	(2.62)		
Sores	10.46	-3.14	-2.81	-1.37	1323	0.73
	(30.65)	(2.44)	(2.30)	(2.36)		
Indigestion	15.38	-4.41*	0.50	-0.23	1323	1.41
	(36.14)	(2.64)	(3.03)	(2.93)		
Tiredness	5.85	-2.80	-1.43	-2.51	1323	1.06
	(23.50)	(1.73)	(1.82)	(1.83)		
Paleness	9.23	-2.83	-3.64*	-1.66	1323	1.11
	(28.99)	(2.20)	(2.12)	(2.24)		

Table 43: ORS and Diarrhoea

	Control Mean	Difference from Control			N	F-
		NE	IS	GS		
Knows all steps (correct)	0.00	28.35***	22.35***	29.70***	1323	72.2
	(0.00)	(3.23)	(2.93)	(3.29)		
Knows between 2 and 4 steps (partially correct)	20.31	21.46***	29.99***	21.51***	1323	29.4
	(40.29)	(3.74)	(3.54)	(3.46)		
Knows less than 2 steps (incorrect)	27.38	-20.07***	-20.03***	-18.29***	1323	16.4
	(44.66)	(3.04)	(3.04)	(3.05)		
Diarrhea in past two weeks	2.46	0.59	-0.40	-0.64	1323	0
	(15.52)	(1.51)	(1.23)	(1.21)		

Impacts on difference aspects of hygiene and sanitation are very much mixed. There seems to be some impact on hand washing and certain illness's, but its unclear if it would survive adjustment for multiple hypothesis testing. There seems to be a marked improvement in knowledge of how to ORS, but no related reduc-

Figure 25: Sanitation Score Density



tion in diarrhoea prevalence (although from a low base).

To measure the underlying sanitation practices in the home we conduct a IRT of the above questions[6], producing the following distribution (figure 25). This is clearly not an ideal outcome and the distribution is very multi-modal, but hopefully serves as some sort of discriminator for sanitation practices. Whilst there is a visible difference between treatment and control this is not significantly different.

We then create a dummy for if a family is in the 60th percentile of the distribution or above, and interact that with the treatment variable in a regression for height for age Z score and weight for age Z score. The results of this are shown in table 44, but show no difference in treatment effect by sanitation status. Whether or not this is due to the poor measurement of sanitation or if there is really no effect is not clear.

[6] All of table 41 and a dummy for if the PCG is at least partially correct on ORS use

Table 44: Treatment Effect for first three components all items

	(1) Length-age Z score	(2) Weight-age Z score
Control × high_san=1	0.0984	0.0442
	(0.106)	(0.0967)
NE × high_san=0	0.104	0.0341
	(0.143)	(0.129)
NE × high_san=1	0.106	-0.00388
	(0.178)	(0.184)
GS+NE × high_san=0	0.0563	0.0546
	(0.143)	(0.131)
GS+NE × high_san=1	0.0931	0.167
	(0.198)	(0.177)
IS+NE × high_san=0	0.191	0.119
	(0.155)	(0.137)
IS+NE × high_san=1	-0.361*	-0.0209
	(0.203)	(0.170)
Constant	-2.023***	-1.247***
	(0.112)	(0.106)
Observations	1295	1301
Adjusted R^2	0.004	-0.003