

Analysing the effect of sibling number on input and output in the first 18 months:

Supplementary Data

Catherine Laing¹ & Erika Bergelson²

¹ University of York, York, UK

² Duke University, Durham, NC, USA

Author Note

Correspondence concerning this article should be addressed to Catherine Laing,
Department of Language and Linguistic Science, University of York, Heslington, York, UK,
YO10 5DD. E-mail: catherine.laing@york.ac.uk

Abstract

Prior research suggests that across a wide range of cognitive, educational, and health-based measures, first-born children outperform their later-born peers. Expanding on this literature using naturalistic home-recorded data and parental vocabulary report, we find that early language outcomes vary by number of siblings in a sample of 43 English-learning U.S. children from mid-to-high socioeconomic status homes. More specifically, we find that children in our sample with two or more - but not one - older siblings had smaller productive vocabularies at 18 months, and heard less input from caregivers across several measures than their peers with two or more siblings. We discuss implications regarding what infants experience and learn across a range of family sizes in infancy.

Keywords: Siblings, Lexical Development, Input Effects, Language Acquisition

Word count: X

Analysing the effect of sibling number on input and output in the first 18 months:
Supplementary Data

S1: Effect of siblings on infants’ input - audio recordings

Our main results analyse input data from the hour-long video recordings taken in the home on a monthly basis. We re-ran our analysis using the home-recorded audio data, which captures a snapshot from a daylong recording using a LENA device (<https://www.lena.org/>) worn in a vest. In this case, age but not sex was included in both models.

Outputs from model comparisons and full model outputs including estimates are shown in Tables 1 and 2. Results were consistent with the video data for object presence, but not overall input.

Table 1

Full model output from linear mixed effects regression models comparing our two input measures in relation to sibling group in the audio data. Age in months was included as a fixed effect; subject was included as a random effect.

Model	Df	Chisq	p value
Caregiver input	2	3.00	0.22
Object presence	2	23.83	0.00

Table 2

Full model output from linear mixed effects regression models comparing our two input measures (object words produced in caregiver input and object presence) over time in relation to sibling group, for the audio data. Age in months was included as a fixed effect in both models and subject was included as a random effect.

Variable	Effect	Estimate	Std. Error	df	t value	p value
Caregiver input	Intercept	6.15	0.13	83.68	46.58	<0.001
	SibGroupOne	0.13	0.18	42.99	0.73	0.469
	SibGroup2+	-0.26	0.20	42.99	-1.26	0.213
	month	-0.04	0.01	472.00	-6.01	<0.001
Object presence	Intercept	0.42	0.03	222.79	16.82	<0.001
	SibGroupOne	-0.08	0.03	43.02	-3.20	0.003
	SibGroup2+	-0.16	0.03	43.01	-5.46	<0.001
	month	0.02	0.00	472.05	9.59	<0.001

S2: Speakers in the dataset

Here we draw on data from the two caregivers who produced the most words in each recording. This is usually the mother and/or the father.

A full breakdown of speakers is shown in Table 3.

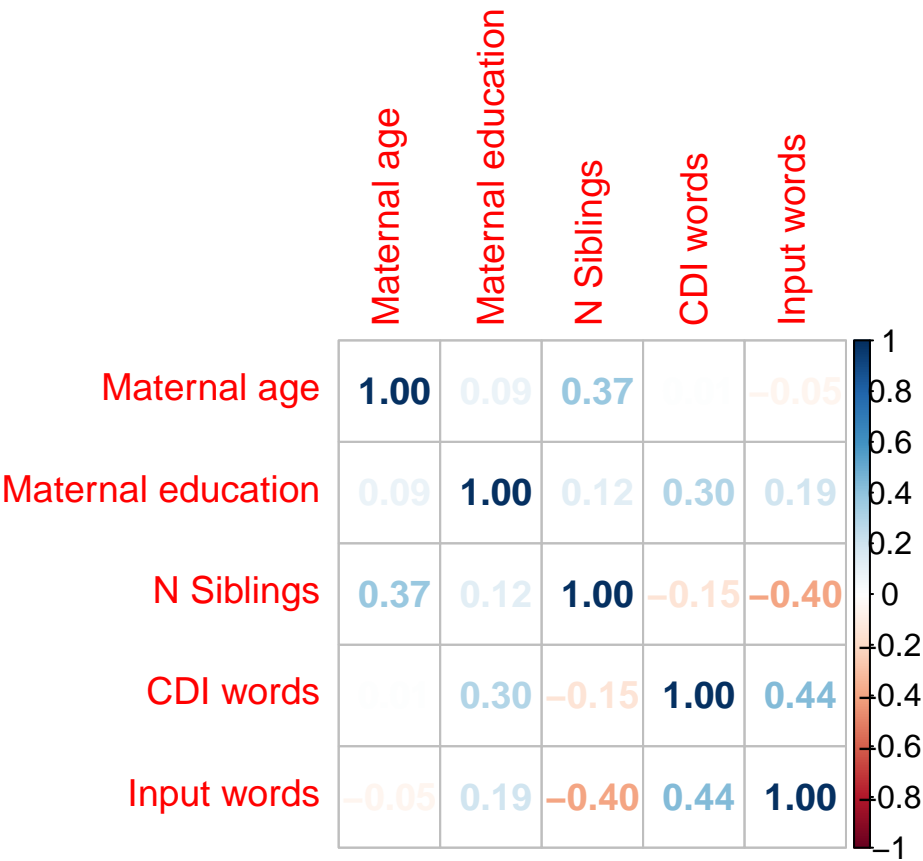
Table 3

Numer of sessions in which each adult speaker is one of the two main caregivers. There were 344 hour-long video recordings in total across the 43 infants in the dataset. In the case where the infant had two mothers, their data was aggregated into the 'Mother' category; in all but two sessions from this family, the two mothers were the two main caregivers.

Speaker	Caregiver 1	Caregiver 2
Mother	425	26
Father	62	121
Grandmother	21	32
Babysitter	5	4
Grandfather	1	9
Other Adult	1	11
Aunt	0	8
Uncle	0	3

S3: Correlations with maternal factors

There were no correlations between sibling number or child word production and maternal age/education. See Figure ??.



41 **S4: Discrete sibling number**

42 As discrete sibling number, as well as sibling group, revealed siblings to be a
 43 significant predictor of vocabulary size at 18 months, we re-ran our input models using
 44 discrete sibling number as a fixed effect, instead of sibling group. Outputs from model
 45 comparisons and full model outputs including estimates are shown in Tables 4 and 5.
 46 Results were consistent with those reported for sibling group.

Table 4

Full model output from linear mixed effects regression models comparing our two input measures in relation to discrete sibling number. Age in months was included as a fixed effect; subject was included as a random effect.

Model	Df	Chisq	p value
Caregiver input	1	7.02	0.01
Object presence	1	22.98	0.00

Table 5

Full model output from linear mixed effects regression models comparing our two input measures (object words produced in caregiver input and object presence) over time in relation to discrete sibling number. Age in months was included as a fixed effect in both models and subject was included as a random effect.

Variable	Effect	Estimate	Std. Error	df	t value	p value
Caregiver input	Intercept	4.83	0.13	79.35	38.53	<0.001
	Siblings6	-0.16	0.06	42.99	-2.76	0.008
	month	0.04	0.01	472.04	6.38	<0.001
	sexM	-0.20	0.13	43.00	-1.51	0.138
Object presence	Intercept	0.59	0.03	218.11	22.45	<0.001

Siblings6	-0.07	0.01	43.00	-5.51	<0.001
month	0.01	0.00	472.13	2.87	0.004

S5: Possible outlier for input data removed

One infant heard substantially more input words and words with object presence than all the other infants in the main sample (n=43) in four of their recording sessions. We retain this infant in the main analysis, but results were consistent when this child was removed from the analysis of input data (n=42). Results were consistent with those reported for sibling group. See Tables 6 and 7.

Table 6

Full model output from linear mixed effects regression models comparing our two input measures in relation to sibling group, with one infant removed who was identified as hearing substantially more input speech (3SDs above the group mean) in 4 recordings (n=42). Age in months was included as a fixed effect; subject was included as a random effect.

Model	Df	Chisq	p value
Caregiver input	2	10.26	0.01
Object presence	2	29.34	0.00

Table 7

Full model output from linear mixed effects regression models comparing our two input measures (object words produced in caregiver input and object presence) over time in relation to sibling group, with one infant removed who was identified as hearing substantially more input speech (3SDs above the group mean) in 4 recordings (n=42). Age in months was included as a fixed effect in both models and subject was included as a random effect.

Variable	Effect	Estimate	Std. Error	df	t value	p value
Caregiver input	Intercept	4.76	0.12	80.32	38.13	<0.001
	SibGroupOne	0.06	0.13	42.03	0.42	0.679

	SibGroup2+	-0.46	0.15	41.99	-3.03	0.004
	month	0.04	0.01	461.06	6.24	<0.001
	sexM	-0.24	0.12	42.01	-2.02	0.049
Object presence	Intercept	0.61	0.03	202.66	22.20	<0.001
	SibGroupOne	-0.11	0.03	42.09	-3.82	<0.001
	SibGroup2+	-0.20	0.03	42.00	-6.28	<0.001
	month	0.01	0.00	461.15	2.91	0.004