The Language ENvironment Analysis system (LENA): A validation study with Italian-learning children

Tamara Bastianello, Irene Lorenzini, Thierry Nazzi, Marinella Majorano

▶ **To cite this version:**

Tamara Bastianello, Irene Lorenzini, Thierry Nazzi, Marinella Majorano. The Language ENvironment Analysis system (LENA): A validation study with Italian-learning children. Journal of Child Language, In press, 10.1017/S0305000923000326 . hal-04197069

HAL Id: hal-04197069
https://hal.science/hal-04197069v1

Submitted on 5 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal of Child Language (2023), 1–21
doi:10.1017/S0305000923000326

The Language ENvironment Analysis system (LENA): A validation study with Italian-learning children

BASTIANELLO Tamara[1,2], LORENZINI Irene[3], NAZZI Thierry[3,4], MAJORANO Marinella[1]

[1] Department of Human Sciences, University of Verona, Italy

[2] Department of Developmental Psychology and Socialisation, University of Padua, Italy

[3] Integrative Neuroscience and Cognition Center, Université Paris Descartes, Sorbonne Paris Cité, Paris, France

[4] CNRS (Integrative Neuroscience and Cognition Center, UMR 8002), Paris, France

Author Note

Tamara Bastianello https://orcid.org/0000-0002-9052-8102

Marinella Majorano https://orcid.org/0000-0002-5033-328X

We have no known conflict of interest to disclose.

Acknowledgment

The study was possible thanks to a collaboration established between the University of Verona (SLD-Lab) and LABEX EFL (ANR-10-LABX-0083) that funded a mobility grant.

The authors are grateful to Marica Saffiotti and Giada Puiatti for their help in transcribing and coding. A special thanks to all the families who participated in the study.

Correspondence concerning this article should be addressed to Tamara Bastianello, Department of Human Sciences, University of Verona

Email: tamara.bastianello@univr.it

Keywords: Language ENvironment Analysis; Italian; Conversational Turns Count; Adult Word Count; Child Vocalisation Count

Abstract

This study is a validation of the LENA system for the Italian language. In Study 1, to test LENA's accuracy, seventy-two 10-minute samples extracted from daylong LENA recordings were manually transcribed for 12 children longitudinally observed at 1;0 and 2;0. We found strong correlations between LENA and human estimates in the number of Adult Word Count (AWC) and Child Vocalisations Count (CVC) and a weak correlation between LENA and human estimates in Conversational Turns Count (CTC). In Study 2, to test the concurrent validity, direct and indirect language measures were considered on a sample of 54 recordings (19 children). Correlational analyses showed that LENA's CVC and CTC were significantly related to the children's vocal production, a parent report measure of prelexical vocalizations and the vocal reactivity scores. These results confirm that the automatic analyses performed by the LENA device are reliable and powerful for studying language development in Italian-speaking infants.

Key-words: Lena system; adult words; child vocalisations; conversational turns; validity.

The Language ENvironment Analysis system (LENA): A validation study with Italian-learning children

Introduction

For decades now, researchers have focused on the characteristics of the language environment and how it shapes language development and knowledge (Hart & Risley, 1995; Hirsh-Pasek et al., 2015; Hoff & Naigles, 2002; Romeo et al., 2018). As reported in several studies, both the quantity and the quality of maternal linguistic input determine language learning's outcomes (Baldwin, 2000; Hoareau, Yeung, & Nazzi, 2019; Hoff & Naigles, 2002; Hoff, 2013; Huttenlocher, et al., 2010; Weizman & Snow, 2001; Weisleder & Fernald, 2013). Moreover, the characteristics of the adult's input are related to language outcomes not only in the home environment but also in educational contexts (Duncan et al., 2020; Majorano, Corsano, & Cigala, 2009).

Many classical studies in the 1980s and 1990s reported quantitative and qualitative descriptions of children's speech production using direct observations (via video- and/or audio-recording) rather than diaries used in previous research (Ferguson, Menn, & Stoel-Gammon, 1992; Oller, Eilers, Neal, & Schwartz, 1999; Vihman, 1991, 1993). Furthermore, phonetic, phonological and lexical descriptions of the linguistic input have been provided for children with typical language development, children with language delays and children with exposition to several languages (Keren-Portnoy, Majorano, & Vihman, 2009; McGillion et al., 2017; VanDam, Ambrose, & Moeller, 2012). Observational studies have also described quantitative and qualitative characteristics of Infant Directed Speech (IDS) and Child Directed Speech (CDS) focusing on mother-child interactions (e.g., Soderstrom, Blossom, Foygel, & Morgan, 2008). However, direct observational studies conducted in naturalistic context (e.g., home) have limitations. Firstly, direct observation (which, by nature, has limited duration) cannot estimate the level of language exposure that a child receives during an entire day or for a period of time longer than the observation; secondly, audio-recordings require a lot of work for language

transcriptions and analyses. Furthermore, reliability assessments across trained transcribers are a critical element and require additional work. Despite these drawbacks, direct observations of speech production are extremely useful to extrapolate measures of preverbal productions (Majorano et al., 2018; Majorano et al., 2020). Moreover, they can also be integrated with standardised parent-report measures, such as the PRISE questionnaire (i.e., parent report measure of prelexical vocalizations; Kishon-Rabin et al., 2004, Italian version by Cuda et al., 2013) or the Infant Behavior Questionnaire – especially the Vocal Reactivity scale (first version, Rothbart, 1981; Italian version by Montirosso et al., 2011). The use of these tools provides an immediate indirect measure of expressive skills in children in the first two years of life.

Researchers have also recently developed automatic systems for speech and language transcriptions and analysis. One of the most important achievement in this domain is the Language ENvironment Analysis system (LENA, LENA Foundation, Boulder, CO, Greenwood, Thiemann-Bourque, Walker, Buzhardt, & Gilkerson, 2011). The LENA system has been used in studies spanning across several languages and countries, in basic as well as in applied research (e.g., intervention programmes), and in both clinical and educational settings (for a recent review, see Greenwood et al., 2018). This recording system is made up of a hardware and a software component. The hardware includes a digital language processor (DLP) that is hidden in a chest pocket, on a special vest, and records the environmental acoustic input around the wearer (the infant) within a six-foot radius. The LENA software, in turn, provides automated measures of the speech heard and produced by children and adults around them. After analysing the audio-recordings, it generates quantitative assessment of a range of linguistic elements recorded (see below) and arranges this information into visual reports, thus allowing data analysis in an easy-to-read interface.

The LENA device provides several pieces of information about the linguistic and auditory characteristics of the environment. In detail, LENA can be used to automatically estimate: 1) basic meaningful speech (clear speech, recorded near the device) and distant speech (distant and not clear speech); 2) basic non-speech sounds: noise (i.e., all noises that are recognized as not coming from a human vocal tract or from an electronic speaker), television/electronic sounds (i.e., sounds from a television, radio, or other electronic media), and silence; 3) linguistic measures: number of words uttered near the child, presumably by adult caregivers (Adult Word Count, AWC); number of vocalisations produced by the child (Child

Vocalization Count, CVC); number of conversational turns (Conversational Turn Count, CTC) between a given child and an adult; Automatic Vocalisation Assessment (AVA) (i.e., a measure of expressive language skills tallied by LENA by comparing the phonemic complexity of the child's output against an adult American English model). The AVA is not as commonly used as the other measures both for English and non-English studies. In addition, for all these measures, LENA can provide 12-hours statistical projections for recordings with at least 10 hours of recording data (i.e., Projected values).

LENA could be a useful research tool, as it allows automatic calculation of input and production language measures on large time windows. However, before using it for research purposes, one would need, first, to establish that the given automatic measures are accurate (that is, reliable), by comparing automatic outputs with hand-coded measures; second, that LENA measures have concurrent validity, by comparing the LENA output with other assessments, such as standardized parental questionnaires. Regarding reliability (or accuracy), LENA has been validated for several languages, through the systematic comparison between the device's automated coding and human transcriptions (Bulgarelli & Bergelson, 2019; Christakis, et al., 2009; Cristia, et al., 2021; Richards et al., 2017). In particular, validation data have been published for: American English (Xu, Yapanel, & Gray, 2009), European French (Canault et al., 2016), Dutch (Bruyneel et al., 2020; Busch et al., 2018), Vietnamese (Ganek & Eriks-Brophy, 2018) Chinese (Gilkerson et al., 2015), Korean (McDonald, Kwon, Kim, Lee, & Ko, 2021), Swedish (Schwarz et al., 2017), Hebrew and Arabic (Levin-Asher et al., 2022), and on data collected from children growing up in a bilingual French–English environment (Orena, Byers-Heinlein, & Polka, 2019). In addition, accuracy measures have also been provided for Spanish (Weisleder & Fernald, 2013). However, that latter study cannot be considered an official validation; indeed, "validation-like" data provided by this study were comprised in an investigation on linguistic input and expressive linguistic skills in families with low socio-economic status. In particular, since LENA had not been completely validated for Spanish, these researchers conducted a small-scale validation analysis for the AWC, based on 60-minute samples taken from 10 recordings. Results showed high correlation between word counts from human transcribers and the automatic LENA estimates (AWC).

This small piece of evidence is relevant to the usage of LENA with Italian participants, given the similarity between the two languages. However, the conclusions that can be drawn are clearly

very limited. Thus, LENA cannot yet be used with reliability to study the Italian population. Moreover, and independently from the language specifically targeted, most of the previously conducted studies have validated the AWC and CVC speech measures, while CTC measures were only validated in Dutch, Chinese, Korean, and Vietnamese (Busch et al., 2018; Ganek & Eriks-Brophy, 2018; Gilkerson et al., 2015; Pae, Yoon, Seol, Gilkerson, Richards, Ma & Topping, 2016). Importantly, a recent literature review based on 33 studies reporting LENA-based accuracy measures (Cristia, Bulgarelli, & Bergelson, 2020) has revealed that only some studies (25 out of 33) provided validity estimates. Furthermore, most of the studies included in Cristia and colleagues's systematic work were found to report only limited information about the methodology used to conduct the validation and the results obtained through the validation process. Using broad definitions of recall (accuracy of the LENA system in detecting an event) and precision (accuracy in defining the event), Cristia et al. (2020) found high accuracy for AWC (13 studies, mean r = .79) and CVC (5 studies, mean r = .77) but lower accuracy for CTC, (note, however, that CTC reliability was computed on a small set of available studies, 6 studies, mean r = .36). More problematic results in the LENA *vs* human estimation of CTC, as compared to the other LENA measures, was also reported using five different corpora (AWC r = .70, CVC r = 65 and CTC r = .36; Cristia et al., 2021; see also the recent study by Ramírez, Jippe, & Kuhl, 2021).

Besides reliability, some attention has been paid to the concurrent validity of LENA measures with other language measurements by comparing the automatic LENA measures with scores from standardised language assessment tools or other direct assessments of language skills. In a recent study validating the LENA technology for Hebrew and Arabic (Levin-Asher et al., 2022), LENA's concurrent validity was tested by comparing its outputs to the PRISE questionnaire, and good concurrent validity was found between the LENA automatic scores (CTC, CVC) and such questionnaire, filled out at the same age of the recording. Finally, a meta-analytic study of 13 papers exploring if LENA measures predict later linguistic outcomes (Wang et al., 2020) showed moderate correlations between both CTC and CVC and standardised language outcomes, as well as a low correlation between AWC and the same language measures.

To the best of our knowledge, no previous study has validated the LENA system for the Italian language. Our aim was thus to fill this gap (Study 1). Additionally, we provide a concurrent validity analysis (Study 2) of the automated CVC and CTC estimates.

The current study

The objective of the present study is twofold.

Study 1 aims to establish the validity of the LENA system among 12 Italian families with children aged 1;0 (at the time of the first meeting) and 2;0 (at the time of the second meeting), by assessing whether or not there are significant relationships between the automatic CVC, AWC and CTC provided by the LENA and those provided by manual transcriptions. Based on the literature, we expected the LENA-counts and the human-counts to be significantly and strongly related for CVC and AWC, while we expected a potentially weaker association for CTC (e.g., Ramírez, Hippe, & Kuhl, 2021).

Study 2 investigates the concurrent validity of the automatic LENA measures that evaluate the children's production abilities (CVC and CTC) by comparing these measures with other direct and indirect measures of language development (the total number of vocal productions including both verbal and preverbal productions from an interaction session that has been video-recorded at the child's home; PRISE; IBQ) (note that Study 2 was conducted across a wider sample of children than Study 1: 19 children longitudinally assessed between the ages of 0;6 months and 2;0 years). We expected to find a positive relationship between automatic LENA counts and the children's vocal production, as manually tallied by considering the number of vocal tokens (i.e., the total number of vocal productions including both verbal and preverbal productions) produced in a direct naturalistic observation, and between CVC and CTC and the scores obtained in the PRISE questionnaire and in the vocal reactivity scale of the IBQ. In particular, we expected to find a significant correlation between the LENA estimates of speech (in terms of CVC) and the child's actual speech video/audio recorded during spontaneous interaction with their mothers and the PRISE scores. Moreover, we also expected to find significant links between the number of conversational turns in which the child is involved during the day, as estimated by LENA, and measures of verbal skills and vocal reactivity. There are good reasons to believe that socio-communicative or pragmatic aspects of language, which are captured by CTC, are linked to the child's expressive skills (Donnelly & Kidd, 2021; Romeo et al., 2018).

Study 1

Methods

Participants

Participants included 12 typically developing children (9 males and 3 females) recorded for 11 hours on average on the same day or on consecutive days, at both 1;0 and 2;0. We chose these two time-points because, respectively, they usually correspond to the beginning of early word production and to the more advanced phase of vocabulary extension. No parents reported developmental delays or problems at the time of their child's birth. Children's mean weight at birth was 3181 grams (*SD* = 511). All infants were born in Italy. Parents' mean years of education were 16.8 (*SD* = 2.77) for the mothers and 16.3 (*SD* = 3.98) for the fathers, broadly corresponding to 1st level degree. At the time of the first data collection (when children were 1;0) mothers were 34;7 on average (*SD* = 5.8) and fathers were 39;8 (*SD* = 7.86) on average. The families were involved in the study through local services for infants and joined the study voluntarily.

Instruments

LENA

Home language environment measures were conducted using the LENA system. The participating children wore the LENA device in a specially designed vest with a chest pocket. This vest is designed to optimise the quality of the recorded sounds (it has low friction properties) and (allowing to keep the recorder on the infant's body) to hear and measure accurately the speech produced by infants and around them. This device was specifically created to assess the child's environment in a typical day and can be used with children in the first three years of life.

Procedure

Language samples collection

Parents of children were asked to use the LENA device on one or more typical days for at least 10 hours. More specifically, on the day of the first meeting, parents were provided with the LENA, and a plasticised sheet containing the instructions for using it. Parents were asked to

switch on the device in the early morning, when the child woke up, and to switch it off after 10 hours had passed, or whenever they needed to have some privacy. If the parents decided to switch off the device before 10 hours had passed, they were asked to switch it on again, until they reached such minimum number of recording hours required. During the day of the recording, parents (or the adult staying with the child) were asked to fill in a form to track the main activities for each recorded hour. In this way, we knew in which moments the adult and the child were carrying out specific interactive activities. Parents were also asked to evaluate how typical the day was for the baby and to tell whether or not the child's speech production that day was in line with what they usually produced. The device was left to the families for a maximum of five days from the day of the visit. Families were asked to record children in natural and spontaneous situations that reflected their child's daily life (e.g. child at home with the parents during the weekend or with other caregivers during the week). For privacy reasons, they were explicitly asked not to use the device when their children were at the day-care center. Moreover, parents were asked to avoid using the recorder during special occasions (i.e., a weekend outside with friends).

As described in more depth in the section below, three samples of 10-minute speech were extrapolated for each child at each age point (1;0 and 2;0) for a total of 72 segments (720 minutes, or 12 hours, in total). The adult and child speech were transcribed independently by two native Italian speakers (two young researchers) and analysed by using the CLAN software from CHILDES (MacWhinney, 2000).

Segments selection

To select 10-minutes samples, we stuck to the following criteria: the chunks of recordings in which we observed no productions, such as silence due to naptime, were excluded; different types of activities were included (e.g., mealtime, bathtime, storytime, playtime, and time outside) and, following Gilkerson et al. (2015), different moments of the day were selected (morning, 8 am-1 pm; afternoon, 1 pm- 4 pm; late afternoon, 4 pm- 9 pm).

Transcriptions

Transcriptions of the 30-min speech samples per child and per age were done manually by two native Italian speakers using CHAT of CHILDES (Codes for the Human Analysis of

Transcripts, MacWhinney, 2000). Transcriptions were done regardless of how the speaker was tagged in the LENA system (i.e., for this reason we cannot include a validation of the speakers tags as done by Xu et al., 2009). The second transcriber was enrolled to independently transcribe 33 out of 72 transcripts, to test reliability (see paragraph below). Since the LENA system defines vocalisations by a "breath-group" criterion (Bruyneel et al., 2020), such that the vocalisation ends each time a 300 ms break occurs, we used ELAN (Version 6.0) [Computer software] (2020) to analyse the exact time in correspondence of the onset of the child's production. If the child produced reduplicated sounds CVCVCVCV or single segments CV, these counted as one vocalisation; if a pause occurred in a sequence of CV (pause > 300 ms), these counted as two vocalisations. Overlapping speech (both for the adult and child) was excluded from the analysis. Non speech sounds, such as vegetative sounds (e.g., burping, sneezing, and breathing), and fixed signals (e.g., crying and laughing) were not transcribed.

After transcribing the child's and the adult's speech, CTC were coded. A conversational turn is a sequence of speech starting from the target child to the adult (occurring within 5 sec) or vice versa. These sequences could be initiated by the child or by the adult and they counted as one CTC if they were in the form "child – adult – child" and two CTC if they were in the form "child – adult – child – adult". Each CTC was coded in CHAT using the coding string ($CTC). CTC were not counted in case of overlapping speech.

To count the number of AWC (Adult Word Count), CVC (Child Vocalisation Count) and CTC (Conversational Turns Count), the CLAN program was used with the function "freq" for speaker tier (speaker tier in CHAT are assigned with the *) and dependent tier (coding tiers in CHAT are assigned with %).

Human coder reliability

Before assessing the LENA's reliability, a reliability index of human transcribers was computed, comparing the transcriptions of the two independent transcribers (Transcriber 1 and 2 in Table 1) on a random sample of 45% transcripts (33 out of 72 10-min segments of speech). To do so, we compared the number of vocal tokens produced by the adult (AWC) and by the child (CVC) and the number of conversational turns (CTC) counted by the two coders. Pearson correlations based on these data were very strong for AWC (r = .99, p <.001), CVC (r = .95, p <.001) and CTC (r = .99, p < .001).

Table 1. *Counts by two transcribers (and difference) for Adult Words (AWC), Child Vocalisations (CVC) and Conversational Turns (CTC) for each 10-min segment*

	Adult Word Count			Child Vocalisation Count			Conversational Turns Count		
	Transcriber 1	Transcriber 2	Difference	Transcriber 1	Transcriber 2	Difference	Coder 1	Coder 2	Difference
1	601	623	-22	21	23	-2	25	24	1
2	520	562	-42	79	70	9	5	5	0
3	715	735	-20	6	6	0	17	24	-7
4	172	178	-6	50	44	6	18	18	0
5	363	369	-6	6	4	2	29	30	-1
6	471	467	4	6	2	4	10	9	1
7	276	284	-8	38	39	-1	11	13	-2
8	436	459	-23	55	56	-1	2	2	0
9	459	454	5	20	21	-1	1	1	0
10	835	913	-78	9	14	-5	1	1	0
11	325	347	-22	48	53	-5	21	27	-6
12	394	411	-17	49	56	-7	23	23	0
13	353	367	-14	10	14	-4	6	7	-1
14	169	185	-16	17	31	-14	8	9	-1
15	827	855	-28	4	5	-1	2	1	1
16	38	48	-10	37	35	2	66	67	-1
17	230	225	5	62	58	4	49	51	-2
18	303	313	-10	26	24	2	16	23	-7
19	315	331	-16	62	63	-1	11	12	-1

20	479	505	-26	17	19	-2	26	26	0
21	726	722	4	23	46	-23	8	8	0
22	400	391	9	92	74	18	22	29	-7
23	609	615	-6	72	74	-2	8	9	-1
24	659	677	-18	69	66	3	46	49	-3
25	36	36	0	84	51	33	10	11	-1
26	301	293	8	204	142	62	6	6	0
27	395	353	42	97	48	49	8	8	0
28	351	354	-3	142	145	-3	38	37	1
29	443	462	-19	112	115	-3	54	57	-3
30	314	322	-8	73	82	-9	16	15	1
31	185	192	-7	125	134	-9	3	4	-1
32	508	537	-29	179	181	-2	39	38	1
33	141	152	-11	171	178	-7	8	7	1

Data analysis

To assess the reliability (or accuracy) of the LENA system for the Italian language, comparisons between AWC, CVC and CTC estimates (LENA Pro - Graduate Version) and human coders were performed for all 72 selected 10-min chunks. Results were generated using Jamovi (Version 1.2, 2020).

In line with previous validation studies (e.g., Bruyneel et al., 2020), we conducted t-tests and Pearson correlations between: the LENA-AWC, CVC, and CTC and the human-AWC, CVC, and CTC. Correlations lower than .30 would reflect poor agreement, correlations between .30 and .50 would reflect low agreement, correlations between .50 and .70 would reflect moderate agreement, and correlations higher than .70 would reflect high agreement (Bruyneel et al., 2020).

Results and Discussion

Each child was recorded for around 11 hours (corresponding to 672 minutes on average, *SD* = 67.8 minutes) at 1;0 and for 12 hours (corresponding to 746 minutes, *SD* = 129 minutes) at 2;0. LENA estimates for AWC, CTC and CVC are reported for each child at 1;0 and 2;0 in Table 2.

Table 2. *LENA estimates for the entire recording in the group of children at 1;0 and 2;0*

	1;0			2;0		
	AWC	CTC	CVC	AWC	CTC	CVC
CHI 1	23645	351	889	23766	745	2439
CHI 2	36660	464	902	18711	184	539
CHI 3	15249	279	846	11550	104	302
CHI 4	10967	211	1081	8052	150	1023
CHI 5	16466	252	1012	5739	263	544
CHI 6	17929	423	1392	8009	249	1648
CHI 7	31326	1052	2808	15486	880	4032
CHI 8	13171	291	988	25811	572	2243
CHI 9	19369	459	1225	19958	232	1019
CHI 10	31027	397	898	28679	86	160
CHI 11	10250	367	1349	9245	352	1559
CHI 12	25184	857	2665	29314	1030	2840

Human Estimates versus LENA estimates (on 72 10-min-long segments)

In order to test the validity of LENA estimates, a series of paired samples t-tests and Pearson product-moment zero-order correlations were computed between those estimates and

Results are presented in Table 3 for the entire sample, together with the means and standard deviations.

Table 3. *Descriptive statistics, and results of the t-tests (and their ps) and correlations coefficients for the human estimates and LENA estimates for AWC, CVC, and CTC*

		Human Estimates	*LENA-Estimates*					
	N	*M (SD)*	*M (SD)*	*t*(71)	*p*	*Cohen's d*	*Mean difference*	*r*
AWC	72	494 (241)	516 (276)	1.068	0.289	.126	20.47	.783***
CVC	72	46.6 (31.9)	43.3 (25.6)	-0.951	0.345	-.112	-3.35	.478***
CTC	72	22.5 (17.3)	18.4 (9.65)	-2.077	0.041*	-.245	-4.11	.327**

Note. AWC = Adult Word Count; CVC = Child Vocalisation Count; CTC = Conversational Turns Count; *p < .05; **p <.001; ***p < .001.

For AWC, the LENA system slightly overestimated the number of words produced by the adults, if compared to the number of words transcribed by the human transcribers, as reported in Table 3. However, this difference was not statistically significant (p = .289), in line with Cristia et al. (2020), D'Apice et al. (2019) and Gilkerson et al. (2015). Pearson's correlations indicated that human counts and LENA estimates, in relation to the number of adult words, were significantly, positively and highly correlated (r = .78, p <.001). The group of children was then divided based on the child's age and correlations were run again for the two ages separately. At both 1;0 and 2;0, correlations between the number of words produced by the adults as reported by the LENA device and as transcribed by the human coder were significant, positive and high (respectively, r = .73, p <.001 at 1;0; r = .83, p <.001 at 2;0). This finding is also in line with other published studies (Busch et al., 2018; D'Apice, Latham, & von Strumm, 2019; Orena et al., 2019; Pae et al., 2016).

For CVC, the LENA system underestimated the child vocalisations, if compared to the number of vocalisations transcribed by the human transcribers, in line with Canault et al. (2016) and Cristia et al. (2020). However, this difference was not statistically significant (p = .345). Pearson's correlations indicated that human counts and LENA estimates, in relation to the number of children's vocalisations, were significantly, positively and weakly correlated (r = .47, p <.001). This is slightly weaker than what most previous studies have observed. Thus, we analysed the data again, at each of the two ages separately. At both 1;0 and 2;0, the correlations between the number of vocalisations produced by the child, as reported by the LENA device and as transcribed by the human coder, were significant, positive and moderate (respectively, r = .66, p < .001 at 1;0 ; r = .51, p = .002 at 2;0). Separate correlation values were closer to the values in the literature (e.g., Cristia et al., 2020).

For CTC, in contrast, significant differences emerged between the number of conversational turns found by human counts and by the LENA system (p = .041, in line with Busch et al., 2018; Cristia et al., 2020). Moreover, we found a low correlation between the two measures (r = .33, p = .005). At both 1;0 and 2;0, the correlations between the number of conversational turns as reported by the LENA device and as coded by the human coder were significant, positive and moderate (respectively, r = .43, p = .008 at 1;0; r = .53, p < .001 at 2;0). This weak finding is in line with what other studies have reported and this index needs to be considered with caution when automatically retrieved from LENA system (Cristia et al., 2020; Ramírez et al., 2022).

Study 2

The objective of the second study was to assess the LENA's concurrent validity against other measures of vocal production. The data analysed in this study were part of a wider longitudinal research on mother-child communication, involving children in the first two years of life. We extracted and analysed the automatic measures of fifty-four speech samples collected from longitudinal recording sessions conducted within a group of 19 children. Each recording session lasted around 12 hours (M = 711 minutes, SD = 93.0). In particular, speech samples were selected for analysis if the recording session was longer than 10 hours, and children were in the age range 0;6 months - 2;0 years.

Participants

As described above, the 54 speech samples retained for analysis came from 19 children (12 males, 7 females). As reported in Table 4, fifteen out of 19 children were recorded on several longitudinal appointments (range: 2-5 appointments, i.e., coded as a categorical variable, hereafter called "Time") between 0;6 and 2;0 years (range: 0;6 months and 11 days – 2;0 years and 27 days; M = 13. 5, SD = 5.86). Four out of 19 children only completed one recording session.

Table 4. Study 2 participants characteristics (Gender and Age (in months) at which we collected the LENA data from each child). For each child, available data were reported.

CHI	SEX	AGE T1	AGE T2	AGE T3	AGE T4	AGE T5
1	M	7 (1;2)				
2	M	6 (1)	9 (1;2)	16 (2)	24 (2)	
3	M	12				
4	M	19				
5	F	7 (1)	9 (1)	15 (2)	24 (2)	
6	M	6 (1)	9 (1)	12 (1;2)	15 (2)	24 (2)
7	F	10 (1)	12 (1;2)	16 (2)	24(2)	
8	M	6 (2)	15 (2)			
9	M	6 (1;2)	13 (2)			
10	M	6 (1;2)	8 (1;2)	13(1;2)	22(2)	
11	M	9 (2)	18(2)			
12	M	9 (1;2)	12 (1;2)	15 (2)	22 (2)	
13	M	6 (2)	9 (2)			
14	F	18 (2)	22 (2)			
15	F	11 (1)	13 (1)	15	18 (2)	
16	F	6 (1)	10 (1)			
17	M	9 (1)	12 (1,2)	21 (2)	24 (2)	
18	F	6 (1)				
19	F	11 (1)	13 (1)	15 (2)	19 (2)	23 (2)

Note. For all children at each time point we considered the automatic measures (CVC and CTC) from at least 10 hours of audio-recording (LENA) and the PRISE questionnaire. For sessions marked (1), we orthographically transcribed the child's speech produced during the mother-child interactions in the video-recording made at that age; for sessions marked (2), we collected the Vocal Reactivity Scale of the Infant Behaviour Questionnaire (Italian version).

No parents reported developmental delays or problems at the time of their child's birth. At the time of the first meeting, children's mean weight at birth was 3158 grams (*SD* = 451.4). All infants were born in Italy. Parents' mean years of education were 16.72(*SD* = 3.31) for the mothers and 15.29 (*SD* = 4.51) for the fathers, broadly corresponding to 1st level degree. Mothers age were 33.44 on average (*SD* = 3.88) and fathers were 37.55 (*SD* = 6.68) on average. The families were involved in the study through local services for infants and joined the study voluntarily.

Procedure

Each family who participated in the study was provided with a LENA device on the day of each home visit (see Procedure section of the Study 1). During this visit, the researcher provided the family with an instruction form to switch the device on/off and obtained informed consent. On the same appointment, the principal caregiver (the mother for all children) and the child were video-recorded in interaction for around 20 minutes. Then, the caregiver was asked to fill two questionnaires regarding the child's phonological and vocal development (PRISE, IBQ). Each family could keep the LENA device for a maximum of five days from the day of the visit, thus carrying out the recording in this period.

Measures

The LENA Device

An in-depth description of the tool is provided in Study 1. Around 12 hours of recordings from 54 speech samples (*M* = 711 minutes, *SD* = 93.0) were considered for the purposes of the present study.

Mother-child naturalistic interaction (video-recording)

Infants were video-recorded for around 20 minutes during spontaneous interaction with their caregiver (i.e., the mother for all participants) while playing with toys provided by the experimenter (duration of the video, *M* = 20.4, *SD* = 2.43). In each play session, four sets of toys were provided to the mothers with the aim of stimulating as many spontaneous productions as possible: 1) a food set, 2) a farm set, 3) a transport set, and 4) a nurturing set. Mothers were

asked to interact with their children as they usually do, to make the situation as natural and spontaneous as possible. The video-recordings were conducted at the infant's home, a familiar context suitable for supporting spontaneous production and reducing distractions.

Only child's speech was transcribed. In particular, children's number of vocal tokens (i.e., the total number of vocal productions including both verbal and preverbal productions) using CHAT of CHILDES (Codes for the Human Analysis of Transcripts, MacWhinney, 2000), and transcriptions were performed using the same criteria as in Study 1 (see Transcriptions paragraph from Study 1). The onset time of each production was annotated on ELAN (Version 6.0) [Computer software] (2020). Crying, vegetative sounds and shouts were not transcribed. Note that, in this second study, we were not able to estimate LENA validity concerning Adult Word Count, since we did not have any concurrent measure of comparison (i.e., no measure of adult speech).

Production of Infant Scale Evaluation (PRISE)

The Italian version of the PRISE questionnaire was provided to parents (Kishon-Rabin L. et al, 2005, adapted by Cuda et al. 2013) during each observation session (see Table 4). PRISE is a parental questionnaire that evaluates a child's preverbal skills (production of vowels, simple vocalization, babbling and words). The questionnaire is made up of 11 questions and each question can have a score from 0 to 4, based on the percentage of time children show that specific behavior (0 is never, 4 is 100% of the time, always). The maximum score is 44. Cronbach's alpha is of .87 in the Italian validation (2013) and of .88 in our sample, thus it can be considered very good.

Infant Behaviour Questionnaire (Vocal Reactivity Scale)

The IBQ-R (Italian version by Montirosso et al., 2011) is a parent-based questionnaire that measures 6 domains of the infant's temperament (activity level, soothability, fear, distress to limitations, smiling and laughter, and duration of orienting). For the present research, we only asked parents to fill the scale related to the child's 'Vocal Reactivity', which refers to the amount of vocalization exhibited by the baby in daily activities (four subscales in the Italian version; Feeding, Bathing and Dressing, Play, Daily Activities). In the Vocal Reactivity scale, parents are

asked to rate the frequency of some specific behaviour shown by their child during the last week. The scale is overall made up by 12 items; each of which have to be rated from 1 (never) to 7 (always); when an item is not applicable, it is not considered for the final score. Cronbach's alpha is of .78 (on average) in the Italian validation (2011).

Data Analysis

To test for concurrent validity, partial Pearson's correlations controlling for age (as a continuous variable) and time (as categorical variable, in terms of repeated measures, for those children having more than one observation) were run between the automatic LENA measures (CVC and CTC) and the direct and indirect language measures, respectively taken from video-recordings and from the from PRISE and IBQ questionnaires. Results were analysed using Jamovi (Version 1.2, 2020).

Results and Discussion

Correlations between LENA estimates and direct language measures (see Table 5)

Children's vocal tokens retrieved from the transcriptions of the mother-child interactions did significantly, positively correlate with the number of CVC as measured through LENA in a typical day ($r = .564$, $p < .01$). However, the number of human-retrieved tokens produced by the children during naturalistic interaction (video-recorded) did not correlate with the CTC as measured by the LENA device (Table 5). These results establish the validity of automatic LENA measurements in describing linguistic skills in terms of tokens children spontaneously produce in daily interactions, regardless of age and the repeated measure effects. The number of tokens expresses a quantitative score that can be strictly linked to the quantity of vocal production as recorded and extrapolated from LENA device (in terms of CVC). Thus, this finding suggests that the LENA device could be an extremely useful tool when the aim is to determine the quantity of speech produced in a typical day. However, we failed to find any concurrent relationship between the human-retrieved tokens produced by the children and the estimate of LENA CTC.

Correlations between LENA estimates and indirect language measures (see Table 5)

Children's PRISE scores significantly, positively, but weakly correlated with CVC as measured by the LENA device ($r = .279$, $p < .05$). Although this correlation is low, it indicates a tendency for those children scoring higher on the PRISE questionnaire to produce more

vocalisation during a typical day, in a spontaneous context. Moreover, we found a significant, positive and low correlation between the CVC and vocal reactivity during play (r = .384, p < .05); and we found a significant, positive and low correlation between the CTC and vocal reactivity during play (r = .422, p < .05) (Table 5). However, no significant relationships were found between the other sub-scales of the Vocal Reactivity Scale and the automatic outputs of the LENA system.

Taken together, these results establish the concurrent validity of LENA with spontaneous measures retrieved in a spontaneous setting and with parent-report tools for providing an estimation of the child's speech.

Table 5. Partial correlation table (controlling for the effects of age and time) showing the link between the automatic measure retrieved by the LENA device (CTC, CVC) and direct (tokens) and indirect (Prise, Vocal Reactivity – VR Scale of the IBQ) language measures.

	1	2	3	4	5	6	7	8
1. CTC	—							
2. CVC	.907***	—						
3. Child's Tokens (Video-rec)	.564**	.385	—					
4. Prise	.279*	.254	.688***	—				
5. IBQ VR Feeding	-.191	-.145	.132	.236	—			
6. IBQ VR Wash and Dress	-.113	-.230	.308	.297	.540***	—		
7. IBQ VR Play	.384*	.422*	.452	.304	.093	.105	—	
8. IBQ VR Daily Activities	-.140	-.189	.244	.380*	.547***	.721***	.305	—

Note. * p < .05, ** p < .01, *** p < .001

General discussion

In the present paper, we report about both reliability (Study 1) and concurrent validity (Study 2) of the LENA tool for a sample of Italian children aged between 0;6 months and 2;0 years. No previous study had investigated such issues in the Italian context.

As for validation of the LENA system, results for the Italian language are in line with most of the validation studies previously conducted for other languages (see Cristia, Bulgarelli, & Bergelson, 2020 or Cristia et al., 2021 for some recent reviews of the literature). They establish the reliability of the LENA device for research conducted with Italian speakers.

More specifically, regarding AWC, the degree of correlation found in our study between the LENA outcomes and the human annotations is very high (r = .78), and this holds for both the joint analysis (all ages considered together) and for analyses conducted in single age-groups (1;0 and 2;0). This result is in line with other published studies that have also reported correlation values of .79 on average (for example, r = .89, Busch et al., 2018; r = .79, D'Apice, Latham, & von Strumm, 2019; r = .77, Orena et al., 2019; r = .72, Pae et al., 2016). Also, in line with previous investigations, we found that LENA slightly, though not significantly, overestimates AWC if compared to human counts (Cristia et al., 2020; D'Apice et al., 2019; Gilkerson et al., 2015).

Regarding CVC, our data are partially in line with what most studies have found. Specifically, we found a low correlation between the LENA and human counts when the analyses were run on all ages pooled together, while other studies found a strong correlation. However, when data were analysed separately based on age subgroups (1;0 and 2;0), the degree of correlation significantly increased, and especially for the group of younger babies, in agreement with Cristia et al. (2020). Additionally, and in line with former reports, we found that LENA slightly, though not significantly, underestimates the number of CVC with respect to human counts (Canault et al., 2016; Cristia et al., 2020).

Regarding CTC, significant differences emerged between the LENA and human estimates, revealing a tendency towards underestimation by the LENA, a finding which is also in line with previous reports (Busch et al., 2018; Cristia et al., 2020). These significant differences were also confirmed by the significant but weak correlations found between the LENA and the human estimates, both in the joint analysis and for analyses conducted in single age-groups (1;0 and 2;0). This second result is also in line with other validation studies which, on average, have found a correlation power of .36 (Cristia et al., 2020), where we found a correlation of .327. Indeed, Ramírez and colleagues (2021) considered the relation between LENA's CTC estimates and human CTC estimates in a wider sample of 70 families, with children longitudinally recorded at 0;6, 0;10, 1;2, 1;6, and 2;0. Results showed that LENA CTC and human CTC are not interchangeable measures and that CTC need to be considered with caution when used as an automatically retrieved measure. Moreover, they found that automatic CTC measures were always higher than manual CTC measures. This specific result contrasts with the findings of the present study (lower *vs* higher estimates), which might be due to differences in the composition

of the samples (different ages, or different individual characteristics of adults and children involved) and/or to the characteristics of the Italian *vs* English language. At any rate, only a few studies have reported validation of the CTC (6 studies out of 33, as to Cristia et al., 2020).

The present results show that, in a sample of Italian recordings, LENA was a reliable/accurate tool for the estimation of both AWC and CVC. This is an important point, as AWC can be considered an important index of language input, being strongly correlated to the child's language outcomes (see Hart & Risley, 1995; Hoff & Naigles, 2002; Rowe, 2008, 2012; Hoareau, Yeung & Nazzi, 2018). A potential implication is that AWC automatically calculated by the LENA could be included in assessments of the risk and protective factors for child language development. At the same time, the possibility of automatically assessing a child's production (in terms of quantity) using the LENA CVC can give researchers an important index of development, especially for children with delay and special needs. In fact, many studies reported that early vocal production is related to language outcomes. For example, lexical production is an important predictor of language and learning outcomes (Baldwin, 2000; Hoff, 2013; Hoff & Naigles, 2002; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010; Weizman & Snow, 2001), while preverbal production predicts early lexical development (Majorano et al., 2014; McGillion et al., 2017). However, LENA count of CVC does not distinguish preverbal and verbal production, thus cannot allow to describe in detail the qualitative level of children' s production. Note that the AVA (Automatic Vocalisation Assessment) index could be considered in such a case, as this index evaluates the child's vocal maturity in terms of expressive language skills (by comparing the phonemic complexity of the child's output against an adult American English model). However, since it is not as commonly used as the other measures, it cannot be used for our Italian sample.

To test the concurrent validity of the LENA measures with direct and indirect linguistic outcomes, a second study was run, with a wider sample of Italian children aged between 0;6 and 2;0 years of age. To the best of our knowledge, this is the first LENA validation study using direct measures of linguistic skills (i.e., child's vocal production counted from direct video-observation) to test the concurrent validity of LENA's estimations. In addition, in line with other validation studies, we compared parent-reported measures of linguistic skills with automatic LENA measures. The findings reported in this study support the claim that LENA data are comparable to data retrieved from direct observations, conducted by a researcher on the same

week of the recording, and with data from parental questionnaires on linguistic skills. This last result, and especially the relationship between the CVC and the PRISE questionnaire, are in line with Levin-Asher and colleagues' 2022 study (showing a similar correlation index between the same variables). Although Levin-Asher's study considered Arabic (Levin-Asher et al., 2022), their results converge with our finding, in suggesting that, the more children vocalize, the higher they score on the PRISE test. Moreover, our study highlights a relationship not only with the PRISE questionnaire, but also between a specific section of the Vocal Reactivity Scale of the Infant Behaviour Questionnaire, i.e., the Vocal Reactivity Scale during play (i.e., how much the child talk when playing with the caregiver), the CVC and the CTC. Thus, the more children talk or the more they are involved in conversational turns as measured by LENA, the more they exhibit vocalisations in their daily play as reported by parents. This result is consistent with evidence showing links between quality of speech, in terms of turn taking, and the child's language skills at the same age (Ferjan Ramínez et al., 2020), or in their later language development (Donnelly et al., 2021; Romeo et al., 2021). Our finding shows that children who are more involved in conversational turns during a typical day with their main caregiver are the same who were perceived as more talkative and linguistically active in play situations from their caregivers. It is interesting to underline that the number of conversational turns is related only to parent's perception of vocal reactivity during play, not in the other situations included in the IBQ (feeding; washing and dressing; daily activities). This could be related to the child's higher vocal productivity during this kind of activity or to mother's higher focus on conversations during play. However, this is only a speculative hypothesis, since we do not have a measure to demonstrate it.

Finally, our most remarkable result regards the relationship with the data from the naturalistic observation. Concurrent validity is shown between LENA estimations (CVC) and direct naturalistic observations, i.e., analyses of linguistic skills based on video samples recording the spontaneous and ecological interaction between participants and caregivers. Importantly, this means that the LENA can be taken as a reliable and valid tool to automatically provide measures of vocal productions, thus reducing the demanding task to transcribe and analyse video observations for future studies. This point brings a concrete methodological contribution for studies examining language development, showing that data that are

automatically retrieved from the LENA device can be immediately and easily used by both researchers and experts in education to sketch out a child's language skills.

Conclusion

In summary, this study confirms that the LENA recording system is a useful, valid and reliable tool to automatically analyse some aspects of children's environment and of child-adult verbal communication. Increasing interest has emerged, in studies on language development, regarding environmental characteristics considered as important factors for developmental outcomes. The LENA system gives the possibility to easily and reliably assess, in naturalistic settings, quantitative aspects of the child's vocal production and of the linguistic input children are exposed to. Furthermore, LENA gives the possibility to collect data without the presence of the researcher, an aspect which became all the more relevant during the Covid-19 pandemic period, when direct contacts between people were limited. Another advantage is the simple use of the device that makes it adequate also for families with special needs or with low SES, and in varied contexts. However, the hardware and software of the LENA device also come with some limitations. Above all, one can consider the fact that, in the estimation of children's production skills, qualitative features of the recorded samples (e.g., indexes of phonetic or lexical diversity, as measured using token versus type ratios) are not automatically computed. In effect, one automatic LENA assessment giving qualitative information on children's production exists: the Automated Vocal Assessment (AVA), but this index has not been as commonly used as the other measures, thus it is not exploitable in the context of our study on Italian). Furthermore, since LENA counts are extracted using audio recordings, no information is reported about nonverbal communication (e.g., gestures).

The present study offers a first contribution about the validity of the LENA system with Italian children. Our results provide a positive evaluation of the device and encourage further research on the relationship between LENA automatic estimations and direct and indirect language measures. Most notably, analyses on the concurrent validity of the LENA system could be conducted in a longitudinal perspective or extended to different socio-cultural groups of participants.

References

Baldwin, D.A. (2000). Interpersonal understanding fuels knowledge acquisition. *Current Directions in Psychological Science, 9*, 40-45.

Bruyneel, E., Demurie, E., Boterberg, S., Warreyn, P., & Roeyers, H. (2020). Validation of the Language ENvironment Analysis (LENA) system for Dutch. *Journal of Child Language*, 1-27. doi: 10.1017/S0305000920000525

Bulgarelli, F., & Bergelson, E. (2019). Look who's talking: A comparison of automated and human-generated speaker tags in naturalistic daylong recordings. *Behavior Research Methods, 52*, 641-653. doi: https://doi.org/10.3758/s13428-019-01265-7

Busch, T., Sangen, A., Vanpoucke, F., & van Wieringen, A. (2018). Correlation and agreement between Language ENvironment Analysis (LENA™) and manual transcription for Dutch natural language recordings. *Behavior Research Methods, 50*(5), 1921-1932. doi: https://doi.org/10.3758/s13428-017-0960-0

Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of the Language ENvironment Analysis system (LENATM) in European French. *Behavior Research Methods, 48*(3), 1109-1124. doi: 10.3758/s13428-015-0634-8

Christakis, D. A., Gilkerson, J., Richards, J. A., Zimmerman, F. J., Garrison, M. M., Xu, D., Gray, S., & Yapanel, U. (2009). Audible television and decreased adult words, infant vocalizations, and conversational turns: a population-based study. *Archives of Pediatrics & Adolescent Medicine, 163*(6), 554-558. doi: 10.1001/archpediatrics.2009.61

Cristia, A., Bulgarelli, F., & Bergelson, E. (2020). Accuracy of the Language Environment Analysis system segmentation and metrics: A systematic review. *Journal of Speech, Language, and Hearing Research, 63*(4), 1093-1105. doi: https://doi.org/10.1044/2020_JSLHR-19- 00017

Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2021). A thorough evaluation of the Language Environment Analysis (LENA) system. *Behav Res., 53*, 467–486. https://doi.org/10.3758/s13428-020-01393-5

Cuda, D., Guerzoni, L., Mariani, V., Murri, A., Biasucci, G., & Fabrizi, E. (2013). Production of infant scale evaluation (PRISE) in Italian normal hearing children: a validation study. *International Journal of Pediatric Otorhinolaryngology, 77*(12), 1969–1974. https://doi.org/10.1016/j.ijporl.2013.09.014

D'Apice, K., Latham, R. M., & von Strumm, S. (2019). A naturalistic home observational approach to children's language, cognition, and behavior. *Developmental Psychology, 55*(7), 1414–1427. https://doi.org/10.1037/dev0000733

Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational turn- taking and vocabulary growth in early language development. Child Development, 92(2), 609–625. https://doi.org/10.1111/cdev.13511

Duncan, R.J., King, Y. A., Finders, J. K., Elicker, J., Schmitt, S. A., & Purpura, D. J. (2020). Prekindergarten classroom language environments and children's vocabulary skills. *Journal of Experimental Child Psychology, 194*, 104829–104829. https://doi.org/10.1016/j.jecp.2020.104829

ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan

Ferguson, C.A., Menn, L., & Stoel-Gammon, C. (eds.) (1992). Phonological development: Models, research, implications. Timonium, MD: York Press. Pp. xviii 693. *Phonology, 12*(1), 135-146. doi:10.1017/S0952675700002414

Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational turns and advances infant language development. *PNAS Proceedings of the National Academy of Sciences of the United States of America, 117*(7), 3484–3491. https://doi.org/10.1073/pnas.1921653117

Ganek, H. V., & Eriks-Brophy, A. (2018). A concise protocol for the validation of Language ENvironment Analysis (LENA) conversational turn counts in Vietnamese. *Communication Disorders Quarterly, 39*(2), 371-380.

Gilkerson, J., Zhang, Y., Xu, D., Richards, J. A., Xu, X., Jiang, F., Harnsberger, J., & Topping, K. (2015). Evaluating LENA System performance for Chinese: A pilot study in Shanghai. *Journal of Speech, Language, and Hearing Research, 58*(2), 445-452. doi: 10.1044/2015_JSLHR-L-14-0014

Greenwood, C. R., Thiemann-Bourque K., Walker D., Buzhardt J., & Gilkerson J. (2011). Assessing children's home language environments using automatic speech recognition technology. *Communication Disorders Quarterly, 32*, 83–92.

Greenwood, C. R., Schnitz, A. G., Irvin, D., Tsai, S. F., & Carta, J. J. (2018). Automated language environment analysis: A research synthesis. *American Journal of Speech-Language Pathology, 27*(2), 853-867. doi: 10.1044/2017_AJSLP-17-0033

Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children.* Paul H Brookes Publishing.

Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust, P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-income children's language success. *Psychological Science*, *26*(7), 1071–1083. https://doi.org/10.1177/0956797615581493

Hoareau, M., Yeung, H. H., & Nazzi, T. (2019). Infants' statistical word segmentation in an artificial language is linked to both parental speech input and reported production abilities. *Developmental science*, *22*(4), e12803. https://doi.org/10.1111/desc.12803

Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and language minority homes: implications for closing achievement gaps. *Developmental psychology*, *49*(1), 4–14. https://doi.org/10.1037/a0027238

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development, 73*(2), 418–433. https://doi.org/10.1111/1467-8624.00415

Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of variability in children's language growth. *Cognitive psychology*, *61*(4), 343–365. https://doi.org/10.1016/j.cogpsych.2010.08.002

Keren-Portnoy, T., Majorano, M., & Vihman, M. (2009). From phonetics to phonology: The emergence of first words in Italian. *Journal of Child Language, 36*(2), 235-267. doi:10.1017/S0305000908008933

Kishon-Rabin, L., Taitelbaum-Swead, R., Ezrati-Vinacour, R., & Hildesheimer, M. (2005). Prelexical vocalization in normal hearing and hearing-impaired infants before and after cochlear implantation and its relation to early auditory skills. *Ear and hearing*, *26*(4 Suppl), 17S–29S. https://doi.org/10.1097/00003446-200508001-00004

Levin-Asher, B., Segal, O., & Kishon-Rabin, L. (2022). The validity of LENA technology for assessing the linguistic environment and interactions of infants learning Hebrew and Arabic. *Behavior research methods*, Advance online publication. https://doi.org/10.3758/s13428-022-01874-9

MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription format and programs, Volume II: the database. *Comput. Linguist., 26*, 657–657. doi: 10.1162/coli.2000.26.4.657

Majorano, M., Cigala, A., & Corsano, P. (2009). Adults' and children's language in different situational contexts in Italian Nursery and Infant School. *Child Care in Practice, 15(4),* 279-297.

Majorano, M., Guidotti, L., Guerzoni, L., Murri, A., Morelli, M., Cuda, D., & Lavelli, M. (2018). Spontaneous language production of Italian children with cochlear implants and their mothers in two interactive contexts. *International Journal of Language & Communication Disorders, 53*(1), 70–84. https://doi.org/10.1111/1460-6984.12327

Majorano, M., Brondino, M., Morelli, M., Ferrari, R., Lavelli, M., Guerzoni, L., Cuda, D., & Persici, V. (2020). Preverbal Production and Early Lexical Development in Children With Cochlear Implants: A Longitudinal Study Following Pre-implanted Children Until 12 Months After Cochlear Implant Activation. *Frontiers in Psychology, 11,* 591584–591584. https://doi.org/10.3389/fpsyg.2020.591584

Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants' production experience and their processing of speech. *Language Learning and Development, 10*(2), 179–204. https://doi.org/10.1080/15475441.2013.829740

McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren- Portnoy, T., & Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive Value of Babble, Pointing, and Socioeconomic Status. *Child Development, 88*(1), 156–166. https://doi.org/10.1111/cdev.12671

MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription format and programs, Volume II: the database. *Comput. Linguist., 26*, 657–657. doi: 10.1162/coli.2000.26.4.657

McDonald, M., Kwon, T., Kim, H., Lee, Y., & Ko, E.-S. (2021). Evaluating the Language ENvironment Analysis system for Korean. *Journal of Speech, Language, and Hearing Research, 64,* 792-808. doi: https://doi.org/10.1044/2020_JSLHR-20-00489

Montirosso, R., Cozzi, P., Putnam, S., Gartstein, M., & Borgatti, R. (2011). Studying crosscultural differences in temperament in the rst year of life: United States and Italy. *International Journal of Behavioral Development, 35*(1), 27–37.

Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in infancy: the prediction of speech and language disorders. *Journal of communication disorders, 32*(4), 223–245. https://doi.org/10.1016/s0021-9924(99)00013-1

Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019). Reliability of the Language Environment Analysis recording system in analyzing French- English bilingual speech. *Journal of Speech, Language, and Hearing Research, 62*(7), 2491-2500. doi: https://doi.org/10.1044/2019_JSLHR-L-18-0342

Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J. A., Ma, L., & Topping, K. (2016). Effects of feedback on parent-child language with infants and toddlers in Korea. F*irst Language, 36*(6), 549-569. doi: 10.1177/0142723716649273

Ramírez, N. F., Hippe, D.S, & Kuhl, P. K. (2021). Comparing automatic and manual measures of parent-infant conversational turns: A word of caution. *Child Development, 92*(2), 672-681. doi: https://doi.org/10.1111/cdev.13495

Richards, J. A., Gilkerson, J., Xu, D., & Topping, K. (2017). How much do parents think they talk to their child? *Journal of Early Intervention, 39*(3), 163-179. doi: 10.1177/1053815117714567

Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., & Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological Science, 29*(5), 700–710. https://doi.org/10.1177/0956797617742725

Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development, 52*, 569–578.

Rowe M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of child language, 35*(1), 185–205. https://doi.org/10.1017/s0305000907008343

Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-directed speech vocabulary development. *Child Development, 83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Schwarz, I.C., Botros, N., Lord, A., Marcusson, A., Tidelius, H., & Marklund, E. (2017). The LENA^{TM} system applied to Swedish: Reliability of the adult word count estimate. *Proceedings of Interspeech 2017*, 2088-2092. doi: 10.21437/Interspeech.2017-1287

Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and grammatical units in speech to two preverbal infants. *Journal of Child Language, 35*(4), 869–902. https://doi.org/10.1017/S0305000908008763

VanDam, M., Ambrose, S. E., & Moeller, M. P. (2012). Quantity of parental language in the home environments of hard-of-hearing 2-year-olds. *Journal of deaf studies and deaf education, 17*(4), 402–420. https://doi.org/10.1093/deafed/ens025

Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly & M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp. 69–84). Hillsdale, NJ: Erlbaum.

Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics, 21*(1-2), 61–82.

Wang, Y., Williams, R., Dilley, L., & Houston, D. M. (2020). A meta-analysis of the predictability of LENA™ automated measures for child language development. *Developmental review, 57*, 100921. https://doi.org/10.1016/j.dr.2020.100921

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science, 24*(11), 2143-2152. doi: 10.1177/0956797613488145

Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental Psychology, 37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265

Xu, D., Yapanel, U., & Gray, S. (2009). Reliability of the LENA^{TM} language environment analysis system in young children's natural home environment. *Boulder, CO: LENA Foundation.*