



**HAL**  
open science

# 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.

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

3

4 **The Language ENvironment Analysis system (LENA):**

5 **A validation study with Italian-learning children**

6 **BASTIANELLO Tamara<sup>1,2</sup>, LORENZINI Irene<sup>3</sup>, NAZZI Thierry<sup>3,4</sup>, MAJORANO**

7 **Marinella<sup>1</sup>**

8 <sup>1</sup> Department of Human Sciences, University of Verona, Italy

9 <sup>2</sup> Department of Developmental Psychology and Socialisation, University of Padua, Italy

10 <sup>3</sup> Integrative Neuroscience and Cognition Center, Université Paris Descartes, Sorbonne Paris  
11 Cité, Paris, France

12 <sup>4</sup> CNRS (Integrative Neuroscience and Cognition Center, UMR 8002), Paris, France

13 **Author Note**

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

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

16 We have no known conflict of interest to disclose.

17 **Acknowledgment**

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

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

22 **Correspondence** concerning this article should be addressed to Tamara Bastianello,

23 Department of Human Sciences, University of Verona

24 Email: tamara.bastianello@univr.it

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

3

4   **Abstract**

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

17

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

19

20

21

22

23

24

25

26

27

28

29

30

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

**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

1 transcriptions and analyses. Furthermore, reliability assessments across trained transcribers are a  
2 critical element and require additional work. Despite these drawbacks, direct observations of  
3 speech production are extremely useful to extrapolate measures of preverbal productions  
4 (Majorano et al., 2018; Majorano et al., 2020). Moreover, they can also be integrated with  
5 standardised parent-report measures, such as the PRISE questionnaire (i.e., parent report measure  
6 of prelexical vocalizations; Kishon-Rabin et al., 2004, Italian version by Cuda et al., 2013) or the  
7 Infant Behavior Questionnaire – especially the Vocal Reactivity scale (first version, Rothbart,  
8 1981; Italian version by Montirosso et al., 2011). The use of these tools provides an immediate  
9 indirect measure of expressive skills in children in the first two years of life.

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

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

## LENA ITALIAN VALIDATION

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

8 LENA could be a useful research tool, as it allows automatic calculation of input and  
9 production language measures on large time windows. However, before using it for research  
10 purposes, one would need, first, to establish that the given automatic measures are accurate (that  
11 is, reliable), by comparing automatic outputs with hand-coded measures; second, that LENA  
12 measures have concurrent validity, by comparing the LENA output with other assessments, such  
13 as standardized parental questionnaires. Regarding reliability (or accuracy), LENA has been  
14 validated for several languages, through the systematic comparison between the device’s  
15 automated coding and human transcriptions (Bulgarelli & Bergelson, 2019; Christakis, et al.,  
16 2009; Cristia, et al., 2021; Richards et al., 2017). In particular, validation data have been  
17 published for: American English (Xu, Yapanel, & Gray, 2009), European French (Canault et al.,  
18 2016), Dutch (Bruyneel et al., 2020; Busch et al., 2018), Vietnamese (Ganek & Eriks-Brophy,  
19 2018) Chinese (Gilkerson et al., 2015), Korean (McDonald, Kwon, Kim, Lee, & Ko, 2021),  
20 Swedish (Schwarz et al., 2017), Hebrew and Arabic (Levin-Asher et al., 2022), and on data  
21 collected from children growing up in a bilingual French–English environment (Orena, Byers-  
22 Heinlein, & Polka, 2019). In addition, accuracy measures have also been provided for Spanish  
23 (Weisleder & Fernald, 2013). However, that latter study cannot be considered an official  
24 validation; indeed, “validation-like” data provided by this study were comprised in an  
25 investigation on linguistic input and expressive linguistic skills in families with low socio-  
26 economic status. In particular, since LENA had not been completely validated for Spanish, these  
27 researchers conducted a small-scale validation analysis for the AWC, based on 60-minute  
28 samples taken from 10 recordings. Results showed high correlation between word counts from  
29 human transcribers and the automatic LENA estimates (AWC).  
30 This small piece of evidence is relevant to the usage of LENA with Italian participants, given the  
31 similarity between the two languages. However, the conclusions that can be drawn are clearly

1 very limited. Thus, LENA cannot yet be used with reliability to study the Italian population.  
2 Moreover, and independently from the language specifically targeted, most of the previously  
3 conducted studies have validated the AWC and CVC speech measures, while CTC measures  
4 were only validated in Dutch, Chinese, Korean, and Vietnamese (Busch et al., 2018; Ganek &  
5 Eriks-Brophy, 2018; Gilkerson et al., 2015; Pae, Yoon, Seol, Gilkerson, Richards, Ma &  
6 Topping, 2016). Importantly, a recent literature review based on 33 studies reporting LENA-  
7 based accuracy measures (Cristia, Bulgarelli, & Bergelson, 2020) has revealed that only some  
8 studies (25 out of 33) provided validity estimates. Furthermore, most of the studies included in  
9 Cristia and colleagues's systematic work were found to report only limited information about the  
10 methodology used to conduct the validation and the results obtained through the validation  
11 process. Using broad definitions of recall (accuracy of the LENA system in detecting an event)  
12 and precision (accuracy in defining the event), Cristia et al. (2020) found high accuracy for  
13 AWC (13 studies, mean  $r = .79$ ) and CVC (5 studies, mean  $r = .77$ ) but lower accuracy for CTC,  
14 (note, however, that CTC reliability was computed on a small set of available studies, 6 studies,  
15 mean  $r = .36$ ). More problematic results in the LENA vs human estimation of CTC, as compared  
16 to the other LENA measures, was also reported using five different corpora (AWC  $r = .70$ , CVC  
17  $r = .65$  and CTC  $r = .36$ ; Cristia et al., 2021; see also the recent study by Ramírez, Jippe, & Kuhl,  
18 2021).

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

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

**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**



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31

## 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 1<sup>st</sup> 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

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

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

21

### 22 *Segments selection*

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

28

### 29 *Transcriptions*

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

1 Transcripts, MacWhinney, 2000). Transcriptions were done regardless of how the speaker was  
2 tagged in the LENA system (i.e., for this reason we cannot include a validation of the speakers  
3 tags as done by Xu et al., 2009). The second transcriber was enrolled to independently transcribe  
4 33 out of 72 transcripts, to test reliability (see paragraph below). Since the LENA system defines  
5 vocalisations by a “breath-group” criterion (Bruyneel et al., 2020), such that the vocalisation  
6 ends each time a 300 ms break occurs, we used ELAN (Version 6.0) [Computer software] (2020)  
7 to analyse the exact time in correspondence of the onset of the child’s production. If the child  
8 produced reduplicated sounds CVCVCVCV or single segments CV, these counted as one  
9 vocalisation; if a pause occurred in a sequence of CV (pause > 300 ms), these counted as two  
10 vocalisations. Overlapping speech (both for the adult and child) was excluded from the analysis.  
11 Non speech sounds, such as vegetative sounds (e.g., burping, sneezing, and breathing), and fixed  
12 signals (e.g., crying and laughing) were not transcribed.

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

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

23

#### 24 ***Human coder reliability***

25         Before assessing the LENA’s reliability, a reliability index of human transcribers was  
26 computed, comparing the transcriptions of the two independent transcribers (Transcriber 1 and 2  
27 in Table 1) on a random sample of 45% transcripts (33 out of 72 10-min segments of speech). To  
28 do so, we compared the number of vocal tokens produced by the adult (AWC) and by the child  
29 (CVC) and the number of conversational turns (CTC) counted by the two coders. Pearson  
30 correlations based on these data were very strong for AWC ( $r = .99, p < .001$ ), CVC ( $r = .95, p$   
31  $< .001$ ) and CTC ( $r = .99, p < .001$ ).

# LENA ITALIAN VALIDATION

1  
2  
3  
4

**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

## LENA ITALIAN VALIDATION

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

---

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15

### 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

## LENA ITALIAN VALIDATION

1  
 2 Each child was recorded for around 11 hours (corresponding to 672 minutes on average,  
 3  $SD = 67.8$  minutes) at 1;0 and for 12 hours (corresponding to 746 minutes,  $SD = 129$  minutes) at  
 4 2;0. LENA estimates for AWC, CTC and CVC are reported for each child at 1;0 and 2;0 in Table  
 5 2.

6  
 7 **Table 2.** *LENA estimates for the entire recording in the group of children at 1;0 and 2;0*  
 8

	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

9  
 10 **Human Estimates versus LENA estimates (on 72 10-min-long segments)**  
 11 In order to test the validity of LENA estimates, a series of paired samples t-tests and  
 12 Pearson product-moment zero-order correlations were computed between those estimates and

## LENA ITALIAN VALIDATION

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

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

		<i>Human Estimates</i>	<i>LENA- Estimates</i>					
	<i>N</i>	<i>M (SD)</i>	<i>M (SD)</i>	<i>t(71)</i>	<i>p</i>	<i>Cohen's d</i>	<i>Mean difference</i>	<i>r</i>
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**

7  
 8 *Note.* AWC = Adult Word Count; CVC = Child Vocalisation Count; CTC = Conversational Turns Count; \* $p < .05$ ;  
 9 \*\* $p < .001$ ; \*\*\* $p < .001$ .

10  
 11 For AWC, the LENA system slightly overestimated the number of words produced by the  
 12 adults, if compared to the number of words transcribed by the human transcribers, as reported in  
 13 Table 3. However, this difference was not statistically significant ( $p = .289$ ), in line with Cristia  
 14 et al. (2020), D'Apice et al. (2019) and Gilkerson et al. (2015). Pearson's correlations indicated  
 15 that human counts and LENA estimates, in relation to the number of adult words, were  
 16 significantly, positively and highly correlated ( $r = .78$ ,  $p < .001$ ). The group of children was then  
 17 divided based on the child's age and correlations were run again for the two ages separately. At  
 18 both 1;0 and 2;0, correlations between the number of words produced by the adults as reported  
 19 by the LENA device and as transcribed by the human coder were significant, positive and high  
 20 (respectively,  $r = .73$ ,  $p < .001$  at 1;0;  $r = .83$ ,  $p < .001$  at 2;0). This finding is also in line with  
 21 other published studies (Busch et al., 2018; D'Apice, Latham, & von Strumm, 2019; Orena et al.,  
 22 2019; Pae et al., 2016).

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

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

21  
22

## Study 2

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

31



## 1 Participants

2 As described above, the 54 speech samples retained for analysis came from 19 children  
 3 (12 males, 7 females). As reported in Table 4, fifteen out of 19 children were recorded on  
 4 several longitudinal appointments (range: 2-5 appointments, i.e., coded as a categorical variable,  
 5 hereafter called “Time”) between 0;6 and 2;0 years (range: 0;6 months and 11 days – 2;0 years  
 6 and 27 days; M = 13. 5, SD = 5.86). Four out of 19 children only completed one recording  
 7 session.

8  
 9 Table 4. Study 2 participants characteristics (Gender and Age (in months) at which we  
 10 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 <sup>(1;2)</sup>				
2	M	6 <sup>(1)</sup>	9 <sup>(1;2)</sup>	16 <sup>(2)</sup>	24 <sup>(2)</sup>	
3	M	12				
4	M	19				
5	F	7 <sup>(1)</sup>	9 <sup>(1)</sup>	15 <sup>(2)</sup>	24 <sup>(2)</sup>	
6	M	6 <sup>(1)</sup>	9 <sup>(1)</sup>	12 <sup>(1;2)</sup>	15 <sup>(2)</sup>	24 <sup>(2)</sup>
7	F	10 <sup>(1)</sup>	12 <sup>(1;2)</sup>	16 <sup>(2)</sup>	24 <sup>(2)</sup>	
8	M	6 <sup>(2)</sup>	15 <sup>(2)</sup>			
9	M	6 <sup>(1;2)</sup>	13 <sup>(2)</sup>			
10	M	6 <sup>(1;2)</sup>	8 <sup>(1;2)</sup>	13 <sup>(1;2)</sup>	22 <sup>(2)</sup>	
11	M	9 <sup>(2)</sup>	18 <sup>(2)</sup>			
12	M	9 <sup>(1;2)</sup>	12 <sup>(1;2)</sup>	15 <sup>(2)</sup>	22 <sup>(2)</sup>	
13	M	6 <sup>(2)</sup>	9 <sup>(2)</sup>			
14	F	18 <sup>(2)</sup>	22 <sup>(2)</sup>			
15	F	11 <sup>(1)</sup>	13 <sup>(1)</sup>	15	18 <sup>(2)</sup>	
16	F	6 <sup>(1)</sup>	10 <sup>(1)</sup>			
17	M	9 <sup>(1)</sup>	12 <sup>(1;2)</sup>	21 <sup>(2)</sup>	24 <sup>(2)</sup>	
18	F	6 <sup>(1)</sup>				
19	F	11 <sup>(1)</sup>	13 <sup>(1)</sup>	15 <sup>(2)</sup>	19 <sup>(2)</sup>	23 <sup>(2)</sup>

11 Note. For all children at each time point we considered the automatic measures (CVC and CTC) from at least  
 12 10 hours of audio-recording (LENA) and the PRISE questionnaire. For sessions marked <sup>(1)</sup>, we  
 13 orthographically transcribed the child’s speech produced during the mother-child interactions in the video-  
 14 recording made at that age; for sessions marked <sup>(2)</sup>, we collected the Vocal Reactivity Scale of the Infant  
 15 Behaviour Questionnaire (Italian version).

16

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

8

### 9 **Procedure**

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

18

19

### Measures

20

#### 21 ***The LENA Device***

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

25

#### 26 ***Mother-child naturalistic interaction (video-recording)***

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

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

4  
5 Only child's speech was transcribed. In particular, children's number of vocal tokens  
6 (i.e., the total number of vocal productions including both verbal and preverbal productions)  
7 using CHAT of CHILDES (Codes for the Human Analysis of Transcripts, MacWhinney, 2000),  
8 and transcriptions were performed using the same criteria as in Study 1 (see Transcriptions  
9 paragraph from Study 1). The onset time of each production was annotated on ELAN (Version  
10 6.0) [Computer software] (2020). Crying, vegetative sounds and shouts were not transcribed.  
11 Note that, in this second study, we were not able to estimate LENA validity concerning Adult  
12 Word Count, since we did not have any concurrent measure of comparison (i.e., no measure of  
13 adult speech).

14  
15 ***Production of Infant Scale Evaluation (PRISE)***

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

24  
25 ***Infant Behaviour Questionnaire (Vocal Reactivity Scale)***

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

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

## 6 **Data Analysis**

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

## 14 **Results and Discussion**

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

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

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

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

1 vocalisation during a typical day, in a spontaneous context. Moreover, we found a significant,  
 2 positive and low correlation between the CVC and vocal reactivity during play ( $r = .384, p <$   
 3  $.05$ ); and we found a significant, positive and low correlation between the CTC and vocal  
 4 reactivity during play ( $r = .422, p < .05$ ) (Table 5). However, no significant relationships were  
 5 found between the other sub-scales of the Vocal Reactivity Scale and the automatic outputs of  
 6 the LENA system.

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

10

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

14

	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	—

15 Note. \*  $p < .05$ , \*\*  $p < .01$ , \*\*\*  $p < .001$

16

17

18

### General discussion

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

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

1 More specifically, regarding AWC, the degree of correlation found in our study between  
2 the LENA outcomes and the human annotations is very high ( $r = .78$ ), and this holds for both the  
3 joint analysis (all ages considered together) and for analyses conducted in single age-groups (1;0  
4 and 2;0). This result is in line with other published studies that have also reported correlation  
5 values of .79 on average (for example,  $r = .89$ , Busch et al., 2018;  $r = .79$ , D'Apice, Latham, &  
6 von Strumm, 2019;  $r = .77$ , Orena et al., 2019;  $r = .72$ , Pae et al., 2016). Also, in line with  
7 previous investigations, we found that LENA slightly, though not significantly, overestimates  
8 AWC if compared to human counts (Cristia et al., 2020; D'Apice et al., 2019; Gilkerson et al.,  
9 2015).

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

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

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

4         The present results show that, in a sample of Italian recordings, LENA was a  
5 reliable/accurate tool for the estimation of both AWC and CVC. This is an important point, as  
6 AWC can be considered an important index of language input, being strongly correlated to the  
7 child's language outcomes (see Hart & Risley, 1995; Hoff & Naigles, 2002; Rowe, 2008, 2012;  
8 Hoareau, Yeung & Nazzi, 2018). A potential implication is that AWC automatically calculated  
9 by the LENA could be included in assessments of the risk and protective factors for child  
10 language development. At the same time, the possibility of automatically assessing a child's  
11 production (in terms of quantity) using the LENA CVC can give researchers an important index  
12 of development, especially for children with delay and special needs. In fact, many studies  
13 reported that early vocal production is related to language outcomes. For example, lexical  
14 production is an important predictor of language and learning outcomes (Baldwin, 2000; Hoff,  
15 2013; Hoff & Naigles, 2002; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010;  
16 Weizman & Snow, 2001), while preverbal production predicts early lexical development  
17 (Majorano et al., 2014; McGillion et al., 2017). However, LENA count of CVC does not  
18 distinguish preverbal and verbal production, thus cannot allow to describe in detail the  
19 qualitative level of children's production. Note that the AVA (Automatic Vocalisation  
20 Assessment) index could be considered in such a case, as this index evaluates the child's vocal  
21 maturity in terms of expressive language skills (by comparing the phonemic complexity of the  
22 child's output against an adult American English model). However, since it is not as commonly  
23 used as the other measures, it cannot be used for our Italian sample.

24         To test the concurrent validity of the LENA measures with direct and indirect linguistic  
25 outcomes, a second study was run, with a wider sample of Italian children aged between 0;6 and  
26 2;0 years of age. To the best of our knowledge, this is the first LENA validation study using  
27 direct measures of linguistic skills (i.e., child's vocal production counted from direct video-  
28 observation) to test the concurrent validity of LENA's estimations. In addition, in line with other  
29 validation studies, we compared parent-reported measures of linguistic skills with automatic  
30 LENA measures. The findings reported in this study support the claim that LENA data are  
31 comparable to data retrieved from direct observations, conducted by a researcher on the same

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

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



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

3

#### 4 **Conclusion**

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

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

30

31

## LENA ITALIAN VALIDATION

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30

### **References**

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

- 1 Bruyneel, E., Demurie, E., Boterberg, S., Warreyn, P., & Roeyers, H. (2020). Validation of the  
 2 Language ENvironment Analysis (LENA) system for Dutch. *Journal of Child Language*,  
 3 1-27. doi: 10.1017/S0305000920000525
- 4 Bulgarelli, F., & Bergelson, E. (2019). Look who's talking: A comparison of automated and  
 5 human-generated speaker tags in naturalistic daylong recordings. *Behavior Research*  
 6 *Methods*, 52, 641-653. doi: <https://doi.org/10.3758/s13428-019-01265-7>
- 7 Busch, T., Sangen, A., Vanpoucke, F., & van Wieringen, A. (2018). Correlation and agreement  
 8 between Language ENvironment Analysis (LENA™) and manual transcription for Dutch  
 9 natural language recordings. *Behavior Research Methods*, 50(5), 1921-1932. doi:  
 10 <https://doi.org/10.3758/s13428-017-0960-0>
- 11 Canault, M., Le Normand, M. T., Foudil, S., Loundon, N., & Thai-Van, H. (2016). Reliability of  
 12 the Language ENvironment Analysis system (LENATM) in European French. *Behavior*  
 13 *Research Methods*, 48(3), 1109-1124. doi: 10.3758/s13428-015-0634-8
- 14 Christakis, D. A., Gilkerson, J., Richards, J. A., Zimmerman, F. J., Garrison, M. M., Xu, D.,  
 15 Gray, S., & Yapanel, U. (2009). Audible television and decreased adult words, infant  
 16 vocalizations, and conversational turns: a population-based study. *Archives of Pediatrics*  
 17 *& Adolescent Medicine*, 163(6), 554-558. doi: 10.1001/archpediatrics.2009.61
- 18 Cristia, A., Bulgarelli, F., & Bergelson, E. (2020). Accuracy of the Language Environment  
 19 Analysis system segmentation and metrics: A systematic review. *Journal of Speech,*  
 20 *Language, and Hearing Research*, 63(4), 1093-1105. doi:  
 21 [https://doi.org/10.1044/2020\\_JSLHR-19-00017](https://doi.org/10.1044/2020_JSLHR-19-00017)
- 22 Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., &  
 23 Bergelson, E. (2021). A thorough evaluation of the Language Environment Analysis  
 24 (LENA) system. *Behav Res.*, 53, 467–486. <https://doi.org/10.3758/s13428-020-01393-5>
- 25 Cuda, D., Guerzoni, L., Mariani, V., Murri, A., Biasucci, G., & Fabrizi, E. (2013). Production of  
 26 infant scale evaluation (PRISE) in Italian normal hearing children: a validation study.  
 27 *International Journal of Pediatric Otorhinolaryngology*, 77(12), 1969–1974.  
 28 <https://doi.org/10.1016/j.ijporl.2013.09.014>
- 29 D'Apice, K., Latham, R. M., & von Strumm, S. (2019). A naturalistic home observational  
 30 approach to children's language, cognition, and behavior. *Developmental Psychology*,  
 31 55(7), 1414–1427. <https://doi.org/10.1037/dev0000733>

- 1 Donnelly, S., & Kidd, E. (2021). The longitudinal relationship between conversational  
 2 turn- taking and vocabulary growth in early language development. *Child Development*,  
 3 92(2), 609–625. <https://doi.org/10.1111/cdev.13511>
- 4 Duncan, R.J., King, Y. A., Finders, J. K., Elicker, J., Schmitt, S. A., & Purpura, D. J. (2020).  
 5 Prekindergarten classroom language environments and children’s vocabulary skills.  
 6 *Journal of Experimental Child Psychology*, 194, 104829–104829.  
 7 <https://doi.org/10.1016/j.jecp.2020.104829>
- 8 ELAN (Version 6.0) [Computer software]. (2020). Nijmegen: Max Planck Institute for  
 9 Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- 10 Ferguson, C.A., Menn, L., & Stoel-Gammon, C. (eds.) (1992). Phonological development:  
 11 Models, research, implications. Timonium, MD: York Press. Pp. xviii  
 12 693. *Phonology*, 12(1), 135-146. doi:10.1017/S0952675700002414
- 13 Ferjan Ramírez, N., Lytle, S. R., & Kuhl, P. K. (2020). Parent coaching increases conversational  
 14 turns and advances infant language development. *PNAS Proceedings of the National*  
 15 *Academy of Sciences of the United States of America*, 117(7), 3484–  
 16 3491. <https://doi.org/10.1073/pnas.1921653117>
- 17 Ganek, H. V., & Eriks-Brophy, A. (2018). A concise protocol for the validation of Language  
 18 ENvironment Analysis (LENA) conversational turn counts in Vietnamese.  
 19 *Communication Disorders Quarterly*, 39(2), 371-380.
- 20 Gilkerson, J., Zhang, Y., Xu, D., Richards, J. A., Xu, X., Jiang, F., Harnsberger, J., & Topping,  
 21 K. (2015). Evaluating LENA System performance for Chinese: A pilot study in  
 22 Shanghai. *Journal of Speech, Language, and Hearing Research*, 58(2), 445-452. doi:  
 23 10.1044/2015\_JSLHR-L-14-0014
- 24 Greenwood, C. R., Thiemann-Bourque K., Walker D., Buzhardt J., & Gilkerson J.  
 25 (2011). Assessing children's home language environments using automatic speech  
 26 recognition technology. *Communication Disorders Quarterly*, 32, 83–92.
- 27 Greenwood, C. R., Schnitz, A. G., Irvin, D., Tsai, S. F., & Carta, J. J. (2018). Automated  
 28 language environment analysis: A research synthesis. *American Journal of Speech-*  
 29 *Language Pathology*, 27(2), 853-867. doi: 10.1044/2017\_AJSLP-17-0033
- 30 Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young*  
 31 *American children*. Paul H Brookes Publishing.

- 1 Hirsh-Pasek, K., Adamson, L. B., Bakeman, R., Owen, M. T., Golinkoff, R. M., Pace, A., Yust,  
 2 P. K. S., & Suma, K. (2015). The contribution of early communication quality to low-  
 3 income children’s language success. *Psychological Science*, 26(7), 1071–  
 4 1083. <https://doi.org/10.1177/0956797615581493>
- 5 Hoareau, M., Yeung, H. H., & Nazzi, T. (2019). Infants' statistical word segmentation in an  
 6 artificial language is linked to both parental speech input and reported production  
 7 abilities. *Developmental science*, 22(4), e12803. <https://doi.org/10.1111/desc.12803>
- 8 Hoff, E. (2013). Interpreting the early language trajectories of children from low-SES and  
 9 language minority homes: implications for closing achievement gaps. *Developmental*  
 10 *psychology*, 49(1), 4–14. <https://doi.org/10.1037/a0027238>
- 11 Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*,  
 12 73(2), 418–433. <https://doi.org/10.1111/1467-8624.00415>
- 13 Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of  
 14 variability in children's language growth. *Cognitive psychology*, 61(4), 343–365.  
 15 <https://doi.org/10.1016/j.cogpsych.2010.08.002>
- 16 Keren-Portnoy, T., Majorano, M., & Vihman, M. (2009). From phonetics to phonology: The  
 17 emergence of first words in Italian. *Journal of Child Language*, 36(2), 235-267.  
 18 doi:10.1017/S0305000908008933
- 19 Kishon-Rabin, L., Taitelbaum-Swead, R., Ezrati-Vinacour, R., & Hildesheimer, M. (2005).  
 20 Prelexical vocalization in normal hearing and hearing-impaired infants before and after  
 21 cochlear implantation and its relation to early auditory skills. *Ear and hearing*, 26(4  
 22 Suppl), 17S–29S. <https://doi.org/10.1097/00003446-200508001-00004>
- 23 Levin-Asher, B., Segal, O., & Kishon-Rabin, L. (2022). The validity of LENA technology for  
 24 assessing the linguistic environment and interactions of infants learning Hebrew and  
 25 Arabic. *Behavior research methods*, Advance online publication.  
 26 <https://doi.org/10.3758/s13428-022-01874-9>
- 27 MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription  
 28 format and programs, Volume II: the database. *Comput. Linguist.*, 26, 657–657. doi:  
 29 10.1162/coli.2000.26.4.657

- 1 Majorano, M., Cigala, A., & Corsano, P. (2009). Adults' and children's language in different  
2 situational contexts in Italian Nursery and Infant School. *Child Care in Practice, 15*(4),  
3 279-297.
- 4 Majorano, M., Guidotti, L., Guerzoni, L., Murri, A., Morelli, M., Cuda, D., & Lavelli, M. (2018).  
5 Spontaneous language production of Italian children with cochlear implants and their  
6 mothers in two interactive contexts. *International Journal of Language &*  
7 *Communication Disorders, 53*(1), 70–84. <https://doi.org/10.1111/1460-6984.12327>
- 8 Majorano, M., Brondino, M., Morelli, M., Ferrari, R., Lavelli, M., Guerzoni, L., Cuda, D., &  
9 Persici, V. (2020). Preverbal Production and Early Lexical Development in Children  
10 With Cochlear Implants: A Longitudinal Study Following Pre-implanted Children Until  
11 12 Months After Cochlear Implant Activation. *Frontiers in Psychology, 11*, 591584–  
12 591584. <https://doi.org/10.3389/fpsyg.2020.591584>
- 13 Majorano, M., Vihman, M. M., & DePaolis, R. A. (2014). The relationship between infants'  
14 production experience and their processing of speech. *Language Learning and*  
15 *Development, 10*(2), 179–204. <https://doi.org/10.1080/15475441.2013.829740>
- 16 McGillion, M., Herbert, J. S., Pine, J., Vihman, M., dePaolis, R., Keren- Portnoy, T., &  
17 Matthews, D. (2017). What Paves the Way to Conventional Language? The Predictive  
18 Value of Babble, Pointing, and Socioeconomic Status. *Child Development, 88*(1), 156–  
19 166. <https://doi.org/10.1111/cdev.12671>
- 20 MacWhinney, B. (2000). The CHILDES project: tools for analysing talk: volume I: transcription  
21 format and programs, Volume II: the database. *Comput. Linguist., 26*, 657–657. doi:  
22 10.1162/coli.2000.26.4.657
- 23 McDonald, M., Kwon, T., Kim, H., Lee, Y., & Ko, E.-S. (2021). Evaluating the Language  
24 ENvironment Analysis system for Korean. *Journal of Speech, Language, and Hearing*  
25 *Research, 64*, 792-808. doi: [https://doi.org/10.1044/2020\\_JSLHR-20-00489](https://doi.org/10.1044/2020_JSLHR-20-00489)
- 26 Montirosso, R., Cozzi, P., Putnam, S., Gartstein, M., & Borgatti, R. (2011). Studying  
27 crosscultural differences in temperament in the first year of life: United States and Italy.  
28 *International Journal of Behavioral Development, 35*(1), 27–37.
- 29 Oller, D. K., Eilers, R. E., Neal, A. R., & Schwartz, H. K. (1999). Precursors to speech in  
30 infancy: the prediction of speech and language disorders. *Journal of communication*  
31 *disorders, 32*(4), 223–245. [https://doi.org/10.1016/s0021-9924\(99\)00013-1](https://doi.org/10.1016/s0021-9924(99)00013-1)

- 1 Orena, A. J., Byers-Heinlein, K., & Polka, L. (2019). Reliability of the Language Environment  
 2 Analysis recording system in analyzing French- English bilingual speech. *Journal of*  
 3 *Speech, Language, and Hearing Research*, 62(7), 2491-2500. doi:  
 4 [https://doi.org/10.1044/2019\\_JSLHR-L-18-0342](https://doi.org/10.1044/2019_JSLHR-L-18-0342)
- 5 Pae, S., Yoon, H., Seol, A., Gilkerson, J., Richards, J. A., Ma, L., & Topping, K. (2016). Effects  
 6 of feedback on parent-child language with infants and toddlers in Korea. *First Language*,  
 7 36(6), 549-569. doi: 10.1177/0142723716649273
- 8 Ramírez, N. F., Hippe, D.S, & Kuhl, P. K. (2021). Comparing automatic and manual measures of  
 9 parent-infant conversational turns: A word of caution. *Child Development*, 92(2), 672-  
 10 681. doi: <https://doi.org/10.1111/cdev.13495>
- 11 Richards, J. A., Gilkerson, J., Xu, D., & Topping, K. (2017). How much do parents think they  
 12 talk to their child? *Journal of Early Intervention*, 39(3), 163-179. doi:  
 13 10.1177/1053815117714567
- 14 Romeo, R. R., Leonard, J. A., Robinson, S. T., West, M. R., Mackey, A. P., Rowe, M. L., &  
 15 Gabrieli, J. D. E. (2018). Beyond the 30-million-word gap: Children’s conversational  
 16 exposure is associated with language-related brain function. *Psychological Science*,  
 17 29(5), 700–710. <https://doi.org/10.1177/0956797617742725>
- 18 Rothbart, M. K. (1981). Measurement of temperament in infancy. *Child Development*, 52, 569–  
 19 578.
- 20 Rowe M. L. (2008). Child-directed speech: relation to socioeconomic status, knowledge of child  
 21 development and child vocabulary skill. *Journal of child language*, 35(1), 185–205.  
 22 <https://doi.org/10.1017/s0305000907008343>
- 23 Rowe, M. L. (2012). A longitudinal investigation of the role of quantity and quality of child-  
 24 directed speech vocabulary development. *Child Development*, 83(5), 1762–1774.  
 25 <https://doi.org/10.1111/j.1467-8624.2012.01805.x>
- 26 Schwarz, I.C., Botros, N., Lord, A., Marcusson, A., Tidelius, H., & Marklund, E. (2017). The  
 27 LENA™ system applied to Swedish: Reliability of the adult word count estimate.  
 28 *Proceedings of Interspeech 2017*, 2088-2092. doi: 10.21437/Interspeech.2017-1287
- 29 Soderstrom, M., Blossom, M., Foygel, R., & Morgan, J. L. (2008). Acoustical cues and  
 30 grammatical units in speech to two preverbal infants. *Journal of Child Language*, 35(4),  
 31 869–902. <https://doi.org/10.1017/S0305000908008763>

- 1 VanDam, M., Ambrose, S. E., & Moeller, M. P. (2012). Quantity of parental language in the  
2 home environments of hard-of-hearing 2-year-olds. *Journal of deaf studies and deaf*  
3 *education*, 17(4), 402–420. <https://doi.org/10.1093/deafed/ens025>
- 4 Vihman, M. M. (1991). Ontogeny of phonetic gestures: Speech production. In I. G. Mattingly &  
5 M. Studdert-Kennedy (Eds.), *Modularity and the motor theory of speech perception* (pp.  
6 69–84). Hillsdale, NJ: Erlbaum.
- 7 Vihman, M. M. (1993). Variable paths to early word production. *Journal of Phonetics*, 21(1-2),  
8 61–82.
- 9 Wang, Y., Williams, R., Dilley, L., & Houston, D. M. (2020). A meta-analysis of the  
10 predictability of LENA™ automated measures for child language development.  
11 *Developmental review*, 57, 100921. <https://doi.org/10.1016/j.dr.2020.100921>
- 12 Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience  
13 strengthens processing and builds vocabulary. *Psychological Science*, 24(11), 2143-2152.  
14 doi: 10.1177/0956797613488145
- 15 Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary  
16 acquisition: Effects of sophisticated exposure and support for meaning. *Developmental*  
17 *Psychology*, 37(2), 265–279. <https://doi.org/10.1037/0012-1649.37.2.265>
- 18 Xu, D., Yapanel, U., & Gray, S. (2009). Reliability of the LENA™ language environment  
19 analysis system in young children's natural home environment. *Boulder, CO: LENA*  
20 *Foundation*.

21  
22  
23  
24