1   Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of

2   later vocabulary than their use of referential gestures

Abstract

Variation in how frequently caregivers engage with their children is associated with variation in children's later language outcomes. One explanation for this link is that caregivers use both verbal behaviors, such as labels, and non-verbal behaviors, such as gestures, to help children establish reference to objects or events in the world. However, few studies have directly explored whether language outcomes are more strongly associated with referential behaviors that are expressed verbally, such as labels, or non-verbally, such as gestures, or whether both are equally predictive. Here, we observed caregivers from 42 Spanish-speaking families in the US engage with their 18-month-old children during 5-min lab-based, play sessions. Children's language processing speed and vocabulary size were assessed when children were 25 months. Bayesian model comparisons assessed the extent to which the frequencies of caregivers' referential labels, referential gestures, or labels and gestures together, were more strongly associated with children's language outcomes than their total numbers of words, or overall talkativeness. The best-fitting models showed that children who heard more referential labels at 18 months were faster in language processing and had larger vocabularies at 25 months. Models including gestures, or labels and gestures together, showed weaker fits to the data. Caregivers' total words predicted children's language processing speed, but predicted vocabulary size less well. These results suggest that the frequency with which caregivers of 18-month-old children use referential labels, more so than referential gestures, is a critical feature of caregiver verbal engagement that contributes to language processing development and vocabulary growth.

Keywords: communicative reference, gestures, labels, word learning, language processing, vocabulary size

Word count: X

Spanish-speaking caregivers' use of referential labels with toddlers is a better predictor of later vocabulary than their use of referential gestures

Research highlights

- We examined the frequency of caregivers' referential labels, referential gestures, and total words spoken to their 18-month-old children during a 5-min lab-based play interaction.

- We assessed the predictive power of referential labels, gestures, the combination of both, and total words spoken to 25-month-old children's processing speed and vocabulary growth.

- Bayesian model comparisons showed that best-fitting models included caregivers' referential labels at 18 months, though total words spoken also predicted later processing speed.

- Caregivers' use of referential labels, more so than referential gestures, are a critical linguistic feature linked to children's later vocabulary learning.

Introduction

Children learn language through interactions with others. Studies of caregiver-child interactions have documented extensive variability in the frequency with which caregivers use verbal behaviors (e.g., words) and nonverbal behaviors (e.g., gestures) when they engage with their children. Individual differences among caregivers have been noted in studies of families across diverse linguistic, cultural, and socioeconomic status (SES) backgrounds (Casillas, Brown, & Levinson, 2019; Hart & Risley, 1995; Hoff, 2003; Weber, Fernald, & Diop, 2017). Moreover, variability in the frequency of caregivers' use of verbal behaviors (Gilkerson et al., 2018; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Rowe, 2012; Shneidman & Goldin-Meadow, 2012; Walker, Greenwood, Hart, & Carta, 1994; Weisleder & Fernald, 2013)

[51] and nonverbal behaviors (Cartmill et al., 2013; Pan, Rowe, Singer, & Snow, 2005; Rowe,

[52] 2008; Rowe & Goldin-Meadow, 2009) has been shown to be positively associated with

[53] children's later language development.

[54]   There are multiple proposals to explain how caregivers' verbal and nonverbal behaviors

[55] support later language learning. Both can be used to refer to objects and events. By using

[56] verbal behaviors, such as labels, in the presence of objects, caregivers support children's

[57] learning of word-referent mappings, a critical step in children's early comprehension and

[58] subsequent word production (Baldwin, 1993; Bohn & Frank, 2019; McMurray, Horst, &

[59] Samuelson, 2012). Nonverbal behaviors, such as gestures, can also be used to refer to and

[60] communicate about the identity of referents (e.g., by pointing to, holding out, or giving a

[61] cup to someone). For example, caregivers' deictic gestures, such as pointing, can help

[62] children disambiguate the referent of a label from other candidate referents (Iverson, Capirci,

[63] Longobardi, & Caselli, 1999; Puccini, Hassemer, Salomo, & Liszkowski, 2010; Rowe, 2000;

[64] Tfouni & Klatzky, 1983; Yuksel & Brooks, 2017; Zukow-Goldring, 1996). Labels and gestures

[65] can also be used together (e.g., saying "give me the cup," while pointing to a cup), providing

[66] the child with two cues to reference in differing modalities. Thus, caregivers' use of labels,

[67] gestures, or both together, can help children to map language onto specific concepts,

[68] strengthening their understanding of how language represents objects or events in their

[69] world. In this study, we compare Spanish-speaking caregivers' use of verbal behaviors (i.e.,

[70] total words and referential labels) and non-verbal behaviors (i.e., referential gestures) during

[71] a play session with their 18-month-old children. We then assess the degree to which these

[72] behaviors are linked to children's language processing efficiency and vocabulary outcomes at

[73] 25 months.

Variation in caregivers' verbal and non-verbal behaviors

[75]   Documenting variability among caregivers in their frequency of communicative

[76] behaviors is critical for establishing links between these behaviors and later child outcomes.

Verbal behaviors have been examined using numerous measures that capture the quantity and quality of caregivers' speech – although they mostly do so ignoring the referential context. Using the LENA technology, Gilkerson et al. (2017) collected day-long recordings of the speech children heard in 329 American families with 2- to 48-month-old English-speaking children from varying SES backgrounds. Speech recognition software provided automated estimates of the quantity of caregivers' speech, i.e., adult word counts (AWC), revealing that children were exposed to as few as 8,000 and as many as 17,000 words in a 12-hour day. Bergelson et al. (2019) collected LENA daylong recordings with 3- to 20-month-old children in 61 American families. Instead of total adult-word counts, they assessed variation in caregiver talk by measuring the amount of time each child was exposed to child-directed speech (CDS). The authors found that children were exposed to CDS for 11.36 min per hour, on average, with a standard deviation over a third of the mean (SD = 4.24 min). Studies of caregiver-child interactions in different sociocultural contexts, such as subsistence farming communities, have found that, children were exposed to far less speech, on average, than in other communities; however, there was still substantial variability among families (Bunce et al., 2020; Casillas et al., 2019, 2021; Shneidman & Goldin-Meadow, 2012; Yuksel & Brooks, 2017). Other studies have specifically examined caregivers' use of nouns in verbal labels and noted variability among caregivers in multiple languages, including English, Italian, French, Spanish, Turkish, Mandarin, and Korean (Altınkamış, Kern, & Sofu, 2014; Bergelson et al., 2019; S. Choi, 2000; Rosemberg et al., 2020; Tardif, Shatz, & Naigles, 1997).

Substantial variability among caregivers in their use of nonverbal gestures is also well documented. Studies examining caregivers' use of gestures have primarily focused on gestures that are symbolic or representational to some degree (Rowe, Wei, & Salo, 2022), such as iconic gestures (e.g., flapping hands for a bird), conventional gestures (e.g., nodding one's head to mean 'yes' in the US), and referential gestures (e.g., holding out objects or deictic gestures such as pointing). For example, Rowe, Özçalışkan, and Goldin-Meadow (2008) videotaped 90-min interactions in 53 English-speaking families with children from 14

to 34 months. They found that caregivers produced, on average, 100-115 symbolic,

conventional, and deictic gestures, with values ranging from only a few gestures to over 400.

Other studies have examined deictic gesture use in families speaking non-English languages

and living in different sociocultural contexts, e.g., in families speaking Yucatec Mayan in

Mexico (Salomo & Liszkowski, 2013) and Lazuri in Turkey (Yuksel & Brooks, 2017), also

noting extensive variability among caregivers in both groups.

Variability among caregivers in their use of verbal behaviors and gestures has been

linked to child language outcomes. In some studies, language samples are used to capture

variation in the frequency of young children's production of recognizable words during

interactions with their caregiver (Huttenlocher et al., 1991). In older school-age children,

researchers have also reported links between frequency of caregiver verbal engagement and

children's scores on standardized tests of language, such as vocabulary (Gilkerson et al.,

2018). When children are infants and toddlers, many studies rely on parent-reports

assessments of children's vocabulary size, such as the MacArthur-Bates Communicative

Developmental Inventories (CDI, Fenson et al., 2007), which ask parents to indicate which

words their child "understands and says" from among several hundred words on a checklist

(e.g., Weisleder & Fernald, 2013). Still other studies have explored links between caregivers'

verbal behaviors and children's performance in tasks that capture skill at processing

language in real time, such as the Looking-While-Listening task (Fernald, Zangl, Portillo, &

Marchman, 2008). For example, in a sample of 27 Spanish-speaking caregiver-child dyads,

Hurtado, Marchman, and Fernald (2008) reported that children who experienced more

speech from their caregivers during a lab-based play session were reported both to know

more words on the CDI and were more efficient at recognizing spoken words in real time.

Weisleder and Fernald (2013) reported similar findings based on estimates of caregivers'

child-directed word counts during day-long recordings. In both studies, mediation models

explored possible pathways among caregiver talk, vocabulary size, and processing efficiency.

Results suggested that frequent engagement with caregivers may be "tuning up" children's

131 abilities to map real-time spoken language onto referents in the world around them, allowing

132 for more efficient use of the input to support language learning.

133   Links between caregivers' use of gesture and children's later vocabulary abilities have

134 also been reported (Iverson et al., 1999; Pan et al., 2005; Rowe, 2008). Rowe and

135 Goldin-Meadow (2009) examined socioeconomically-diverse caregivers and children in the

136 home across multiple visits, beginning when children were 14 months. They found that

137 variation among children in their gesture use at 14 months was related to their vocabulary

138 skills at 54 months, measured using a standardized test. Importantly, this study and others

139 have found that the frequency of caregivers' gesture use is related to the frequency of

140 children's gesture use. In particular, caregivers' use of deictic gestures, such as pointing, has

141 been viewed as a potential means of influencing children's own use of deictic gestures, an

142 important prelinguistic skill (Goodwyn, Acredolo, & Brown, 2000; Matthews, Behne, Lieven,

143 & Tomasello, 2012; Rowe, 2000; Rowe & Leech, 2019). Other studies propose that caregivers'

144 use of deictic gestures can support word learning by bringing attention to an object and

145 reducing spatial ambiguity, thus allowing children to attend more effectively to the referent

146 and/or the auditory signal (Iverson et al., 1999; Puccini et al., 2010; Rowe, 2000; Tfouni &

147 Klatzky, 1983; Yuksel & Brooks, 2017; Zukow-Goldring, 1996).

Labels, gestures, or both?

149   Taken together, there is substantial evidence that how frequently caregivers use

150 communicative behaviors is associated with children's language learning. However, few

151 studies have directly contrasted the predictive relations to children's outcomes from verbal

152 versus non-verbal behaviors that establish reference. This referential function of labels and

153 gestures is important because it serves as a means to support children's early label-referent

154 associations. Additionally, it is critical to remember that these behaviors frequently occur

155 together in real time (Iverson et al., 1999; Pan et al., 2005; Puccini et al., 2010; Rowe &

156 Goldin-Meadow, 2009; Tfouni & Klatzky, 1983; Yuksel & Brooks, 2017; Zukow-Goldring,

157 1996). Thus, it is difficult to address whether links between caregiver verbal or nonverbal

158 behaviors and children's outcomes may in fact be better explained by caregivers' combined

159 use of labels and gestures. For example, Rowe (2000) proposed that there may be a shared

160 construct underlying caregivers' use of verbal behaviors and gestures, such as

161 communicativeness. This hypothesis is supported by evidence of a small to moderate

162 positive correlation between the frequency of caregivers' verbal behaviors and gestures; those

163 caregivers who used more total words also gestured more frequently than caregivers who

164 used fewer words (Pan et al., 2005; Rowe, 2000, 2008; Rowe & Goldin-Meadow, 2009; Salo,

165 Reeb-Sutherland, Frenkel, Bowman, & Rowe, 2019). In the present study, we ask if the

166 predictive power of caregivers' communicative use of reference may be captured more fully

167 by measures that reflect the combined use of referential labels and gestures, rather than each

168 measure taken alone.

169   How caregivers combine labels and gestures in real time has been widely discussed in

170 the experimental literature on early word learning (Gogate, Bahrick, & Watson, 2000;

171 Tincoff, Seidl, Buckley, Wojcik, & Cristia, 2019; Villiers Rader & Zukow-Goldring, 2012;

172 Zukow-Goldring, 1996). For example, Kalagher and Yu (2006) found that novel word

173 learning was more successful when caregivers introduced words while narrating a story and

174 pointing to the objects than when narrating a story without pointing. Gogate et al. (2000)

175 examined European American and Hispanic American families residing in a major

176 metropolitan area in the United States. They found that when they were teaching novel

177 labels to young infants, caregivers were more likely to use labels while moving objects.

178 Moreover, caregivers of linguistically less-advanced infants, compared to more-advanced

179 infants, were those who were more likely to synchronize labels with object motion. These

180 findings suggest that caregivers are sensitive to children's level of language skills when using

181 labels and gestures together to highlight new label-referent associations.

The Current Study

In this longitudinal study, we observed 42 Spanish-speaking caregivers during play interactions with their 18-month-old children. We coded the frequency and duration of caregivers' referential labels to objects and referential gestures to objects. At 25 months, children's language skills were assessed using an on-line language processing task and caregiver reports of productive vocabulary size. Bayesian methods were used to construct different models of the frequency of caregivers' use of labels, gestures, and both in combination, as predictors of child outcomes. We predicted that if children's later language abilities are best predicted by the frequency of caregivers' use of labels or gestures taken independently, this would suggest a primary role for learning based on either modality. However, if language learning is supported more by the frequency of caregivers' use of reference across verbal and nonverbal modalities, then one or more models including both labels and gestures would be stronger predictors of our measures of language outcomes (Cartmill et al., 2013). We also included a model capturing the total number of words spoken by caregivers to explore the specificity of caregivers' use of referential labels, in contrast to overall talkativeness. By comparing these models, we asked what is the smallest set of caregiver's communicative behaviors that best predicts children's language outcomes at 25 months.

Methods

Participants

Participants were 42 primarily Spanish-speaking children[1] (42 females) and their caregivers who were participating in a longitudinal study examining language development in

---

[1] As seen in our pre-registration, we determined a sample size of $n = 50$ based on a priori frequentist power analyses, but stopped at n = 42 because at the time of analysis there were no more available families to include in the study.

Table 1

Participant age and SES.

	M	SD	Range
Age (pre-test)	18.54	0.84	17.1 - 19.8
Age (post-test)	25.46	0.68	24.2 - 26.8
SES (pre-test)	26.44	11.82	8 - 62

Note. SES was calculated based on the

Hollingshead Index.

primarily monolingual Spanish-speaking families in the US. Families were recruited from birth records or community contacts in Northern California and were excluded if the child was born preterm, had a known neurodevelopmental disorder, or loss of hearing or vision. As shown in Table 1, children were approximately 18 months at the start of the study and approximately 25 months when we assessed language processing speed and vocabulary size. We calculated SES using the Hollingshead Index, which reflects education and occupation for both mothers and fathers. SES was included as a covariate based on prior studies (Daneri, Blair, & Kuhn, 2018; Hoff, 2003; Huttenlocher, Waterfall, Vasilyeva, Vevea, & Hedges, 2010), to examine the unique role of caregiver behaviors on children's language skills over and above potential confounding variables.

Families represented a diverse range of SES backgrounds. All mothers reported that they were native Spanish speakers. All families lived in the US but the mothers were primarily born in Mexico (33), with a few from Central America (5).

Procedure

Native Spanish-speaking research staff met with the caregiver to explain study protocol, and all caregivers gave their informed consent prior to study participation.

Caregivers participated in a videotaped lab-based play session with their 18-month-old children at a community laboratory. Each caregiver was asked to engage with her child using a standard set of toys (e.g., plates, pretend food, cutlery, pots, doll) for approximately 5 min. During the session, the child wore a LENA recorder placed inside a specially-designed vest to capture the adult speech spoken during the play session (Marchman, Weisleder, Hurtado, & Fernald, 2021). At 18 and 25 months, children participated in the Looking-While-Listening task to assess spoken language understanding (Fernald et al., 2008). At both time points, caregivers completed parent-report assessments of their child's productive vocabulary size (Jackson-Maldonado, Thal, & Fenson, 2003).

Measures

Coding of caregiver referential gestures and labels. A native Spanish-speaker using ELAN (version 5.0, Wittenburg, Brugman, Russel, Klassmann, & Sloetjes, 2006) coded all caregivers' referential gestures and labels from the video recordings of the play sessions. Gestures were coded first without audio. Referential gestures were defined as those gestures used to attract infants' attention to the toys or other objects in the environment. Gestures included holding out objects/giving, pointing, descriptive or iconic gestures (e.g., making a chopping motion with their hand), and touching with an open hand. Physically playing with toys was not included as a gesture (e.g., holding the knife and pretending to cut vegetables in front of the child). A standardized protocol used to define the onset and offset of each gesture is available in our full codebook (https://osf.io/xqpsy/). Frequency of gestures was derived for each caregiver, and the onset and offset of gestures were used for our overlap measure below.

Caregivers' use of object labels was then coded by the same coder. The coder listened to the video and marked the onset and offset of all object labels that referred to objects in the play session. Frequency counts of label tokens were derived for each caregiver. Successive repetitions of a single label were counted as individual tokens. General category terms (e.g.,

²⁴⁶ "comida" [food], or "juguetes" [toys]) were excluded because our goal was to focus on specific

²⁴⁷ labels rather than category names for available objects. All English labels were excluded,

²⁴⁸ given that we were assessing children's later Spanish language outcomes.

²⁴⁹   Finally, we determined the number of times that each caregiver produced an object

²⁵⁰ label while using a gesture (overlaps: labels + gestures). An R script used the duration

²⁵¹ coding of each label and gesture in the ELAN output to identify the number of labels that

²⁵² occurred within a 1-sec window before or after a gesture (Cartmill et al., 2013).

²⁵³   Figure 1 depicts examples of the final label and gesture coding for three caregivers over

²⁵⁴ the 5-min observation window. These examples illustrate variation among caregivers in the

²⁵⁵ overall frequency of labels and gestures, as well as variation in the number of overlapping
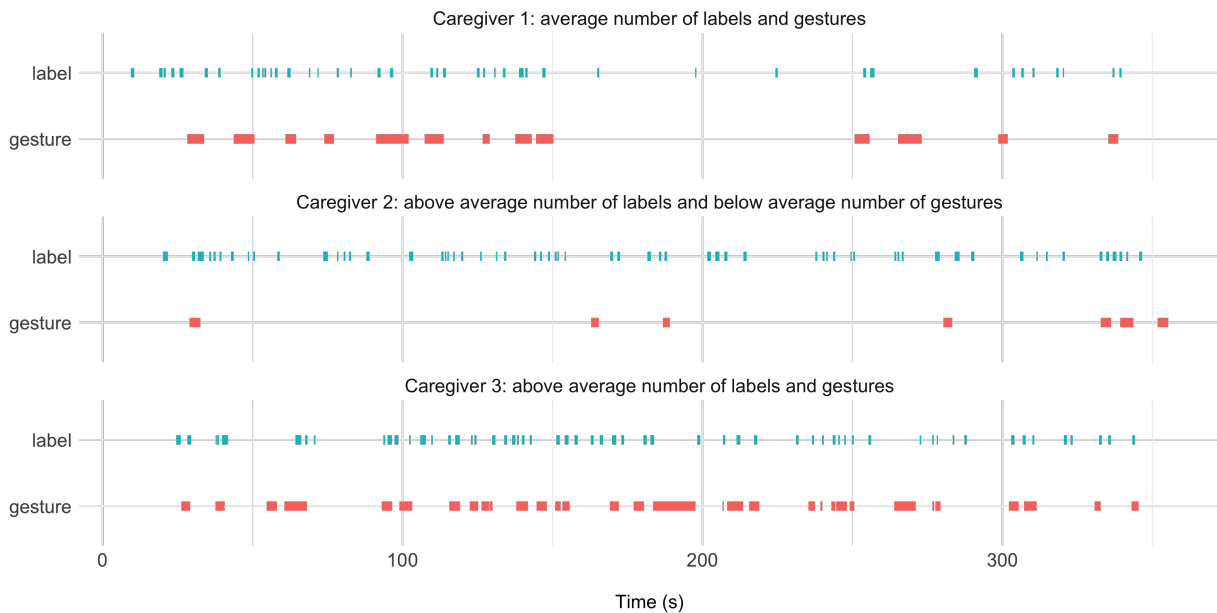
²⁵⁶ labels and gestures.

Figure 1. Examples from three participants chosen to illustrate the variability in frequency and duration of label and gesture use. Ticks represent each instance and the size depicts the duration. Caregiver 1 provided an average number of labels and gestures, Caregiver 2 provided an above average number of labels and a below average number of gestures, and Caregiver 3 provided an above average number of labels and gestures.

257   **Reliability Coding.**   A second native Spanish-speaking coder coded labels and

258   gestures for approximately 20% of the families (n = 8). The second coder was blind to the

259   study hypotheses and to the coding by the first coder. Intraclass correlations (ICC)

260   suggested strong reliability for number of labels (ICC = .996, 95% CI [.96, 1]), gestures (ICC

261   = .89, 95% CI [.54, .98]), and overlaps (ICC = .99, 95% CI [.98, 1]).

262   **Caregiver verbal engagement during play session.**   During the play session, a

263   LENA audio recorder was used to provide an estimate of the number of adult word counts

264   (AWC) produced during the session. The AWC measures generated by the LENA speech

265   recognition software were converted to a rate per hour based on the 5-min sample, to

266   account for minor differences in the duration of play sessions. This measure was included in

267   the models as an estimate of overall caregiver talkativeness.

268   **Spoken language processing.**   At each time point, the child participated in the

269   Looking-While-Listening task (LWL, Fernald et al., 2008). In this task, the child sits on

270   their caregiver's lap while viewing pictures of two familiar objects on a screen. After 2 sec, a

271   voice of a female, native-Spanish speaker names one of the objects (e.g., "¿Dónde está el

272   perro?", Where's the doggy?), followed by an attention-getter phrase (e.g., "¿Te gustan las

273   fotos?, Do you like the pictures?). On each trial, the pictures were presented in fixed pairs,

274   matched for salience, and the target words were matched in grammatical gender. At 18

275   months, auditory stimuli consisted of eight familiar words presented 6 times each as target

276   and distracter. At 25 months, auditory stimuli consisted of twelve familiar words presented 4

277   times each as target and distracter. Each word in the pair served an equal number of times

278   as target and distracter, for a total of 48 trials, with target picture counterbalanced across

279   side across trials.

280   After a brief calibration session, trials were presented in two fixed pseudo-random

281   orders such that the target picture was not presented on the same side for more than two

282   trials in a row. Patterns of children's eye-gaze were captured at 60 frames/sec by a Tobii

283   X60 eye-tracker, mounted to the bottom of the monitor. A video camera attached to the top

of the monitor also provided a record of children's eye gaze across the full session. All video-recordings of the testing sessions were prescreened to exclude trials when the child was inattentive or if there was any concern that the caregiver was biasing the child. Based on which picture the child was fixated at target noun onset, trials were defined as distracter or target initial. Trials on which the child was not looking at either picture at target noun onset were not analyzed. Trials were also later removed on a child-by-child basis if the parent reported that the child did not know the target word. Due to calibration failures or experimental error, some portion of the sessions (11/42, 26%) were hand-coded by trained coders following standard protocols (Fernald et al., 2008). Processing speed was calculated on all distracter-initial trials as the mean reaction time (RT) in milliseconds to shift from the distracter to the target picture measured from the onset of the target noun. Trials were excluded if shifts were faster than 300 ms or slower than 1800 ms from target noun onset, since these shifts are not likely to be in response to the target word. Given that children could have different numbers of distracter-initial trials, the mean number of trials per child varied (M = 9.81, SD = 4.70), however, all children had at least 2 trials contributing to the computation of RT (range = 2 − 21).

Vocabulary size. Children's vocabulary size in Spanish was assessed at each time point by parent report with the MacArthur-Bates Inventarios del Desarollo de Habilidades Communicativas (CDI, Jackson-Maldonado et al., 2003). These instruments ask parents to indicate what words their child can "understand and say" from a list of hundreds of items. At 18 months, some parents completed the Inventario I form and others completed Inventario II form, due to slight changes in protocol over time. For those children whose parents completed Inventario I, scores were converted to proportions based on the number of items on the Inventario II form. At 25 months, all parents completed Inventario II. Vocabulary size was the number of words chosen (680 words maximum). Due to missing data, 37 families are included for analyses with the CDI.

Analysis Strategy

³¹¹ We first present descriptive statistics of all variables at 18 and 25 months. We then

³¹² present a series of Bayesian model comparisons that allowed direct comparisons of

³¹³ non-nested models to examine the predictive roles of labels, gestures, or their combination

³¹⁴ (i.e., overlaps), on child outcomes (Donnellan, Bannard, McGillion, Slocombe, & Matthews,

³¹⁵ 2020; Mahr & Edwards, 2018). This approach contrasts with prior studies that seek to

³¹⁶ isolate unique contributions of caregivers' verbal behaviors or gestures to outcomes using

³¹⁷ nested hierarchical regression (Iverson et al., 1999; Pan et al., 2005). We compared seven

³¹⁸ independent models, each representing a different hypothesis about how caregivers'

³¹⁹ communicative behaviors contribute to children's language processing speed and vocabulary

³²⁰ size at 25 months. These models assessed the independent contributions of labels and

³²¹ gestures, the conditional relation between labels and gestures, as well as the overlapping use

³²² of labels and gestures (overlap). We also tested a model including AWC, to rule out the

³²³ effects of caregiver talkativeness. All models controlled for SES and 18-month vocabulary

³²⁴ size and processing speed as appropriate, depending on the model. By including 18-month

³²⁵ language skills, we can ask the more specific question of which input variable(s) best predict

³²⁶ gains in language processing or vocabulary size over and above SES and children's earlier

³²⁷ language skills.

³²⁸ For each dependent variable (dv), we compared the same set of models[2]: (1) `dv ~`

³²⁹ `labels`; (2) `dv ~ gestures`; (3) `dv ~ overlaps`; (4) `dv ~ adult_words_per_hour`, which

³³⁰ considers all speech using AWC; (5) `dv ~ labels + gestures`, which assumes that both

³³¹ labels and gestures contribute independently; (6) `dv ~ labels * gestures`, which assumes

³³² that the contribution of labels and gestures are conditional on one another, and (7) `dv ~`

---

[2] The preregistration did not include a) the adult word count model and b) the baseline model. We added

these models later a) to see if the number of labels was simply an indicator of overall caregiver talkativeness

and b) to be able to judge if the inclusion of predictors improved predictions at all.

333 `covariates` is the baseline model. If a model performs at or worse than the baseline, its

334 predictor(s) do not contribute to predicting gains in processing or vocabulary over and above

335 the covariates.

336   All models were fit in a Bayesian framework as linear models in `R` (Team, 2021) via the

337 function `brm` from the R-package `brms` (Bürkner, 2017) using default priors for all model

338 parameters. All caregiver behavior variables were scaled to have a mean of 0 and a standard

339 deviation of 1. Following McElreath (2020), we compared models using WAIC (widely

340 applicable information criterion) scores and weights, an indicator of the model's predictive

341 accuracy for out-of-sample data; models with lower scores are preferred. Roughly speaking,

342 WAIC scores reflect the model's predictive accuracy with a penalty for the number of

343 effective parameters. As such, model comparisons favor simpler models and thereby guard

344 against overfitting. WAIC weights are an estimate of the probability that each model

345 (compared to all models considered) will make the best predictions on new data. We next

346 inspected the posterior distributions of the model predictors in the best models via their

347 means and 95% credible intervals (CI) to inform the nature (positive or negative) and

348 strength of the influence of the respective caregiver engagement variable on the dependent

349 variable.

Results

Descriptive statistics

352   Figure 2A provides descriptives for the four measures of caregiver communication.

353 Caregivers produced approximately 3500 words per hour (M = 3,447.26, SD = 1,491.97,

354 range = 531.94 - 6,683.38), on average, based on the automated LENA counts. Caregivers

355 produced just over 40 labels (M = 43.42, SD = 25.55, range = 0 - 120) and about 18

356 gestures (M = 17.93, SD = 8.11, range = 2 - 41). When considering overlaps, caregivers

357 produced about 15 labels that were also accompanied with a referential gesture, (M = 16.05,
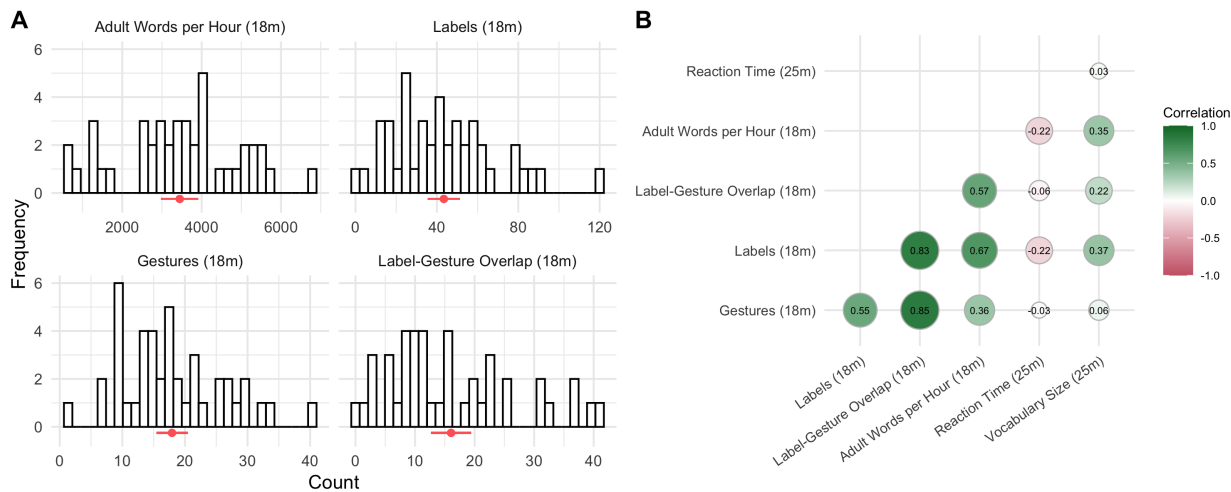
Figure 2. A) Descriptive distribution of independent variables with mean and 95% CI (in red), B) Zero-order correlations between dependent variables and input variables. Circle size and color intensity increase with the absolute magnitude of correlation.

SD = 10.89, range = 0 - 41).

Figure 2B shows the zero-order correlations among all variables. As expected, the three measures capturing caregivers' language (AWC per hour, labels, overlaps) were significantly correlated. Numbers of referential gestures also correlated with verbal behavior variables, suggesting some shared underlying variance. However, none of the correlations indicated that any two measures were redundant (i.e., all r < .90), which justifies assessing their independent predictive relation to the dependent variable in the model comparison.

Spanish language processing

Table 2 shows WAIC scores and weights for each model predicting children's language processing speed (RT). Only two models outperformed the baseline model: labels and AWC per hour, with both models similar in their weights (model weights: 0.23 labels; 0.18 AWC per hour). None of the models that included gestures, either as the only test predictor or in combination with labels, made better predictions compared to the baseline model than models that included labels. Thus, children's language processing speed at 25 months was

Table 2

WAIC scores and weights for models predicting language processing speed.

Model	waic	se_waic	weight
Labels	554.55	9.99	0.23
Adult words per hour	555.04	10.05	0.18
Baseline (covariates only)	555.23	10.22	0.16
Labels + gestures	555.90	9.96	0.12
Label-gestures overlap	556.72	9.98	0.08
Gestures	557.01	9.94	0.07
Labels * gestures	557.17	9.82	0.06

best predicted by models that included some form of caregivers' verbal behavior as predictors.

Figure 3A-i shows the posterior distribution of the model estimates for number of labels to be negative (β = -39.96) and largely different from 0 (95% credible interval (CrI) = -91.91 - 12.11). This speaks for a positive relation: the more labels the caregiver used at 18 months, the more the child improved in their reaction time from 18 to 25 months. However, the fact that the 95% CrI included zero, cautions against an overly strong interpretation. A similar pattern was found when investiagting the estimate for adult word count in the respective model: more adult talk was related to gains in reaction time – with considerable uncertainty (β = -27.88, 95% CrI = -80.57 - 25.19). The effect of SES was also similar. Children from families higher in SES tended to have greater developmental gains in reaction time, however, this effect was weak in magnitude (β = -27.67, 95% CrI = -79.96 - 24.31). Finally, children with a slower reaction time at 18 months were also slower at 25 months (β = 52.69, 95% CrI = 0.12 - 105.42). Figure A-ii shows the observed vs. predicted values from

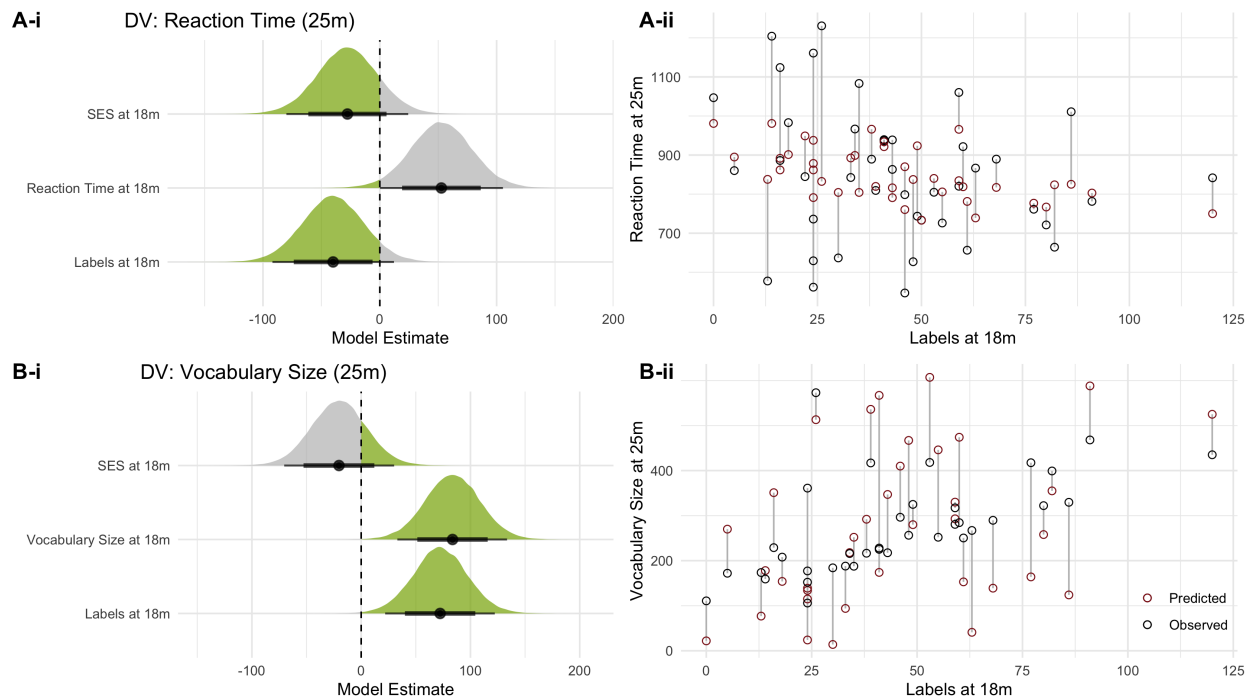386 the model with labels as the test predictor.

Figure 3. Left: Posterior distributions for model estimates, right: model predictions. On the left, the green area denotes the section of the distribution that is supportive (i.e. faster reaction time and larger vocabulary). Points below each distribution show means, and error bars show 80% (thick) and 95% (thin) CrIs. A-i shows the posterior distribution of all model estimates in the labels model for reaction time. B-i shows the same in the model predicting vocabulary size. On the right, A-ii and B-ii contrast the observed (black) values with the values predicted by the model (red) for reaction time (A) and for vocabulary size (B).

Vocabulary size

388 Table 3 shows the model comparisons for vocabulary size. All predictor models made
389 better predictions compared to the baseline model. As with RT, the model including the
390 number of labels produced by the caregiver made the best predictions – this time, however,
391 it clearly outperformed all the other models (model weight = 0.38). Models including
392 gestures were given more weight only when they also included labels.

Table 3

WAIC scores and weights for models predicting vocabulary size.

Model	waic	se_waic	weight
Labels	480.08	7.61	0.38
Labels + gestures	482.46	7.69	0.12
Adult words per hour	482.55	5.96	0.11
Label-gestures overlap	482.99	6.68	0.09
Labels * gestures	484.81	7.62	0.04
Gestures	486.45	6.57	0.02
Baseline (covariates only)	486.62	6.98	0.01

As shown in Figure 3B-i, the posterior distribution for the model estimate for labels was positive, large and reliably different from 0 (β = 72.29, 95% CrI = 21.95 - 122.26). Children who heard more labels at 18 months increased more in their reported vocabulary size from 18 to 25 months. SES had a weak effect (β = -20.34, 95% CrI = -70.46 - 30.14). Finally, children who had a larger reported vocabulary at 18 months also had a larger reported vocabulary at 25 months (β = 83.57, 95% CrI = 33.10 - 133.49). Figure 3B-ii shows the observed versus predicted values from the model with labels as the test predictor.

Comparing the contribution of labels and gestures

The model comparisons suggested that including the number of gestures as a predictor did not contribute to a model's predictive accuracy above baseline for RT, although gestures performed better than baseline for vocabulary size. Nevertheless, it is still interesting to see how the number of gestures related to the dependent variable in the different models. Thus, we compared the posterior distributions of the model estimates for labels and gestures across

the models that included them. Figure 4 shows this comparison. Looking first at labels,

regardless of model, the supportive contribution of labels was stable whether tested as the

only predictor or together with gestures for both reaction time and vocabulary size. In

contrast, gestures only supported the outcome of vocabulary growth when considered as the

sole test predictor. When combined with labels, the model estimates were essentially zero.

This pattern affirms the conclusion based on the model comparisons, i.e., that knowing the

number of gestures in the input – in addition to the number of labels - did not improve
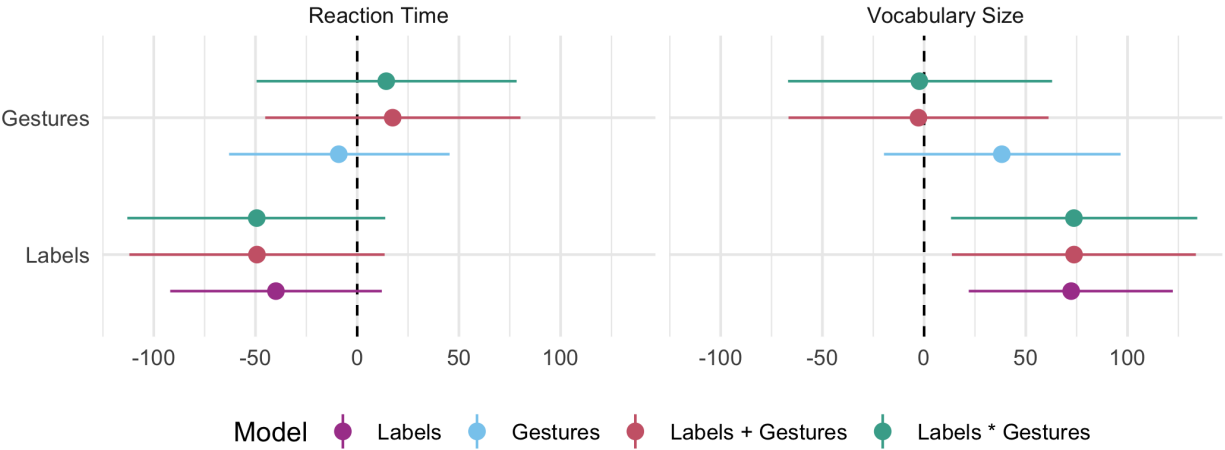
predictions.

Figure 4. Comparing estimates for labels and gestures across models. Points show means of the posterior distribution (95% CrIs) for the estimates. Estimates were extracted from all models that included one or both of the predictors.

Discussion

Our goal was to compare variation among Spanish-speaking caregivers in the number of words, labels, gestures, and combined labels and gestures used when interacting with their toddlers, in order to determine the smallest set of caregivers' communicative behaviors that best predicted children's language outcomes at 25 months. We found that over and above SES and children's earlier language skills, variability in caregivers' use of referential labels

420 was the strongest predictor of children's processing speed and vocabulary, when pitted

421 against variability in referential gestures or in different combinations of labels and gestures.

422 Caregivers' total words predicted children's later language processing speed but not their

423 vocabulary. We discuss two questions raised by the results: Why might caregivers' use of

424 referential labels predict children's later language processing efficiency and vocabulary size?

425 Why are labels more predictive than gestures?

426 Why might caregivers' use of referential labels predict children's language

427 processing efficiency and vocabulary size?

428   Those caregivers who used more labels also used more words overall (Figure 2B),

429 reflecting an r2 of 45% shared variance and demonstrating a strong relation between these

430 measures. However, while both measures of talk predicted reaction time, only caregivers' use

431 of labels better predicted both outcomes of children's language processing and their

432 vocabulary size. One possibility is that the frequency of caregiver labels is more closely

433 linked to children's understanding of word meaning, which is reflected in outcome measures

434 of both language processing and vocabulary size. Labels themselves are symbols that refer to

435 the objects, ideas, or events they represent (Acredolo & Goodwyn, 1988; Bates, Thal,

436 Whitesell, Fenson, & Oakes, 1989; Colonnesi, Stams, Koster, & Noom, 2010), and both the

437 mapping of a label to a referent and the learning of a label for a referent are directly assessed

438 in both of our outcome measures. Language processing speed reflects children's ability to

439 map a spoken object name in real time onto one of two familiar pictures, assessed only on

440 trials when the child demonstrates a clear shift from the distracter to the target picture.

441 Thus, this task taps into children's familiar knowledge of these everyday objects where

442 children who are faster at processing the object label may have stronger conceptual and

443 linguistic representations than those who are slower. Vocabulary size, as reported by parents

444 on the CDI, reflects children's abilities to produce the names of objects and concepts.

445 Therefore, variation among caregivers in the frequency of specific use of referential labels may

446 provide a closer link to individual differences in children's linguistic knowledge about objects

447 or events. While caregivers' use of total words use may help 'tune' up children's language

448 processing speed, and provide children with the practice of hearing language, caregivers' use

449 of labels, in particular, specifically provides the linguistic information that enables early

450 word learning. These results suggest that during early stages of language learning, repeated

451 and varied exposure to labels embedded within day-to-day conversations may help children

452 associate, prune, and strengthen these links (McMurray et al., 2012), quickly process how

453 labels map onto objects in real time (Fernald, Perfors, & Marchman, 2006), and build a

454 vocabulary that reflects their understanding about the world (Weisleder & Fernald, 2013).

Why are labels more predictive than gestures?

456   Caregivers who used more referential labels were those who used more referential

457 gestures, ($r = .55$; Figure 2B). The strength of this association is within expectations based

458 on prior studies of children across a broad age range (i.e., 8 to 36 months), in spite of slightly

459 different operationalizations of total words, labels, and gestures (e.g., Pan et al., 2005: $rs =$

460 .35 - .54; Rowe, 2000: $r = .58$; Rowe & Goldin-Meadow, 2009: $r = .67$; Salo et al., 2019: $r =$

461 .30; Salomo & Liszkowski, 2013: $r = .63$). However, we did not find support for our

462 hypothesis that an underlying shared characteristic of caregivers' communicative reference

463 across referential labels and gestures was predictive of children's language skills (Rowe, 2000;

464 Rowe et al., 2008). Instead, it was the frequency in caregivers' use of labels that best

465 predicted later language outcomes. Rather than the shared referential function that both

466 labels and gestures serve, there is information in the linguistic signal specifically associated

467 with label use that supports children's later vocabulary outcomes.

468   It is important to note that as in previous studies, our measures of referential labels

469 and gestures were not mutually exclusive. Labels may have occurred alone in an utterance or

470 embedded in a multi-word utterance, with each instance co-occurring with a variety of

471 socio-pragmatic behaviors such as eye-gaze, facial expressions, body movement, in addition

to referential and non-referential gestures. Our findings suggest that variability in caregivers' use of referential labels, regardless of how these labels are combined with nonverbal behaviors, is most strongly associated with later vocabulary in 25-month-old children.

These results should not be taken as evidence that caregivers' gesture use plays a less influential role in children's language learning. In exploratory analyses, we found that caregivers' use of referential gestures predicted vocabulary growth when included as the only test predictor, although not in combination with labels. These links are in line with those of prior studies showing that variation in caregiver gestures or nonverbal behaviors predicted children's later vocabulary, although those studies differed in whether or not they controlled for children's earlier language skills (Cartmill et al., 2013; Rowe & Goldin-Meadow, 2009). By directly contrasting the use of referential labels and gestures in the same context, our study demonstrated that knowing the number of referential gestures did not improve our predictions for growth in children's language processing or vocabulary size, if the number of labels was already known (Iverson et al., 1999 ; Pan et al., 2005).

It is also possible that caregivers' use of referential labels and gestures are of different importance at different phases of children's communicative development. Children in our study were 17 to 19 months old, whereas prior studies linking caregivers' gesture use to later outcomes examined gestures when children were around 14 to 16 months old (Iverson et al., 1999 ; Pan et al., 2005; Rowe & Goldin-Meadow, 2009). At earlier ages more children are in an early pre-linguistic stage, and thus may benefit more from the support for learning provided by caregivers' use of referential gestures. Children who produce more gestures early in life have been found to have stronger vocabulary later on (e.g., Colonnesi et al., 2010). Caregivers' gestures may be particularly supportive of children's prelinguistic gestures and short-term language outcomes (Rowe & Leech, 2019), an effect that is less evident as children become more linguistically advanced. It is also important to note that the current study focused specifically on referential gestures, whereas prior work has considered a larger set of caregivers' communicative behaviors, including symbolic gestures (e.g., cutting motion

499 with hands) and conventional gestures (e.g., nodding to mean 'yes' in the United States).

500 Therefore, at any given moment, caregivers can use both referential and non-referential

501 gestures to direct children's attention to the label-object link, support visual object

502 recognition, and resolve ambiguity of the intended referent (Tincoff et al., 2019; Villiers

503 Rader & Zukow-Goldring, 2012; Zukow-Goldring, 1996), all of which are likely to provide a

504 foundation for stronger language learning.

Limitations

506   While our results shed light on which specific features of caregiver communicative

507 behaviors may be important for language learning, we are unable to establish definitively the

508 direction of any causal link between caregivers' verbal behaviors and children's language

509 skills. Though we included a covariate of children's initial language skills on the respective

510 outcome measure to assess caregivers' contribution to children's growth in language skills, we

511 cannot rule out the possibility that caregivers who use more labels do so because their

512 children are more verbal. Correlational links represent average effects, with much still left

513 unexplained (Bailey, Duncan, Watts, Clements, & Sarama, 2018). Rather than a causal

514 pathway of caregivers influencing children, correlations may represent relatively stable

515 individual differences among children and families with shared genes and/or environments.

516 Correlations may also be attributable to individual differences in children's propensity or

517 ability to elicit engagement from others or in children's ability to effectively process

518 information (Pace, Luo, Hirsh-Pasek, & Golinkoff, 2017; Weisleder & Fernald, 2013).

519 Though there is growing research examining whether intervening with caregivers in their use

520 of verbal and nonverbal behaviors can influence children's early language development

521 (Matthews et al., 2012; Rowe & Leech, 2019; Suskind et al., 2016), findings to date are

522 mixed. Ongoing research should continue to explore the effectiveness of such interventions

523 for children's short- and long-term outcomes, as well as potential moderators that influence

524 which families are likely to benefit the most (Rowe & Leech, 2019).

⁵²⁵   Moreover, the potential for short- or long-term causal impacts of caregivers' verbal or

⁵²⁶ nonverbal behaviors for children's language outcomes should be considered within the

⁵²⁷ context of broader socioeconomic and political systems that underlie families' day-to-day

⁵²⁸ experiences (Rowe & Weisleder, 2020). This work examined caregiver behaviors in a

⁵²⁹ lab-based interaction, which may be consistent with caregivers' densest periods of

⁵³⁰ interactions in the home; however, testing children in a lab still differs from the ebb and flow

⁵³¹ of interactions over the course of a day, when children may engage with multiple individuals

⁵³² (Bergelson, Amatuni, Dailey, Koorathota, & Tor, 2019; Reynolds, Vernon-Feagans,

⁵³³ Bratsch-Hines, Baker, & Investigators, 2019). Our study also included children with typical

⁵³⁴ development from one unique cultural context, primarily Spanish-speaking families raising

⁵³⁵ their children in an English-dominant community in the United States. More work is needed

⁵³⁶ to understand if these links are seen in comparative studies across cultures, languages, and in

⁵³⁷ populations which include neurodiverse children (Bang, Adiao, Marchman, and Feldman

⁵³⁸ (2019) ;B. Choi, Shah, Rowe, Nelson, and Tager-Flusberg (2020); Salomo and Liszkowski

⁵³⁹ (2013)]. Across contexts, children and parenting practices may vary widely (Rowe &

⁵⁴⁰ Weisleder, 2020), likely influencing how frequently children are exposed to labels and

⁵⁴¹ gestures during direct engagement with caregivers. There is still much to understand for

⁵⁴² what processes may be shared, but also what may very well be different pathways that

⁵⁴³ support language acquisition in different populations.

⁵⁴⁴ **Conclusion**

⁵⁴⁵   Children who engage more frequently with their caregivers tend to have stronger

⁵⁴⁶ language outcomes. Here, we explored one possible explanation of that relation, namely, that

⁵⁴⁷ caregiver engagement is more supportive of learning because caregivers use a variety of

⁵⁴⁸ verbal and non-verbal behaviors to help children establish reference to objects and events in

⁵⁴⁹ the world. Specifically, we investigated how caregivers' use of referential labels and gestures

⁵⁵⁰ predicted children's later vocabulary skills, rather than focusing on a single form of reference.

Contrary to our predictions, we found that the frequency of caregivers' use of referential labels when communicating with children at 18 months, but less so their frequency of labels and gestures in combination, best predicted growth in children's language processing and vocabulary skills at 25 months. Caregivers' overall talkativeness was also associated with children's later processing speed, suggesting that overall experience with language supports skill in real-time language comprehension. However, later vocabulary development was best predicted by caregivers' use of labels, more strongly than overall talkativeness, suggesting that it is the use of labels, per se, that provides important cues to vocabulary learning. Taken together, these findings reveal that specific properties of caregiver verbal engagement may support different aspects of language learning, providing important insights into the pathways through which caregiver engagement supports children's learning.

References

Acredolo, L., & Goodwyn, S. (1988). Symbolic gesturing in normal infants. *Child Development*, *59*(2), 450–466. https://doi.org/10.2307/1130324

Altınkamış, N. F., Kern, S., & Sofu, H. (2014). When context matters more than language: Verb or noun in French and Turkish caregiver speech. *First Language*, *34*(6), 537–550. https://doi.org/10.1177/0142723714560179

Bailey, D. H., Duncan, G. J., Watts, T., Clements, D. H., & Sarama, J. (2018). Risky business: Correlation and causation in longitudinal studies of skill development. *American Psychologist*, *73*(1), 81–94. https://doi.org/10.1037/amp0000146

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *Journal of Child Language*, *20*(2), 395–418. https://doi.org/10.1017/S0305000900008345

Bang, J. Y., Adiao, A. S., Marchman, V. A., & Feldman, H. M. (2019). Language nutrition for language health in children with disorders: A scoping review. *Pediatric Research*, *87*(2), 300–308. https://doi.org/10.1038/s41390-019-0551-0

Bates, E., Thal, D., Whitesell, K., Fenson, L., & Oakes, L. (1989). Integrating language and gesture in infancy. *Developmental Psychology*, *25*(6), 1004–1019. https://doi.org/10.1037/0012-1649.25.6.1004

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. https://doi.org/10.1111/desc.12715

Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do north American babies hear? A large-scale cross-corpus analysis. *Developmental Science*, *22*(1), e12724. https://doi.org/10.1111/desc.12724

Bohn, M., & Frank, M. (2019). The pervasive role of pragmatics in early language. *Annual Review of Developmental Psychology*, *1*, 223–249. https://doi.org/10.1146/annurev-devpsych-121318- 085037

Bunce, J., Soderstrom, M., Bergelson, E., Rosemberg, C., Stein, A., Alam, F., . . . Casillas,

M. (2020). *A cross-cultural examination of young children's everyday language experiences.* https://doi.org/10.31234/osf.io/723pr

Bürkner, P. C. (2017). Brms: An r package for bayesian multilevel models using stan. *Journal of Statistical Software, 80*(1), 1–28. https://doi.org/10.18637/jss.v080.i01

Cartmill, E. A., Armstrong, B. F., Gleitman, L. R., Goldin-Meadow, S., Medina, T. N., & Trueswell, J. C. (2013). Quality of early parent input predicts child vocabulary 3 years later. *Proceedings of the National Academy of Sciences, 110*(28), 11278–11283. https://doi.org/10.1073/pnas.1309518110

Casillas, M., Brown, P., & Levinson, S. C. (2019). Early language experience in a Tseltal Mayan village. *Child Development, 91*(5), 1819–1835. https://doi.org/10.1111/cdev.13349

Casillas, M., Brown, P., & Levinson, S. C. (2021). Early language experience in a Papuan community. *Journal of Child Language, 48*(4), 792–814. https://doi.org/10.1017/S0305000920000549

Choi, B., Shah, P., Rowe, M. L., Nelson, C. A., & Tager-Flusberg, H. (2020). Gesture development, caregiver responsiveness, and language and diagnostic outcomes in infants at high and low risk for Autism. *Journal of Autism and Developmental Disorders, 50*(7), 2556–2572. https://doi.org/10.1007/s10803-019-03980-8

Choi, S. (2000). Caregiver input in English and Korean: Use of nouns and verbs in book-reading and toy-play contexts. *Journal of Child Language, 27*(1), 69–96. https://doi.org/10.1017/S0305000999004018

Colonnesi, C., Stams, G. J. J. M., Koster, I., & Noom, M. J. (2010). The relation between pointing and language development: A meta-analysis. *Developmental Review, 30*(4), 352–366. https://doi.org/10.1016/j.dr.2010.10.001

Daneri, M. P., Blair, C., & Kuhn, L. J. (2018). Maternal language and child vocabulary mediate relations between socioeconomic status and executive function during early childhood. *Child Development, 90*(6), 1–18. https://doi.org/10.1111/cdev.13065

Donnellan, E., Bannard, C., McGillion, M. L., Slocombe, K. E., & Matthews, D. (2020). Infants' intentionally communicative vocalizations elicit responses from caregivers and are the best predictors of the transition to language: A longitudinal investigation of infants' vocalizations, gestures and word production. *Developmental Science*, *23*(1), e12843. https://doi.org/10.1111/desc.12843

Fenson, L., Marchman, V. A., Thal, D. J., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: User's guide and technical manual* (2nd editio). Baltimore: Paul H. Brookes.

Fernald, A., Perfors, A., & Marchman, V. A. (2006). Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental Psychology*, *42*(1), 98–116. https://doi.org/10.1037/0012-1649.42.1.98

Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. (2008). Looking while listening: Using eye movements to monitor spoken language comprehension by infants and young children. In *Developmental Psycholinguistics: Online methods in children's language processing* (I. A. Sekerina, E. M. Fernández, & H. Clahsen (Eds.), Vol. 44, pp. 97–135). Amsterdam, The Netherlands: John Benjamins.

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., ... Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American Journal of Speech-Language Pathology*, *26*(2), 248–265. https://doi.org/10.1044/2016_AJSLP-15-0169

Gilkerson, J., Richards, J. A., Warren, S. F., Oller, D. K., Russo, R., & Vohr, B. (2018). Language experience in the second year of life and language outcomes in late childhood. *Pediatrics*, *142*(4), e20174276. https://doi.org/10.1542/peds.2017-4276

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. *Child Development*, *71*(4), 878–894. https://doi.org/10.1111/1467-8624.00197

Goodwyn, S. W., Acredolo, L. P., & Brown, C. A. (2000). Impact of symbolic gesturing on

643   early language development. *Journal of Nonverbal Behavior, 24*(2), 81–103.

644   https://doi.org/10.1023/A:1006653828895

645   Hart, B., & Risley, T. (1995). *Meaningful differences in the everyday experience of young*

646   *American children.* Paul H Brookes Publishing.

647   Hoff, E. (2003). The specificity of environmental influence: Socioeconomic status affects

648   early vocabulary development via maternal speech. *Child Development, 74*(5), 1368–1378.

649   https://doi.org/10.1111/1467-8624.00612

650   Hurtado, N., Marchman, V. A., & Fernald, A. (2008). Does input influence uptake? Links

651   between maternal talk, processing speed and vocabulary size in Spanish-learning children.

652   *Developmental Science, 11*(6), 31–39. https://doi.org/10.1111/j.1467-7687.2008.00768.x

653   Huttenlocher, J., Haight, W., Bryk, A., Seltzer, M., & Lyons, T. (1991). Early vocabulary

654   growth: Relation to language input and gender. *Developmental Psychology, 27*(2),

655   236–248. https://doi.org/10.1037/0012-1649.27.2.236

656   Huttenlocher, J., Waterfall, H., Vasilyeva, M., Vevea, J., & Hedges, L. V. (2010). Sources of

657   variability in children's language growth. *Cognitive Psychology, 61*(4), 343–365.

658   https://doi.org/10.1016/j.cogpsych.2010.08.002

659   Iverson, J., M., Capirci, O., Longobardi, E., & Caselli, M. C. (1999). Gesturing in

660   mother-child interactions. *Cognitive Development, 14*(1), 57–75.

661   https://doi.org/10.1016/S0885-2014(99)80018-5

662   Jackson-Maldonado, D., Thal, D., J., & Fenson, L. (2003). *MacArthur Inventarios del*

663   *Desarrollo de Habilidades Comunicativas: User's guide and technical manual.* Brookes

664   Publishing.

665   Kalagher, H., & Yu, C. (2006). The effects of deictic pointing in word learning. *Proceedings*

666   *of the 5th International Conference of Development and Learning.* Bloomington, IN.

667   Mahr, T., & Edwards, J. (2018). Using language input and lexical processing to predict

668   vocabulary size. *Developmental Science, 21*(6), e12685.

669   https://doi.org/10.1111/desc.12685

Marchman, V. A., Weisleder, A., Hurtado, N., & Fernald, A. (2021). Accuracy of the Language Environment Analyses (LENA) system for estimating child and adult speech in laboratory settings. *Journal of Child Language*, *48*(3), 605–620. https://doi.org/10.1017/S0305000920000380

Matthews, D., Behne, T., Lieven, E., & Tomasello, M. (2012). Origins of the human pointing gesture: A training study. *Developmental Science*, *15*(6), 817–829. https://doi.org/10.1111/j.1467-7687.2012.01181.x

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan.* Chapman; Hall/CRC.

McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, *119*(4), 831–877. https://doi.org/10.1037/a0029872

Pace, A., Luo, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2017). Identifying pathways between socioeconomic status and language development. *Annual Review of Linguistics*, *3*, 285–308. https://doi.org/10.1146/annurev-linguistics-011516-034226

Pan, B. A., Rowe, M. L., Singer, J. D., & Snow, C. E. (2005). Maternal correlates of growth in toddler vocabulary production in low-income families. *Child Development*, *76*(4), 763–782. https://doi.org/10.1111/j.1467-8624.2005.00876.x

Puccini, D., Hassemer, M., Salomo, D., & Liszkowski, U. (2010). The type of shared activity shapes caregiver and infant communication. *Gesture*, *10*(2-3), 279–296. https://doi.org/10.1075/gest.10.2-3.08puc

Reynolds, E., Vernon-Feagans, L., Bratsch-Hines, M., Baker, C., & Investigators, T. F. L. P. K. (2019). Mothers' and fathers' language input from 6 to 36 months in rural two-parent-families: Relations to children's kindergarten achievement. *Early Childhood Research Quarterly*, *47*(2), 385–395. https://doi.org/10.1016/j.ecresq.2018.09.002

Rosemberg, C. R., Alam, F., Audisio, C. P., Ramirez, M. L., Garber, L., & Migdalek, M. J. (2020). Nouns and verbs in the linguistic environment of Argentinian toddlers:

Socioeconomic and context-related differences. *First Language*, *40*(2), 192–217. https://doi.org/10.1177/0142723719901226

Rowe, M. L. (2000). Pointing and talk by low-income mothers and their 14-month-old children. *First Language*, *20*(60), 305–330. https://doi.org/10.1177/014272370002006005

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 in vocabulary development: Child-directed speech and vocabulary. *Child Development*, *83*(5), 1762–1774. https://doi.org/10.1111/j.1467-8624.2012.01805.x

Rowe, M. L., & Goldin-Meadow, S. (2009). Differences in early gesture explain SES disparities in child vocabulary size at school entry. *Science*, *323*(5916), 951–953. https://doi.org/10.1126/science.1167025

Rowe, M. L., & Leech, K. A. (2019). A parent intervention with a growth mindset approach improves children's early gesture and vocabulary development. *Developmental Science*, *22*(4), e12792. https://doi.org/10.1111/desc.12792

Rowe, M. L., Özçalışkan, Ş., & Goldin-Meadow, S. (2008). Learning words by hand: Gesture's role in predicting vocabulary development. *First Language*, *28*(2), 182–199. https://doi.org/10.1177/0142723707088310

Rowe, M. L., Wei, R., & Salo, V. C. (2022). Early gesture predicts later language development. In A. Morgenstern & S. Goldin-Meadow (Eds.), *Gesture in language: Development across the lifespan* (A. Morgenstern & S. Goldin-Meadow (Eds.), pp. 93–111). Boston: De Gruyter Mouton. https://doi.org/10.1037/0000269-004

Rowe, M. L., & Weisleder, A. (2020). *Language development in context. 2*, 201–223. https://doi.org/10.1146/annurev-devpsych-042220-121816

Salo, V. C., Reeb-Sutherland, B., Frenkel, T. I., Bowman, L. C., & Rowe, M. L. (2019). Does intention matter? Relations between parent pointing, infant pointing, and

developing language ability. *Journal of Cognition and Development*, *20*(5), 635–655. https://doi.org/10.1080/15248372.2019.1648266

Salomo, D., & Liszkowski, U. (2013). Sociocultural settings influence the emergence of prelinguistic deictic gestures. *Child Development*, *84*(4), 1296–1307. https://doi.org/10.1111/cdev.12026

Shneidman, L. A., & Goldin-Meadow, S. (2012). Language input and acquisition in a Mayan village: How important is directed speech? *Developmental Science*, *15*(5), 659–673. https://doi.org/10.1111/j.1467-7687.2012.01168.x

Suskind, D. L., Leffel, K. R., Graf, E., Hernandez, M. W., Gunderson, E. A., Sapolich, S. G., . . . Levine, S. C. (2016). A parent-directed language intervention for children of low socioeconomic status: A randomized controlled pilot study. *Journal of Child Language*, *43*(2), 366–406. https://doi.org/10.1017/S0305000915000033

Tardif, T., Shatz, M., & Naigles, L. (1997). Caregiver speech and children's use of nouns versus verbs: A comparison of English, Italian, and Mandarin. *Journal of Child Language*, *24*, 535–565. https://doi.org/10.1017/S030500099700319X

Team, R. C. (2021). *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Tfouni, L. V., & Klatzky, R. L. (1983). A discourse analysis of deixis: Pragmatic, cognitive and semantic factors in the comprehension of "this,""that,"'here'and "there." *Journal of Child Language*, *10*(1), 123–133.

Tincoff, R., Seidl, A., Buckley, L., Wojcik, C., & Cristia, A. (2019). Feeling the way to words: Parents' speech and touch cues highlight word-to-world mappings of body parts. *Language Learning and Development*, *15*(2), 103–125. https://doi.org/10.1080/15475441.2018.1533472

Villiers Rader, N. de, & Zukow-Goldring, P. (2012). Caregivers' gestures direct infant attention during early word learning: The importance of dynamic synchrony. *Language*

Sciences, *34*(5), 559–568. https://doi.org/10.1016/j.langsci.2012.03.011

Walker, D., Greenwood, C., Hart, B., & Carta, J. (1994). Prediction of school outcomes based on early language production and socioeconomic factors. *Child Development*, *65*(2), 606–621. Retrieved from https://doi.org/10.1111/j.1467-8624.1994.tb00771.x

Weber, A., Fernald, A., & Diop, Y. (2017). When cultural norms discourage talking to babies: Effectiveness of a parenting program in rural senegal. *Child Development*, *88*(5), 1513–1526. https://doi.org/10.1111/cdev.12882

Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, *24*(11), 2143–2152. https://doi.org/10.3399/096016407782317928

Wittenburg, P., Brugman, H., Russel, A., Klassmann, A., & Sloetjes, H. (2006). ELAN: A professional framework for multimodality research. *5th international conference on language resources and evaluation (LREC 2006)*, 1556–1559.

Yuksel, P., & Brooks, P. J. (2017). Encouraging usage of an endangered ancestral language: A supportive role for caregivers' deictic gestures. *First Language*, *37*(6), 561–582. https://doi.org/10.1177/0142723717713502

Zukow-Goldring, P. (1996). Sensitive caregiving fosters the comprehension of speech: When gestures speak louder than words. *Early Development and Parenting: An International Journal of Research and Practice*, *5*(4), 195–211. https://doi.org/10.1002/(SICI)1099-0917(199612)5:4%3C195::AID-EDP133%3E3.0.CO;2-H

Conflict of Interest Disclosure

The authors declare no potential conflicts of interest.

Data Availability Statement

Pre-registration of study design and analyses are available on the Open Science Framework: https://osf.io/s2jqy. The coding protocol is publicly

776 available:https://osf.io/fmvyc/. All data and reproducible code are available on GitHub:

777 [masekd for peer-review]