

RISK FACTORS ASSOCIATED WITH SMALL EXPRESSIVE VOCABULARIES IN PORTUGUESE-SPEAKING TODDLERS

IRENE CADIME¹, ANA LÚCIA SANTOS², FERNANDA LEOPOLDINA VIANA³

¹Psychology Research Center, School of Psychology, University of Minho, Braga, Portugal

²University of Lisbon, School of Arts and Humanities, Center of Linguistics, Lisbon, Portugal

³Research Center on Child Studies, Institute of Education, University of Minho, Braga, Portugal

Corresponding author: Irene Cadime, e-mail: irenecadime@psi.uminho.pt

Received: 31.10.2025

Accepted: 04.01.2026

ORIGINAL RESEARCH ARTICLE

UDK: 81'23: 811.134.3

DOI: <https://doi.org/10.31299/hrri.62.si.10>

Abstract: *This study investigates the individual and contextual factors associated with small expressive vocabularies in Portuguese-speaking toddlers. Data were drawn from the validation study of the European Portuguese MacArthur-Bates Communicative Development Inventory: Words and Sentences, which included 1,042 children between the ages of 24 and 30 months. Boys were nearly three times more likely to have a small expressive vocabulary than girls. Later entry into childcare and lower maternal education were also significant predictors. In contrast, birth order had no significant effect. The probability of having a small expressive vocabulary increased with the number of co-occurring risk factors, thus, underscoring the multifactorial and ecological nature of early language development.*

Keywords: *MacArthur-Bates Communicative Development Inventory, vocabulary, risk factors*

INTRODUCTION

Some children develop language at a slower pace compared to their peers, demonstrating delays in expressive vocabulary acquisition despite the absence of underlying neurological, sensory, or cognitive deficits (Singleton, 2018). In order to identify children with such delays in language development, it is common to use caregiver-completed questionnaires, such as the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 2007). Research suggests that a number of factors can contribute to linguistic development. Sex has consistently been identified as a contributing factor: the male sex has been recognised as a significant predictor of late language emergence and smaller lexicons (Collisson et al., 2016; Zubrick et al., 2007). However, contextual factors, such as those related to family or caregiving environments, may also play a role. Variables such as birth order and maternal education have also been shown to influence the linguistic input received by children within the family context and, subsequently, their vocabulary development

(Desmarais et al., 2008). Nevertheless, these findings remain inconsistent, since other studies have not been able to confirm the significant effects of these variables on children's language development (e.g., Almohammadi et al., 2025; Zubrick et al., 2007). Attendance at childcare centres has also been identified as a protective factor in some research studies (Collisson et al., 2016; Lekhal et al., 2011), however, further investigation is needed to clarify its influence across different contexts. Hence, the present study aimed to examine the effects of sex, maternal education, birth order, and age of entry into a daycare centre as potential risk factors for small expressive vocabularies, while also considering their cumulative effect on a large sample of Portuguese-speaking children.

METHODS

Participants

The data used in the present study were drawn from the database developed for the validation and norming of the European Portuguese version

of the CDI: Words and Sentences (PT-CDI:W&S). Participants were selected from all seven regions of Portugal, using a proportional stratified sampling approach to ensure that the number of children from each region reflected the population distribution (a detailed description of the sample can be found in Silva et al., 2017). In the present study, we included only children between the ages of 24 and 30 months for whom data were available on sex, maternal education, birth order, and age at entry into a childcare centre. The 10th percentile on the CDI word production subscale was used as the cutoff score. A total of 93 children scored below this threshold and they were classified as children with a small expressive vocabulary. The remaining children from the validation study who scored above the cutoff ($n = 949$) served as the comparison group with typical language development. The present study used the following exclusion criteria that were identical to those used in the validation study of the PT-CDI:W&S – prematurity and low birth weight (< 1500 grams), no exposure to European Portuguese in the family context, and the presence of severe medical conditions that could result in language impairment.

Measures

The PT-CDI:W&S (Silva et al., 2017) was developed to assess language in children aged 16 to 30 months. In the present study, the PT-CDI:W&S was used to collect data on expressive vocabulary (word production) in Portuguese-speaking tod-

dlers. The vocabulary checklist comprises 639 words organised into 22 semantic categories (for details, see Silva et al., 2017). Parents were asked to go through the checklist and indicate which words their child could produce. Regarding its psychometric properties, the validation study of the PT-CDI:W&S reported excellent internal consistency for this subscale (Cronbach's $\alpha = .99$; Silva et al., 2017). A demographic questionnaire was also administered to collect information on demographic characteristics of the children and their families. Data collection was authorised by the Portuguese National Committee of Data Protection, and all parents provided informed consent.

RESULTS

Table 1 presents the demographic and background characteristics of children with typical language development (TLD) and those with a small vocabulary (SV). Both groups were similar in terms of chronological age ($t(1040) = 0.79$, $p = .430$). The groups differed in sex distribution: there was a higher proportion of boys in the SV group compared to the TLD group ($X^2(1) = 19.96$, $p < .001$). Maternal education levels also differed between the two groups ($X^2(1) = 5.47$, $p = .019$); in this case, a larger percentage of mothers of children from the SV group had lower educational levels. The groups were similar with respect to birth order ($X^2(1) = 1.80$, $p = .180$). However, children from the SV group entered childcare at a later age than those with TLD ($t(1017) = -5.38$, $p < .001$).

Table 1. Demographic characteristics of participants

	Group	
	TLD	SV
Sex		
Female	454 (47.8%)	22 (23.7%)
Male	495 (52.2%)	71 (76.3%)
Maternal education		
Up to secondary school	547 (58.4%)	64 (71.1%)
Higher education	389 (41.6%)	26 (28.9%)
First-born child		
No	401 (42.3%)	46 (49.5%)
Yes	548 (57.7%)	47 (50.5%)
Mean age in months (SD)	26.9 (1.9)	26.8 (1.9)
Mean age (in months) at entry to childcare (SD)	10.6 (6.7)	14.7 (7.4)

Note: TLD refers to the group of children with typical language development, SV refers to the group of children with a small vocabulary; SD, standard deviation.

A binary logistic regression was conducted to examine the extent to which each of the four variables of interest (sex, maternal education, birth order, and age at entry to childcare) predicted group membership. Although interaction effects were tested, they were not significant and thus, were not included in the final model. The overall model was significant ($\chi^2(4) = 47.15, p < .001$; see Table 2).

Sex, maternal education, and age at entry to childcare were significant predictors of group

status. Specifically, boys were 2.77 times more likely to have a small vocabulary than girls. Children whose mothers had completed higher education were less likely to be part of the SV group compared to those whose mothers had completed a secondary school education. With each one-month increase in age at entry to childcare, children were 1.08 times more likely to be part of the SV group. Birth order did not significantly predict vocabulary development.

Table 2. Results of regression analysis to predict group status

Variables	Beta	Standard error	p	Odds ratio	95% Confidence interval for Odds ratio	
					Lower	Upper
Sex	1.019	.264	< .001	2.770	1.651	4.649
Maternal education	-.523	.257	.041	.593	.358	.980
First-born child	-.385	.236	.102	.680	.429	1.080
Age at entry to childcare (months)	.073	.015	< .001	1.076	1.044	1.108

Note: The outcome variable was group (0 = typical language development, 1 = small vocabulary). Categorical predictors were coded as follows: sex, 0 = girl, 1 = boy; maternal education, 0 = up to secondary school, 1 = higher education; first-born child, 0 = no, 1 = yes.

Cox & Snell $R^2 = .046$; Nagelkerke $R^2 = .104$.

In the next step, we examined the cumulative effect of the identified risk factors associated with a small vocabulary, including being male, having a mother with an educational level up to secondary school, and entering childcare after 9 months of age (9 months was the median age of childcare entry in the full sample). The distribution of risk factors across groups is presented in Table 3. All children in the SV group had at least one risk factor. Additionally, nearly half (47.8%) of the children in the SV group presented two risk factors, compared to 39.1% in the TLD group. The proportion of children with a small vocabulary who had three risk factors (32.2%) was more than double the proportion observed in the TLD group (14.5%).

Table 3. Number of risk factors in each group

Number of risk factors	Group	
	TLD	SV
0	116 (12.4%)	0 (0%)
1	318 (34%)	18 (20%)
2	366 (39.1%)	43 (47.8%)
3	136 (14.5%)	29 (32.2%)

Note: TLD refers to children with typical language development, and SV refers to children with a small vocabulary.

A binary logistic regression was conducted to examine the effect of cumulative risk, using the number of risk factors as a predictor of group membership. The results indicate that the number of risk factors significantly predicted group membership (Beta = 0.79, SE = 0.14, $p < .001$). With each one-unit increase in the number of risk factors, the odds of being part of the SV group were 2.20 times higher (95% CI [1.67, 2.91]).

CONCLUSION

The present study examined individual and contextual factors associated with a small expressive vocabulary in a large sample of Portuguese-speaking toddlers. The findings revealed that boys were nearly three times more likely to have a small expressive vocabulary than girls. This difference may reflect the influence of biological and environmental factors (e.g., parents or other caregivers may engage more often in verbal interaction with girls). Children who entered childcare later were also at increased risk, with each additional month of delay corresponding to a higher likelihood of being identified as part of the

small vocabulary group. Given the observational design of the study, this finding should be understood as correlational, rather than causal. Later entry into childcare may reflect reduced early exposure to diverse linguistic input, but it may also result from parental decisions influenced by early concerns about the child's language development or by broader familial, socioeconomic, or cultural factors. Maternal education emerged as a modest, but significant predictor. On the one hand, lower maternal education levels may be associated with environments that provide fewer opportunities for linguistic stimulation. On the other hand, mothers who have a speech or language impairment may be less likely to attain university-level education, and a family history of language impairment is a known predictor of poor language outcomes in children. Birth order, however, did not significantly contribute to predicting group membership. Taken together, these results underscore the multifactorial nature of early language development, reflecting the interplay between individual predispositions and environmental opportunities. The findings align with previous evidence high-

lighting that sex is a robust biological risk factor (Zubrick et al., 2007; Collisson et al., 2016) and early childcare attendance has a potential protective influence (Lekhal et al., 2011). Importantly, the cumulative analysis demonstrated that the likelihood of a smaller expressive vocabulary increases with the number of co-occurring risk factors, reinforcing the need for a comprehensive and ecological perspective when identifying children at risk for language delays.

Acknowledgment

This work was financially supported by Portuguese national funds through the FCT (Foundation for Science and Technology) within the frameworks of the Research Centre on Child Studies of the University of Minho, reference UID/00317/2025, the Center of Linguistics of the University of Lisbon, reference UID/214/2025 (<https://doi.org/10.54499/UID/00214/2025>) and the Psychology Research Center of the University of Minho, reference UID/01662/2025 (<https://doi.org/10.54499/UID/01662/2025>).

REFERENCES

- Almohammadi, A., Alaslani, K., Alroqi, H., Aljahlan, Y., Alsulaiman, R., Albeeshi, A., Murad, A., & Alnemary, F. (2025). Late talking in young children in Saudi Arabia: Identifying key risk factors. *First Language, 45*(3), 323–375. <https://doi.org/10.1177/01427237251313562>
- Collisson, B. A., Graham, S. A., Preston, J. L., Rose, M. S., McDonald, S., & Tough, S. (2016). Risk and protective factors for late talking: An epidemiologic investigation. *Journal of Pediatrics, 172*, 168-174.e1. <https://doi.org/10.1016/j.jpeds.2016.02.020>
- Desmarais, C., Sylvestre, A., Meyer, F., Bairati, I., & Rouleau, N. (2008). Systematic review of the literature on characteristics of late-talking toddlers. *International Journal of Language & Communication Disorders, 43*(4), 361–389. <https://doi.org/https://doi.org/10.1080/13682820701546854>
- Fenson, L., Marchman, V., Thal, D., Dale, P. S., Reznick, J. S., & Bates, E. (2007). *MacArthur-Bates Communicative Development Inventories: Users guide and technical manual, 2nd edition*. Paul H Brookes.
- Lekhal, R., Zachrisson, H. D., Wang, M. V., Schjølberg, S., & von Soest, T. (2011). Does universally accessible child care protect children from late talking? Results from a Norwegian population-based prospective study. *Early Child Development and Care, 181*(8), 1007–1019. <https://doi.org/10.1080/03004430.2010.508558>
- Silva, C., Cadime, I., Ribeiro, I., Santos, S., Santos, A. L., & Viana, F. L. (2017). Parents' reports of lexical and grammatical aspects of toddlers' language in European Portuguese: Developmental trends, age and gender differences. *First Language, 37*(3), 267–284. <https://doi.org/10.1177/0142723716689274>
- Singleton, N. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics of North America, 65*(1), 13–29. <https://doi.org/https://doi.org/10.1016/j.pcl.2017.08.018>
- Zubrick, S. R., Taylor, C. L., & Rice, M. L. (2007). Late language emergence at 24 months: An epidemiological study of prevalence, predictors, and covariates. *Journal of Speech, Language and Hearing Research, 50*(6), 1562–1592. [https://doi.org/10.1044/1092-4388\(2007/106\)](https://doi.org/10.1044/1092-4388(2007/106))