

## Development of a figurative language comprehension instrument in Brazilian Portuguese: A pilot study with preschoolers

Caroline Girardi Ferrari  
Jagiellonian University, Poland

Maity Siqueira  
Federal University of Rio Grande do Sul, Brazil

**Abstract:** This study pilot-tests a figurative language comprehension instrument in Brazilian Portuguese, evaluating its applicability and sensitivity with preschoolers aged 3 to 6 years. The instrument comprehends four types of figurative phenomena (metaphors, metonymies, idioms, and proverbs) through previously developed individual tasks. This is the first study to combine all these tasks into a single standardized instrument and examine its applicability as a comprehensive psycholinguistic tool. Specifically, the study investigates whether the order of item presentation (grouped by phenomenon versus randomized) affects participants' performance and explores early comprehension patterns of different figurative phenomena within this age range. A total of 34 preschoolers participated in the study. Overall, results indicate that the instrument is appropriate and sensitive to analyze figurative language comprehension. Quantitative analyses showed no significant differences between the versions, whereas qualitative analyses indicated practice effects in the version grouped by phenomenon. Consequently, the randomized presentation was selected, as it balances the different phenomena and reduces potential practice effects. Findings also show that children as young as three years already demonstrate emerging understanding of metaphors and metonymies, while idioms and proverbs are not yet comprehended across the 3- to 6-year-old sample. Taken together, results support the hypothesis of gradual acquisition of figurative phenomena, beginning with greater sensitivity to metonymies and metaphors and subsequently extending to idioms and proverbs. Overall, this study provides preliminary validity evidence for the COMFIGURA instrument, demonstrating that even in a small-scale pilot study, it is sensitive to developmental differences in preschoolers' figurative language comprehension.

**Keywords:** figurative language comprehension; metaphor-related figurative phenomena; psycholinguist instrument; Conceptual Metaphor Theory

**Resumo:** Este trabalho apresenta um estudo-piloto de um instrumento de compreensão de linguagem figurada em português brasileiro, avaliando sua aplicabilidade e sensibilidade com crianças em idade pré-escolar, com idades entre 3 e 6 anos. O instrumento abrange quatro tipos de fenômenos figurados (metáforas, metonímias, expressões idiomáticas e provérbios), por meio de tarefas individuais previamente desenvolvidas. Este é o primeiro estudo a reunir todas essas tarefas em um único instrumento padronizado e a examinar sua aplicabilidade como uma ferramenta psicolinguística abrangente. Especificamente, o estudo investiga se a ordem de apresentação dos itens (agrupados por fenômeno ou randomizados) afeta o desempenho dos participantes e explora padrões iniciais de compreensão dos diferentes fenômenos figurados nessa faixa etária. Participaram do estudo 34 crianças em idade pré-escolar. De modo geral, os resultados indicam que o instrumento é adequado e sensível para analisar a compreensão da linguagem figurada. As análises quantitativas não revelaram diferenças significativas entre as ordens de aplicação, enquanto as análises qualitativas indicaram efeitos de prática na versão com itens agrupados por fenômeno. Conseqüentemente, optou-se pela ordem de apresentação

randomizada, por equilibrar os diferentes fenômenos e reduzir possíveis efeitos de prática. Os resultados também mostram que crianças a partir dos três anos já demonstram uma compreensão emergente de metáforas e metonímias, enquanto expressões idiomáticas e provérbios ainda não são compreendidos pela amostra. Em conjunto, os achados sustentam a hipótese de aquisição gradual dos fenômenos figurados, iniciando-se com maior sensibilidade a metonímias e metáforas e estendendo-se posteriormente a expressões idiomáticas e provérbios. De modo geral, o estudo fornece evidências preliminares de validade ao instrumento, demonstrando sua sensibilidade a diferenças desenvolvimentais na compreensão de linguagem figurada em crianças em idade pré-escolar, mesmo em um estudo-piloto de pequena escala.

**Palavras-chave:** compreensão de linguagem figurada; fenômenos figurados relacionados à metáfora; instrumento psicolinguístico; Teoria da Metáfora Conceitual

**Corresponding author(s):** Caroline Girardi Ferrari, Jagiellonian University, Faculty of Philology, Institute of Romance Studies, Department of Portuguese Studies, al. Mickiewicza 9, 31-120 Kraków, Poland. Email: [caroline.girardi.ferrari@uj.edu.pl](mailto:caroline.girardi.ferrari@uj.edu.pl).

**ORCID ID(s):** Caroline Girardi Ferrari <https://orcid.org/0000-0003-3853-3673>; Maity Siqueira <https://orcid.org/0000-0002-8775-4563>

**Citation:** Ferrari, C. G., & Siqueira, M. (2026). Development of a figurative language comprehension instrument in Brazilian Portuguese: a pilot study with preschoolers. *Language Development Research*, 6(1), 21–50. <http://doi.org/10.34842/ldr2026-1022>

## Introduction

Imagine you are a psychologist assessing a child's language development. You notice that the child struggles with non-literal expressions, and you look for standardized tools to evaluate figurative language comprehension. In Brazil, several widely used instruments exist to assess language comprehension, yet a closer examination reveals a limitation: the sections designed to test figurative language often combine multiple types of figurative expressions, such as metaphors, proverbs and idioms, stating that they're testing only one phenomenon, as metaphors. While related, these phenomena differ in their core characteristics and typical ages of acquisition (Gibbs & Colston, 2012). Consequently, a child might appear to fail at metaphor comprehension not because they lack the ability, but because the test contains items that they are not expected to understand along with those they should already master. This illustrates the current challenges in assessing figurative language comprehension in Brazilian Portuguese.

To address this gap, a figurative language comprehension instrument (COMFIGURA – Instrument for the Assessment of Figurative Language Comprehension) has been developed in Brazilian Portuguese, following the theoretical principles of Cognitive Linguistics and Conceptual Metaphor Theory. This instrument aims to evaluate the comprehension of metaphors, metonymies, idioms, and proverbs in a standardized and sensitive way, respecting the particularities of each phenomenon. Unlike other tools, COMFIGURA treats these phenomena separately while acknowledging their interconnections within a continuum of figurative language (Lakoff & Johnson, 1980; Gibbs & Colston, 2012). This continuum ranges from cognitively grounded and early acquired phenomena such as primary metaphors and metonymies, to culturally embedded and late-acquired constructions such as idioms and proverbs.

Previous research suggests that children's comprehension of figurative phenomena follows different developmental trajectories. Primary metaphors, for instance, are observable in speech by ages 2 and 3 and tend to reach adult status by age 7 (Gaskins & Rundblad, 2023; Özçalışkan, 2005; Siqueira & Gibbs, 2007; Siqueira & Lamprecht, 2007). Metonymy has also been shown to emerge early and may even precede metaphors in some contexts (Nerlich et al., 1999; Rundblad & Annaz, 2010a; Siqueira, Melo et al., 2023). Idioms and proverbs, in contrast, depend more heavily on cultural knowledge and sociopragmatic skills, showing later acquisition, extending into adolescence and even adulthood (Cain et al., 2009; Ferrari, 2024; Ferrari & Siqueira, *submitted*; Nippold et al., 1997). In the case of clinical populations, many conditions present later acquisition when compared to their neurotypical pairs (e.g., Chahboun et al., 2021; Rundblad & Annaz, 2010b; Siqueira et al., 2016; Vulchanova et al., 2015).

Given these conceptual and developmental differences, it is important to have an assessment tool that can distinguish between these phenomena in a systematic way,

and COMFIGURA was designed for this purpose. It is composed of five tasks: one non-verbal metaphor task and four verbal tasks targeting metaphors, metonymies, idioms, and proverbs. Its development started with the verbal and non-verbal metaphor tasks (Siqueira, 2004; Siqueira & Gibbs, 2007; Siqueira & Lamprecht, 2007), later with idioms (Marques, 2018; Siqueira & Marques, 2018), proverbs (Ferrari, 2020; Ferrari & Siqueira, 2020), and most recently with metonymies (Siqueira, Melo et al., 2023). These tasks have already been tested with typical and clinical populations (Afonso, 2012; Cony, 2025; De Leon, 2008; Lopes, 2019; Marques, 2018; Marques et al., 2025), and present converging validity evidence consistent with literature about figurative language comprehension (Ferrari, 2020; Ferrari & Siqueira, 2020; *submitted*; Siqueira, Ferrari et al., 2023, Siqueira & Marques, 2018; Siqueira, Melo et al., 2023).

Until now, however, all these tasks had only been tested separately, analysing the comprehension of each phenomenon individually. There were no guidelines or evidence about their combined use as a single figurative language instrument. This raised important methodological questions: if the tasks are put together, would the instrument still be effective? Could the integration of the tasks be considered a real standardized instrument? Should the instrument follow the canonical order of the five tasks, grouping items by phenomenon and gradually increasing complexity? Or should all items be randomized, mixing different figurative phenomena and complexity levels in the instrument?

Motivated by these questions, the present pilot study aims to evaluate the applicability and sensitivity of the instrument with preschoolers, with the expectation that task structure would influence performance patterns across items and figurative phenomena. More specifically, it aims to investigate whether the order of item presentation (by phenomenon vs. randomized) affects performance, and to explore early comprehension patterns of different figurative phenomena in children aged 3 to 6 years.

### **Theoretical Background**

In Cognitive Linguistics, it is assumed that figurative language phenomena are distinct but interrelated, being organized in a continuum of figures of speech, with different levels of complexity (Gibbs & Colson, 2012; Siqueira et al., *in press*). This continuum includes phenomena located at the level of a single word, such as metaphor or metonymy, as well as phenomena that take place at the level of a whole sentence construction, such as idioms and proverbs. Within this continuum, metaphor is the central phenomenon, described by the theory not only as a linguistic device, but also as a cognitive and conceptual one: the understanding of one conceptual domain in terms of another conceptual domain (Lakoff & Johnson, 1980). In other words, it is the understanding of abstract, difficult, complex, or less well-delineated concepts (such as emotions) in terms of more accessible and familiar ideas (such as heat)

(Gibbs, 1994). When one says *The meeting was on fire*, one is instantiating a metaphor, motivated by the mapping INTENSITY OF EMOTION IS HEAT, not to talk about literal fire, but to say that the meeting had intense discussions, for example. Other expressions that instantiate this same metaphorical mapping are *She was boiling with anger*, *The heated debate lasted long*, *The player is fuming after the match*, *His words left me cold*, and many others. Metaphors as such, that rely on embodied experiences and have the potential for universality are called primary metaphors (Grady, 1997). GOOD IS UP (*My spirits are high today*, *She has high hopes this year*, *Our mood rises as the class starts*) and MORE IS UP (*The prices went up again*, *The number of applicants is climbing*, *She's rising for fame*) are other examples of primary metaphors.

Studies suggest that neurotypical children already show early signs of metaphor understanding between the ages of 2 and 3, in spontaneous speech and naturalistic corpora (Gaskins & Rundblad, 2023; Gaskins et al., 2024), with rapid progress at ages 4 and 5 and near-adult performance around ages 7 to 12 (Ferrari, 2024; Ferrara et al., 2025; Özçalışkan, 2005; Siqueira, 2004; Siqueira & Gibbs, 2007; Siqueira & Lamprecht, 2007). Specifically with regard to primary metaphors, consistent results in Portuguese, English, and Turkish, obtained through verbal and pictorial tasks, reinforce the potential for universality of the phenomenon. This is characterized by a continuous developmental trajectory: coherent responses already in early childhood, significant advances up to age 6, and near-adult status between ages 7 and 8 (Özçalışkan, 2005; Rundblad & Annaz, 2010a; Siqueira, 2004; Siqueira & Lamprecht, 2007; Siqueira & Gibbs, 2007; Stites & Özçalışkan, 2013).

Metonymy is another highly recurrent phenomenon in discourse, defined as a figure of language and thought in which one entity is used to refer to, and thus provide access to, another related entity (Littlemore, 2015). It is often used when, for example, the name of a writer is used to refer to their work, an object to refer to what it contains, or a place to refer to the people who work in that place. By establishing these relations, one entity (the one mentioned) provides access to another (the one intended to be referred to), carrying out a process in which one aspect of something is used to refer to another aspect, or to a whole related to that aspect (Gibbs, 1994; Langacker, 1993; Littlemore, 2015). Like metaphors, metonymies are also motivated by metonymic mappings, such as AUTHOR FOR WORK (*I'm reading Shakespeare*, meaning Shakespeare's books, not the person), CONTAINER FOR CONTENT (*The kettle is boiling*, meaning the water, not the kettle itself), PLACE FOR PERSON (*Hollywood is obsessed with remakes*, meaning the people in the film industry, not the district), etc. These are associations between related aspects within the same domain, which enable the intended meaning of expressions to be communicated in a direct and efficient way, with the minimum possible cognitive effort (Panther & Radden, 1999).

The comprehension of metonymies also seems to emerge very early in children. Studies indicate that metonymies are already understood at the age of 2 (Nerlich et al.,

1999; Siqueira, Melo et al., 2023) and tend to be more easily grasped than metaphors in childhood (Rundblad & Annaz, 2010a). By the ages of 4 and 5, children are already able to recognize specific mappings (Zhu, 2021), although comprehension only reaches adult status around the ages of 7 and 10, or even after 12 (Ferrari, 2024; Nerlich et al., 1999; Rundblad & Annaz, 2010a). Some authors further argue for a U-shaped curve, with better performance at age 3 than at ages 4 and 5, possibly linked to the development of metalinguistic awareness (Falkum et al., 2017; Miorando & Siqueira, 2024).

While metaphor and metonymy are more central phenomena, idioms and proverbs are more complex, involving stronger cultural influences. Idioms are defined here as relatively fixed and institutionalized constructions, consisting of two or more words, whose meaning cannot be derived solely from the compositionality of their parts (Bosers, 2014; Langlotz, 2006). Typically formed by combinations involving verbs and complements, idioms often require a higher level of abstraction in discourse and reflect cultural aspects to a greater extent (Gibbs & Colston, 2012). Their meanings, often opaque in relation to sentential constructions, are socially constructed, revealing values that are appreciated and conventionalized within specific linguistic communities (Nayak & Gibbs, 1990). They are non-compositional constructions, with varying degrees of opacity, and are frequently motivated by metaphorical and metonymic mappings (Gibbs, 1994). The expression *To make a tempest in a teapot*, for instance, is motivated by the conceptual mappings REACTIONS ARE NATURAL PHENOMENA, since exaggerated reactions are described in terms of a storm; IMPORTANCE IS SIZE, since a teacup, presumably small, is used to represent something that should not be a concern; and CIRCUMSTANCES ARE CONTAINERS, as the circumstances of a reaction are represented as being contained in a teacup.

Regarding idiom comprehension, research across different languages shows that, from the age of 5, children are already sensitive to the figurative nature of more decomposable idioms (Caillies & Le Sourn-Bissaoui, 2006; Gibbs, 1991). However, comprehension remains largely literal until the age of 6, developing significantly between ages 8 and 9, especially in narrative contexts (Gibbs, 1991; Vulchanova et al., 2011). Development accelerates between 7 and 10 years and reaches near-adult levels around ages 11 to 12 (Cain et al., 2009; Ferrari, 2024; Hamdan & Smadi, 2021; Siqueira et al., 2017a; Siqueira & Marques, 2018). Moreover, Ferrari (2024) indicates a continuous progression until about age 18, suggesting that idiom comprehension is a late-acquired phenomenon, possibly modulated by factors such as decomposability, context, schooling, and linguistic exposure.

Finally, proverbs are well-known popular sayings that express social knowledge and wisdom in simple sentences, reflecting collective experiences and cultural values. They convey lessons about life in a concise manner, which contributes to such sayings enduring across generations. In the theoretical framework adopted here, proverbs

are fixed, non-literal, and familiar sentences that express social norms, morals, and widely recognized truths within a linguistic community (Gibbs & Beitel, 1995). In addition to conveying morals, proverbs are also motivated by metaphorical and metonymic mappings (Gibbs & Beitel, 1995). Moreover, Lakoff and Turner (1989) suggest that, beyond specific mappings, there is a single underlying conceptual mapping across all proverbs: GENERIC IS SPECIFIC. This metaphor allows a general concept (such as a situation) to be understood through a specific concept (such as a proverb). According to the authors, this schema is part of a larger framework of understanding (the Great Chain of Being), which postulates a structured hierarchy of human existence. Consider the proverb *A closed mouth catches no flies*, which is a simple sentence, typically used in a fixed form, and conveys the moral lesson that remaining silent is wiser than speaking unnecessarily. When analysing its underlying motivations, Siqueira, Pereira et al. (2017) suggest that the proverb is motivated by the conceptual metaphor UNPLEASANT CONSEQUENCES ARE INSECTS, in which the fly represents the negative consequences of what is said. Ferrari and Siqueira (2023) further suggest the conceptual metonymy THE STATE FOR ITS EFFECT, since the state of having a closed mouth stands for its effect, silence. Beyond these specific mappings, through the GENERIC IS SPECIFIC metaphor and the Great Chain of Being, the proverb allows to move from a specific scenario (speaking unnecessarily) to a general moral inference (the consequences of speaking unnecessarily), in line with the hierarchy proposed by the framework. In other words, the abstract moral is not explicit but is inferred from the specific situation in the proverb.

Studies in different languages show that the comprehension of proverbs in neurotypical individuals is limited until around the age of 5, when literal interpretations predominate (Ferrari, 2024; Ferrari & Siqueira, 2020; *submitted*). Between 6 and 11 years, a gradual development is observed, especially for more concrete and familiar proverbs (Nippold et al., 1988; Nippold & Haq, 1996). From 12 years onward, there is significant progress, with greater capacity for abstraction (Duthie et al., 2008) and overall better performance (Ferrari, 2024; Ferrari & Siqueira, *submitted*; Nippold et al., 1988; 1998). In late adolescence (15 to 17 years), comprehension tends to approach adult levels (Ferrari & Siqueira, *submitted*; Nippold et al., 1997; 1998; Nippold & Haq, 1996), accompanied by mastery of abstract vocabulary (Nippold et al., 2000). Nevertheless, development can continue into adulthood, with decline after 60 years (Nippold et al., 1997; Uekermann et al., 2008). Besides age, variables such as vocabulary, reasoning, reading, and schooling have a strong influence on the comprehension of the phenomenon (Fernandes, 2018; Ferrari, 2020; 2024; Ferrari & Siqueira, *submitted*; Nippold et al., 1997; 2000; 2001).

## Method

### Participants

In this study, a sample of preschool children was interviewed in two private early childhood education schools in the city of Porto Alegre, Brazil. As an inclusion criterion, all participants were required to be native speakers of Brazilian Portuguese. Clinical diagnoses of psychological or neurological conditions, as reported by parents or teachers, were used as an exclusion criterion. In total, 34 children were interviewed, with ages ranging from 3 years and 2 months to 6 years and 6 months ( $M = 4.9$  years;  $SD = 1$  year). Of these, 13 were enrolled in nursery-level classes and 21 in kindergarten-level classes.

Among the demographic data collected were parental education level, children's exposure to reading, and the languages spoken by both parents and children. Regarding parental education, 5.88% had completed high school, 55.88% held a university degree, and 38.24% had completed postgraduate studies. As for reading exposure, 11.76% of the participants were considered to have low exposure, 50% moderate exposure, and 38.24% high exposure to reading. It is important to note that among the 34 participants, only two children enrolled in kindergarten-level classes were reported as literate. In terms of language background, all children were native speakers of Portuguese. Of the 34 children, 76.47% were considered learners of English as an additional language. Since both schools offer English classes for young children, it was assumed that all participants had some contact with the language.

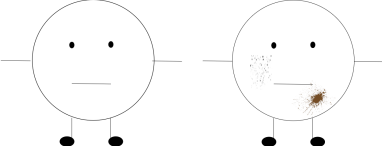
To test for potential effects of stimulus presentation in our instrument, children were randomly assigned to two groups: one with items organized by figurative phenomenon (hereafter, OP), and another with randomized items (hereafter, OR). In the OP group, the mean age was 4.8 years ( $SD = 1$  year), with 64.71% of the participants enrolled in kindergarten-level classes and 35.29% in nursery-level classes. In the OR group, the mean age was 4.4 years ( $SD = 1.1$  years), with 58% of participants in kindergarten and 41.18% in nursery.

### Materials

Two instruments were used here: COMFIGURA and a demographic questionnaire. COMFIGURA is an instrument for assessing figurative language comprehension, composed of five tasks: two for metaphors (one verbal and one non-verbal); one for metonymies; one for idioms; and one for proverbs. All tasks follow the same pattern: an example item, followed by six test items. Each item consists of a figurative stimulus

sentence (or picture), an open-ended question, and a closed-ended question with two answer options.

**Table 1. Examples of COMFIGURA items**

STIMULI	QUESTIONS	EXPECTED ANSWERS
<p><b>Non-verbal metaphor</b></p> 	<p>a) Aponte para o Duni perfeito. [Point to the perfect one.] a') Por que ele é o perfeito? [Why is it the perfect one?]</p>	<p>a) O que está limpo. [The one that's clean.] a') Porque está limpo. [Because it's clean.]</p>
<p><b>Verbal metaphor</b> Sujou para o lado do Renato [Renato's situation stinks.]</p>	<p>a) Como estão as coisas para Renato? [How are things going for Renato?] a') As coisas deram certo ou errado para ele? [Are things going wrong or right for him?]</p>	<p>a) Ruins, erradas, difíceis, complicadas, feias, más. [Bad, wrong, difficult, complicated, ugly, evil.] a') Errado. [Wrong.]</p>
<p><b>Metonymy</b> O prédio todo foi à festa de aniversário. [All the building went to the birthday party.]</p>	<p>a) Quem foi à festa de aniversário? [Who went to the birthday party?] a') Quem foi à festa foram as pessoas que moram no prédio ou o prédio? [Was it the building or the people who live in it who went to the birthday party?]</p>	<p>a) As pessoas, os moradores do prédio, os vizinhos, os amigos, os colegas, algo que remeta à ideia de pessoas. [People, the residents of the building, neighbors, friends, colleagues, anything that refers to the idea of people.] a') As pessoas. [The people.]</p>
<p><b>Idiom</b> Ana e Lia saíram como um par de vasos. [Ana and Lia went out like a pair of vases]</p>	<p>a) Como elas estavam? [How were they?] a') As roupas delas eram diferentes ou iguais? [Were their clothes different or the same?]</p>	<p>a) Estavam se vestindo igual, da mesma forma. [They were dressing the same, in the same way.] a') As roupas eram iguais. [Their clothes were the same.]</p>
<p><b>Proverb</b> A mentira tem perna curta [Lies have short legs]</p>	<p>a) O que isso quer dizer? [What does that mean?] a') O ditado quer dizer que uma mentira demora ou não demora para ser descoberta? [Does the saying mean that a lie takes time or not to be discovered?]</p>	<p>a) As mentiras são logo descobertas. [The lies are soon discovered.] a') Não demora. [It does not take time.]</p>

In general, the procedure begins with the open-ended question, and the closed-ended question is subsequently administered if the participant does not reach the expected answer in the open-ended question. In the Non-verbal Metaphor Task, however, the procedure is reversed: the closed-ended question is asked first, followed by the open-ended to justify the closed-ended response. Table 1 illustrates one example item for each task addressed by COMFIGURA. Note that literal translations from Portuguese to English are provided in square brackets.

Participants' scores on the task range from 0 to 2 points per item, with each question being worth 1 point. A participant receives 0 points if they do not provide the expected answers and 1 point if they reach the expected answer. In verbal items, the participant may receive 2 points straightaway if they answer the open-ended question correctly, demonstrating full comprehension of the figurative expression, and dismissing the closed-ended question. The total score ranges from 0 to 60 points, with each task contributing up to 12 points.

As this was the first pilot application including all the tasks comprised by COMFIGURA, the stimuli were initially gathered from the latest versions of each individual task. They were then analyzed and standardized to ensure consistency across the whole instrument. Necessary controls included the length of stimulus sentences and questions, using words easily understood by children, and employing similar syntactic structures. Minor adjustments of this kind were made for items in the metaphor and idiom tasks, which had been developed earlier than the proverb and metonymy tasks. More substantial adjustments were carried out in two tasks. In the non-verbal metaphor task, all illustrations were updated to higher-quality vector images while maintaining the same mappings and intended meanings. Additionally, one item was modified in the idiom task. After conducting a familiarity task with 210 adult participants ( $M = 41.07$  years;  $SD = 15.7$ ), it was observed that one of the original idioms was outdated, being the only item to receive less than 70% of "highly familiar" responses on the familiarity scale. Consequently, this item was replaced with a new idiom that received 98% of "highly familiar" responses on the same scale, receiving similar rates to the other idioms in the task. The metonymy and proverb tasks did not require any adjustments. After adjustments, the items were considered standardized in the format of the instrument.

The presentation of the items on the instrument was then organized in two formats: one with the items organized by phenomena (OP), and another with all items in randomized order (OR). In the OP group, the Instrument began with the Non-Verbal Metaphor task, followed by the verbal Metaphor, Metonymy, Idiom, and Proverb tasks, always in this order, without mixing phenomena. In this order, each example item was presented as the first in its block. That is, when testing non-verbal metaphors, testing began with an example item, followed by six test items; this was followed by a verbal metaphor example and six verbal metaphor test items. The same order was

used for all phenomena. In the OR group, the Instrument began with five example items: non-verbal metaphor, followed by verbal metaphor, metonymy, idiom, and proverb. Next, the testing items began, interleaving items from each phenomenon. To randomize the items, the five tasks were grouped, and all their items were sequentially numbered from 1 to 30. These numbers were then randomized using a free digital randomization tool (available at: <https://www.random.org/lists/>). After obtaining the randomized list from the tool, the results were checked to ensure that items from the same task did not appear consecutively. When this occurred, they were swapped with an item from a different task. In the end, a list of 30 items interleaving the phenomena was obtained, ensuring the items' randomness. The randomized order was the same for all participants.

### **Procedure**

Data collection was conducted between November and December 2022. The interviews took place in two nursery schools selected by convenience in the city of Porto Alegre, Rio Grande do Sul, Brazil. For the most part, the interviews were conducted individually, but for children who showed any signs of discomfort, a school monitor accompanied without interfering with the participants' responses. Participants were welcomed with a playful animal-painting activity. During this activity, the researcher explained that she would say some sentences and ask a few questions. After participants consented, the researcher started the interview with the COMFIGURA instrument. When the participant showed signs of discomfort or fatigue, testing was paused and resumed at a later time. The interviews lasted approximately 20 minutes per participant, including the welcoming, the playful activity, the testing session, and the conclusion of the interview.

After the interviews, each participant's data were anonymized, and their responses were transcribed into a database. At the end of data collection, the responses were evaluated by the researchers and classified as expected (1) or not expected (0), according to the guidelines proposed by the own instrument (see Table 1 for examples of expected answers).

### **Results**

This section presents the results of the statistical, descriptive, and qualitative analyses conducted. Data were analyzed using R (version 4.4.1), through RStudio. The significance level adopted for all analyses was set at  $\alpha = 0.05$ . Note that the quantitative results should be interpreted with caution, given the pilot nature of the study and the small sample size. For this reason, the statistical analyses presented here are

complemented by descriptive and qualitative analyses, which allow a better examination of participants' responses and contribute to the interpretation of the findings.

Analysis started aiming at the examination of differences between the two testing groups, OR and OP. A Wilcoxon test was performed to examine statistically significant differences between the OR and OP groups, considering the scores of open- and closed-ended questions for all items, as well as the total scores by phenomenon and in the whole instrument. Given the small sample size and the multiple comparisons performed across items, groups, phenomena, types of questions, and scores, *p*-values were adjusted using the Benjamini–Hochberg false discovery rate procedure. Results (Table 2) show that the comparisons were not statistically significant at any level (item-level, task-level, or overall instrument scores). This suggests that performance was similar regardless of the order in which items were presented, with slightly higher percentages of expected responses in the OP group.

**Table 2. Wilcoxon test results organized by phenomenon**

Task	Item	Presentation order		Type of question	Percentage of expected answers		W	p-value
		OP	OR		OP	OR		
Non-verbal metaphors	HAPPINESS IS UP	1	27	Closed-ended	70%	76%	153	.844
				Open-ended	29%	41%	161.5	.745
	GOOD IS LIGHT	2	3	Closed-ended	94%	88%	136	.796
				Open-ended	58%	41%	119	.745
	EMOTIONAL INTIMACY IS PROXIMITY	3	20	Closed-ended	82%	76%	136	.844
				Open-ended	52%	64%	161.5	.745
	INTENSITY OF EMOTION IS HEAT	4	7	Closed-ended	11%	47%	195.5	.496
				Open-ended	5%	41%	195.5	.484
	DIFFICULTY IS WEIGHT	5	14	Closed-ended	70%	70%	144.5	1
				Open-ended	52%	64%	161.5	.745
	IMPORTANCE IS SIZE	6	30	Closed-ended	82%	82%	144.5	1
				Open-ended	64%	58%	136	.844
TASK TOTAL				Closed-ended	68%	73%	167.5	.550
				Open-ended	44%	51%	170.5	.550
				<b>General</b>	56%	62%	670.5	.610
Verbal metaphors	HAPPINESS IS UP	7	25	Open-ended	29%	17%	127.5	.745
				Closed-ended	64%	35%	102	.569
	GOOD IS LIGHT	8	16	Open-ended	35%	23%	127.5	.745
				Closed-ended	88%	47%	85	.484
	INTIMITY IS PROXIMITY	9	5	Open-ended	5%	11%	153	.796
				Closed-ended	82%	76%	136	.844
	INTENSITY OF EMOTION IS HEAT	10	23	Open-ended	29%	11%	119	.745
				Closed-ended	70%	41%	102	.569
	DIFFICULTY IS WEIGHT	11	12	Open-ended	29%	11%	119	.745
				Closed-ended	88%	100%	161.5	.745

Task	Item	Presentation order		Type of question	Percentage of expected answers		W	p-value	
		OP	OR		OP	OR			
Metonymies	IMPORTANCE IS SIZE	12	18	Open-ended	35%	23%	127.5	.745	
				Closed-ended	52%	47%	136	.844	
	<b>TASK TOTAL</b>				Open-ended	27%	16%	107.5	.550
					Closed-ended	74%	57%	68	.063
					<b>General</b>	50%	37%	427.5	.310
	CONTAINER FOR ITS CONTENT	13	4	Open-ended	29%	47%	170	.745	
				Closed-ended	76%	100%	178.5	.523	
	PART FOR THE WHOLE	14	2	Open-ended	5%	0%	136	.745	
				Closed-ended	70%	70%	144.5	1	
	AUTHOR FOR THE BOOK	15	28	Open-ended	0%	0%	144.5	NaN	
Closed-ended				70%	76%	153	.844		
INSTRUMENT FOR THE ACTION	16	13	Open-ended	17%	11%	136	.844		
			Closed-ended	76%	88%	161.5	.745		
CLOTHES FOR THE PERSON	17	26	Open-ended	35%	35%	144.5	1		
			Closed-ended	64%	82%	170	.745		
RECIPIENT FOR ITS CONTENT	18	22	Open-ended	47%	58%	161.5	.745		
			Closed-ended	88%	82%	136	.844		
<b>TASK TOTAL</b>				Open-ended	22%	25%	166.5	.550	
				Closed-ended	74%	83%	186	.550	
				<b>General</b>	48%	54%	646.5	.663	
Idioms	COMPRAR GATO POR LEBRE [To buy a cat instead of a hare]	19	24	Open-ended	0%	5%	153	.745	
				Closed-ended	70%	41%	102	.569	
	METER OS PÉS PELAS MAOS [To put one's feet through one's hands]	20	21	Open-ended	0%	17%	170	.569	
				Closed-ended	76%	64%	127.5	.745	
	FAZER TEMPESTADE EM COPO D'ÁGUA [To make a storm in a glass of water]	21	10	Open-ended	5%	0%	136	.745	
				Closed-ended	76%	52%	110.5	.745	
	QUEBRAR UM GALHO [To break a stick]	22	1	Open-ended	5%	0%	136	.745	
				Closed-ended	41%	47%	153	.844	
	SER A METADE DA LARANJA [To be the half of the orange]	23	19	Open-ended	0%	17%	170	.569	
				Closed-ended	47%	58%	161.5	.745	
TOMAR UM CHÁ DE CADEIRA [To drink a chair tea]	24	8	Open-ended	0%	0%	144.5	NaN		
			Closed-ended	41%	64%	178.5	.745		
<b>TASK TOTAL</b>				Open-ended	1%	6%	163.5	.550	
				Closed-ended	58%	54%	131	.710	
				<b>General</b>	30%	30%	588	.903	

Task	Item	Presentation order		Type of question	Percentage of expected answers		W	p-value	
		OP	OR		OP	OR			
Proverbs	EM BOCA FECHADA NÃO ENTRA MOSCA [A closed mouth catches no flies]	25	9	Open-ended	5%	0%	136	.745	
	Closed-ended			64%	47%	119	.745		
	FILHO DE PEIXE PEIXINHO É [Son of fish is little fish]	26	6	Open-ended	0%	0%	144.5	NaN	
	Closed-ended			29%	29%	144.5	1		
	QUEM VÊ CARA NÃO VÊ CORAÇÃO [Those who see the face don't see the heart]	27	29	Open-ended	0%	0%	144.5	NaN	
	Closed-ended			64%	64%	144.5	1		
	ONDE HÁ FUMAÇA HÁ FOGO [Where there's smoke there's fire]	28	11	Open-ended	0%	0%	144.5	NaN	
	Closed-ended			47%	58%	161.5	.745		
	QUEM NÃO CHORA NÃO MAMA [Those who don't cry don't get breastfed]	29	17	Open-ended	0%	5%	153	.745	
	Closed-ended			52%	42%	127.5	.745		
	CACHORRO QUE LATE NÃO MORDE [Dogs who bark don't bite]	30	15	Open-ended	0%	0%	144.5	NaN	
	Closed-ended			58%	47%	127.5	.745		
	<b>TASK TOTAL</b>				Open-ended	1%	1%	144.5	1
					Closed-ended	52%	48%	119	.550
				<b>General</b>	26%	24%	553	.903	
<b>OVERALL INSTRUMENT TOTAL</b>				Open-ended	19%	20%	155	.729	
				Closed-ended	65%	63%	121	.729	
				<b>General</b>	42%	41%	137	.808	

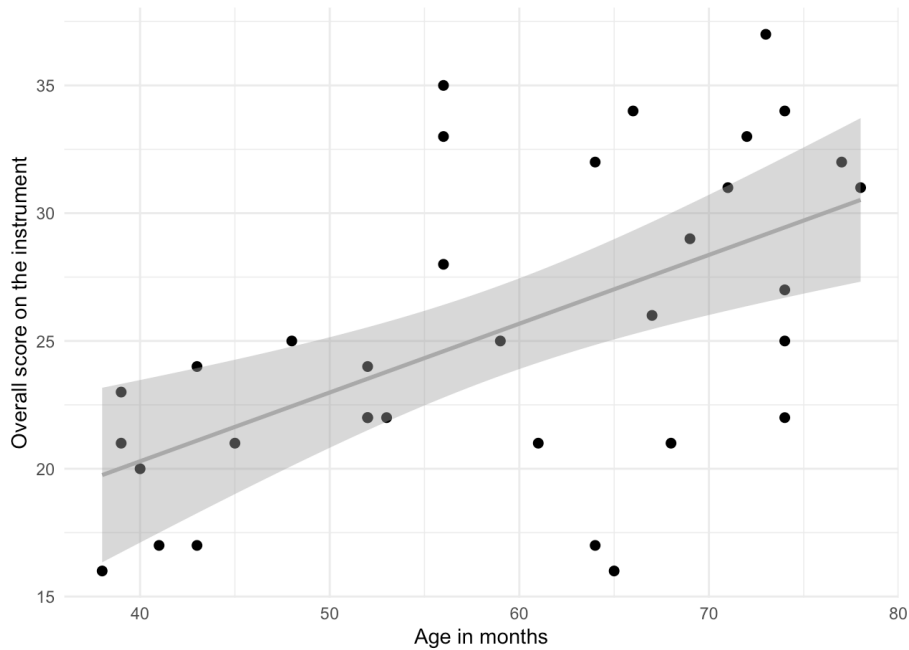
Note. The “Item” column presents the underlying conceptual metaphors and metonymies, as well as the idioms and proverbs included in each item, rather than the items themselves (see Table 1 for example items structure). Literal translations of idioms and proverbs from Brazilian Portuguese into English are provided in square brackets. The “Presentation order” column indicates item numbering across the two conditions: OP = items organized by figurative phenomenon; OR = items presented in randomized order. In the rows, “Task total” represents the overall results for each task, without considering individual items, whereas “Overall instrument total” represents the overall results for the complete instrument, without considering items or tasks individually.

Participants’ performance was also analyzed according to school grade (nursery and kindergarten). Age and school grade are acknowledged here as correlated variables, with school grade often overlapping with age in research with children. However, this analysis was conducted for exploratory purposes, given that, in the Brazilian educational context, school grade is often associated with differences in early literacy exposure and formal educational practices, which may influence language

performance even at an early age (see Borges et al., 2023; Nascimento, 2023). Of note, at this early developmental stage, differences in school grade should be interpreted with caution, as both nursery and kindergarten correspond to pre-literate levels of education. Applying the Wilcoxon test to the total Instrument score, no statistically significant differences were found between the educational groups (nursery:  $W = 781$ ,  $p = .325$ ; kindergarten:  $W = 193$ ,  $p = .461$ ). Nursery children achieved 34% expected responses in the OR order and 37% in the OP order. Kindergarten children achieved 48% expected responses in the OR order and 46% in the OP order. As no significant differences were found between task orders, a second exploratory Wilcoxon test was conducted to analyze differences between the performances of the two educational grades, without considering task order. In this analysis with merged sample, a statistically significant difference ( $W = 434$ ,  $p < .01$ ) was found between the school grades, with kindergarten children outperforming those in nursery. Children in the nursery group achieved 35% expected responses, whereas those in kindergarten achieved 46%.

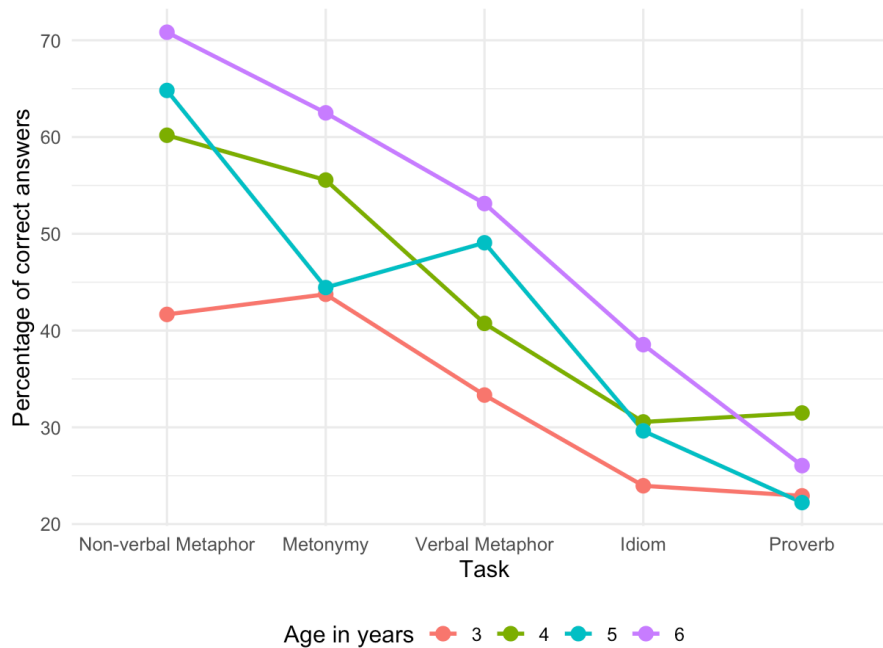
To further explore early comprehension patterns of different figurative phenomena in pre-school children, participants' performance was also analyzed according to age. Since no significant differences were observed between task application orders, data from both orders were grouped for this analysis. Participants were organized into four age groups: 3-year-olds ( $N = 8$ ), 4-year-olds ( $N = 9$ ), 5-year-olds ( $N = 9$ ), and 6-year-olds ( $N = 8$ ). A non-parametric regression test using Rank-Based Estimation was applied to examine the relationship between age and the overall COMFIGURA score. Results revealed a significant effect of age ( $\beta = 3.33$ ,  $t = 3.02$ ,  $p = .004$ ), indicating that with each additional year, there was an average increase of 3.33 points in the COMFIGURA total score. The regression model was also significant, explaining approximately 22.9% of the variance in the data ( $R^2 = 0.229$ ;  $p = .004$ ). Figure 1 illustrates the dispersion of results by participants' chronological age in months.

Although there is variability, the central gray line in Figure 1 represents the regression line identified in the previous analysis, together with its confidence intervals, reinforcing the trend of increasing scores. Around age 3, scores concentrate between 15 and 25 points, corresponding to approximately 30–40% of the expected responses in the instrument. At ages 4 and 5, some children still perform similarly to the 3-year-olds; however, higher performances also emerge, with some children reaching approximately 30 points, or 40–50% of the task score. Finally, in the 6-year-old group, another increase in scores is observed, with most children reaching more than 30 points, that is, above 50% of expected responses. In summary, when considering the instrument's total score, there is evidence of a continuous progression in participants' performance with age, indicating a gradual development in figurative language comprehension.



**Figure 1. Overall scores by participants' age in months.** The horizontal axis presents age in months: children between 38 and 47 months were classified as 3-year-olds, those between 48 and 59 months as 4-year-olds, those between 60 and 71 months as 5-year-olds, and those between 72 and 78 months as 6-year-olds. The vertical axis represents the overall score obtained on the instrument, which could range from 0 to 60 points. In the present sample, however, observed scores ranged from 15 to 35 points, which is why only this interval appears in the graph.

Participants' performance on each phenomenon, according to age, was analyzed through its descriptive results. As in the previous analyses, the results obtained in the two task versions were combined. Figure 2 displays participants' performance by age group and phenomenon/task. Figure 2 presents a descending line across all age groups, starting with a higher percentage of expected answers for simpler phenomena and ending with a lower percentage for more complex ones. As expected, the highest scores were observed in the non-verbal metaphor task, with children aged 4, 5, and 6 achieving an average of 60% or more of the expected responses. These were followed by metonymies and verbal metaphors, while idioms and proverbs are the least comprehended phenomena for all age groups. It should be noted, however, that the small number of participants in this pilot study limits the possibility of carrying out and interpreting statistical analyses. Hence, these findings should be interpreted with caution. Nevertheless, there is descriptive evidence suggesting that the comprehension of figurative phenomena develops gradually, in line with the complexity of each figure of speech and the child's level of maturity.



**Figure 2. Percentage of expected responses by age and phenomenon.** The horizontal axis presents the five figurative language tasks, and the vertical axis presents the average percentage of correct answers achieved by the participants. Separate lines indicate the performance of children aged 3, 4, 5, and 6 years.

## Discussion

This paper reports a pilot study using a figurative language comprehension instrument in Brazilian Portuguese, assessing metaphors, metonymies, idioms and proverbs in 34 preschoolers aged 3 to 6 years. The study focused on evaluating the presentation format of the instrument's items (whether organized by phenomenon or randomized) and the applicability and sensitivity of the instrument with preschoolers, exploring early comprehension patterns across different figurative phenomena in this age group.

Data were first analyzed using a Wilcoxon test to examine differences in participants' performance according to the presentation format of the instrument. Statistically significant differences were not found in the overall scores of the instrument nor in the total task scores. Considering the question types (open and closed-ended), only the verbal metaphor task showed a marginal difference in the closed-ended questions, with a tendency for higher scores in the OP group. However, this was not a significant result and should be interpreted with caution. Nevertheless, this pattern may be considered informative when considered alongside qualitative observations collected during testing, raising the possibility of a practice effect on that specific task. In the

OP group, testing began with the non-verbal metaphor items, followed by the verbal metaphor items. It is worth noting that both tasks use the same metaphorical mappings, instantiated either pictorially or verbally. For example, the non-verbal item instantiating the GOOD IS UP mapping presents two characters, one at ground level and another positioned higher, with the instruction *Point to the happiest one*, and the question *Why is it the happiest one?* The verbal item presents the sentence *Lúcia is feeling uplifted*, with the questions *How is she feeling?* and *Is she feeling happy or sad?* During testing, some participants in the OP order noticed a common element across items, commenting to the researcher that they had answered that question before, even though the question was not identical, but was related to the same mapping. While such comments are meaningful to the instrument, as they demonstrate the children's awareness and a solid understanding of the phenomenon across two different types of tasks, they also suggest a potential consequence of testing in the OP order: participants may have been more aware of what was being tested, leading to a trained and possibly better performance, an effect that did not occur in the OR order.

The Wilcoxon test was also conducted at the item level, examining differences in participants' performance on each of the two questions administered for each of the 30 items. This analysis supported the overall results, with no significant differences according to the testing order. Such findings were complemented by both descriptive and qualitative analyses. For metaphors and metonymies, higher levels of comprehension were observed for the age group under study, as indicated by larger proportions of expected responses and qualitative evidence suggesting figurative understanding. This occurs particularly in open-ended questions, given that closed-ended questions allow a 50% chance of guessing. These results are consistent with previous literature on metaphor and metonymy comprehension in preschool-aged children, indicating that children aged 3 to 6 can already grasp figurative language (Ferrari, 2024; Ferrara et al., 2025; Özçalışkan, 2005; Rundblad & Annaz, 2010a; Siqueira, 2004; Siqueira & Gibbs, 2007; Siqueira & Lamprecht, 2007).

Within the same tasks, item-level analysis also allowed the identification of items with higher and lower proportions of expected responses, which, together with pilot study insights, highlighted one specific item that required closer examination. Specifically, the non-verbal metaphorical item based on the conceptual mapping INTENSITY OF EMOTION IS HEAT elicited lower-than-expected correct responses in closed-ended questions (11% for the OP group and 47% for the OR group, compared to over 70% for other non-verbal items). This is a sensory item in which two water bags are presented to the participant, one hot and one cold, representing two characters. Participants were asked to touch the bags, indicate which character felt more emotions, and explain why, with the expected response being that the character associated with the hot bag would be the one feeling more emotions due to the higher temperature. During testing, it was observed that many children did not understand the word *emotion* and that may have confused results. Besides, the interviews were conducted

during the summer, in warm testing conditions, which also may have influenced the results for this item. In many cases, participants responded that the correct answer would be the cold character due to personal preference or because heat could cause burns, as illustrated below. Therefore, the observed descriptive and qualitative differences for this item may reflect the misunderstanding of a complex word, an environmental interference (ambient temperature) in a sensory item involving temperatures, or even the metalinguistic difficulty to explain a metaphorical association in the item for younger participants.

Part. O11, 5,3 years old, Kindergarten – *Porque tá gelado. É bom.* [Because it's icy. It's good.]

Part. O10, 5,6 years old, Kindergarten – *Porque eu gosto de ficar no frio.* [Because I like being in the cold.]

Part. A2, 4,9 years old, Kindergarten – *Porque fica mais fresco. O quente fica um pouco calor.* [Because it is cooler. The hot one is a bit warm.]

Part. A5, 5,9 years old, Kindergarten – *Porque o quente vai queimar e vai doer.* [Because the hot one will burn and hurt.]

After noticing that the word emotion could be harming results, a follow up question was added for some participants, aiming to reframe the question in future versions of the instrument. Then, after answering the testing questions, participants were asked to touch the water bags, indicate which character was angrier, and explain their choice. The expected response was that the character associated with the hot bag would be the angrier one. Indeed, when doing this follow up question, participants used to reach the expected association more frequently, even commenting that they felt the same way when they themselves were angry (e.g., Part. A10, 6,1 years old, Kindergarten – *Quando eu tô brabo, eu tô quente.* [When I'm angry, I'm hot]). For the final version of the instrument, then, this item was updated, no longer asking about who feels more emotion, but instead about who is feeling angrier.

Idioms and proverbs did not show statistically significant differences either in the analyses of overall scores or in the item-level analysis. This result was expected, since the age group analyzed does not yet present sufficient age to understand these phenomena. Therefore, it was expected that there would be no difference in phenomena that are still scarcely understood. Such a result, by demonstrating low comprehension of the phenomenon in both testing formats, shows that the instrument is sensitive and in line with previous findings in the literature (e.g., Caillies & Le Sourn-Bissaoui, 2006; Cain et al., 2009; Ferrari, 2024; Gibbs, 1991; Hamdan & Smadi, 2021; Nippold et al., 1988; Nippold & Haq, 1996; Siqueira, Duarte et al., 2017; Siqueira & Marques, 2018). On the other hand, it also points to a limitation of the pilot study, since the age group considered here is not yet sensitive to the phenomenon, which makes it difficult to obtain more relevant results about the functioning of the instrument.

Overall, the results of the first analyses suggest that the instrument's outcomes are not influenced by the presentation format of the stimuli. However, qualitatively, testing effects were more frequently observed in the OP order, which favors applying the instrument in the OR order, thereby ensuring fewer testing effects. Based on this, the adoption of the randomized version of the instrument for future testing is supported, as it minimizes testing effects. It is possible that the OP format, which groups items by phenomenon, facilitates the concentration on a single type of cognitive processing at a time but, at the same time, may only reflect a result of repeated practice in the application of the instrument. The OR format, in turn, seems less mechanical, as it requires constant alternation between different types of figurative reasoning. This factor may help reduce practice effects, making the assessment process more effective. At the same time, the limitations on the use of the randomized version of the Instrument are acknowledged, such as a possible effect of cognitive overload on participants. By alternating between items that require different levels of cognitive effort, the overall cognitive demand of the task may increase, requiring more resources of attention, working memory, and inference, which could impact participants' performance. Still, the possible effects of cognitive overload in the OR format seem to represent a lower risk to the validity of the results than the potential impact of testing effects in the OP format. Moreover, no significant differences were found when analyzing performance on the last ten items in the OR format, nor in any of the tested conditions, suggesting that any potential fatigue or cognitive overload effects do not appear to have affected participants' performance.

The second stage of the analyses aimed to examine the effects of schooling and age on the results obtained. First, a Wilcoxon test indicated that there were no differences between the overall scores of the two versions of the instrument according to school grade. Based on this result, performance across both versions of the instrument was grouped and analyzed in relation to age and schooling to obtain preliminary evidence of task validity. As expected, a Wilcoxon test revealed significant differences between school levels, with kindergarten children outperforming those in nursery. Although both groups are still at pre-literacy stages, this result suggests that schooling already provides children with broader opportunities for linguistic and cognitive development, which may facilitate the comprehension of figurative language. Such findings can be interpreted as preliminary evidence of construct validity for COMFIGURA.

A regression test was then conducted on participants' scores and their relationship with age. Significant effects were found, suggesting that, between 3 and 6 years, there is an increase of approximately 3 points in the Instrument each year. This finding corroborates previous research conducted with the instrument's tasks around this age (Ferrari & Siqueira, *submitted*; Siqueira, 2004; Siqueira, Ferrari et al., 2023) as well as the general literature, which indicates an increase in figurative language comprehension with age (e.g., Gibbs & Colston, 2012), showing that the comprehension of figurative language follows a progressive developmental course starting as early as 3 years

old. Although comprehension still reaches low levels at this age, phenomena as metaphors and metonymies already show some understanding, demonstrating that children's cognitive and linguistic abilities are sufficiently developed to grasp basic figurative relations. This result can also be related to other findings which show that, as vocabulary, working memory, and inferential skills develop, children become increasingly able to understand non-literal language (e.g., Falkum & Köder, 2020; Gibbs & Colston, 2012). In other words, age or schooling alone are probably not the only variables responsible for figurative understanding. Despite the linear effect, the regression model previously presented suggests that most of the variability in scores is not explained by age alone. That is, factors such as language exposure, sociocultural background, early schooling, metalinguistic skills, and individual differences may also directly affect the results obtained (e.g., Falkum & Köder, 2020; Gibbs & Colston, 2012). Future studies, with a larger sample, should take these variables into consideration, examining their potential effects and covariances. Even so, the results suggest that, even in preschool ages, systematic effects can already be detected, highlighting the sensitivity of COMFIGURA to small developmental variations. Therefore, this finding can also be considered preliminary evidence of construct validity, as it indicates that the task is able to capture an ability that improves with age since and early age.

Finally, one last descriptive analysis was conducted to observe participants' performance according to age and the phenomena included in the instrument. Note that, due to sample limitations, the results discussed here are only descriptive and indicate possible tendencies for future studies conducted with a larger sample. First, the average of expected responses was very similar for metaphors (when tested non-verbally) and metonymies in the group of 3-year-olds. Even though the average scores are still low, around 40–45% of correct answers, this already suggests that both phenomena are under development in this age range, with metonymies showing relatively higher scores than metaphors. Considering the verbal tasks, which are naturally more difficult for children, the phenomenon with the highest level of comprehension is metonymy, corroborating the hypothesis that metonymies may be understood even earlier than metaphors (Rundblad & Annaz, 2010a). This also stands out in the performance of the other age groups for metonymy, with means already above chance level for the 4- and 6-year-olds.

Another highlight is the trajectory of performance across phenomena, which seems similar for all age groups. The highest scores appear for non-verbal metaphors, followed by metonymies, verbal metaphors, idioms, and finally proverbs. In all groups, performance is consistently better for metaphors and metonymies, suggesting earlier developmental phenomena, compared to the more culture-dependent phenomena, which are still poorly understood at the ages analyzed here. This finding is in line with the literature, showing that metaphor-related phenomena, besides presenting an increasing number of dimensions and complexity (see Siqueira et al., *in press*), are

possibly developed in language acquisition in that same order. This result also provides preliminary evidence of construct validity for COMFIGURA, showing that it is sensitive to assessing the comprehension of different phenomena across early ages.

It is also important to mention that the graph does not show a strictly linear comprehension pattern across age. In some phenomena, such as metonymy, 5-year-olds scored lower than 4-year-olds. These results should, once again, be interpreted with caution, given the limited statistical power of such a small sample. The small number of participants in this pilot study further limits the possibility of conducting robust statistical analyses, preventing any strong generalizations. Nevertheless, these preliminary findings provide initial evidence that comprehension of figurative phenomena develops gradually, depending both on the complexity of each figure of speech and on the participant's developmental stage.

### **Conclusions**

In conclusion, this study pilot-tested a figurative language comprehension instrument in Brazilian Portuguese, evaluating its applicability and sensitivity with preschoolers aged 3 to 6 years. More specifically, it investigated whether the order of item presentation (by phenomenon vs. randomized) affected performance and explored early comprehension patterns of different figurative phenomena in this age group. Overall, results indicated that the instrument is appropriate for use with preschoolers and sensitive to the expected variations in figurative language comprehension for this age. Regarding the order of item presentation, the analyses did not reach significant statistical differences between the two versions. However, qualitative analyses suggested practice effects in one of the versions (the one with items organized by phenomenon). Therefore, the randomized presentation was chosen, as it better balances the different phenomena and minimizes practice effects. The findings also aligned with previous literature on figurative language development, suggesting that children as young as three already show some ability to understand metaphors and metonymies. Idioms and proverbs, on the other hand, were not yet consistently comprehended, which corroborates prior evidence that such skills typically emerge later. Taken together, these results support the hypothesis of a gradual acquisition of figurative phenomena, beginning with a greater sensitivity to metonymies and metaphors, and later extending to idioms and proverbs. Future studies with larger and more diverse samples, in terms of both age and educational background, are needed to further explore the patterns observed here. Overall, the findings provide preliminary validity evidence for COMFIGURA, which, even in this small-scale pilot study, showed sensitivity to developmental differences in figurative language comprehension.

## References

- Afonso, N. C. P. (2012). *Compreensão de metáforas primárias e deficiência auditiva* [Master's thesis, Universidade de Aveiro]. Aveiro, Portugal. <http://hdl.handle.net/10773/9837>
- Boers, F. (2014). Idioms and phraseology. In J. Littlemore & J. R. Taylor (Eds.), *The Bloomsbury Companion to Cognitive Linguistics* (pp. 185–202). Bloomsbury. <https://doi.org/10.5040/9781472593689.ch-011>
- Borges, É. P. K., Koltermann, G., Minervino, C. A. D. S. M., & de Salles, J. F. (2023). The role of emergent literacy assessment in Brazilian Portuguese literacy acquisition during COVID-19. *Behavioral Sciences*, 13(6), 510. <https://doi.org/10.3390/bs13060510>
- Caillies, S., & Le Sourn-Bissaoui, S. (2006). Idiom comprehension in French children: A cock-and-bull story. *European Journal of Developmental Psychology*, 3(2), 189–206. <https://doi.org/10.1080/17405620500412325>
- Cain, K., Towse, A. S., & Knight, R. S. (2009). The development of idiom comprehension: An investigation of semantic and contextual processing skills. *Journal of Experimental Child Psychology*, 102(3), 280–298. <https://doi.org/10.1016/j.jecp.2008.08.001>
- Chahboun, S., Kvello, Ø., & Page, A. G. (2021). Extending the field of extended language: A literature review on figurative language processing in neurodevelopmental disorders. *Frontiers in Communication*, 6, 661528. <https://doi.org/10.3389/fcomm.2021.661528>
- Cony, I. (2025). Atravessando a literalidade no espectro: Avaliação da compreensão metafórica no autismo. *ReVEL*, 23(44), 97–123. <http://www.revel.inf.br/files/63ca6861768f10bc7de89320b47569e0.pdf>
- De Leon, V. C. (2008). *A compreensão e a produção de enunciados metafóricos em crianças com transtornos globais do desenvolvimento* [Doctoral dissertation, Universidade Federal do Rio Grande do Sul]. Porto Alegre, Brazil. <http://hdl.handle.net/10183/17822>
- Duthie, J. K., Nippold, M. A., Billow, J. L., & Mansfield, T. C. (2008). Mental imagery of concrete proverbs: A developmental study of children, adolescents, and adults. *Applied Psycholinguistics*, 29(1), 151–173. <https://doi.org/10.1017/S0142716408080077>

- Falkum, I. L., & Köder, F. (2020). The acquisition of figurative meanings. *Journal of Pragmatics*, 164, 18-24. <https://doi.org/10.1016/j.pragma.2020.04.007>
- Falkum, I. L., Recasens, M., & Clark, E. V. (2017). “The moustache sits down first”: On the acquisition of metonymy. *Journal of Child Language*, 44(1), 87-119. <https://doi.org/10.1017/S0305000915000720>
- Fernandes, D. P. (2018). *Compreensão e interpretação de provérbios: o papel de familiaridade, idade e escolaridade* [Master's thesis, Universidade Católica Portuguesa]. Lisbon, Portugal. <http://hdl.handle.net/10400.14/28207>
- Ferrara, S., Aguert, M., & Declercq, C. (2025). Not early, not late, but developing: Children's “good-enough” understanding of metaphors. *Language Development Research*, 5(2), 67–92. <https://doi.org/10.34842/ldr2025-806>
- Ferrari, C. G. (2020). *Evidências de validade de uma tarefa de compreensão de provérbios* [Master's thesis, Universidade Federal do Rio Grande do Sul]. Porto Alegre, Brazil. <http://hdl.handle.net/10183/230076>
- Ferrari, C. G. (2024) *Compreensão de linguagem figurada em português e inglês: adaptação e validação de um instrumento interlinguístico*. [Doctoral dissertation, Universidade Federal do Rio Grande do Sul]. Porto Alegre, Brazil. <http://hdl.handle.net/10183/292104>
- Ferrari, C. G., & Siqueira, M. (2020). Água mole em pedra dura tanto bate até que fura: uma comparação entre a compreensão de provérbios por crianças e adultos. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 36(2), 1–29. <http://dx.doi.org/10.1590/1678-460X2020360204>
- Ferrari, C. G., & Siqueira, M. (2023). Where there's a proverb, there are many conceptual mappings. *Crossroads. A Journal of English Studies*, (43), 57-81. <https://doi.org/10.15290/CR.2023.43.4.04>
- Ferrari, C. G. & Siqueira, M. (submitted) *With age comes wisdom: Development of proverb comprehension from childhood to adulthood*. Unpublished manuscript.
- Gaskins, D., & Rundblad, G. (2023). Metaphor production in the bilingual acquisition of English and Polish. *Frontiers in Psychology*, 14, 1162486. <https://doi.org/10.3389/fpsyg.2023.1162486>

- Gaskins, D., Falcone, M., & Rundblad, G. (2024). A usage-based approach to metaphor identification and analysis in child speech. *Language and Cognition*, 16(1), 32-56. <https://doi.org/10.1017/langcog.2023.17>
- Gibbs Jr, R. W. (1991). Semantic analyzability in children's understanding of idioms. *Journal of Speech, Language, and Hearing Research*, 34(3), 613-620. <https://doi.org/10.1044/jshr.3403.613>
- Gibbs Jr, R. W. (1994). *The poetics of mind: Figurative thought, language, and understanding*. Cambridge University Press.
- Gibbs Jr, R. W., & Beitel, D. (1995). What proverb understanding reveals about how people think. *Psychological Bulletin*, 118(1), 133. <https://doi.org/10.1037/0033-2909.118.1.133>
- Gibbs Jr, R. W., & Colston, H. L. (2012). *Interpreting figurative meaning*. Cambridge University Press. <https://doi.org/10.1017/CBO9781139168779>
- Grady, J. E. (1997). *Foundations of meaning: Primary metaphors and primary scenes* [Doctoral dissertation, University of California, Berkeley]. <https://escholarship.org/uc/item/3g9427m2>
- Hamdan, J. M., & Smadi, A. M. (2021). Comprehension of idioms by Jordanian Arabic-speaking children. *Journal of Psycholinguistic Research*, 50(5), 985-1008. <https://doi.org/10.1007/s10936-021-09773-4>
- Lakoff, G. & Johnson, M. (1980). *Metaphors we live by*. University of Chicago Press.
- Lakoff, G., & Turner, M. (1989). *More than cool reason: A field guide to poetic metaphor*. University of Chicago Press. <https://doi.org/10.7208/chicago/9780226470986.001.0001>
- Langacker, R. (1993). Reference-point constructions. *Cognitive Linguistics*, 4(1), 1-38. <https://doi.org/10.1515/cogl.1993.4.1.1>
- Langlotz, A. (2006). *Idiomatic creativity: A cognitive-linguistic model of idiom-representation and idiom-variation in English*. John Benjamins. <https://doi.org/10.1075/hcp.17>
- Littlemore, J. (2015). *Metonymy: Hidden Shortcuts in Language, Thought and Communication*. Cambridge University Press. <https://doi.org/10.1017/CBO9781107338814>
- Lopes, N. A. S. (2019). *Compreensão de linguagem figurada por crianças com TEA e desenvolvimento típico: um estudo de caso com irmãos gêmeos*. [Undergraduate thesis,

Universidade Federal do Rio Grande do Sul]. Porto Alegre, Brazil. <http://hdl.handle.net/10183/242012>

Marques, D. (2018). *É possível ser todo ouvidos após engolir um sapo? Contribuições para o estudo da compreensão da linguagem figurada por deficientes auditivos oralizados* [Doctoral dissertation, Universidade Federal do Rio Grande do Sul]. Porto Alegre, Brazil. <http://hdl.handle.net/10183/193566>

Marques, D., Baiocco, L., & Siqueira, M. (2025). How do oral deaf individuals comprehend primary metaphors and idioms? Let's begin to dot the i's and cross the t's. (2025). *Revista De Estudos Da Linguagem*, 33(3), 114-131. <https://doi.org/10.17851/2237-2083.33.3.114-131>

Miorando, R. B., & Siqueira, M. (2024). Uma análise da compreensão de metonímia em fase de aquisição da linguagem. *Signo*, 49(95), 71-85. <https://doi.org/10.17058/signo.v49i95.18989>

Nascimento, L. (2023, September 8th). Literacy among Brazilian kids below 50% in 2021. *Agência Brasil*. <https://agenciabrasil.ebc.com.br/en/educacao/noticia/2023-09/literate-kids-brazil-down-alarming-494-2021>.

Nayak, N. P., & Gibbs, R. W. (1990). Conceptual knowledge in the interpretation of idioms. *Journal of Experimental Psychology: General*, 119(3), 315-330. <https://doi.org/10.1037/0096-3445.119.3.315>

Nerlich, B., Clarke, D. D., & Todd, Z. (1999). Mummy, I like being a sandwich. Metonymy in language acquisition. In K. U. Panther & G. Radden (Eds.), *Metonymy in Language and Thought* (pp. 361–384). John Benjamins. <https://doi.org/10.1075/hcp.4>

Nippold, M. A., Allen, M. M., & Kirsch, D. I. (2000). How adolescents comprehend unfamiliar proverbs: the role of top-down and bottom-up processes. *Journal of Speech, Language, and Hearing Research*, 43(3), 621-630. <https://doi.org/10.1044/jslhr.4303.621>

Nippold, M. A., Allen, M. M., & Kirsch, D. I. (2001). Proverb comprehension as a function of reading proficiency in preadolescents. *Language, Speech, and Hearing Services in Schools*, 32(2), 90-100. [https://doi.org/10.1044/0161-1461\(2001/009\)](https://doi.org/10.1044/0161-1461(2001/009))

Nippold, M. A., & Haq, F. S. (1996). Proverb comprehension in youth: The role of concreteness and familiarity. *Journal of Speech, Language, and Hearing Research*, 39(1), 166-176. <https://doi.org/10.1044/jshr.3901.166>

- Nippold, M. A., Hegel, S. L., Uhden, L. D., & Bustamante, S. (1998). Development of proverb comprehension in adolescents: Implications for instruction. *Journal of Children's Communication Development*, 19(2), 49-55. <https://doi.org/10.1177/152574019801900206>
- Nippold, M. A., Martin, S. A., & Erskine, B. J. (1988). Proverb comprehension in context: A developmental study with children and adolescents. *Journal of Speech, Language, and Hearing Research*, 31(1), 19-28. <https://doi.org/10.1044/jshr.3101.19>
- Nippold, M. A., Uhden, L. D., & Schwarz, I. E. (1997). Proverb Explanation Through the Lifespan: A Developmental Study of Adolescents and Adults. *Journal of Speech, Language, and Hearing Research*, 40(2), 245-253. <https://doi.org/10.1044/jslhr.4002.245>
- Özçalışkan, Ş. (2005). On learning to draw the distinction between physical and metaphorical motion: Is metaphor an early emerging cognitive and linguistic capacity?. *Journal of Child Language*, 32(2), 291-318. <https://doi.org/10.1017/S0305000905006884>
- Panther, K. U. & Radden, G. (1999). *Metonymy in Language and Thought*. John Benjamins Publishing Company. <https://doi.org/10.1075/hcp.4>
- Rundblad, G., & Annaz, D. (2010a). Development of metaphor and metonymy comprehension: Receptive vocabulary and conceptual knowledge. *British Journal of Developmental Psychology*, 28(3), 547-563. <https://doi.org/10.1348/026151009X454373>
- Rundblad, G., & Annaz, D. (2010b). The atypical development of metaphor and metonymy comprehension in children with autism. *Autism*, 14(1), 29-46. <https://doi.org/10.1177/1362361309340667>
- Siqueira, M. (2004). *As metáforas primárias na aquisição da linguagem: um estudo interlingüístico* [Unpublished Doctoral dissertation, Pontifícia Universidade Católica do Rio Grande do Sul]. Porto Alegre, Brazil.
- Siqueira, M., Duarte, S., Pereira, L. B., Ferrari, C. G., & Lopes, N. (2017). Compreensão de expressões idiomáticas em período de aquisição da linguagem. *Letras de Hoje*, 52(3), 391-400. <https://doi.org/10.15448/1984-7726.2017.3.29371>
- Siqueira, M., Ferrari, C. G., de Carvalho Rodrigues, J., Baiocco, L., de Oliveira, T. M., Duarte Jr, S., & Marques, D. (2023). Evidências de validade de Tarefas de Compreensão de Metáforas Primárias: uma revisão da literatura. *Letrônica*, 16(1), e44362-e44362. <https://doi.org/10.15448/1984-4301.2023.1.44362>

Siqueira, M., Ferrari, C. G., Silva Tavares, V. R., & Miorando, R. (in press) The great(er and greater) chain of metaphor-related phenomena. *Metaphor & Symbol*.

Siqueira, M., & Gibbs, R. (2007). Children's acquisition of primary metaphors: a crosslinguistic study. *Organon*, 21(43). <https://doi.org/10.22456/2238-8915.39590>

Siqueira, M., & Lamprecht, R. R. (2007). As metáforas primárias na aquisição da linguagem: um estudo interlingüístico. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 23(2), 245-272. <https://doi.org/10.1590/S0102-44502007000200004>

Siqueira, M., & Marques, D. F. (2018). Desenvolvimento e validação do instrumento de compreensão de expressões idiomáticas. *Revista de Estudos da Linguagem*, 26(2) 571-591. <https://doi.org/10.17851/2237-2083.26.2.571-591>

Siqueira, M., Marques, D. F., & Gibbs Jr, R. (2016). Metaphor-related figurative language comprehension in clinical populations: a critical review. *Scripta* 20(40), 36-60. <https://doi.org/10.5752/P.2358-3428.2016v20n40p36>

Siqueira, M., Melo, T., Duarte Jr, S., Baiocco, L., Ferrari, C. G., & Lopes, N. (2023). Many hands on this study: Development of a metonymy comprehension task. *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada*, 39, 202339350607. <https://doi.org/10.1590/1678-460X202339350607>

Siqueira, M., Pereira, L. B, Ferrari, C. G., & Lopes, N. (2017). Mapeamentos metafóricos e metonímicos em provérbios do português brasileiro. *ReVEL*, 15(29), 159-175. <http://www.revel.inf.br/files/3cd6f1594c564f8a72950c7b79a87996.pdf>

Stites, L. J., & Özçalışkan, Ş. (2013). Developmental changes in children's comprehension and explanation of spatial metaphors for time. *Journal of Child Language*, 40(5), 1123-1137. <https://doi.org/10.1017/S0305000912000384>

Uekermann, J., Thoma, P., & Daum, I. (2008). Proverb interpretation changes in aging. *Brain and Cognition*, 67(1), 51-57. <https://doi.org/10.1016/j.bandc.2007.11.003>

Vulchanova, M., Saldaña, D., Chahboun, S., & Vulchanov, V. (2015). Figurative language processing in atypical populations: The ASD perspective. *Frontiers in Human Neuroscience*, 9, 24. <https://doi.org/10.3389/fnhum.2015.00024>

Vulchanova, M., Vulchanov, V., & Stankova, M. (2011). Idiom comprehension in the first language: a developmental study. *Vigo International Journal of Applied Linguistics*, (8), 206-234.

Zhu, R. (2021). Preschoolers' acquisition of producer-product metonymy. *Cognitive Development*, 59, 101075. <https://doi.org/10.1016/j.cogdev.2021.101075>

### **Data, Code and Materials Availability Statement**

Quantitative data and R scripts are available at <https://osf.io/mxvta/>. The study materials cannot be shared publicly due to copyright restrictions. For more information, contact the authors.

### **Ethics Statement**

To ensure the integrity and ethics of the research, all stages of the study were approved by the Research Ethics Committee of the Federal University of Rio Grande do Sul, under approval numbers 2.469.701 and 5.260.845. Written informed consent was obtained from all parents or legal guardians before participation, and oral assent was obtained from all children at the time of the interview.

### **Authorship and Contributorship Statement**

All authors conceived and designed the study. **Caroline Girardi Ferrari** collected and analyzed the data and wrote the manuscript. **Maity Siqueira** supervised the study, contributed to data analysis and revised the manuscript. All authors read and approved the final version of the manuscript.

### **Acknowledgements**

This study is derived from the first author's doctoral dissertation, supervised by the second author. The research was funded by the National Council for Scientific and Technological Development (CNPq, Brazil) through a PhD scholarship. We thank the reviewers and editors for their suggestions, which helped improve the study. We are also grateful to Cristiano Sulzbach for his assistance with data analysis, and to Vinícius da Rosa da Silva Tavares, Rafaeli Miorando, Felipe Tota, and Isabel Cony for their contributions to data classification.

### **License**

*Language Development Research* (ISSN 2771-7976) is published by TalkBank and the Carnegie Mellon University Library Publishing Service. Copyright © 2026 The Author(s). This work is distributed under the terms of the Creative Commons Attribution-Non-commercial 4.0 International license (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits any use, reproduction and distribution of the work for non-commercial purposes without further permission provided the original work is attributed as specified under the terms available via the above link to the Creative Commons website.