Not early, not late, but developing: Children's "good-enough" understanding of metaphors

Sarah Ferrara
Université de Reims Champagne-Ardenne, C2S, Reims, France

Marc Aguert
Normandie Univ, UNICAEN, LPCN, France

Christelle Declercq
Université de Reims Champagne-Ardenne, C2S, Reims, France

Abstract: To date, the debate over the age at which children begin to understand metaphors remains unresolved. Do children begin to acquire comprehension early, around age 3 to 4, or later, around age 8? One way to answer this question is to use the notion of "good-enough" comprehension proposed by Ferreira et al. (2002) and to hypothesize that young children understand metaphors in a "good enough" manner while older children understand them in a more precise and accurate manner. This hypothesis was tested using a task where children were asked to assess the extent to which more or less precise and relevant rephrasings resembled nominal metaphors. We therefore sought to (1) differentiate between "good-enough" and "good" (precise and accurate) understanding in school-age children and (2) show that the former appears earlier than the latter during development. Data collected from 300 children aged 5 to 11 suggest that both goals were reached. These results suggest that, while metaphorical abilities emerge early, comprehension processes then evolve during childhood, with a refinement of understanding between 7 and 9 years old. These results may open a path to reconcile the proponents of early acquisition with those of later acquisition.

Keywords: cognitive development; metaphor comprehension; language development

Corresponding author(s): Sarah Ferrara, Laboratoire C2S (Cognition, Santé, Société), Université de Reims Champagne-Ardenne, 57 rue Pierre Taittinger, 51571 Reims Cedex, France. Email: sarah.ferrara@univ-reims.fr

ORCID ID(s): Sarah Ferrara: https://orcid.org/0000-0002-3953-0661; Marc Aguert: https://orcid.org/0000-0002-0484-0328; Christelle Declercq: https://orcid.org/0000-0002-3242-644X

Citation: 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. http://doi.org/10.34842/ldr2025-806

Introduction

Metaphor comprehension in children has been the subject of research for decades, but the field lacks a coherent account of how metaphor comprehension abilities develop throughout childhood due to the diversity of theoretical and methodological approaches. This study aimed to examine the development of metaphor understanding in children aged 5 to 11 years old using the theory of "good-enough" comprehension (Ferreira et al., 2002). We hypothesised that understanding could fall outside of the dichotomy of being either "correct" or "incorrect": some understanding could be correct "enough" to ensure the conversational flow, while still being incomplete and imprecise.

Metaphors are widely used utterances describing one thing using another element that is conceptually different. For instance, in the metaphor "cold air is a needle", the speaker communicates something about the topic (e.g., cold air) by means of the vehicle (e.g., a needle) in a non-literal way to describe the hurtful nature of coldness (Di Paola et al., 2019). In the developmental literature, interest has long been focused on the age at which children begin to understand metaphors, with mixed findings. On one hand, some studies have concluded that children begin to understand metaphors rather late, around 8 to 10 years of age, with possible further developments in adolescence, thus supporting the notion that metaphors are complex statements to comprehend (Billow, 1977; Carriedo et al., 2016; Cometa & Eson, 1978; Deckert et al., 2019; Willinger et al., 2019). On the other hand, alternative studies have argued that metaphors are not inherently complex statements to understand. These studies have empirically demonstrated an earlier comprehension, around 3 to 4 years of age, with a methodology adapted to young children, including facilitating contexts and age-appropriate vocabulary (Vosniadou, 1987 for a review; Pouscoulous et al., 2011).

This divergence has long been explained with reference to the method used to assess understanding. In particular, the commonly used method of asking children to verbally explain the meaning of the metaphor has been criticised as being too cognitively costly, falling more within the realm of metalinguistic abilities (Vosniadou, 1987; Pouscoulous, 2011). Other methods that do not involve verbalisation have been developed to address this criticism, including selecting from multiple rephrasings of the metaphor (Nippold et al., 1984; Nippold & Sullivan, 1987; Winner et al., 1980) matching sentences with metaphorically related pictures (Kogan, 1980) , choosing appropriate metaphorical completions for sentences (Gardner et al., 1975), or performing an action such as an enactment (Vosniadou & Orthony, 1986). Recently, Pouscoulous and Tomasello (2020) found that children as young as 4 years old showed an understanding of referential metaphors by being able to choose the right puppet when asked questions such as, "Show me the dog with brown socks" (e.g. the dog with brown feet, while also being presented with a dog with a brown bow).

The variety of methodologies in the literature likely reflects different conceptions of what it means to understand a metaphor, which extends from describing the elementary processes enabling the understanding of a basic metaphor to portraying what children understand about the variety of metaphors that surround them in their natural environment. In addition, multiple research aims are associated with this question, such as the description, mostly through correlational studies, of factors influencing children's understanding of metaphors. Some factors have been considered as specific to metaphorical abilities (e.g., classificatory skills or analogical reasoning are precisely what enable metaphor comprehension, Billow, 1977; Johnson & Pascual-Leone, 1989; Gentner, 1988), while others have been thought to be more broadly related to cognitive development (e.g. vocabulary, executive functions or theory of mind abilities, Białecka-Pikul, 2010; Del Sette et al., 2020; Huang et al., 2015; Pierce et al., 2010; Rundblad & Annaz, 2010; Tonini et al., 2023; Willinger et al., 2019). Overall, these results depict metaphors as utterances that are most often intricate, and understanding them involves multiple social cognitive constructs.

Based on the previous data, one could consider that metaphor comprehension emerges early in some simple situations but continues to progress during childhood. This would offer a means to reconcile findings that demonstrate early metaphorical abilities with those indicating the influence of various factors associated with cognitive development. However, there is an intensive body of research on the development of metaphor comprehension and conclusions often refers to understanding as a process ending in an understanding or an incorrect one. Understanding is then considered as something perfect or described as attained by some children if not all of an age group. Some authors have implied an intermediate level without really seeking theoretical implications (Deckert, 2019 ; Lecce at al., 2019). An alternative stance would be to consider that, between an undeniable understanding or an understanding in progress, one might find a formalized intermediate level of comprehension. Children could strive to grasp the overall meaning of what is being said by relying on the whole situation, rather than undertaking a detailed analysis of the metaphor's meaning. It is this stance that we aimed to explore further, assuming that metaphor comprehension falls on a spectrum rather than a binary categorisation of good or bad. This hypothesis is grounded in the “good-enough” theory introduced by Ferreira et al. (2002).

Ferreira et al. (2002) posit the existence of different levels of comprehension within their "good-enough" hypothesis. They define “good-enough comprehension” as the creation of a vague and superficial yet largely satisfactory understanding of the meaning of a statement. An illustrative example is the Moses illusion (Erickson & Mattson, 1981, cited by Ferreira et al., 2002). When people are asked, "How many animals of each sort did Moses put on the ark?" they usually answer "two" instead of the correct answer, "zero," as it was Noah who placed the animals on the ark. This example demonstrates how adults, despite their capacity for precise comprehension, can

construct an imprecise and shallow representation of a statement. Ferreira et al. (2002) proposed that good-enough comprehension is commonplace, as it reduces cognitive effort in everyday interactions where people communicate without investing the time needed for precise comprehension. In this view, relying on a good-enough mode of comprehension could be more efficient than exerting the effort to achieve precise comprehension.

Good-enough comprehension could be especially effective in processing complex language forms such as metaphors by economizing resources. Metaphor processing requires linguistic skills (knowing which semantic features of the topic and the vehicle are salient), pragmatic skills (using language in an unconventional way which is contextually relevant), executive resources (inhibiting literal meaning, switching meanings with flexibility), and conceptual and classificatory skills (finding common ground between the topic and the vehicle). Adults possess advanced linguistic skills, but children experience a considerable evolution in their abilities through childhood and even adolescence, often presenting them with language-related difficulties (see Gervain, 2020 for a review). Considering these inherent difficulties and the importance of verbal communication in daily life, good-enough understanding could be a significant advantage and even a preferred path to more precise and complete "good" understanding. Moreover, school-aged children also possess limited cognitive resources, and their executive functions (i.e. working memory, flexibility and inhibition) are still in development (Diamond, 2020). Good-enough understanding could also be advantageous by reducing working memory load and thus providing a possible economy in cognitive resources. In addition, studies have shown a complementary role played by vocabulary skills and working memory on understanding, thus supporting the complex relationships between these skills (Chiappe & Chiappe, 2007; Mashal, 2013; Stamenković et al., 2019). Good-enough understanding could therefore represent an advantage for children regarding these limitations.

The concept of good-enough understanding could also help make sense of conflicting findings in the literature. In regards of the theoretical and methodological divergences, we could hypothesize that authors had accounted for different levels of understanding without intending it. For instance, studies asking children to explain precisely what they understood by the metaphor could be seen as demonstrating a "good" (i.e. precise and complete) understanding of the metaphor, while studies using an action criterion for comprehension may show what could be defined as good-enough understanding. Thus, the difference observed in the results, especially regarding the emergence of comprehension, could depend not only on the nature of their methodologies but also on the type of comprehension authors were actually accounting for. Good-enough understanding could have been considered "good" by early-understanding proponents (even if this entailed glossing over certain nuances in understanding) while it would have been considered "bad" by later-understanding proponents (even if this meant overlooking early signs of understanding), in accordance

with a dichotomic consideration of comprehension. At that stage, authors did not explore the theory of different levels of understanding, as their existence had not yet been hypothesised. However, certain studies have operationalised intermediate levels of understanding. For instance, in a verbal explication task of metaphors, Lecce et al. (2019) assigned an intermediate score to children who gave an "incomplete" response ("They are very good" for the metaphor "Soldiers are lions") or referred to "a non-salient feature of the metaphor vehicle" ("They run as fast as lions"). However, these intermediate levels were not correlated with any clear theoretical framework as is the case with Ferreira's good-enough understanding.

Following Ferreira et al. (2002), three levels of understanding of metaphors were distinguished in the current study: good understanding and poor understanding, as traditionally described in the literature, with the addition of an in-between level, good-enough understanding. Good understanding is defined as accurate and comprehensive, incorporating the specific reasons why the vehicle has been chosen to qualify the topic. Good-enough understanding relies on the whole situation, avoiding the cost of a fine-grained construction of the metaphor's meaning: it enables the listener to understand the intended meaning of the metaphor, that is the communicative intent relative to the topic, in coherence with the context, but not necessarily the specific reasons why the vehicle has been chosen to represent the topic. Poor understanding does not account for the meaning of the metaphor; this type of understanding could be diverse, as children can produce multiple misinterpretations of a metaphor's meaning, using a variety of clues from the material presented within and outside of the metaphor. Previous studies have for example presented children with literal interpretations or irrelevant context-based interpretations in multiple choice tasks (Declercq et al., 2010).

The current study aimed to test whether children showed evidence of good-enough understanding of metaphors with an innovative task involving metaphor rephrasing judgement. In each trial, participants were presented with a metaphor and a rephrasing of this metaphor varying in quality. They evaluated the quality of this rephrasing compared to the metaphor on a 5-point scale. Their evaluation of the different rephrasings would determine whether they had a good, good-enough, or poor understanding of the presented metaphor. Based on previous literature, we expected that by the age of 5, a good-enough understanding would be observed, and supplemented later in development by a good understanding. Our study aimed to (1) differentiate between good-enough understanding and good understanding in school-aged children and (2) establish a developmental pattern of both levels of understanding.

Method

Participants

We planned to include 300 children in the study, aiming to recruit between 30 and 50 participants of each age. The experimental design, which involved three crossed random variables (participants, schools, and items, analyzed using a linear mixed-effects model), did not allow for a conventional a priori power analysis (e.g., using G*Power) or a simulation-based approach (e.g., using R, *MixedPower* package). Rather than providing a power analysis that would lack interpretability, we opted for a large sample size (N = 300), which is substantial relative to the standards observed in studies on children's metaphor comprehension (Almohammadi et al., 2024 ; Pouscoulous & Tomasello, 2020).

Three hundred participants (52.7% girls), whose mean age was 8 years and 3 months (SD = 2 years, range [5 years 1 month; 11 years 12 months]) were recruited (See Figure 1 for the distribution depending on age). Participants were recruited from nine French public schools, situated in both rural and urban areas to ensure a high level of representativeness. All children were met with after their caregiver had given informed consent to their participation. They spoke fluent French and had not been diagnosed with any language impairment. The experiment began after the child also gave their informed consent.

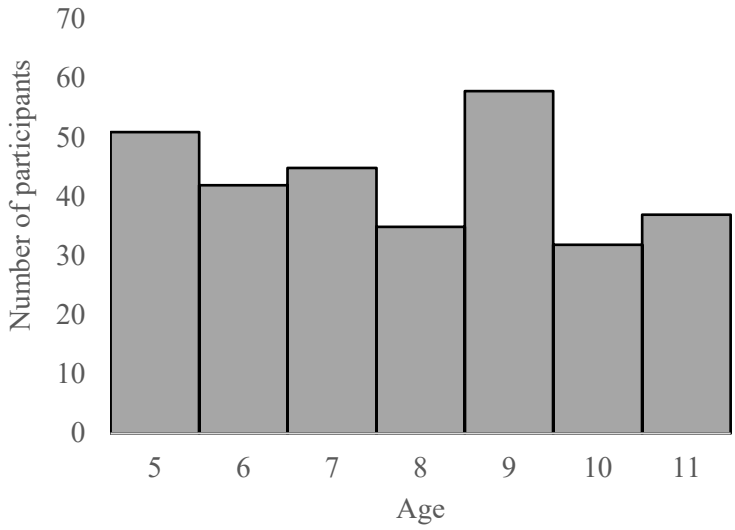

Figure 1. ***Distribution of Participants by Age***

Material

Metaphor understanding task

In the task used to assess metaphor understanding, children were presented with 10 metaphors, each preceded by a short context. Each metaphor was followed by a

rephrasing to be judged on a 5-point scale for resembling quality, that is, representing how closely the rephrasing resembled the metaphor. A similarity judgment on a scale between the metaphor and the reformulation was used, rather than selecting among multiple rephrasings, due to the large number of rephrasings proposed.
Answers were compiled as a resembling score, from –2 (poor meaning resemblance) to 2 (very good meaning resemblance). Rephrasings were randomly selected from one of 4 types (See Table 1 for an example). The complete set of material (contexts, metaphors, and rephrasings) is available on OSF. Hereafter, we refer to an item as a metaphor within its context and a rephrasing of this metaphor.

Table 1. ***Level of understanding reflected by each type of rephrasing when judged as highly resembling the metaphor "Victor collects photos of butterflies in an album. For him, butterflies are rainbows".***

Good understanding	**Good-enough understanding**	**Poor understanding**			**Absence of understanding**
Metaphorical	Situational	Contextual	Literal	Vehicle	Incongruous
Victor loves butterflies because they are full of colours.	Victor thinks that butter-flies are very beautiful insects.	Victor cuts out pictures of butter-flies from magazines.	A butterfly is a colourful half-circle that we see in the sky after it rains.	After it rains, Victor looks for rainbows in the sky.	Victor's father bought a new car.

The different rephrasings were created to reflect different levels of understanding. The first type of rephrasings was labelled as metaphorical rephrasings. When judged as highly resembling the target metaphor, this type reflected a nuanced and thorough comprehension of the metaphor by precisely mirroring the characteristics of the vehicle in relation to the topic. The second type of rephrasings, situational rephrasings, demonstrated children's good-enough understanding of a metaphor when they received high ratings. Situational rephrasings were relevant in the communication context but did not precisely convey the link between the relevant semantic features of the vehicle and the topic. In the example presented in the Table 1, the situational rephrasing "Victor thinks that butterflies are very beautiful insects" convey the positive, esthetical aspect of butterfly without linking it to their colorful aspect. The third type of rephrasings, poor rephrasings, reflected a poor understanding of the metaphor when they were rated as highly resembling the metaphor. Given the various ways one can misunderstand a metaphor, we chose to present participants with different subtypes of poor rephrasings within this category. Specifically, for each metaphor, a participant could choose (1) the literal subtype which literally translated the metaphor, without regard for the context, resulting in a semantic incongruity; (2) the vehicle-

oriented subtype which used the vehicle in its literal sense within a non-semantically incongruous sentence; (3) the contextual subtype, which was compatible with the context but did not convey the metaphor's figurative meaning. Finally, the fourth type of rephrasings, incongruous rephrasings, had no connection with the context of the metaphor, its figurative meaning, or the literal meaning of the words that made it up. When judged as highly resembling the target metaphor, this type of rephrasing would reflect a complete absence of understanding or consideration of the material, allowing for a methodological check to ensure that children correctly used the 5-point scale.

Metaphors, their embedding context and their rephrasings were pretested on multiple criteria, starting with an initial 28 items. First, we conducted two pretests on the metaphors themselves with adults to establish their comprehensibility and metaphoricity. We then conducted a pretest on the rephrasings with 295 adults (via an online questionnaire) to confirm that they did not all reformulate the metaphor equally well. Specifically, this pretest allowed us to retain only the items for which the metaphorical rephrasing was judged significantly better than the situational rephrasing, as intended in the construction of the material. A fourth pretest was conducted with 43 adults on the contexts accompanying the metaphors to ensure that situational rephrasings matched the context better than the contextual subtype of poor rephrasings, which in turn matched the context better than incongruous rephrasings. Finally, the 143 words from the initial pool of 28 items judged most difficult by the three authors of this article were included in two questionnaires designed to assess the knowledge of these words by 5- to 7-year-old children. These questionnaires consisted of 104 questions each and were administered to 67 young children (M_{age} = 6.5; SD_{age} = 0.89), allowing us to identify problematic words, i.e., those known by less than 70% of the children in the sample. When self-evident, a simpler word replaced a problematic one. More often, items containing problematic words were discarded. After these pretests, the initial number of items was reduced from 28 to 10 (refer to OSF documents for a detailed presentation of all pretests and results).

The final metaphor comprehension task was computerized with OpenSesame software (Mathôt et al., 2012). The interface allowed children to use a Likert scale represented by five coloured circles, blue to green, from left to right: non resembling, slightly resembling, moderately resembling, very resembling, highly resembling. Pictures of two similar cartoon characters were associated with the resembling side of the scale while two different characters were associated with the non-resembling side of the scale (see Figure 2). Once the whole item was heard, children could ask to replay either the context, metaphor or rephrasing to hear them again if needed. The children participated in a short training session on 6 items (5 literal and one metaphoric) before moving on to the test items to ensure their understanding of the task and answering scale. This training session was not the object of any analysis as its objective was primarily to allow the experimenter to explain the scale use to the

participants. On each of the 6 items, the experimenter precise the meaning of each circle by rephrasing the choice of the participants (e.g. when the participant chose the "-2" circle: "so you mean that [target sentence] is not at all resembling to the [rephrasing]?").

The selection of rephrasings was pseudo-randomised by OpenSesame, ensuring that, on average, participants encountered an equal number of acceptable rephrasings (i.e., metaphoric and situational rephrasings) and unacceptable ones (i.e., literal, contextual, and vehicle-oriented subtypes of poor rephrasings and incongruous rephrasings), approximately 5 of each. Specifically, the probability of selecting a metaphorical rephrasing was 2 out of 8, that of a situational rephrasing was also 2 out of 8, that of a poor rephrasing was 3 out of 8 (one for each subtype), and that of an incongruent rephrasing was 1 out of 8. OpenSesame also made it possible to randomly reverse the response scale.

The final task with 10 items was presented to 50 adults (M_{age} = 36.10; SD_{age} = 11.74; 29 women; 21 men) before being used with children to ensure that the task yielded consistent results in adults. A robust mixed model analysis showed that, as expected, participants judged metaphorical rephrasings as more resembling to the metaphor compared to situational ones (b = 0.848, CI_{95} [0.437, 1.258], $t(9.8)$ = 4.050, p = .002, η^2 = 0.626), those two better than poor rephrasings (b = 2.716, $CI95$ [2.346, 3.086], $t(9)$ = 14.382, p < .001, η^2 = 0.958) and finally incongruous as less resembling compared to all other rephrasings (b = 2.114, CI_{95} [1.833, 2.395], $t(28.78)$ = 14.76, p < .001, η^2 = 0.883) (See data and complete analyses on OSF)

In summary, the computerized metaphor comprehension task consisted of 10 items, each comprising a metaphor, a brief context, and a rephrasing, pseudo-randomly chosen from 4 possible types. For each item, the participant's task was to judge the extent to which the meaning of the rephrasing accurately conveyed the presented metaphor. Participants responded on a 5-point scale, allowing OpenSesame to collect a resembling score. Like the traditional forced-choice tasks of selecting the best metaphor rephrasing, the present task did not require participants to verbalise anything. However, unlike forced-choice tasks, participants were presented with only a single rephrasing for each metaphor, which (1) reduced the cognitive cost of the task by sparing them from having to maintain multiple competing rephrasings in working memory and (2) minimised the possibility of choosing a rephrasing by eliminating other options.

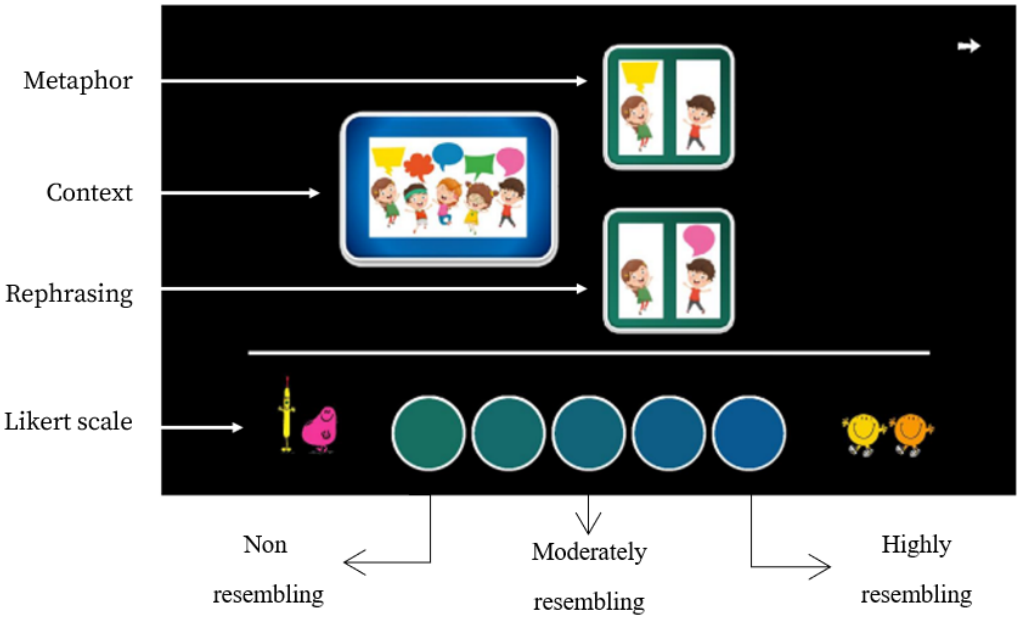

Figure 2. ***Opensesame Interface***

Peabody Picture Vocabulary Test (French Version)

The Peabody Picture Vocabulary Test (French version) (Dunn & Dunn, 1997) measured the receptive vocabulary level of French children by presenting four images and asking the participant to find the one representing a word. An age-adjusted standard score was calculated for each child, and this standard score was used for the statistical analyses (see the results section). Given the strong correlation between age and raw vocabulary score (r = .79), no young participant had a high raw vocabulary score, just as no older participant had a low raw vocabulary score, which prevented any satisfactory statistical modelling that included both age and raw vocabulary score.

Procedure

Children were met individually in a quiet room in their schools after their informed consent was obtained. They were first given the Peabody Picture Vocabulary Test. This test lasted approximately 10 minutes and was directly followed by the metaphor understanding task. This second task took around 15 minutes to be completed by participants.

Results

Overall, we collected 3000 observations (300 participants * 10 items), but 38 observations (1.27%) were discarded because we established that the participants did not hear either the metaphor or its rephrasing. All analyses and visualizations were conducted using the following R packages: lme4 (Bates et al., 2015), lmerTest (Kuznetsova et al., 2017), dplyr (Wickham et al., 2021), ggplot2 (Wickham et al., 2016), and robustlmm (Koller, 2016). All scripts and data can be found on OSF. We used a linear mixed model to analyse the resembling scores provided by participants depending on (1) their age, a continuous variable ranging from 5 to 11 years old, centred around the age 5; (2) the type of rephrasing presented, a four-level factor coded with a Helmert contrast: metaphorical rephrasings, situational rephrasings, poor rephrasings and incongruous rephrasings; and, most importantly, (3) the interaction between these two variables. We hypothesised that the resembling score would vary with age, depending on the type of rephrasing: it would increase with age for the metaphorical and situational rephrasing and decrease with age for poor rephrasings. Furthermore, we expected the resembling score for metaphorical rephrasings to show a greater increase with age compared to situational rephrasings. The participants' standard score on the Peabody Picture Vocabulary Test (a continuous variable centred around its mean) was included in the model as a covariate (4) to account for between-participant variability in vocabulary within each age bracket, as the standard score of vocabulary was adjusted for age. Additionally (5), since the interaction between the vocabulary score and the type of rephrasing was found to be significant, this interaction was also included in the model. This model was selected as the best fit for the data, as determined by the "step" function in the lmerTest package. The model's random structure, also determined with the "step" function, included by-schools, by-participants and by-items as random intercepts. The variance covariance structure was assumed to be unstructured. The degrees of freedom were calculated using the Satterthwaite method. The model assumptions were checked, and as the residuals deviated slightly from normal distribution, a robust analysis was conducted with the robustlmm package. Effect sizes were computed using the "tback" method described in Corell et al. (2022). Note that according to these authors, usual rule of thumb does not apply, η^2 being structurally smaller in mixed models.

Analysis of the Resembling Scores

As the variable age was centred around 5, this analysis indicates how children judged the different rephrasings at exactly 5 years old, independently of their vocabulary score (see Table 3 in Appendix A). Five-year-old children did manage to grasp the task, as evidenced by the fact that the resembling score they assigned to incongruous rephrasings was significantly lower than that assigned to other rephrasings (b = 1.69, *Wald* CI_{95} [1.33, 2.05], $t(2821) = 9.14$, $p < .001$, $\eta^2 = 0.029$); see Figure 3. However, no significant difference was found, either between the metaphorical and the

situational rephrasings (b = -0.15, *Wald* CI_{95} [-0.49, 0.19], $t(2830) = -0.87$, $p = .38$), or between these two acceptable rephrasings taken together and the poor rephrasings (b = 0.08, *Wald* CI_{95} [-0.17, 0.34], $t(2824) = 0.68$, $p = .50$), which prevents us from concluding that 5-year-old children understand metaphors.

The pattern observed at age 5 evolved over time, as shown by significant partial interactions between age and the type of rephrasing; see Figure 3. First, children's ability to differentiate between poor rephrasings on the one hand and situational and metaphorical (i.e., acceptable rephrasings) on the other hand increased significantly and rapidly with age (b = 0.28, *Wald* CI_{95} [0.21, 0.34], $t(2807) = 8.31$, $p < .001$, = 0.024). The resembling score for poor rephrasings decreased while it increased for acceptable rephrasings, indicating an improving comprehension of metaphors over time. An examination of Figure 3 suggests that by age 6, children were able to distinguish between these two types of rephrasings. More crucially and as expected, over time, participants increasingly judged metaphorical rephrasings as resembling the metaphor better than situational rephrasings. Accordingly, the difference in resembling scores between metaphorical and situational rephrasings became significantly more pronounced with age (b = 0.12, *Wald* CI_{95} [0.04, 0.21], $t(2824) = 2.76$, $p < .001$, $\eta^2 = 0.0027$). An examination of Figure 3 suggests that the difference was well established around age 9. In summary, between the ages of 5 and 11, the children not only advanced in their understanding of metaphors by rejecting poor rephrasings, but they also appeared to evolve from good-enough understanding (with no clear difference between situational and metaphorical rephrasings) to fine understanding, where they rated metaphorical rephrasings higher than situational ones.

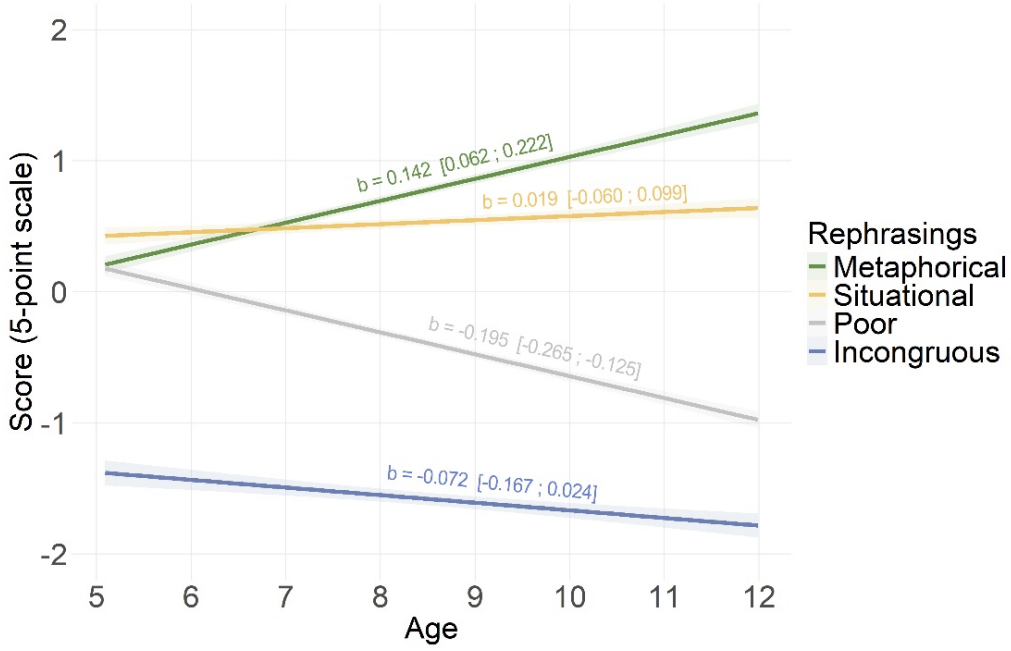

Figure 3. ***Resembling score for the four types of rephrasings as a function of age. Note: The translucent coloured areas around the lines represent the standard error. The b values reported inside the figure indicate the slope of each line [and its 95% confidence interval].***

In addition, a partial interaction between the standard score of vocabulary and the type of rephrasing was also observed. Specifically, as vocabulary level increased, the difference in resembling scores between metaphorical and situational rephrasings on the one hand and poor rephrasings on the other became more pronounced (b = 0.28, *Wald* CI_{95} [0.21, 0.34], $t(2807) = 8.31$, $p < .001$, $\eta^2 = 0.024$). In other words, at all brackets of age, children with the highest vocabulary proficiency were those who distinguished the best between acceptable and poor rephrasings. No other comparisons involving the standard score of vocabulary were found to be significant.

Exploratory Analyses on Poor Rephrasings

An original aspect of this study was to employ various subtypes of poor rephrasings of the metaphors: literal, contextual, and vehicle-oriented subtypes. We grouped these three subtypes of poor rephrasings together in the main analysis since our main hypotheses focused on the comparison between metaphorical, situational, and poor rephrasings. However, a post-hoc analysis of the resembling score differences between these different subtypes seemed valuable given the limited data in the literature comparing different forms of poor comprehension. This analysis, conducted using robust mixed-effects models with standard score of vocabulary as a covariate,

included three pairwise comparisons between literal, contextual, and vehicle-oriented rephrasings at age 5 and their three interactions with age. The Holm method was used to adjust for multiple comparisons (see Table 2).

Examination of Figure 4 reveals that at age 5, the contextual subtype was easier to identify as a poor rephrasing than the literal and vehicle subtypes, which shared a literal interpretation of the metaphor's vehicle. Over development, the resemblance score assigned to all three subtypes decreased, more notably for the vehicle subtype than the other two. Consequently, it was the literal subtype that was the most challenging to identify as a poor rephrasing for older children. However, with the exception of the difference between the contextual and vehicle subtypes at age 5, Holm's correction applied to limit the risk of type I error prevents concluding significant statistical differences (see Table 2).

Table 2. ***Post-hoc comparisons between the three subtypes of poor rephrasings at age 5 and throughout development***

Comparisons	**Estimate**	***SE***	**Lower 95% CI**	**Upper 95% CI**	***t***	***df***	**Holm *p* (adjusted)**	**η^2**
Age	-0.180	0.043	-0.264	-0.096	-4.200	14.64	0.006*	0.547
Literal vs. Vehicle	-0.155	0.261	-0.667	0.357	-0.594	1085	1	0.000
Literal vs. Contextual	0.595	0.258	0.090	1.100	2.308	1090	0.085	0.005
Vehicle vs. Contextual	-0.750	0.259	-1.258	-0.241	-2.891	1079	0.023*	0.008
Literal vs. Vehicle *Age	0.158	0.066	0.030	0.286	2.412	1085	0.080	0.005
Literal vs. Contextual *Age	0.006	0.068	-0.127	0.140	0.090	1087	0.929	0.000
Vehicle vs. Contextual*Age	0.152	0.066	0.022	0.282	2.285	1073	0.068	0.005

Note: * indicates statistical significance

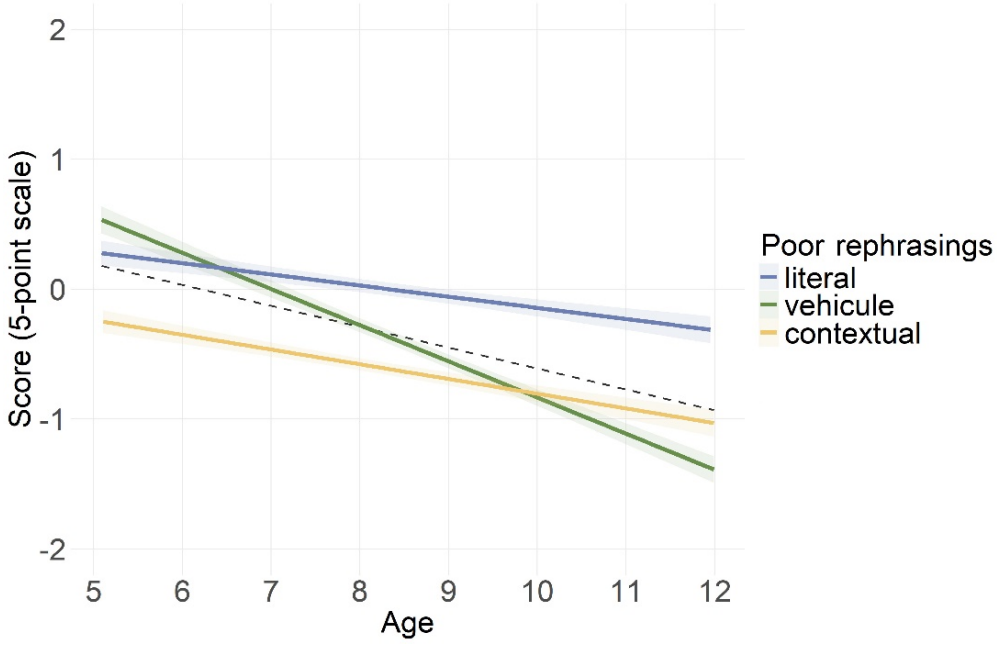

Figure 4. ***Resembling score for the three subtypes of poor rephrasings as a function of age. Note: the translucent coloured areas around the lines represent the standard error; the dashed ghost-line represents the resembling score for all three poor rephrasings, see also Figure 3.***

Discussion

The current study was designed to investigate the development of metaphor understanding in 5- to 11-year-old children within the theory of "good-enough" comprehension. An innovative metaphor rephrasing judgement task was used which was suitable for young children and did not involve verbal explanation or forced choices between several rephrasings of the metaphor. Our goals were to discover whether children demonstrated "good-enough" comprehension of metaphors, as opposed to "good" (precise and accurate) comprehension, and to test the hypothesis that good-enough comprehension precedes good understanding in development. We thus expected children to judge metaphorical rephrasings better than situational ones, and that their ability would become increasingly accurate with age.

Firstly, our data showed a vigorous and continuous development of metaphor comprehension between the ages of 5 and 11. This result was shown by a continuous increase in the resembling score assigned to the acceptable rephrasings, the metaphorical and situational rephrasings, and, simultaneously, a continuous decrease in the score assigned to the poor rephrasings. Our data thus support a substantial development in the understanding of metaphors throughout childhood which aligns with

most of the previous literature (Di Paola et al, 2019; Deckert al., 2019; Ortony et al., 1978; Pouscoulous et al., 2011; Vosniadou, 1987; Willinger et al., 2019). The most original contribution of our work is to show not only that the understanding of metaphors evolves but also how it evolves. Our results show that children around the age of 7 understand metaphors, as they clearly distinguish between poor and acceptable rephrasings. However, they do not differentiate between metaphorical rephrasings, which formulate the link between the topic and the vehicle, and situational rephrasings, which do not. Their lack of distinction between situational and metaphorical rephrasings likely reflects the children's good-enough understanding at that stage that is vague yet satisfactory. Later in development, around the age of 9, children are able to distinguish between metaphorical and situational rephrasings, judging the former to be more accurate than the latter. This result suggests that their understanding has evolved to become more precise and accurate – good understanding – by taking into account the very precise link between the topic and the vehicle.

Karimi and Ferreira (2016) argue that two routes can lead to these types of understandings: 1) a heuristic route, applying straightforward guidelines that can generate a rapid, general representation of the information being processed, providing a benefit in efficiency in terms of cognitive exertion; 2) an algorithmic route, processing with precise and unambiguous syntactical procedures to compute accurate representations from all the provided linguistic input. Our data support the idea that children primarily access the meaning of metaphors through the heuristic route around the age of 7. When processing complex statements such as metaphors, it is likely that the algorithmic route is not sufficiently effective or is too costly at this age. As development progresses, the algorithmic route develops to become children's preferred route for understanding metaphors by around 9 years old in the context of experimental tasks like those used in this study. Consistent with Ferreira's good-enough model, good comprehension via the algorithmic route does not replace good-enough comprehension via the heuristic route at some point between 7 and 9 years old. Instead, children seem to expand their range of strategies to cope with complex statements to understand as metaphors. In a real-life situation with low stakes or limited integration time, 9- to 10-year-olds can settle for good-enough understanding, just like adults. However, in the tasks typically used to measure metaphor comprehension in children, good comprehension is favoured over good-enough comprehension: Indeed, elements such as the setting of the experiment in a laboratory or school, the presence of the experimenter, or the academic formatting of the task can make it difficult to measure good-enough comprehension in an experimental setting as it leads children to produce a good understanding. This should be a point of concern for future research: if children have different treatment modes, based on different route, resulting in different levels of understanding, does the task used bias responses towards one treatment mode rather than another?

In addition to the observation of vigorous and continuous development of metaphor

comprehension between the ages of 5 and 11, the current study found that 5-year-old children struggled to distinguish between good and poor rephrasings, in particular those reflecting a literal interpretation of the metaphor's vehicle (literal and vehicle-oriented subtypes). We interpret the inability to distinguish between these rephrasings as a difficulty in understanding the meaning of the presented metaphors. This difficulty certainly cannot be attributed to a task unsuitable for children of this age, as they managed to judge incongruous rephrasings accurately. In addition, the vocabulary used in the task was carefully chosen to be understood by 5-year-old children. How do we explain the difficulties encountered by 5-year-old children in our study when previous research has shown that children can understand metaphors as early as at age 3 (e.g., Pouscoulous & Tomasello, 2020)? An explanation for this divergence in findings could lie in the methodology employed. The notion of metaphor corresponds to a large variety of utterance types that can also occur in a variety of situations. For instance, while Pouscoulous and Tomasello exclusively used simple metaphors (e.g., "The tower with the hat") referring to a presented object, the current study employed classic "an X is a Y" nominal metaphors embedded in a linguistic context and presented outside any situational context. Studying the development of metaphor comprehension in these different frameworks is valuable since it is the reality of what children experience every day.

Regarding the study of the evolution of metaphor comprehension, a contribution of the current study is related to the methodological choice to include measurements of poor understanding and not require the child to choose between different rephrasings of the metaphor. Asking children to evaluate each rephrasing on a scale provided a genuine method to assess the evolution of various potential interpretations of metaphors throughout the entire developmental period under consideration. In particular, our data suggest that literal rephrasings may pose difficulties even for older children. Although this exploratory observation needs to be confirmed by future research, such a tendency to retain the literal meaning has already been found in the comprehension of other forms of figurative language, such as idioms, metonyms or irony, and termed "literal bias" (Köder & Falkum, 2021). The two aspects presented, better assessing the impact of the intrinsic characteristics of metaphors on their comprehension throughout development, and better evaluating the developmental impact of literal bias on comprehension, appear to be important avenues for future research.

Although different from tasks that ask the child to explain the metaphors presented to them or tasks that ask them to choose among several options the one that best paraphrases the metaphor, our task of judging the similarity between the metaphor and a rephrasing on a 5-point scale necessarily involves metalinguistic skills. This is a clear limitation for many experiments that assess metaphor comprehension in children, because the development of metalinguistic skills after 5-year-old (Melogno et al., 2022) may explain part of the children's success or failure in the task. Measuring

children's metalinguistic skills, with CELF-V subtests per example (Coret & McCrimmon, 2015) and using this measure as a covariate in analyses is a direction to explore for future research.

In addition, future research should be aware of the risk of bias in the use of Likert scales with young children. Although scales have been used in experimental tasks with children as young as five years old, they have been criticized for a potential tendency among younger children to favor the extreme ends of the scale—a tendency that decreases with age (Chambers & Craig, 1998; Chambers & Johnstone, 2002; von Baeyer et al., 1997). This bias toward the scale's endpoints was observed in the present data. However, this bias alone cannot fully explain our results. To explore the impact of the bias, we performed a mixed logistic regression after recoding the original 5-point scale into a binary response system: scores of -2 and -1 were recoded as "no" (i.e., the metaphor and its rephrasing do not convey the same meaning), 9.8% of "0" scores were treated as N/A, and scores of +1 and +2 were recoded as "yes" (i.e., the two sentences convey the same meaning). This binary recoding of participants' responses eliminated both the scale's gradation and the potential bias. The regression yielded the same results as those presented in this article: we found an identical pattern of changes in resemblance scores for each type of rephrasing across age groups (see the "dichotomous answers" analysis and results on OSF). These additional analyses support the conclusion that, although a bias toward the scale's endpoints was present in young children, it does not account for the primary findings of this study. Nonetheless, future research using Likert scales with young children should remain mindful of this potential bias and exercise caution when employing similar methodologies.

These findings allow for a reconsideration of the classical debate between early vs. later understanding. Metaphor understanding appears to be best described through a developmental lens, beginning with good-enough understanding that allows children to process metaphors they hear in everyday contexts in a manner that can be difficult to measure in demanding, experimental settings. Precise and accurate understanding was shown here to emerge later, between the ages of 7 and 9 years old, which aligns with previous literature showing a late understanding of metaphors (Billow, 1977; Deckert et al., 2019; Cometa & Eson, 1978). Metaphor comprehension may even undergo further refinement during adolescence (Carriedo et al., 2016; Nippold, 2006). According to our statistical model, we observed a gap of 0.71 between the resembling scores assigned to situational and metaphorical rephrasings at precisely 12-years-old. Data from the pilot study with adults (see Methodology) indicated a gap of 0.85. Although merely descriptive, this statistic suggests that development in understanding might continue into adolescence. Future work would benefit from extending the developmental period both into preschool as well as into adolescence.

In conclusion, this study conducted with an original task over a significant developmental period paves the way for a new conceptualisation of metaphor understanding.

We have shown that good (precise and accurate) understanding of metaphors is a slow process extending throughout childhood and possibly beyond, and that before being able to fully understand metaphors, children go through a phase where their understanding is "good enough". New theoretical questions arise for the future: how does good-enough understanding emerge? In what way is good-enough understanding a lever for the development of good understanding? How and under what conditions do older children, adolescents, and adults transition from one type of understanding to another? There are also methodological questions, since it has become apparent that the tasks used to measure metaphor understanding inherently induce one level of understanding or another. Ultimately, considering and describing these different levels of comprehension can be used to support children in learning metaphors and perhaps other forms of language as well, both figurative and literal.

References

Almohammadi, A., Gaskins, D. K., & Rundblad, G. (2025). Metaphor comprehension in the acquisition of Arabic. *Journal of Child Language, 52*(2), 334–354. https://doi.org/10.1017/S0305000923000740

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. https://doi.org/10.18637/jss.v067.i01

Białecka-Pikul, M. (2010). Teaching children to understand metaphors as a path leading to theory of mind development. *European Journal of Developmental Psychology, 7*(5), 529-544. https://doi.org/10.1080/17405620802643692

Billow, R. M. (1977). Metaphor: A review of the psychological literature. *Psychological Bulletin, 84*(1), 81-92. https://doi.org/10.1037/0033-2909.84.1.81

Carriedo, N., Corral, A., Montoro, P. R., Herrero, L., Ballestrino, P., & Sebastián, I. (2016). The development of metaphor comprehension and its relationship with relational verbal reasoning and executive function. *PLoS One, 11*(3), e0150289. https://doi.org/10.1371/journal.pone.0150289

Chambers, C. T., & Craig, K. D. (1998). An intrusive impact of anchors in children's faces pain scales. *Pain, 78*(1), 27-37. https://doi.org/10.1016/S0304-3959(98)00112-2

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology, 27*, 27–36. https://doi.org/10.1093/jpepsy/27.1.27

Chiappe, D. L., & Chiappe, P. (2007). The role of working memory in metaphor production and comprehension. *Journal of Memory and Language, 56*(2), 172-188. https://doi.org/10.1016/j.jml.2006.11.006

Cometa, M. S., & Eson, M. E. (1978). Logical operations and metaphor interpretation: A Piagetian model. *Child Development, 49*(3), 649-659. https://doi.org/10.2307/1128232

Coombes, L., Bristowe, K., Ellis-Smith, C., Aworinde, J., Fraser, L. K., Downing, J., ... & Harding, R. (2021). Enhancing validity, reliability and participation in self-reported health outcome measurement for children and young people: a systematic review of recall period, response scale format, and administration modality. *Quality of Life Research, 30*, 1803-1832. https://doi.org/10.1007/s11136-021-02814-4

Coret, M. C., & McCrimmon, A. W. (2015). Test Review: Wiig, E. H., Semel, E., & Secord, W. A. (2013). Clinical Evaluation of Language Fundamentals–Fifth Edition (CELF-5). https://doi.org/10.1177/0734282914557616

Correll, J., Mellinger, C., & Pedersen, E. J. (2022). Flexible approaches for estimating partial eta squared in mixed-effects models with crossed random factors. *Behavior Research Methods, 54*(4), 1626–1642. https://doi.org/10.3758/s13428-021-01687-2

Deckert, M., Schmoeger, M., Schaunig-Busch, I., & Willinger, U. (2019). Metaphor processing in middle childhood and at the transition to early adolescence: The role of chronological age, mental age, and verbal intelligence. *Journal of Child Language, 46*(2), 334-367. https://doi.org/10.1017/S0305000918000491

Declercq, C., Baltazart, V., & Didon, D. (2010). Les nuages sont un troupeau de moutons: À quel âge les enfants comprennent-ils qu'il n'y pas de moutons dans le ciel? *CanadianJournal of Experimental Psychology, 64*(2), 142–152. https://doi.org/10.1037/a0020175

Del Sette, P., Bambini, V., Bischetti, L., & Lecce, S. (2020). Longitudinal associations between theory of mind and metaphor understanding during middle childhood. *Cognitive Development, 56*, 100958. https://doi.org/10.1016/j.cogdev.2020.100958

Diamond, A. (2020). Executive functions. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 225-240). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Di Paola, S., Domaneschi, F., & Pouscoulous, N. (2019). Metaphorical developing minds: The role of multiple factors in the development of metaphor

comprehension. *Journal of Pragmatics, 156,* 235-251. https://doi.org/10.1016/j.pragma.2019.08.008

Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test (4th ed.)*. [Measurement instrument]. NCS Pearson.

Erickson, T. D., & Mattson, M. E. (1981). From words to meaning: A semantic illusion. *Journal of Verbal Learning and Verbal Behavior*, *20*(5), 540-551. https://doi.org/10.1016/S0022-5371(81)90165-1

Falkum, I. L. (2022). The development of non-literal uses of language: Sense conventions and pragmatic competence. *Journal of Pragmatics*, *188*, 97-107. https://doi.org/10.1016/j.pragma.2021.12.002

Ferreira, F., Bailey, K. G. D., & Ferraro, V. (2002). Good-enough representations in language comprehension. *Current Directions in Psychological Science, 11*(1), 11–15. https://doi.org/10.1111/1467-8721.00158

Gardner, H., Kircher, M., Winner, E., & Perkins, D. (1975). Children's metaphoric productions and preferences. *Journal of Child Language*, *2*(1), 125-141. https://doi.org/10.1017/S0305000900000921

Gentner, D. (1988). Metaphor as structure mapping: The relational shift. *Child Development, 59*(1), 47-59. https://doi.org/10.2307/1130388

Gervain, J. (2020). Typical language development. In A. Gallagher, C. Bulteau, D. Cohen, & J. L. Michaud (Eds.), *Handbook of Clinical Neurology* (Vol. 173, pp. 171-183). Elsevier. https://doi.org/10.1016/B978-0-444-64150-2.00016-2

Huang, S. F., Oi, M., & Taguchi, A. (2015). Comprehension of figurative language in Taiwanese children with autism: The role of theory of mind and receptive vocabulary. *Clinical Linguistics & Phonetics*, *29*(8-10), 764-775. https://doi.org/10.3109/02699206.2015.1027833

Johnson, J., & Pascual-Leone, J. (1989). Developmental levels of processing in metaphor interpretation. *Journal of Experimental Child Psychology*, *48*(1), 1-31. https://doi.org/10.1016/0022-0965(89)90038-6

Karimi, H., & Ferreira, F. (2016). Good-enough linguistic representations and online cognitive equilibrium in language processing. *Quarterly Journal of Experimental Psychology*, *69*(5), 1013-1040. https://doi.org/10.1080/17470218.2015.1053951

Köder, F., & Falkum, I. L. (2021). Irony and perspective-taking in children: The roles of norm violations and tone of voice. *Frontiers in Psychology*, *12*, 624604. https://doi.org/10.3389/fpsyg.2021.624604

Kogan, N., Connor, K., Gross, A., & Fava, D. (1980). Understanding visual metaphor: Developmental and individual differences. *Monographs of the Society for Research in Child Development, 45*(1), 1-78. https://doi.org/10.2307/1165832

Koller, M. (2016). robustlmm: An R Package for Robust Estimation of Linear Mixed-Effects Models. *Journal of Statistical Software, 75*(5), 1-24. https

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1-26. https://doi.org/10.18637/jss.v082.i13

Lecce, S., Ronchi, L., Del Sette, P., Bischetti, L., & Bambini, V. (2019). Interpreting physical and mental metaphors: Is Theory of Mind associated with pragmatics in middle childhood? *Journal of Child Language*, *46*(2), 393-407. https://doi.org/10.1017/S030500091800048X

Mashal, N. (2013). The role of working memory in the comprehension of unfamiliar and familiar metaphors. *Language and Cognition, 5*(4), 409-436. https://doi.org/10.1515/langcog-2013-0024

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behavior Research Methods*, *44*(2), 314-324. https://doi.org/10.3758/s13428-011-0168-7

Nippold, M. A. (2006). Language development in school-age children, adolescents, and adults. In K. Brown (Ed.), *Encyclopedia of Language & Linguistics* (2nd ed., pp. 368-373). Elsevier. https://doi.org/10.1016/B0-08-044854-2/00852-X

Nippold, M. A., Leonard, L. B., & Kail, R. (1984). Syntactic and conceptual factors in children's understanding of metaphors. *Journal of Speech and Hearing Research*, *27*(2), 197–205. https://doi.org/10.1044/jshr.2702.197

Nippold, M. A., & Sullivan, M. P. (1987). Verbal and perceptual analogical reasoning and proportional metaphor comprehension in young children. *Journal of Speech, Language, and Hearing Research, 30*(3), 367-376. https://doi.org/10.1044/jshr.3003.367

Ortony, A., Schallert, D. L., Reynolds, R. E., & Antos, S. J. (1978). Interpreting metaphors and idioms: Some effects of context on comprehension. *Journal of Verbal Learning and Verbal Behavior, 17*(4), 465-477. https://doi.org/10.1016/S0022-5371(78)90283-9

Pierce, R. S., MacLaren, R., & Chiappe, D. L. (2010). The role of working memory in the metaphor interference effect. *Psychonomic Bulletin & Review, 17*, 400-404. https://doi.org/10.3758/PBR.17.3.400

Petit, N., Bambini, V., Bischetti, L., Prado, J., & Noveck, I. (2024). How do theory of mind and formal language skills impact metaphoric reference comprehension during children's school-age years. *Journal of Experimental Psychology: Learning, Memory, and Cognition 51*(3), 460–477. https://psycnet.apa.org/doi/10.1037/xlm0001381

Pouscoulous, N. (2011). Metaphor: For adults only? *Belgian Journal of Linguistics, 25*(1), 51-79. https://doi.org/10.1075/bjl.25.04pou

Pouscoulous, N., & Tomasello, M. (2020). *Early birds: Metaphor understanding in 3-year-olds. Journal of Pragmatics, 156,* 160-167. https://doi.org/10.1016/j.pragma.2019.05.021

R Core Team (2021). *R: A Language and environment for statistical computing*. (Version 4.1) [Computer software]. Retrieved from https://cran.r-project.org. (R packages retrieved from MRAN snapshot 2022-01-01).

Rundblad, G., & Annaz, D. (2010). 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

Stamenković, D., Ichien, N., & Holyoak, K. J. (2019). Metaphor comprehension: An individual-differences approach. *Journal of Memory and Language, 105*, 108-118. https://doi.org/10.1016/j.jml.2018.12.003

Tonini, E., Bischetti, L., Del Sette, P., Tosi, E., Lecce, S., & Bambini, V. (2023). The relationship between metaphor skills and Theory of Mind in middle childhood: Task and developmental effects. *Cognition, 238*, 105504. https://doi.org/10.1016/j.cognition.2023.105504

von Baeyer, C., Carlson, G., & Webb, L. (1997). Underprediction of pain in children undergoing ear piercing. *Behavioral Research and Therapy, 35*, 399–404. https://doi.org/10.1016/S0005-7967(96)00127-1

Vosniadou, S. (1987). Contextual and linguistic factors in children's comprehension of nonliteral language. *Metaphor and Symbolic Activity, 2*(1), 1-11. https://doi.org/10.1207/s15327868ms0201_1

Vosniadou, S., & Ortony, A. (1986). Testing metaphoric competence of the young child: Paraphrase vs. enactment. *Human Development, 29*, 226–230. https://www.jstor.org/stable/26764522

Wickham, H., Chang, W., Henry, L., Pedersen, T. L., Takahashi, K., Wilke, C., & Woo, K. (2016). *ggplot2: Elegant graphics for data analysis*. Springer-Verlag. https://ggplot2.tidyverse.org

Wickham, H., François, R., Henry, L., & Müller, K. (2021). dplyr: A grammar of data manipulation. R package version 1.0.7. https://doi.org/10.32614/CRAN.package.dplyr

Willinger, U., Deckert, M., Schmöger, M., Schaunig-Busch, I., Formann, A. K., & Auff, E. (2019). Developmental steps in metaphorical language abilities: The influence of age, gender, cognitive flexibility, information processing speed, and analogical reasoning. *Language and Speech, 62*(2), 207–228. https://doi.org/10.1177/0023830917746552

Winner, E., Engel, M., & Gardner, H. (1980). Misunderstanding metaphor: What's the problem? *Journal of Experimental Child Psychology, 30*(1), 22-32. https://doi.org/10.1016/0022-0965(80)90072-7

Data, Code and Materials Availability Statement

The material, data and analysis syntaxes that support the findings of this study are openly available on OSF at: https://osf.io/9pw7u/?view_only=None.

Ethics Statement

All the procedures contributing to this study complied with the ethical standards of the relevant national and institutional committees on human experimentation and with the 1975 Declaration of Helsinki, as revised in 2008. No information other than age, gender, first language, grade and oral consent were collected on behalf of the children. All children were met with after their caregiver had given informed consent

to their participation. The experiment began after the child also gave their informed consent.

Authorship and Contributorship Statement

Sarah Ferrara, **Marc Aguera** and **Christelle Declercq** designed the study and created the material. **Sarah Ferrara** collected and analysed data and wrote the first draft of the manuscript. **Marc Aguera** collected and analysed data and revised the manuscript along with **Christelle Declercq**. All authors approved the final version of the manuscript and agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Acknowledgements

The authors would like to express their gratitude to all participants and their families, as well as the students and schools involved in the data collection process. Additionally, we extend our thanks to colleagues who supported this work with their valuable advice, especially on analysis and statistical software.

License

Appendix A

Table 3: ***Fixed Effects Parameter Estimates***

	Estimate	***SE***	***t***	***p***	**η2**
(Intercept)	-0.063	0.141	-0.449	0.658	0.0099
Age	-0.026	0.031	-0.863	0.395	0.0251
Vocabulary	-0.003	0.003	-0.817	0.415	0.0025
Incongruous - (Poor, Situational, Metaphorical)	**1.688**	**0.185**	**9.140**	**0.000**	**0.0288**
Poor - (Situational, Metaphorical)	0.087	0.129	0.676	0.499	0.0002
Situational - (Metaphorical)	-0.150	0.172	-0.871	0.384	0.0003
Incongruous - (Poor, Situational, Metaphorical) ⁎ Age	0.061	0.045	1.354	0.176	0.0006
Poor - (Situational, Metaphorical) ⁎ Age	**0.276**	**0.033**	**8.314**	**0.000**	**0.0240**
Situational - Metaphorical ⁎ Age	**0.123**	**0.044**	**2.761**	**0.006**	**0.0027**
Incongruous - (Poor, Situational, Metaphorical) ⁎ Vocabulary	0.012	0.007	1.880	0.060	0.0013
Poor - (Situational, Metaphorical) ⁎ Vocabulary	**0.020**	**0.005**	**4.113**	**0.000**	**0.0060**
Situational - Metaphorical ⁎ Vocabulary	0.005	0.006	0.722	0.470	0.0002

Note: Significant results are presented in bold.