Acta Psychologica 267 (2026) 107138

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Numerical precision or inhibitory control? Exploring latent profiles in preschoolers' numerosity judgments

Irene Oeo Morín [a,b,*], Flore Maricau [b,c], Fien Depaepe [b,c], Bert Reynvoet [a,b]

[a] Brain and Cognition, KU Leuven, Leuven, Belgium
[b] Faculty of Psychology and Educational Sciences, KU Leuven @Kulak, Kortrijk, Belgium
[c] ITEC, IMEC Research Group at KU Leuven, Kortrijk, Belgium

ARTICLE INFO

Keywords:
Individual differences
Numerical precision
Inhibitory control
Cognitive profiles

ABSTRACT

Children show significant individual differences in their performance on non-symbolic numerosity comparison tasks. While some studies attribute this variability to differences in numerical precision, others highlight the importance of inhibitory control in filtering out irrelevant information. Rather than interpreting these findings as contradictory, the current study investigates whether distinct performance profiles underlie such variability. Specifically, it examines whether preschool children exhibit individual differences in their reliance on numerical versus non-numerical cues when comparing dot arrays. 256 preschoolers completed a dot comparison task. First, a confirmatory factor analysis was used to confirm the task's factor structure, corroborating a multidimensional model capturing both numerical precision and inhibitory control processes. As a novel approach, a Latent Profile Analysis was then applied to identify subgroups of children based on task performance. The results revealed three profiles: (1) children with good numerical precision and good inhibitory control who performed well across all trials ($N = 119$); (2) children with good numerical precision but poor inhibitory control who showed a strong congruency effect ($N = 40$); and (3) children with low numerical precision who performed poorly in all conditions ($N = 97$). These findings integrate previous conflicting results on the cause of individual variability by showing that performance differences may reflect differences in numerical precision in one group of children, and in inhibitory control in another group, indicating distinct cognitive profiles. This study enhances understanding of individual differences in numerosity perception and helps identify specific challenges children may be encountering.

1. Introduction

Numerosity perception is the ability to understand quantities and approximate the number of objects in a set without explicitly counting (Dehaene, 2001; Gilmore et al., 2018). This cognitive ability is a crucial skill in daily life and serves as a foundation for later, more advanced mathematical skills (see Schneider et al., 2017 for a metanalysis). Despite its importance, individuals sometimes encounter difficulties in perceiving numerosities (Gilmore, 2023).

A widely supported assumption is that numerosity perception is based on the Approximate Number System (ANS). This system allows individuals to extract number directly from a set. Despite its popularity, the specific factors underlying numerosity perception are still debated (see Lourenco & Aulet, 2023 for a review), with some of the theories suggesting that accurate numerosity perception requires also other cognitive abilities such as inhibitory control (Leibovich et al., 2017; Leibovich & Henik, 2013). These theories argue that when perceiving numerosity, individuals also need to inhibit irrelevant, distracting information from other continuous magnitudes such as size or spacing of the stimulus that is being processed in parallel with number.

1.1. Numerosity perception and its underlying cognitive processes

The numerosity comparison task is the most widely used instrument for studying numerosity perception, as it is considered one of the purest measures of this cognitive ability (Dietrich et al., 2015). In this task, participants are presented with two sets of dots and indicate which array is the more numerous one (e.g., a group of 12 dots vs. a group of 18

* Corresponding author at: Faculty of Psychology and Educational Sciences, KU Leuven @Kulak, Etienne Sabbelaan 51, 8500, Kortrijk, Belgium.
E-mail addresses: irene.oeomorin@kuleuven.be (I.O. Morín), flore.maricau@kuleuven.be (F. Maricau), fien.depaepe@kuleuven.be (F. Depaepe), bert.reynvoet@kuleuven.be (B. Reynvoet).

https://doi.org/10.1016/j.actpsy.2026.107138
Received 10 October 2025; Received in revised form 21 April 2026; Accepted 25 May 2026
Available online 29 May 2026

dots). Performance on this task is characterized by a *ratio effect*: participants perform better with larger ratios, that is when the distance between two numerosities is larger (10 vs. 20 dots is easier than 10 vs. 14 dots) (Gebuis & Reynvoet, 2012b; Halberda et al., 2012). This ratio effect can also be expressed with the Weber fraction, which represents the numerical difference between two sets of objects required for reliable discrimination between them. A smaller Weber fraction indicates a more precise numerical representation in the brain (Halberda et al., 2008).

However, numerosity perception is not only based on numerical information. Numerous studies observed that individuals' performance is also affected by non-numerical magnitude properties such as size or density (Abalo-Rodríguez et al., 2022; Defever et al., 2013; Gebuis & Reynvoet, 2011, 2012, 2012b; Reynvoet et al., 2021; Rodríguez & Ferreira, 2023; Soltész et al., 2010). This idea arises from the observation of a *congruency effec*t in comparison tasks, where performance improves when numerical and non-numerical properties are congruent (e.g., the most numerous group of dots also has the largest total surface area) than when they are incongruent (e.g., the most numerous group of dots has the smallest total surface area) (e.g., Gebuis & Reynvoet, 2012, 2012b; Oeo-Morín et al., 2025; Smets et al., 2015). It is argued that in such incongruent trials, performance also depends on a successful inhibitory control of the interfering non-numerical properties of the items (Gilmore et al., 2013; Piazza et al., 2018; Rodríguez & Ferreira, 2023; Szücs et al., 2013).

The role of inhibition in numerosity comparison is also supported by studies showing a relation between performance on typical inhibitory control tasks, such as the Stroop task, and performance on incongruent trials (Fuhs et al., 2018; Merkley et al., 2016). Further evidence in favour of the involvement of inhibition comes from studies on the congruency sequence effect. This effect was first described by Gratton et al. (1992) using a flanker task, another typical inhibitory control task. The researchers observed better performance in incongruent trials when they were preceded by another incongruent trial rather than a congruent one, suggesting a dynamic adjustment of cognitive control (Gratton et al., 1992). This congruency sequence effect has also been observed in numerosity comparison tasks (Roquet et al., 2020; Viarouge et al., 2019; Viarouge et al., 2023), where the congruency effect is also reduced following an incongruent trial. After encountering interference on an incongruent trial, participants appear to prepare themselves to process conflicting information more efficiently on subsequent trials. In contrast, after a congruent trial, inhibitory control is less engaged, leading to a larger congruency effect on the next trial. A stronger congruency sequence effect is observed in individuals with more efficient inhibitory control, such as adults and older children, but it might be absent in young children or older adults with less efficient inhibitory control abilities (Roquet et al., 2020).

Neuroimaging data further support the interaction between the nature of the trial congruency and inhibitory control. Several studies have shown that incongruent trials on numerosity comparison tasks elicit stronger neural responses in specific frontal brain regions (e.g., the inferior frontal gyrus) than congruent trials. This suggests that additional control mechanisms are recruited to manage the conflict between numerical and non-numerical properties (Leibovich et al., 2017; Leibovich & Ansari, 2017; Wilkey & Price, 2019). Interestingly, activation in these specific areas appears to predict children's mathematical competence (Wilkey & Price, 2019). Specifically, the study observed that children with higher mathematical scores showed a more efficient inhibitory control, meaning there was less difference in neural activation between incongruent and congruent trials (Wilkey & Price, 2019).

To sum up, the performance on a numerosity comparison task is likely not only based on how precisely number is extracted from the set, but also on the ability to suppress irrelevant information.

1.2. Individual differences in numerosity perception

Heterogeneity in numerosity perception performance is well documented (DeWind & Brannon, 2016; Guillaume & Van Rinsveld, 2018; Halberda et al., 2012). For example, there are age-related improvements, with older participants performing significantly better and showing a reduced congruency effect compared to younger individuals (Oeo-Morín et al., 2025). These developmental differences are also reflected in the Weber fraction values (Piazza & Izard, 2009), which show an exponential decrease with age. This decrease indicates an increasingly precise representation of number (Halberda & Feigenson, 2008).

Beyond age-related variation, individual differences in numerosity perception have also been linked to mathematical performance. For example, Starr et al. (2017) showed that variability in mathematical performance in preschoolers and adults was explained only by differences in numerical precision and not by inhibitory control of irrelevant non-numerical cues. Similarly, the study by Keller and Libertus (2015) found that performance of preschoolers on numerosity comparison tasks was significantly correlated with math achievement, even after controlling for inhibitory control, suggesting that inhibitory control does not drive differences in performance. Furthermore, in clinical populations, children with developmental dyscalculia (DD) or mathematical learning disabilities (MLD) tend to have overall higher Weber fractions, reflecting less accurate numerosity perception compared to typically developing children (Mazzocco et al., 2011; Piazza et al., 2018). In particular, Dolfi et al. (2024) showed a weaker numerical precision in children aged eight to fourteen years old with dyscalculia compared to a control group with matched average mathematical skills.

Alternatively, evidence suggests that the reason some children show a poor performance in numerosity comparison tasks may result from differences in inhibitory control abilities (Bugden & Ansari, 2016; Cappelletti et al., 2014; Gilmore et al., 2013; Rodríguez & Ferreira, 2023; Wilkey et al., 2020). In typically developing preschool children, research shows that inhibitory control abilities predict performance on the incongruent trials of a numerosity comparison task (Rodríguez & Ferreira, 2023). In this study, preschoolers with greater inhibitory control were better able to inhibit the non-numerical information of the fully incongruent trials (Rodríguez & Ferreira, 2023). Cappelletti et al. (2014) observed that older adults show reduced performance specifically on incongruent trials in a numerosity comparison task, due to a decrease inhibitory control ability typical of healthy aging. Regarding clinical populations, in the study of Bugden and Ansari (2016), it was observed that children aged between nine and thirteen with DD had higher Weber fractions, but only in incongruent trials. This finding suggests that poorer performance in numerosity comparison tasks is due to a difficulty inhibiting the non-numerical properties. Similarly, Wilkey et al. (2020) found that differences in performance between children with DD and low and typical achieving peers was limited to incongruent trials. This shows that children with DD may not have a generally impaired numerosity perception, but instead have difficulties with numerical magnitudes under additional executive function demands (Wilkey et al., 2020).

These previous results are not necessarily contradictory. Instead, they may reflect substantial heterogeneity in numerosity performance, with individuals relying to different degrees on the main cognitive mechanisms involved in solving the task (i.e., numerical precision and inhibitory control). In other words, individual differences in performance could arise from variation in both numerical precision and the efficiency of inhibitory control processes. Consequently, difficulties in numerosity comparison task performance could arise from impairments in one or both of these cognitive systems: a deficient numerical precision system or a weakened inhibitory control (Cowan & Powell, 2014; Gilmore et al., 2013; Szücs et al., 2013; Wilkey et al., 2022).

Building on this idea of heterogeneous individual performance patterns, this study aims to explore whether different performance profiles can be identified in preschool children when completing a numerosity comparison task. These profiles may reflect the extent to which children rely on numerical precision versus inhibitory control when solving the task. To our knowledge, no prior research has examined individual profiles in this context. By using a Latent Profile Analysis (LPA) to capture different performance patterns, this study aims to better understand the origins of individual variability and clarify why the existing literature reports mixed findings.

1.3. The present study

Preschool children were selected for this study because at this stage many children are being introduced to mathematical concepts and activities, including numerosity perception (Anders et al., 2013). In addition, at this age both executive functions and number sense are bidirectionally related (Ernst et al., 2025).

The primary objective of the current study is to conduct a LPA to categorise children into different profiles based on their performance in a non-symbolic numerosity comparison task. Unlike traditional methods such as ANOVA, LPA accounts for heterogeneity in performance that is not observable a priori and does not rely on assumptions such as linearity, normality, or homogeneity. This makes it a robust alternative to conventional clustering techniques (Magidson & Vermunt, 2002; Spurk et al., 2020). The study expects to identify at least three different profiles: (1) children with good numerical precision and good inhibitory control; these children are expected to show a high performance on both congruent and incongruent trials; (2) children with good numerical precision but poor inhibitory control; these children are expected to show an acute congruency effect, performing well on congruent trials but poorly on incongruent trials; and (3) children with poor numerical precision; these children perform poorly in all conditions (including large ratio and congruent trials). Differentiating a profile characterized by poor numerical precision but high inhibitory control from one with both poor numerical precision and poor inhibitory control is difficult within this paradigm. When numerical precision is low, performance on both congruent and incongruent trials is expected to be poor, regardless of the inhibitory ability, as the child cannot reliably extract the relevant numerical information. Consequently, the current task may not allow inhibitory differences to manifest behaviourally in a way that allows a distinct latent profile to emerge.

To conduct a LPA, it is first necessary to confirm the underlying factor structure of the task. The study hypothesises that participants' performance will be driven by numerical precision and inhibitory control. Therefore, a multidimensional model would provide a better fit to the data than a unidimensional model. This hypothesis is also consistent with the involvement of both numerical precision and inhibitory control processes in numerosity perception (Fuhs & McNeil, 2013; Leibovich et al., 2017; Leibovich & Henik, 2013; Lourenco & Aulet, 2023; Soltész et al., 2010).

Understanding the mechanisms underlying success and failure in numerosity comparison tasks may provide valuable insights into the origins of learning differences in the context of numerosity perception (Hart et al., 2016). By identifying distinct performance profiles, educators can better recognize specific challenges students may be encountering, leading to customized educational interventions aimed at improving concrete cognitive abilities involved in numerosity perception.

2. Methods

2.1. Participants

A total of 262 children completed the study. Data from six participants were removed (see Section 3.1 on data preparation), leading to a final sample of 256. The sample included preschool children between the ages of four and seven[1] ($M = 5.49$ years, $SD = 0.65$, 46.9% girls), recruited in ten preschools in Belgium and one preschool in Spain. No significant differences between schools were found in terms of the children's average performance ($p = .55$).

Of the ten preschools in Belgium, five were international preschools, with a total of eight different classes. In these schools, the task instructions were provided in English, and the teachers made sure that language was not a barrier for the children. The other five preschools were Flemish schools, data was recruited from ten different classes and instructions were given in Dutch. Finally, the data from one preschool in Spain was recruited from two different classes and instructions were given in Spanish. In all cases, an information letter explaining the aims of the study, the methodology, data protection measures and assurances of anonymity was given to the children's legal representatives before the task was carried out. Written informed consent was then obtained from them. Teachers and children were clearly informed that participation was optional and they could withdraw from the study at any time.

2.2. Procedure

The data was collected in two rounds, one at the end of the academic year (i.e., between April and May) and the other at the beginning of the academic year (i.e., between September and October). All participants completed the task on an iPad tablet (85%) or Samsung Galaxy Android tablet (15%). The preschool children worked in groups of no more than seven in a separate room. Following the recommendations of Krajcsi et al. (2024), and to ensure that the children understood the instructions, they were first given two printed examples with specific feedback before starting with the practice trials of the task. The instructions were given verbally and repeated after the practice trials.

2.3. Materials

2.3.1. Numerosity comparison task

Participants were asked to compare two sets of dots displayed side by side on a grey background. The dots in the set on the left were always blue, and the dots in the set on the right were always yellow (see Fig. 1). Following the instructions of previous studies, participants were asked to decide, as quickly as possible and without counting, which set of dots was the more numerous one (Oeo-Morín et al., 2025).

As mentioned in the introduction, numerosity is confounded by the non-numerical visual properties of the objects. To ensure that participants were making judgments on the basis of numerosity, the non-numerical cues were controlled with the CUSTOM algorithm (De Marco & Cutini, 2020). This algorithm is available for Matlab (version R2018b, The MathWorks Inc., Natick, Massachusetts, USA) and produces paired dot arrays. The CUSTOM algorithm enables the manipulation of the numerosities and the control of different non-numerical visual properties of the stimuli, including: *Convex Hull*, which refers to the area enclosed by an imaginary boundary connecting all the external dots; *Total surface area*, the sum of all dots surface areas; *Total contour*, the combined perimeter of all dots; and the *Average Diameter* of the dots. Density is also manipulated indirectly, as it corresponds to the *Total*

[1] Preschool normally includes children up to age six. However, data were collected in international schools, and therefore, some children had to complete an extra year of preschool due to language barriers, curriculum differences, or variations in the educational system of their home countries.

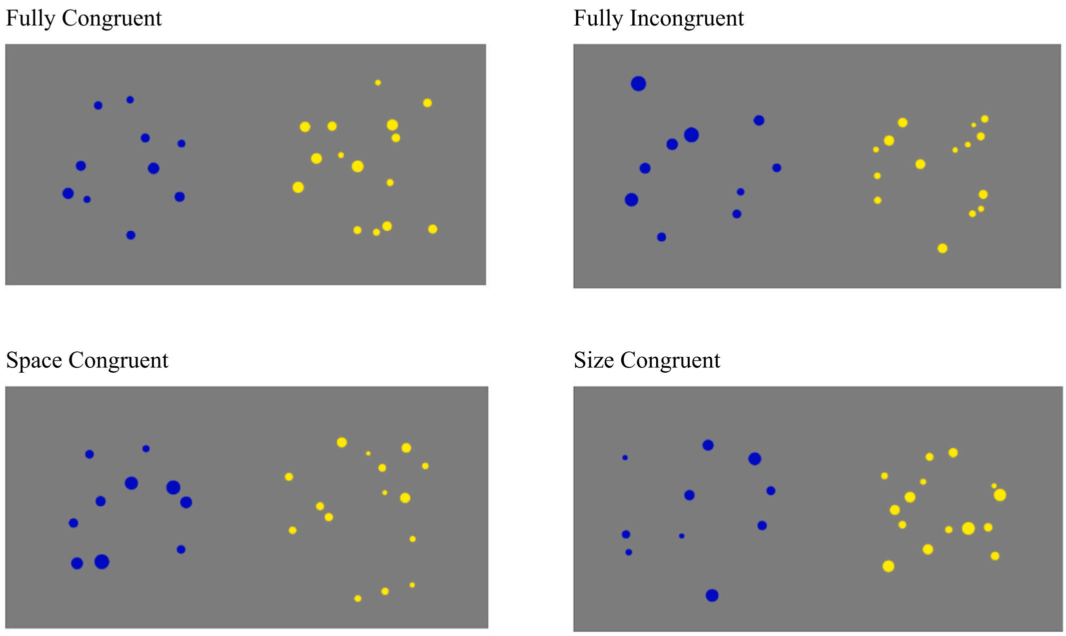

Fig. 1. Examples of dot arrangements for each congruency condition.

surface area divided by the *Convex Hull* (De Marco & Cutini, 2020).

In this study, the CUSTOM algorithm parameters were set in the same way as in the developmental study of Oeo-Morín et al. (2025): numerosities ranging from 10 to 40, Convex Hull values of 11,000 and 75,000, and Total Surface Area values of 4100 and 6000, applying a tolerance ratio of .001. Using this settings, the algorithm produced pairs of dot arrays that were either congruent or incongruent with respect to *Convex Hull* and/or *Total surface area* (Oeo-Morín et al., 2025).

The stimuli set was subsequently converted into the *Number*, *Size* and *Spacing* parameters using the mathematical model developed by DeWind et al. (2015). In this framework, *Number*, *Size* and *Spacing* are considered independent dimensions, with each dot array characterized by a distinct combination of values across these features. The *Size* parameter is derived from the *Total Surface Area* and *Total Contour*, while the *Spacing* parameter is computed from the *Convex Hull* and *Sparsity* (see equations in DeWind et al., 2015). These mathematical computations provided a methodological advantage. In this way, *Size* and *Spacing* are orthogonally independent of *Number* and of each other (see Table 1 Supplementary Material). This allows the assessment of the specific contribution of each feature (i.e., *Number*, *Size* and *Spacing*) in the responses of the children (DeWind et al., 2015).

Using the new *Size* and *Spacing* parameters, and as previously done (Oeo-Morín et al., 2025), the stimuli were categorised into four conditions: (1) Fully Congruent, where the most numerous set of dots is also larger in *Size* and *Space* (upper left in Fig. 1); (2) Fully Incongruent, where the most numerous set of dots is not larger in *Size* and *Space* (upper right in Fig. 1); (3) Space Congruent, where the most numerous set of dots is larger in *Space* but not in *Size* (lower left in Fig. 1); and (4) Size Congruent, where the most numerous set of dots is larger in *Size* but not in *Space* (lower right in Fig. 1).

Four different ratios (1.17, 1.25, 1.5 and 2) were selected to avoid ceiling and/or flooring effects and to create different levels of difficulty, in line with developmental studies (Halberda & Feigenson, 2008; Oeo-Morín et al., 2025). For each ratio, participants were presented four distinct pairs of numerosities (e.g., for ratio 1.5: 10 vs. 15, 14 vs. 21, 22 vs. 33, and 26 vs. 39 on both sides). The design and counterbalancing procedures followed those used in a previous study by Oeo-Morín et al. (2025).

The task had a total of 128 trials. Each item began with a fixation cross presented for 1000 ms, after which an image containing two dot arrays appeared (1200 × 600 pixels). The display remained visible until the child tapped the array they judged as more numerous. Keeping the stimuli on screen until a response was made helped minimise working memory demands and reduced the likelihood of errors due to lapses of attention or missed stimuli (Krajcsi et al., 2024).

To make sure children understood the task, participants first completed a series of easier practice trials with larger ratios (i.e., 3.5 and 4) and received immediate trial-by-trial feedback on their performance, as suggested by Negen and Sarnecka (2015). These practice trials were also equally representing the four congruency conditions. Participants had to complete at least eight practice trials with feedback. They could only proceed to the test trials if they achieved an accuracy rate of at least 75%. Accuracy during the practice phase was calculated cumulatively across all completed practice items. If a participant did not reach the 75% accuracy threshold, additional practice trials with feedback were provided until the criterion was met, up to a maximum of 32 practice trials. Children who failed to achieve this criterion could still proceed with the task, but their data were excluded from the final analyses (Oeo-Morín et al., 2025). The task was administered in the online software of PsychoPy, Pavlovia (Peirce et al., 2019). There was a short break in the middle, and the task took about 10 min to complete.

3. Results

3.1. Data preparation

Children who achieved the maximum number of practice trials with feedback but still failed to answer correctly in 75% of the trials were not included in the final sample ($N = 6$). Therefore, the final analysis was conducted with a total of 256 participants. To identify potential outliers, reaction-time data were visualized using graphical methods, and values exceeding three standard deviations above or below the mean were examined. Reaction times were inspected solely for the purpose of detecting excessively long or short responses that could indicate counting or random answering. However, no significant cases were found. Reaction times were not included in the subsequent analyses due to the high variability caused by children's distractions. Moreover, no behaviours suggesting counting strategies, such as sequential pointing or finger use, were observed among the preschool children.

For the analyses involving the calculation of the mean Weber fraction per performance profile, an additional data-cleaning step was applied. Specifically, negative Weber fraction values and positive values exceeding 2.5 were excluded (Oeo-Morín et al., 2025). Such extreme values arise when participants rely heavily on non-numerical cues, respond inconsistently, or exhibit heavy random responding. As a result, 225 participants were included in these specific analyses.

3.2. Ratio and congruency effects

A Generalized Linear Mixed Model (GLMM) was used to explore the main ratio and congruency effects in the comparison task. This analysis was used to account for clustering of data within students and items. The *lme4* (Bates et al., 2015) and the *lmerTest* (Kuznetsova et al., 2017) packages in R were used for these analyses. Pairwise post hoc Tukey comparisons were calculated for significant congruency interactions using the *emmeans* package (Lenth, 2025). In addition, following DeWind's mathematical model (DeWind et al., 2015), the individual beta weights for the three predictors *Number*, *Size*, and *Spacing* were derived from a cumulative-Gaussian psychometric model, on each participant's data. In this model, the β coefficients quantify the weight that each visual cue contributes to the latent decision variable. These coefficients therefore provide interpretable indices of numerical acuity as well as systematic biases in how participants use non-numerical stimulus dimensions (DeWind et al., 2015). To compare possible differences between participant's strategies when performing a comparison task, a repeated measures ANOVA was performed with the three regression coefficients as the within-subjects variables.

The GLMM revealed that the best-fitting model included participant and item as random effects, and congruency type, ratio difficulty, and age as fixed effects (see Table 2 in the Supplementary Material). The effect of congruency, ratio, and age on accuracy was tested using Type III Wald chi-square test, due to the categorical nature of the accuracy variable. The results showed a significant ratio effect ($\chi^2(1) = 54.18$, $p < .001$), with participants showing higher accuracy on larger ratios (see Fig. 1 in the Supplementary Material), as well as a significant congruency effect ($\chi^2(3) = 52.2$, $p < .001$), with students showing higher accuracy scores on *Fully Congruent* ($M = 0.75$) and *Space Congruent* ($M = 0.74$) trials compared to *Fully Incongruent* ($M = 0.52$) and *Size Congruent trials* ($M = 0.56$) (see Fig. 2 in the Supplementary Material). Post hoc analysis revealed no significant differences between *Fully Congruent* and *Space Congruent* ($p = .92$) and between *Fully Incongruent* and *Size Congruent* items ($p = .26$). In addition, a significant interaction effect was found between congruency and ratio ($\chi^2(3) = 14.84$, $p < .01$) (see Fig. 3 in the Supplementary Material). Finally, age had a significant effect ($\chi^2(1) = 27.56$, $p < .001$, with older participants having better overall accuracy (see Fig. 4 in the Supplementary Material).

The computation of the regression coefficients for the numerical (i.e., *Number*) and non-numerical predictors (i.e., *Size* and *Spacing*) showed significant positive values for the three regression coefficients (see Table 1), suggesting that participants' performance was influenced by these three predictors (DeWind et al., 2015). *Number* emerged as the main predictor of performance, showing the largest β value (see Table 1). Higher positive βNumber values indicate that subjects tended to focus more on, or demonstrated stronger sensitivity to, numerical information. This allowed them to discriminate between more difficult ratios. In addition, a positive βSpacing value suggests that more spread-out dots were perceived as more numerous, leading to better performance when this feature was congruent with number (i.e., the set of more dots is also the set with more spread out dots) (DeWind et al., 2015). βSize was very close to zero but still significantly different (see Table 1), indicating that the bigger the size of dots the more numerous the group is perceived (DeWind et al., 2015). Significant differences were observed between the *Number*, *Size* and *Spacing* predictors (F (1.22, 311.05) = 392, $p < .001$, $\eta^2 = 0.606$). Pairwise comparisons revealed differences between the three predictors ($p < .001$). A Pearson correlation coefficient was calculated to evaluate the relationship between age and the three regression coefficients. There was a significant, but weak, positive relationship only between age and the predictor *Number* ($r = 0.23$, $p \leq 0.001$), with older participants showing a higher βNumber. Interestingly there was no significant effect of age in the βSpacing ($r = 0.055$) nor βSize ($r = 0.012$) suggesting that these biases are consistent across the different ages. The mean Weber fraction was computed as a transformation of the βNumber (DeWind et al., 2015). As previously mentioned, additional data cleaning was performed for these analyses. Values of the Weber fraction that were negative or exceeded 2.5 were

Table 1
Descriptives of the three regression coefficients.

	Mean	SD	Minimum	Maximum	T-score	Cohen's d
βNumber	1.18	0.96	−0.31	9.64	19.6**	0.96
βSpacing	0.31	0.41	−0.36	4.62	12.09**	0.41
βSize	0.03	0.19	−0.63	1.98	2.87*	0.19

Note: SD, standard deviation.
** $p < .001$.
* $p < .01$.

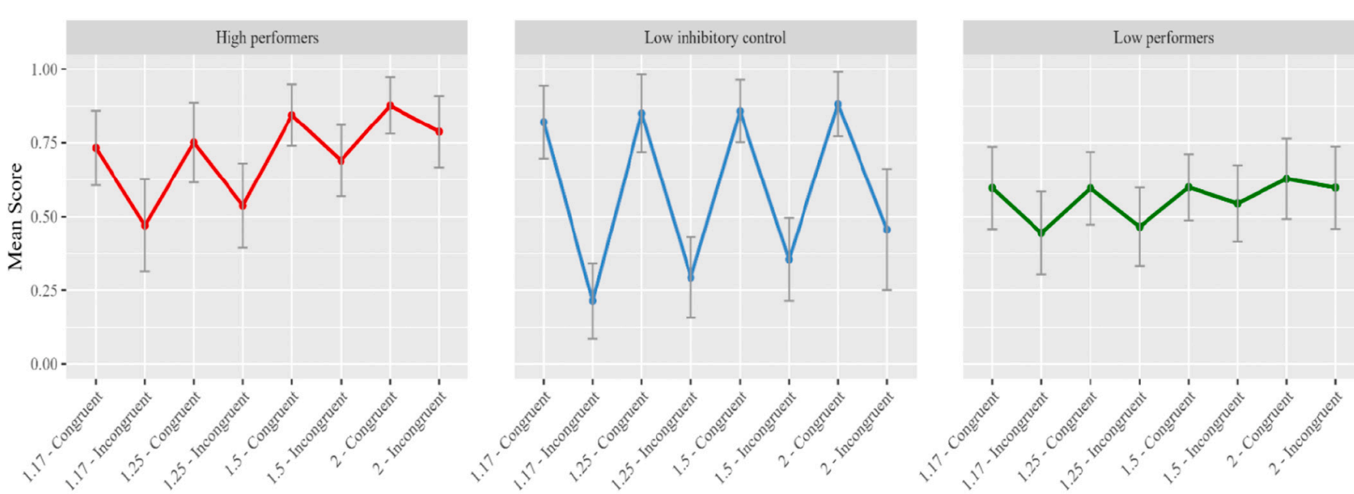

Fig. 2. Latent Profile Analysis results.
Note. The "Mean Score" on the y-axis reflects the mean score on the numerosity comparison task. The number before "Congruent" or "Incongruent" indicates the ratio of the dot pair (i.e., 1.17, 1.25, 1.5 or 2).
*Additional analysis shows a significant ratio effect in all three profiles ($p = .001$), indicating that performance significantly improves as the ratio increases.

excluded (Oeo-Morín et al., 2025). As a result, these specific analyses included data from 225 participants. The mean Weber fraction was 0.74 ($SD = 0.45$; $Median = 0.60$; $N = 225$). A Weber fraction of 0.74 indicates that children, on average, require the larger set to contain approximately 74% more items than the smaller set to reliably identify which has more.

3.3. Underlying factor structure of the task

The primary aim of the study was to detect different performance profiles of children in a numerosity comparison task. As already mentioned, to do this, confirmatory multidimensional item response theory (MIRT) with a Rash model was first performed to confirm the underlying factor structure within the data. Analysis were conducted using the *mirt* package in the R software (version 4.3.2) (Chalmers, 2012). Based on previous literature on numerosity comparison, five hypothesised models were tested. The first tested model assumed a unidimensional factor structure, with all items falling into one latent variable. The second model assumed a ratio-only factor structure, where the items fall into four categories based on ratio and performance is assumed to be based only on the numerical properties of the stimuli. The third model assumed a congruency-only factor structure, with items falling into two congruency conditions, and assumes that performance is based only on the non-numerical visual cues. Based on previous literature (Oeo-Morín et al., 2025) and the preliminary analysis of the current study, it was hypothesised that *Space Congruent* and *Fully Congruent* items will fall into the same congruency category (i.e., Congruent) and *Fully Incongruent* and *Size Congruent* items will be the other congruency category (i.e., Incongruent). A fourth model assumed a 4 × 2 factor structure in which performance on the task is guided by both the ratio and congruency type (i.e., the numerical and non-numerical visual cues). Finally, a last model assumed a 4 × 4 factor structure to test the original structure of the task, in which congruency conditions were not clustered into two but were treated as separate factors. All models assumed covariance between the different factors. The fourth model was hypothesised to provide the best fit to the data, as it takes into account the numerical and non-numerical properties of the stimuli (e.g., Lourenco & Aulet, 2023), and merges the congruency conditions into two (Oeo-Morín et al., 2025). Acceptable model fit was determined based on the root mean square error of approximation (RMSEA) and the standardized root mean square residual (SRMR), both below.08, and the Comparative Fit Index (CFI) and Tucker Lewis Index (TLI), both above 0.90 (Hooper et al., 2008; Hu & Bentler, 1999). To compare the hypothesised models, Cai and Monroe's (2014) C2 index was calculated, with the smaller value being preferred (Said-Metwaly et al., 2020).

Table 2 shows the model fit indices of the different hypothetical models. None of the models met the threshold requirements for the CFI and TLI to be considered an acceptable fit (>0.90), but they did meet the requirements for the RMSEA and SRMR (<0.08). In fact, methodologists suggest that CFI may be more appropriate for exploratory contests, whereas RMSEA indices is considered to be more informative for confirmatory factor analysis as this one (Rigdon, 1996). In addition, it is important to note that different fit indices reflect different aspects of model fit (Hu & Bentler, 1999). The RMSEA measures absolute fit, whereas the CFI is an incremental index that compares the target model to a null model. In multidimensional data, null models usually fit poorly, making the CFI appear artificially low even when absolute fit indices like the RMSEA (<0.05) indicate good fit (Kenny, 2015). Furthermore, simulation studies show that CFI and TLI can fall below 0.90 in complex models even when correctly specified (Marsh et al., 2004). This further suggests that the CFI and TLI should not be interpreted in isolation.

Importantly, of all the models tested, the 4 × 2 factor structure showed the best fit (see Table 2), especially compared to the unidimensional and the ratio-only models. This was indicated by values within the satisfactory threshold of the RMSEA and SRMSR, as well as higher CFI and TLI indices, and the lowest C2 value. Therefore, considering both the theoretical assumptions and the fit requirements, the 4 × 2 correlated model appeared to be the best option.

Table 2
Model fit indices of the MIRT analyses of the hypothesised models.

Model	*df*	C2	CFI	TLI	RMSEA	SRMSR
Unidimensional model	8127	13,539.03	0.33	0.33	0.05	0.08
Ratio only correlated model (4 levels)	8118	13,347.38	0.35	0.35	0.05	0.08
Congruency only correlated model (2 levels)	8125	10,402.13	0.72	0.72	0.03	0.07
Eight-factor correlated model: ratio (4 levels) × congruency (2 levels)	**8092**	**10,164.75**	**0.74**	**0.74**	**0.03**	**0.07**
Sixteen-factor correlated model ratio (4 levels) × congruency (4 levels)	7992	10,242.98	0.72	0.72	0.03	0.07

Note. The chosen model is in **bold**.

Empirical reliability was assessed in the MIRT analysis to evaluate the internal consistency of the best-fitting model. Empirical reliability is specific to item response theory and estimates the squared correlation between observed scores and true latent trait levels (Brown & Maydeu-Olivares, 2011). Unlike classical reliability indices that provide a single overall estimate (e.g., Cronbach's alpha), empirical reliability varies across the latent trait continuum. This measure provides a more detailed understanding of measurement precision at different ability levels. The empirical reliability of the 4 × 2 model was found to be high for the eight specific factors, with values ranging from 0.80 to 0.85.

3.4. Performance profiles

The eight factors of the best fitting model were then used as dependent variables in the LPA to classify individuals who share similar performance patterns. In this analysis, children are assigned to different profiles with varying probabilities. At least three different profiles were hypothesised (see Section 1.3). However, as this is the first time to our knowledge that a profile analysis has been done in a numerosity comparison task, an exploratory approach was adopted. Several models were fitted to the data, each increasing the number of profiles from two to nine, in order to determine the optimal number of distinct profiles. The guide of Spurk et al. (2020) on LPA was used for the interpretation and selection of the most appropriate number of profiles. This selection was based, firstly, on evaluating the fit of the different models and, secondly, on considering the most theoretically meaningful and plausible solution. Several indices of goodness of fit were used, including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), entropy and the posterior probabilities. These indices were selected because they capture complementary aspects of model evaluation. While the BIC and AIC assess the relative statistical fit across competing models, entropy and posterior probabilities specifically evaluate classification precision (Clark & Muthén, 2009). Using both types of criteria together with a theoretically relevant choice is recommended to avoid selecting solutions that fit well statistically but produce poorly separated or theoretically unclear profiles (Spurk et al., 2020). Lower values of the AIC and BIC indicated a better fit (Spurk et al., 2020). However, the AIC favours more complex latent profile solutions systematically, so it should not be used as the primary criterion (Morgan, 2015; Spurk et al., 2020). In addition, the entropy level, which indicates how well the models classify individuals, was expected to be as close to one as possible with a cut of score of 0.80 (Clark & Muthén, 2009). Posterior probabilities of belonging to each profile were also considered, selecting the profile with posterior probabilities closer to one. Finally, it is also highly relevant that each profile has a significant number of participants (i.e., covering at least 1% of the total sample or more than 25 cases) (Lubke & Neale, 2006). The LPA analysis was performed using the *tidyLPA* (Rosenberg et al., 2018) and *mclust* (Scrucca et al., 2016) packages in R. For the validation of the identified profiles, significant differences between the relevant variables were analysed.

Although the six profile solution had the lowest BIC value (see Table 3), additional diagnostic criteria indicated that this model was not optimal. Specifically, the six-profile solution produced a very small profile of only 13 participants, which is generally considered too small for reliable interpretation (Lubke & Neale, 2006). Furthermore, it did not provide theoretically meaningful distinctions. Instead, it generated sub profiles representing variations of the three broader profiles (e.g., slightly weaker or slightly stronger inhibitory control or numerical acuity). The literature emphasises that model selection should consider not only statistical fit indices but also theoretical coherence and substantive interpretability (Spurk et al., 2020).

Taking the balance of fit indices and the theoretical relevance into account, the three profile model emerged as the most appropriate solution. As shown in Table 3, the three profiles also resulted in the highest entropy (i.e. the certainty of the model to provide well-separated profiles) and the highest posterior probabilities (i.e. the probability that individuals are correctly classified). Furthermore, all profiles were of sufficient size to allow meaningful interpretation.

Once the best fitting profile classification was selected, the accuracies of each of the eight factors or blocks of items (see Fig. 2), as well as the regression coefficients of the numerical and non-numerical predictors (see Fig. 3) were compared between the different profiles. The profiles are described as follows:

Profile one: High Performers. High numerosity perception and high inhibitory control. The first profile was the largest with more than half of the sample falling into this profile (N = 119). Children in this profile were characterized by a higher performance on both congruent and incongruent trials at all ratio levels (see Fig. 2). This group showed a high βNumber (M = 1.60; SD = 0.071) and the lowest βSpacing (M = 0.29; SD = 0.03) (see Fig. 3). The Weber fraction of this profile ranged from 0.15 to 0.94 (M = 0.50; SD = 0.17; N = 119).

Profile two: Low Inhibitory Control. The second profile had the fewest children (N = 40). Children in this profile showed high performance on congruent trials, but a very poor performance on incongruent trials (see Fig. 2). These children showed a strong congruency effect. The regression coefficients also showed a high βNumber (M = 1.72; SD = 0.12), but they had a significantly higher βSpacing (M = 0.87; SD = 0.05) compared to the other two profiles (see Fig. 3). This group of children appeared to rely significantly more on the non-numerical cues (i.e., how spread the dots are) to solve the comparison task. This reliance impaired their performance on incongruent trials. However, on congruent trials with the two smallest ratios, children in this profile showed significantly higher accuracy compared to children in profile one or high performers (see Fig. 2). The Weber fraction of this profile ranged from 0.07 to 1.34 (M = 0.56; SD = 0.31; N = 39), and was not significantly different from the High Performers profile.

Profile three: Low Performers. 97 children were classified in this profile. Children in this profile showed low overall performance on both congruent and incongruent trials, with accuracies close to chance levels in all of the eight blocks. Children in this group also showed the lowest βNumber (M = 0.45; SD = 0.08) and βspacing (M = 0.11; SD = 0.03). In this profile there was a significant variability in the Weber fraction ranging from 0.52 to 2.47 (M = 1.26; SD = 0.43; N = 67).

Overall there was a statistical difference between the total accuracy of the three profiles (F (2, 253) = 206.28, p < .001, η^2 = 0.62). In addition, there was a significant interaction effect between the performance profiles and the regression coefficients of *Number* (F (2, 87.35) = 139.99, p < .001, η^2 = 0.36) and *Spacing* (F (2, 89.60) = 42.17, p < .001, η^2 = 0.38). Post hoc pairwise comparison of the regression coefficients of *Number* revealed that there was no significant difference between profiles one (i.e., High Performers) and two (i.e., Low Inhibitory Control), meaning that both rely on this perceptual cue equally (see Fig. 3). However, the two profiles clearly differ in the regression coefficient of *Spacing*, with profile two (i.e., Low Inhibitory Control) showing the highest value (see Fig. 3).

Additionally, a one-way ANOVA was calculated to examine differences in age across the three profiles. There was a small but significant effect of age (F (2, 253) = 13.08; p < .001; η^2 = 0.09), older children appeared to be more represented in profile one (M = 5.70; SD = 0.64) than in profiles two and three. However, no significant differences in age were observed between profiles two (M = 5.40; SD = 0.63) and three (M = 5.49; SD = 0.65). Chi-squared test of independence showed no association between the profile of the child and gender (χ^2 (2) = 4.65; p = .098) or school class (χ^2 (38) = 4.65; p = .15).

4. Discussion

Previous research has shown significant individual variability in performance on numerosity comparison tasks. Some studies attribute this heterogeneity primarily to differences in numerical precision (Dolfi et al., 2024; Starr et al., 2017), whereas others emphasise differences in inhibitory control (Bugden & Ansari, 2016; Rodríguez & Ferreira, 2023; Wilkey et al., 2020). Rather than viewing these findings as contradictory, these perspectives may *both* be valid. The current study proposes that discrepancies in the literature may be due to differences in sample composition, specifically, which cognitive profiles are most prevalent within a given group. The current study hypothesises that children do not form a homogenous group but instead exhibit distinct cognitive profiles: some may show reduced numerical precision, while other may primarily experience difficulties with inhibitory control. As a result, studies with different proportions of these profiles may reach different conclusions.

Following the theoretical accounts of Doebel (2020) and Wilkey (2023), the present study adopts the view that EF are higher-order processes that guide cognition in the service of a particular task goal, rather than as fully separable components applied uniformly across contexts. From this perspective, the inhibition required in numerosity comparison is not the same to the inhibition involved in other domains,

Table 3
Model fit indices of the LPA analyses for the 9 testes models.

N classes	AIC	BIC	Entropy	Min. probability	Max. probability	N min	N max	BLRT
2	−1706.11	−1617.48	0.75	0.92	0.94	0.48	0.52	0.01
3	**−2038.6**	**−1918.07**	**0.86**	**0.91**	**0.96**	**0.16**	**0.46**	**0.01**
4	−2129.01	−1976.56	0.86	0.88	0.95	0.05	0.37	0.01
5	−2168.81	−1984.46	0.83	0.84	0.94	0.05	0.31	0.01
6	−2250.35	−2034.09	0.84	0.86	0.92	0.04	0.28	0.01
7	−2275.91	−2027.75	0.85	0.86	0.96	0.04	0.26	0.01
8	−2298.46	−2018.39	0.86	0.83	0.99	0.04	0.25	0.01
9	−2302.06	−1990.09	0.85	0.79	1	0.04	0.22	0.1

Note. The chosen model is in **bold**.

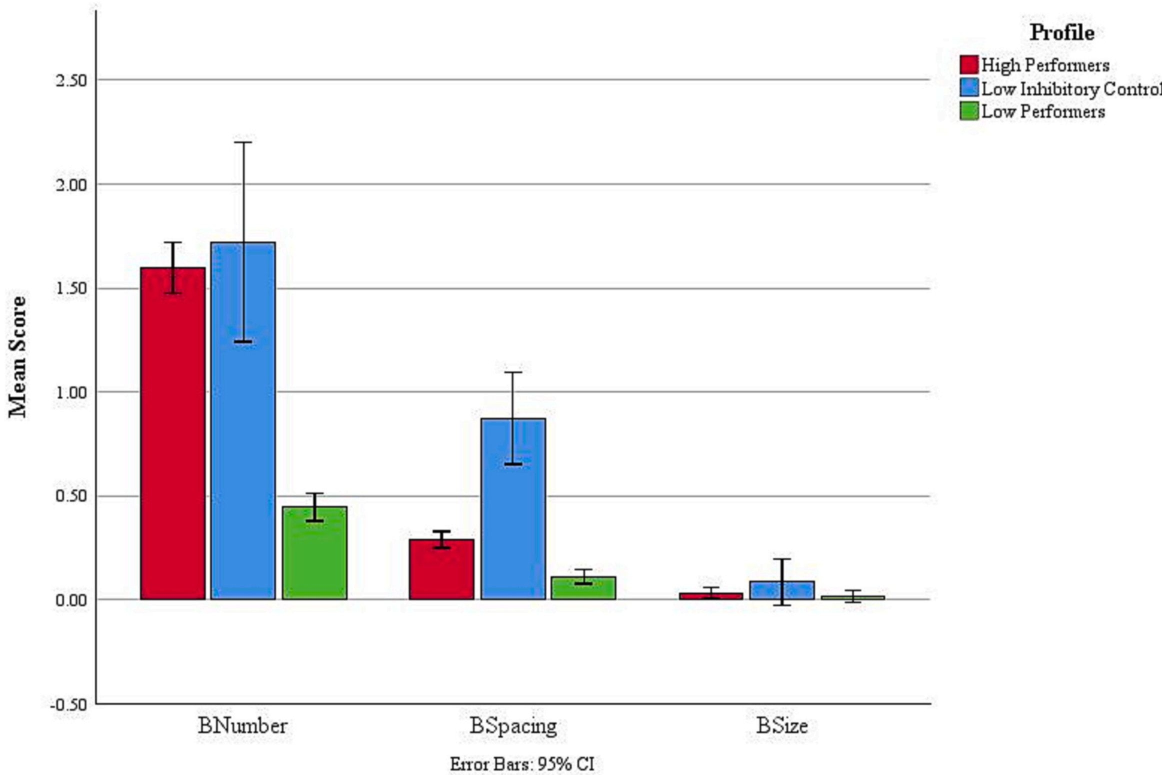

Fig. 3. Regression coefficients of the three profiles.
Note. The "Mean Score" on the y-axis reflects the mean score on the numerosity comparison task. BNumber is the regression coefficient of Number BSpacing regression coefficient of Spacing and BSize is the regression coefficient of Size.

such as theory of mind. Instead, domain general and domain specific processes interact dynamically. In this view, inhibitory control is recruited in the service of task-relevant goals, in this case, selecting numerosity while suppressing competing non-numerical cues such as dot spacing. This interpretation aligns with contemporary views proposing that EF manifest through domain-specific behavioral mechanisms depending on the context in which they are recruited. Examining the recruitment of EF within a numerical task provides a more ecologically valid understanding of their contribution to early mathematical abilities than relying on EF recruited in non-numerical tasks.

The present study aimed to capture the heterogeneity in children's early numerosity processing by identifying distinct performance profiles among preschoolers, based on their performance in a numerosity comparison task. To achieve this objective, the study also explores the factor structure of the numerosity comparison task. The objective was to confirm that a multidimensional structure based on ratio and congruency characteristics describes the data better than a unidimensional model, or a model that only considers numerical ratio. The findings support the view that numerosity comparison task measures not only number sense, but also participants' ability to inhibit irrelevant information (e.g., Leibovich et al., 2017; Leibovich & Henik, 2013).

4.1. Individual differences in the performance of a numerosity comparison task

In general, children in the current study showed a preference for numerical information, which allowed them to accurately perform the task. This finding aligns with previous studies showing that children between the ages of five and nine are increasingly sensitive to numerical information (Starr et al., 2017; Yan et al., 2024). In addition, the present study also found a clear bias towards the *Spacing* parameter, a non-numerical feature indicating that more spread-out groups of dots are perceived as more numerous (DeWind et al., 2015). This aligns with the results of a recent developmental study comparing preschoolers, older children, and young adults. In this study all age groups showed a bias specifically towards the *Spacing* parameter, suggesting that non-numerical factors also play a significant role in children's perception and task performance (Oeo-Morín et al., 2025).

The current study focuses on the individual heterogeneity observed during a numerosity comparison task. The analysis revealed three distinct performance profiles. The first profile, called "High Performers", included children who demonstrated a stronger ability to inhibit distracting cues, such as how spread the dots appeared. This ability allowed them to perform well in both congruent and incongruent trials (Rodríguez & Ferreira, 2023). As expected, their performance improved as the numerical ratio increased, and they achieved higher accuracy in congruent trials. The high numerical precision of children in this profile, combined with their ability to ignore irrelevant information, contributed to their overall success in the task. Interestingly, children in this profile were slightly older on average, suggesting that age-related maturation may play a role in the development of both numerical precision and inhibitory control.

The second profile, named "Low Inhibitory Control", describes children who are not able to inhibit or suppress the irrelevant non-numerical information, specifically the *Spacing* parameter. These children still showed higher values on the *Number* predictor, similar to those of the "High Performers", indicating that numerical information remained the primary decision cue. In line with this, the coefficient for numerical information (βNumber) was larger than the coefficient for spatial information (βSpacing), demonstrating that these children did attend to numerical magnitude but experienced difficulty suppressing spatial cues. If children in this group had relied on an incorrect, space-based conception of numerosity, the opposite pattern would have been expected, with βSpacing emerging as the stronger predictor.

Thus, high attention to numerical information, paired with a reliance on non-numerical cues led to good performance on congruent trials but very poor performance on incongruent ones. Interestingly, despite no significant difference in the attention towards *Number* between "High Performers" and "Low Inhibitory Control", children in this last profile showed better accuracy in small and medium ratio congruent trials. This suggests that non-numerical cues, such as the spacing of dots (i.e., how spread the dots are presented) may boost performance in more difficult conditions by acting as facilitators.

From a theoretical perspective, this pattern is consistent with accounts that propose both numerical and non-numerical magnitudes influence performance in comparison tasks (see Lourenco & Aulet, 2023 for a review). Some frameworks propose that numerosity judgments arise from the combination of non-numerical visual features (Gebuis & Reynvoet, 2012), while others assume that numerical and non-numerical information is processed separately but interacts, requiring the inhibition of irrelevant cues to resolve the conflict (Leibovich & Henik, 2013; Piazza et al., 2018). More recent perspectives propose perceptual interdependence, in which numerical and non-numerical magnitudes are processed holistically from early stages of visual processing (Lourenco & Aulet, 2023). However, the present findings do not allow one to determine whether the observed interference arises during early visual processing (Lourenco & Aulet, 2023), or at a later decisional stage requiring active inhibition of irrelevant information (Leibovich et al., 2017; Leibovich & Henik, 2013; Piazza et al., 2018).

Finally, the study observed a third performance profile "Low Performers", which includes children who exhibited low general accuracy, performing close to chance levels. This could be due to random answering because they found the task difficult, either because their ability is still developing or due to a lack of motivation. It is important to note that this low accuracy is not due to a lack of understanding, as comprehension was ensured through practice trials with immediate trial-by-trial feedback (see Section 2.3.1).

Taking into account the results of the factor analysis and the performance profiles that emerged, the current study support the idea that both inhibitory control and numerical precision are crucial in completing numerosity comparison tasks (Fuhs et al., 2018; Leibovich et al., 2017; Lourenco & Aulet, 2023). This findings challenge the theories suggesting that numerosity perception relies only on the acuity of the number sense of subjects and that non-numerical magnitude visual features do not contribute to the perception of numerosity (Dehaene, 2001).

Furthermore, the present study supports the view that children may struggle with numerosity comparison tasks for different reasons. Some children show difficulties due to weak numerical acuity (i.e., "Low achievers"), while others perform poorly specifically on incongruent trials because they have difficulty inhibiting irrelevant visual cues (i.e., "Low Inhibitory Control"). By applying a person-centred analytic approach, the study provides empirical evidence that preschool children cluster into *distinct* performance profiles that differ in the extent to which numerical precision and inhibitory control shape their behaviour. This profile-based structure shows that individual variability in numerosity performance is not random noise, but reflects meaningful cognitive pathways. This highlights the importance of considering heterogeneity when interpreting performance on early numeracy tasks.

4.2. Limitations

Some limitations should be acknowledged. The results showed a surprisingly large proportion of children in the "Low Performers" group ($N = 97$). However, this is unlikely to reflect a lack of understanding of the task. A more plausible explanation could be that some of this children could show motivational issues, some children may have become bored or distracted, leading them to respond randomly rather than engage with the task. This group showed indeed a lower split-half reliability (odd–even split; Spearman–Brown = 0.52) compared to the reliability when the whole sample is taken into account (0.84). Although accuracy in numerosity comparison tasks is generally reliable and valid (Dietrich et al., 2015; Oeo-Morín et al., 2025), lapses in attention can still affect performance and result in guessing (Krajcsi et al., 2024). That said, the split half is still significantly different from zero, the total number of trials was relatively low, and the stimuli stayed on the screen until a response was given, allowing participants to respond despite momentary distractions.

In addition, some model fit indices (CFI and TLI) did not reach conventional thresholds. There may be several reasons for this. Although a sample size of 256 is generally sufficient, a larger sample might have been required given the complexity of the model (Fujimoto & Neugebauer, 2020; Hu & Bentler, 1999). Furthermore, simulation studies have shown that complex structural models are expected to have lower CFI values even when the model is correctly specified (Kenny & McCoach, 2003; Marsh et al., 2004). Another possible explanation is that children may have performed randomly on incongruent trials with small ratios (i.e., 1.17 and 1.25), due to a higher difficulty level. A confirmatory MRIT analysis considering only the 1.5 and 2 ratios did improve the fit indices (see Table 3 in the Supplementary Material). Other indices such as RMSEA, SRMR and the C2 were within acceptable ranges in the present study, offering stronger support for overall model fit. In line with methodological recommendations, RMSEA measures absolute fit and may be more informative in confirmatory factor analysis, whereas CFI is an incremental index that may be more appropriate for exploratory contexts (Kenny, 2015; Rigdon, 1996).

4.3. Implications and conclusions

The identification of performance profiles provides valuable insights into the origins of individual differences in numerosity comparison tasks. These profiles offer a more nuanced understanding of how children process numerical and non-numerical information, moving beyond the assumption of a homogeneous population. As suggested by Hart et al. (2016), understanding the mechanisms underlying success and failure in numerosity comparison tasks can offer important insights into the origins of learning differences in numerosity perception. These insights can inform the development of more efficient learning environments tailored to the *specific* needs of each individual. For instance, children who may rely heavily on non-numerical cues, could benefit from learning environments designed to reduce this dependence and promote numerical processing strategies. In contrast, children with low general performance may need more foundational support aimed at strengthening their core numerical understanding. Future research should investigate how these profiles evolve with age, and how children in different profiles respond to different educational instruction. Moreover, it remains an open question whether similar performance profiles would emerge in older populations or in clinical groups, such as children with developmental disorders. Exploring these questions could clarify whether the profiles observed in the present study reflect general patterns of numerical cognition or are specific to early developmental stages.

This study highlights individual heterogeneity in numerosity perception performance. By identifying distinct performance profiles, we can better understand the cognitive processes involved in numerical perception and their contribution to individual variability.

CRediT authorship contribution statement

Irene Oeo Morín: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Flore Maricau:** Writing – review & editing, Methodology. **Fien Depaepe:** Writing – review & editing, Supervision. **Bert Reynvoet:** Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used *Microsoft M365 Copilot* to assist with text editing and language refinement to improve the clarity and readability of the text. After using this tool, the author(s) reviewed and revised the content as needed and take full responsibility for the final version of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This research project was funded by a KU Leuven C1-grant (C14/23/057). The ethical committee at KU Leuven (G-2024-7575) reviewed and approved the procedures and methods of the study. We thank all school staff members, parents, and children for their participation in this study. We have no conflicts of interest to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2026.107138.

Data availability

Data on the stimuli parameters, the numerosity task, and the classification codes are available on OSF: https://doi.org/10.17605/OSF.IO/6NA2J. Participant data will be available on request.

References

Abalo-Rodríguez, I., De Marco, D., & Cutini, S. (2022). An undeniable interplay: Both numerosity and visual features affect estimation of non-symbolic stimuli. *Cognition, 222*, 104944. https://doi.org/10.1016/j.cognition.2021.104944

Anders, Y., Grosse, C., Rossbach, H. G., Ebert, S., & Weinert, S. (2013). Preschool and primary school influences on the development of children's early numeracy skills between the ages of 3 and 7 years in Germany. *School Effectiveness and School Improvement, 24*(2), 195–211. https://doi.org/10.1080/09243453.2012.749794

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

Brown, A., & Maydeu-Olivares, A. (2011). Item response modeling of forced-choice questionnaires. *Educational and Psychological Measurement, 71*(3), 460–502. https://doi.org/10.1177/0013164410375112

Bugden, S., & Ansari, D. (2016). Probing the nature of deficits in the "Approximate Number System" in children with persistent developmental dyscalculia. *Developmental Science, 19*(5), 817–833. https://doi.org/10.1111/desc.12324

Cai, L., & Monroe, S. (2014). *A new statistic for evaluating item response theory models for ordinal data (CRESST Report 839)*. Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Cappelletti, M., Didino, D., Stoianov, I., & Zorzi, M. (2014). Number skills are maintained in healthy ageing. *Cognitive Psychology, 69*, 25–45. https://doi.org/10.1016/j.cogpsych.2013.11.004

Chalmers, R. P. (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Clark, S. L., & Muthén, B. (2009). *Relating latent class analysis results to variables not included in the analysis* (pp. 1–55).

Cowan, R., & Powell, D. (2014). The contributions of domain-general and numerical factors to third-grade arithmetic skills and mathematical learning disability. *Journal of Educational Psychology, 106*(1), 214–229. https://doi.org/10.1037/a0034097

De Marco, D., & Cutini, S. (2020). Introducing CUSTOM: A customized, ultraprecise, standardization-oriented, multipurpose algorithm for generating nonsymbolic number stimuli. *Behavior Research Methods, 52*(4), 1528–1537. https://doi.org/10.3758/s13428-019-01332-z

Defever, E., Reynvoet, B., & Gebuis, T. (2013). Task- and age-dependent effects of visual stimulus properties on children's explicit numerosity judgments. *Journal of Experimental Child Psychology, 116*(2), 216–233. https://doi.org/10.1016/j.jecp.2013.04.006

Dehaene, S. (2001). Précis of the number sense. *Mind & Language, 16*(1), 16–36.

DeWind, N. K., Adams, G. K., Platt, M. L., & Brannon, E. M. (2015). Modeling the approximate number system to quantify the contribution of visual stimulus features. *Cognition, 142*, 247–265. https://doi.org/10.1016/j.cognition.2015.05.016

DeWind, N. K., & Brannon, E. M. (2016). Significant inter-test reliability across approximate number system assessments. *Frontiers in Psychology, 7*, 310. https://doi.org/10.3389/fpsyg.2016.00310

Dietrich, J. F., Huber, S., & Nuerk, H. C. (2015). Methodological aspects to be considered when measuring the approximate number system (ANS)—A research review. *Frontiers in Psychology, 6*, 295. https://doi.org/10.3389/fpsyg.2015.00295

Doebel, S. (2020). Rethinking executive function and its development. *Perspectives on Psychological Science, 15*(4), 942–956.

Dolfi, S., Decarli, G., Lunardon, M., De Filippo De Grazia, M., Gerola, S., Lanfranchi, S., … Zorzi, M. (2024). Weaker number sense accounts for impaired numerosity perception in dyscalculia: Behavioral and computational evidence. *Developmental Science, 27*(6), Article e13538. https://doi.org/10.1111/desc.13538

Ernst, J. R., Mazzocco, M. M., & Carlson, S. M. (2025). Concurrent and predictive associations between executive function and numerical skills in early childhood. *Journal of Experimental Child Psychology, 250*, Article 106113. https://doi.org/10.1016/j.jecp.2024.106113

Fuhs, M. W., & McNeil, N. M. (2013). ANS acuity and mathematics ability in preschoolers from low-income homes: Contributions of inhibitory control. *Developmental Science, 16*(1), 136–148. https://doi.org/10.1111/desc.12013

Fuhs, M. W., Nesbitt, K. T., & Connor, D. O. (2018). Approximate number system task performance: Associations with domain-general and domain-specific cognitive skills in young children. *Journal of Numerical Cognition, 4*(3), 590–612. https://doi.org/10.5964/jnc.v4i3.141

Fujimoto, K. A., & Neugebauer, S. R. (2020). A general Bayesian multidimensional item response theory model for small and large samples. *Educational and Psychological Measurement, 80*(4), 665–694. https://doi.org/10.1177/0013164419891205

Gebuis, T., & Reynvoet, B. (2011). Generating nonsymbolic number stimuli. *Behavior Research Methods, 43*, 981–986. https://doi.org/10.3758/s13428-011-0097-5

Gebuis, T., & Reynvoet, B. (2012). The interplay between nonsymbolic number and its continuous visual properties. *Journal of Experimental Psychology: General, 141*(4), 642–648. https://doi.org/10.1037/a0026218

Gebuis, T., & Reynvoet, B. (2012b). The role of visual information in numerosity estimation. *PLoS One, 7*(5), e37426. https://doi.org/10.1371/journal.pone.0037426

Gilmore, C. (2023). Understanding the complexities of mathematical cognition: A multi-level framework. *Quarterly Journal of Experimental Psychology, 76*(9), 1953–1972. https://doi.org/10.1177/17470218231175325

Gilmore, C., Attridge, N., Clayton, S., Cragg, L., Johnson, S., Marlow, N., … Inglis, M. (2013). Individual differences in inhibitory control, not non-verbal number acuity, correlate with mathematics achievement. *PLoS One, 8*(6), Article e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C., Göbel, S. M., & Inglis, M. (2018). Nonsymbolic number. In *An introduction to mathematical cognition* (pp. 6–28). Routledge. https://doi.org/10.4324/9781315684758.

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General, 121*(4), 480–506. https://doi.org/10.1037/0096-3445.121.4.480

Guillaume, M., & Van Rinsveld, A. (2018). Comparing numerical comparison tasks: A meta-analysis of the variability of the Weber Fraction relative to the generation algorithm. *Frontiers in Psychology, 9*, 1694. https://doi.org/10.3389/fpsyg.2018.01694

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The Approximate Number System in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology, 44*(5), 1457–1465. https://doi.org/10.1037/a0012682

Halberda, J., Ly, R., Wilmer, J. B., Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive internet-based sample. *Proceedings of the National Academy of Sciences of the United States of America, 109*(28), 11116–11120. https://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature, 455*(7213), 665–668. https://doi.org/10.1038/nature07246

Hart, S. A., Logan, J. A., Thompson, L., Kovas, Y., McLoughlin, G., & Petrill, S. A. (2016). A latent profile analysis of math achievement, numerosity, and math anxiety in twins. *Journal of Educational Psychology, 108*(2), 181. https://doi.org/10.1037/edu0000045

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Business Research Methods, 6*, 53–60.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Keller, L., & Libertus, M. (2015). Inhibitory control may not explain the link between approximation and math abilities in kindergarteners from middle class families. *Frontiers in Psychology, 6*, 685. https://doi.org/10.3389/fpsyg.2015.00685

Kenny, D. A. (2015). Measuring model fit. http://davidakenny.net/cm/fit.htm.

Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*(3), 333–351. https://doi.org/10.1207/S15328007SEM1003_1

Krajcsi, A., Chesney, D., Cipora, K., Coolen, I., Gilmore, C., Inglis, M., Libertus, M., Nuerk, H.-C., Simms, V., & Reynvoet, B. (2024). Measuring the acuity of the approximate number system in young children. *Developmental Review, 72*, Article 101131. https://doi.org/10.1016/j.dr.2024.101131

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

Leibovich, T., & Ansari, D. (2017). Accumulation of non-numerical evidence during nonsymbolic number processing in the brain: An fMRI study. *Human Brain Mapping, 38*(10), 4908–4921. https://doi.org/10.1002/hbm.23703

Leibovich, T., & Henik, A. (2013). Magnitude processing in non-symbolic stimuli. *Frontiers in Psychology, 4*, 375. https://doi.org/10.3389/fpsyg.2013.00375

Leibovich, T., Katzin, N., Harel, M., & Henik, A. (2017). From "sense of number" to "sense of magnitude": The role of continuous magnitudes in numerical cognition. *Behavioral and Brain Sciences, 40*, e164. https://doi.org/10.1017/S0140525X16000960

Lenth, R. (2025). emmeans: Estimated marginal means, aka least-squares means. R package version 1.11.1-00001. https://rvlenth.github.io/emmeans/.

Lourenco, S. F., & Aulet, L. S. (2023). A theory of perceptual number encoding. *Psychological Review, 130*(1), 155–182. https://doi.org/10.1037/rev0000380

Lubke, G., & Neale, M. C. (2006). Distinguishing between latent classes and continuous factors: Resolution by maximum likelihood? *Multivariate Behavioral Research, 41*(4), 499–532. https://doi.org/10.1207/s15327906mbr4104_4

Magidson, J., & Vermunt, J. K. (2002). A nontechnical introduction to latent class models. In *, Vol. 1. Statistical innovations white paper* (p. 15).

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11* (3), 320–341.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Impaired acuity of the approximate number system underlies mathematical learning disability (dyscalculia). *Child Development, 82*(4), 1224–1237. https://doi.org/10.1111/j.1467-8624.2011.01608.x

Merkley, R., Thompson, J., & Scerif, G. (2016). Of huge mice and tiny elephants: Exploring the relationship between inhibitory processes and preschool math skills. *Frontiers in Psychology, 6*, 1903. https://doi.org/10.3389/fpsyg.2015.01903

Negen, J., & Sarnecka, B. W. (2015). Is there really a link between exact-number knowledge and approximate number system acuity in young children? *British Journal of Developmental Psychology, 33*(1), 92–105. https://doi.org/10.1111/bjdp.12071

Oeo-Morín, I., Depaepe, F., & Reynvoet, B. (2025). Sharpening the number sense: Developmental trends in numerosity perception. *Journal of Experimental Child Psychology, 256*, 106262. https://doi.org/10.1016/j.jecp.2025.106262

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., … Lindeløv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*, 195–203. https://doi.org/10.3758/s13428-018-01193-y

Piazza, M., De Feo, V., Panzeri, S., & Dehaene, S. (2018). Learning to focus on number. *Cognition, 181*, 35–45. https://doi.org/10.1016/j.cognition.2018.07.011

Piazza, M., & Izard, V. (2009). How humans count: Numerosity and the parietal cortex. *The Neuroscientist, 15*(3), 261–273. https://doi.org/10.1177/1073858409333073

Reynvoet, B., Ribner, A. D., Elliott, L., Van Steenkiste, M., Sasanguie, D., & Libertus, M. E. (2021). Making sense of the relation between number sense and math. *Journal of Numerical Cognition, 7*(3), 308–327. https://doi.org/10.5964/jnc.6059

Rigdon, E. E. (1996). CFI versus RMSEA: A comparison of two fit indexes for structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 3*(4), 369–379. https://doi.org/10.1080/10705519609540052

Rodríguez, C., & Ferreira, R. A. (2023). To what extent is dot comparison an appropriate measure of approximate number system? *Frontiers in Psychology, 13*, Article 1065600. https://doi.org/10.3389/fpsyg.2022.1065600

Roquet, A., Poletti, C., & Lemaire, P. (2020). Sequential modulations of executive control processes throughout lifespan in numerosity comparison. *Cognitive Development, 54*, Article 100884. https://doi.org/10.1016/j.cogdev.2020.100884

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Van Lissa, C. J., & Schmidt, J. A. (2018). tidyLPA: An R package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *The Journal of Open Source Software, 3*(30), 978. https://doi.org/10.21105/joss.00978

Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2020). The factor structure of the verbal Torrance test of creative thinking in an Arabic context: Classical test theory and multidimensional item response theory analyses. *Thinking Skills and Creativity, 35*, 100609. https://doi.org/10.1016/j.tsc.2019.100609

Schneider, M., Beeres, K., Coban, L., Merz, S., Schmidt, S., Stricker, J., & De Smedt, B. (2017). Associations of non-symbolic and symbolic numerical magnitude processing with mathematical competence: A meta-analysis. *Developmental Science, 20*(3), e12372. https://doi.org/10.1111/desc.12372

Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal, 8*(1), 205–233. https://journal.r-project.org/archive/2016/RJ-2016-021/RJ-2016-021.pdf.

Smets, K., Sasanguie, D., Szücs, D., & Reynvoet, B. (2015). The effect of different methods to construct non-symbolic stimuli in numerosity estimation and comparison. *Journal of Cognitive Psychology, 27*(3), 310–325. https://doi.org/10.1080/20445911.2014.996568

Soltész, F., Szűcs, D., & Szűcs, L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions, 6*, 1–14. http://www.behavioralandbrainfunctions.com/content/6/1/13.

Spurk, D., Hirschi, A., Wang, M., Valero, D., & Kauffeld, S. (2020). Latent profile analysis: A review and "how to" guide of its application within vocational behavior research. *Journal of Vocational Behavior, 120*, Article 103445. https://doi.org/10.1016/j.jvb.2020.103445

Starr, A., DeWind, N. K., & Brannon, E. M. (2017). The contributions of numerical acuity and non-numerical stimulus features to the development of the number sense and symbolic math achievement. *Cognition, 168*, 222–233. https://doi.org/10.1016/j.cognition.2017.07.004

Szücs, D., Nobes, A., Devine, A., Gabriel, F. C., & Gebuis, T. (2013). Visual stimulus parameters seriously compromise the measurement of approximate number system acuity and comparative effects between adults and children. *Frontiers in Psychology, 4*, 444. https://doi.org/10.3389/fpsyg.2013.00444

Viarouge, A., Houdé, O., & Borst, G. (2019). Evidence for the role of inhibition in numerical comparison: A negative priming study in 7- to 8-year-olds and adults. *Journal of Experimental Child Psychology, 186*, 131–141. https://doi.org/10.1016/j.jecp.2019.05.011

Viarouge, A., Lee, H., & Borst, G. (2023). Attention to number requires magnitude-specific inhibition. *Cognition, 230*, 105285. https://doi.org/10.1016/j.cognition.2022.105285

Wilkey, E. D. (2023). The domain-specificity of domain-generality: Attention, executive function, and academic skills. *Mind, Brain, and Education, 17*(4), 349–361.

Wilkey, E. D., Pollack, C., & Price, G. R. (2020). Dyscalculia and typical math achievement are associated with individual differences in number-specific executive function. *Child Development, 91*(2), 596–619. https://doi.org/10.1111/cdev.13194

Wilkey, E. D., & Price, G. R. (2019). Attention to number: The convergence of numerical magnitude processing, attention, and mathematics in the inferior frontal gyrus. *Human Brain Mapping, 40*(3), 928–943. https://doi.org/10.1111/CDEV.13194

Wilkey, E. D., Shanley, L., Sabb, F., Ansari, D., Cohen, J. C., Men, V., … Clarke, B. (2022). Sharpening, focusing, and developing: A study of change in nonsymbolic number comparison skills and math achievement in 1st grade. *Developmental Science, 25*(3), e13194. https://doi.org/10.1111/desc.13194

Yan, L., Qian, P., & Yan, R. (2024). Developmental changes in numerosity and area perception in school-age children. *Acta Psychologica, 249*, 104466. https://doi.org/10.1016/j.actpsy.2024.104466