



BIROn - Birkbeck Institutional Research Online

Gilligan, K. and Hodgkiss, A. and Thomas, Michael S.C. and Farran, E. (2018) The use of discrimination scaling tasks: a novel perspective on the development of spatial scaling in children. *Cognitive Development* 47 , pp. 133-145. ISSN 0885-2014.

Downloaded from: <https://eprints.bbk.ac.uk/id/eprint/23120/>

Usage Guidelines:

Please refer to usage guidelines at <https://eprints.bbk.ac.uk/policies.html>
contact lib-eprints@bbk.ac.uk.

or alternatively

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

Cognitive Development

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

The use of discrimination scaling tasks: A novel perspective on the development of spatial scaling in children

Katie A. Gilligan^{a,c,*}, Alex Hodgkiss^{a,c,*}, Michael S.C. Thomas^{b,c}, Emily K. Farran^{a,c}

^a Department of Psychology and Human Development, UCL Institute of Education, University College London, 25 Woburn Square, London, WC1H 0AA, UK

^b Department of Psychological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK

^c Centre for Educational Neuroscience, University of London, UK

ARTICLE INFO

Keywords:

Spatial Scaling
Development
Discrimination scaling task
Relative scaling strategies

ABSTRACT

Spatial scaling is the ability to transform distance information between shapes of differing sizes. Research on the developmental trajectories of spatial scaling beyond the pre-school years has been limited by a lack of suitable scaling measures for older children. Here we developed an age-appropriate discrimination scaling task, and demonstrated that children (N = 386) achieve performance gains in spatial scaling skills between 5 and 8-years-of-age, after which no significant improvements were found. Furthermore, the results support the use of relative distance strategies for task completion. These findings contrast to localisation paradigms, where performance reaches a plateau by age 6 and mental transformation strategies are used for scaling. The finding that scaling skills continue to develop until 8 years highlight the potential of scaling interventions in the early primary school years. Such interventions may infer direct benefits on spatial thinking and indirect advantages for science, technology, engineering and maths (STEM) achievement.

1. Introduction

1.1. Overview

To successfully navigate, individuals must be capable of representing their own location with reference to their external environment. Hence, navigation within an environment is inherently dependent on spatial thinking. Furthermore, the use of common navigation aids such as maps or GPS systems to assist navigation requires effective spatial scaling, a particular sub-domain of spatial cognition. Spatial scaling is the ability to transform distance information from one representation to another representation of a different size (Frick & Newcombe, 2012). Scaling requires comprehension of both the symbolic and spatial correspondence between a map (model), and an associated referent space, in addition to the ability to mentally manipulate and transform spatial information between spaces of different sizes. Beyond navigation, spatial scaling is also associated with success in aspects of mathematics. For example, spatial scaling skills explain a significant proportion of the variation in proportional reasoning, above that explained by verbal intelligence (Möhring, Newcombe, & Frick, 2015). Across other domains of spatial thinking, significant age based differences

* Corresponding authors at: Department of Psychology and Human Development, UCL Institute of Education, University College London, 25 Woburn Square, London, WC1H 0AA, UK.

E-mail addresses: katie.gilligan.15@ucl.ac.uk (K.A. Gilligan), alex.hodgkiss.14@ucl.ac.uk (A. Hodgkiss), m.thomas@bbk.ac.uk (M.S.C. Thomas), e.farran@ucl.ac.uk (E.K. Farran).

¹ Joint first authorship.

<https://doi.org/10.1016/j.cogdev.2018.04.001>

Received 3 August 2017; Received in revised form 27 February 2018; Accepted 10 April 2018

0885-2014/ © 2018 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

in performance have been reported for children up to 8 years of age. For example, children's performance on tasks such as perspective taking, mental folding and mental rotation improve until at least 8 years with significant individual variation reported at all ages (for example see, Frick, Möhring, & Newcombe, 2014; Newcombe, Uttal, & Suater, 2013). However, no suitable measure is available to explore the development of spatial scaling throughout the primary school years. Hence, this study aims to develop a spatial scaling measure suitable for the investigation of both developmental differences and individual differences in scaling performance in children aged 5–10 years.

1.2. The development of spatial scaling

Traditional Piagetian theories suggest that the skills required for spatial scaling are not evident until 10–11-years-of-age when children no longer hold topological views of spatial concepts (Piaget and Inhelder, 1948; Piaget & Inhelder, 1948, 1956, 1967). In contrast, more recent evidence suggests that some prerequisite skills required for spatial scaling including symbolic correspondence and metric encoding emerge in early childhood (e.g. Huttenlocher, Newcombe, & Sandberg, 1994; Newcombe, Sluzenski, & Huttenlocher, 2005; Vasilyeva and Huttenlocher, 2004). First, comprehension of *symbolic correspondence*, or the correspondence between a model and a referent space has been reported in children as young as 3-years-old (DeLoache, 1987, 1989). At this age children recognise that features on a map or model represent features in the real world. Second, *metric encoding*, or the ability to encode distances metrically, has been reported in infants as young as 5-months-old, with some infants demonstrating sensitivity to distance differences of just 20 cm (Newcombe, Huttenlocher, & Learmonth, 1999; Newcombe et al., 2005). Similarly, Bushnell, McKenzie, Lawrence, and Connell (1995) reported that 12-month-old infants can locate an object which is hidden in a circular enclosure under one of many randomly placed identical cushions. Given the lack of cues or landmarks and the random arrangement of the cushions, this suggests an ability to use metric encoding relative to the participant, in order to identify the correct cushion. Similar findings from Huttenlocher et al. (1994) propose that metric encoding in children is robust by 16 months of age.

Beyond these prerequisite skills, there is also evidence that the ability to successfully map encoded distances between different sized spaces, spatial scaling, develops at a younger age than suggested by Piaget and Inhelder, with significant development in scaling proficiency between the ages of 3 and 6 years (Piaget and Inhelder, 1948; Piaget & Inhelder, 1948, 1956, 1967). Frick and Newcombe (2012) reported that children's scaling ability, measured using a two-dimensional localisation task, improves with age from 3 to 6-years-of-age, at which time children's accuracy levels are broadly comparable to adult scores. No significant difference in performance between 5- and 6-year-old children was reported. In a similar computer-based study, Möhring, Newcombe, and Frick (2014) demonstrated improvements in spatial scaling across different scaling factors between the ages of 4–5 years. Similar results have also been reported in studies using more naturalistic environments. For example, Vasilyeva and Huttenlocher (2004) reported that 90% of 5-year-old children tested could successfully place objects on a rectangular rug using a two-dimensional map. In comparison, only 60% of 4-year-olds were successful when presented with the same task.

While some studies have reported accurate spatial scaling in children younger than 5 years, these findings may be attributable to the use of simplified tasks in which targets are presented along a single dimension. For example, Huttenlocher, Newcombe, and Vasilyeva (1999) reported accurate spatial scaling for most 3 and 4-year-old participants when tested using a scaling paradigm with a single dimension, the horizontal axis. Similarly, four-year-old children can successfully use a one-dimensional map to locate one of three target bins in a rectangular room (Shusterman, Ah Lee, & Spelke, 2008). In contrast, in more complex studies requiring scaling in two dimensions, 4 year-olds typically struggle. For example, 4-year-old children struggle to correctly place an object in a target location within a room, based on locations learnt from a corresponding map (Uttal, 1996).

To conclude, in contrast to findings from Piaget and Inhelder, 1948; Piaget and Inhelder (1948, 1956, 1967), it appears that children as young as three years demonstrate symbolic correspondence and metric coding, prerequisite skills for spatial scaling. Successful performance on one and two-dimensional scaling tasks is typically evident from three years and five years respectively. Importantly, although developmental differences are evident in spatial scaling abilities, individual differences in performance are also reported for scaling performance in children up to 10-years-old (e.g. Liben and Downs, 1993).

1.3. Cognitive scaling strategies

Comprehending the strategies used for spatial scaling, and identifying how these strategies may change and adapt throughout development, may explain how and why spatial scaling has been associated with other aspects of cognition, including mathematics. Proposed cognitive strategies for spatial scaling include absolute, relative and mental transformation strategies. Absolute strategies as described by Möhring, Newcombe, and Frick (2016) involve absolute encoding of distance in one space with direct mapping of this distance onto a second space. While effective for similarly sized spaces, this strategy leads to reductions in accuracy when scaling between spaces of differing sizes. Differences in response times between scaled and non-scaled spaces are not expected when using this strategy, as the same cognitive mechanism, the direct mapping of absolute distance, is applied regardless of scaling factor. In contrast, relative scaling strategies involve the use of proportional distances (Huttenlocher et al., 1999). For example, when using a map, a target location may be encoded as being one third of the distance between two landmarks. This relative position (one third) can then be located on a referent map of a different size. Applying this scaling strategy to spaces of differing sizes leads to increased accuracy in comparison to absolute strategies. However, relative scaling strategies do not lead to differences in accuracy or response times for scaled compared to non-scaled trials, as a similar strategy is applied regardless of scaling factor. Thirdly, mental transformation strategies require the mental expansion or contraction of one space to match a second differentially sized space (Vasilyeva & Huttenlocher, 2004). This technique requires distance coding, preservation of metric distances and mental transformation. The use

Table 1

Expected patterns of performance with increasing scaling factor across different cognitive scaling strategies.

Strategy	Change in accuracy with increasing scaling factor	Change in response time with increasing scaling factor
Absolute	Reduced	No change
Relative Scaling	No change	No change
Mental Transformation	Reduced	Increased

of mental transformation strategies is expected to reduce accuracy with increasing scaling factor. This is attributable to the fact that as scaling factor increases, the mental expansion (or contraction) required is greater, making an accurate transformation more difficult. Furthermore, in line with findings from mental rotation paradigms, response times are expected to increase with increasing scaling factor (Möhrling et al., 2016).

As summarised in Table 1 each of the aforementioned strategies of spatial scaling are associated with individual patterns of performance, with increasing scaling factor. Recent findings suggest that mental transformation strategies are required for effective spatial scaling (Möhrling et al., 2014, 2016). Evidence outlining linear increases in both error rates and response times with increasing scaling factor have been reported in localisation tasks with children aged 4–5 years (Möhrling et al., 2014) and both localisation and discrimination tasks in adult populations (Möhrling et al., 2016). These performance patterns fit with the proposed mental transformation model of spatial scaling. However, there is also evidence supporting the use of relative scaling strategies in spatial scaling. Frick and Newcombe (2012) reported no linear increase in errors or response time with increasing scaling factor in children aged 3–6 years. Despite reported reductions in performance accuracy for scaled compared to unscaled trials, these results indicate the use of a relative scaling strategy. Taken together, these conflicting findings suggest differences in the use of spatial strategies across contexts and experimental paradigms.

1.4. Features of spatial task design

Scaling tasks typically require participants to use a map or model with a labelled feature to identify a corresponding feature on a referent space. However, it appears that many features of task design may influence performance on scaling tasks. Findings from Frick and Newcombe (2012) demonstrate the positive influence of landmarks and other reference points, as aids to effective spatial scaling. In contrast, reduced scaling performance is seen in paradigms with high working memory demands. For example, accuracy appears to be reduced when participants are required to remember the location of multiple targets, or in cases where stimuli and referent spaces are not presented simultaneously (Uttal, 1996). As previously described, dimensionality also influences success in spatial scaling paradigms. Performance accuracy is higher and the age at which children can successfully complete spatial scaling is lower for maps including targets distributed on one dimension (i.e. targets distributed along a single, horizontal axis) compared to targets distributed on two dimensions (i.e. targets distributed using a horizontal and vertical axis) (Vasilyeva & Huttenlocher, 2004). Finally, limitations introduced by using technology for scaling tasks, such as size limitations imposed by screen size, do not appear to influence scaling performance. For example, Möhrling et al. (2016) reported no significant differences in scaling accuracy based on the absolute size of the stimuli presented. For each scaling factor tested, participants were presented with a map and corresponding referent space. No significant difference in accuracy was reported for trials in which the referent space was smaller than the map compared to trials in which the referent space was larger than the map. This suggests that the absolute size of the map and referent space used do not significantly influence accuracy at a given scaling factor. However, no known studies investigate the influence of level of acuity, on scaling accuracy at different scaling factors. Here, we propose that the degree of sensitivity (scaling precision) required to identify the correct answer is relatively low when the grid on which targets are presented is less dense (fewer, larger squares within the same surface area). This makes scaling accessible for younger children. However, for denser grids, we propose that the sensitivity required is relatively higher. Hence, 10×10 grid trials may highlight variation in the scaling abilities of older children. For example, the inclusion of both 6×6 and 10×10 grids may enable movement beyond the question of whether participants can scale or not, allowing for the generation of a more sensitive metric for the precision of individual's scaling abilities.

Recent developments in spatial scaling paradigms have been led by Frick, Newcombe, Möhrling and colleagues who have produced a series of spatial paradigms aimed at limiting the cognitive load of scaling tasks and reducing the influence of confounding factors (Frick & Newcombe, 2012; Möhrling et al., 2014, 2016). For localisation tasks, participants are shown the position of a target and are asked to find the corresponding position on a referent space (for example see Frick and Newcombe, 2012). Responses are coded as absolute deviation from the correct answer. For tasks using localisation paradigms, ceiling effects are reported by 6 years of age (for example see Frick & Newcombe, 2012). In contrast, one recent study of adults used a discrimination rather than a localisation paradigm (Möhrling et al., 2016). Participants were asked to distinguish whether a referent map was a scaled correspondent of a target map, or not. Within discrimination paradigms participant's responses are encoded as correct or incorrect. This mode of coding is less forgiving and does not afford any marks for responses that are close to being accurate. As such this more stringent scoring method leads to more discrete scores and may allow for the identification of subtler developmental differences in performance in older children. Furthermore, the use of a discrimination paradigm with multiple response options enables the categorisation and analysis of error types. Whilst one might assume similar frequencies of errors along the horizontal and vertical axes, a higher frequency of errors on the horizontal plane for example, might suggest more inaccurate scaling on this, relative to the vertical plane. Analysis of this type offers a novel insight into scaling processes.

1.5. Current study

This study aims to extend previous findings by exploring age-based and individual variation in spatial scaling using a novel, age-appropriate discrimination paradigm. The results will be reported as a developmental profile of scaling ability in childhood from 5 to 10 years. The secondary aim of this study is to explore the cognitive strategies used by children in the completion of discrimination tasks. Previous studies have highlighted both mental transformation and relative encoding as possible cognitive strategies used in localisation scaling paradigms in children under 6 years of age (Frick & Newcombe, 2012; Möhring et al., 2014). However, findings from discrimination paradigms are limited to adult populations (Möhring et al., 2016). It is unknown whether similar developmental patterns in scaling performance and similar trends in cognitive strategy use, are seen for localisation and discrimination tasks in children, in young and middle childhood (age 5–10 years). The findings from this study provide important evidence on the development of spatial scaling by using a novel scaling task in which both scaling factor and required level of visual acuity are manipulated to create a suitable measure of scaling for children aged 5–10 years. The findings presented are vital for informing the design of future spatial scaling interventions aimed at improving both spatial scaling and STEM achievement more generally. Improved information on age-based differences, and indeed individual differences in spatial scaling ability, will enable the identification of age groups for which targeted scaling interventions may lead to greater gains. Furthermore, information pertaining to cognitive strategy use in spatial scaling tasks will guide and inform the approach taken in future spatial scaling interventions.

2. Material and methods

2.1. Participants

This study included 386 participants across 6 age groups. Participants were aged between 5 and 10 years. Approximately equal numbers of males (55.2%) and females (44.8%) participated in the study. All participants had normal or corrected to normal vision. The sample size, mean age and gender ratios of each age group are shown in Table 2. Participants were recruited from a middle-class, suburban London school in the UK.

2.2. Procedure

Participants were tested individually in a quiet room in their school. In this task, participants were required to choose which one of four onscreen referent maps matched a printed model map. Model maps were either the same size as the on-screen referent maps, or were scaled-up versions of the referent maps (further details in materials section). The experimenter sat to the left of the participant while model and referent maps were positioned in front of the participant as shown in Fig. 1. The experimenter introduced the task as a pirate map game explaining that the yellow colouring on the maps represented sand, while the black boxes were targets showing where hidden treasure was buried. Participants were encouraged to respond as quickly and accurately as possible, by manually pressing one of the maps on the screen to indicate their answer. Following each trial a fixation dot appeared on screen, allowing the experimenter time to turn the page on the A3 flip chart and present the next trial. The task was presented as three blocks of six experimental trials preceded by 2 practice trials with a scaling factor of 1. Feedback was given for practice trials. For incorrect practice trials, participants were asked to repeat the trial until the correct referent space was selected. Only participants achieving at least 50% accuracy for practice trials (i.e. correctly answering at least one of the two practice items on their first attempt), continued to the experimental blocks. All participants successfully completed at least one of the practice trials. Between each block the task instructions were repeated. Participants received no feedback on their performance during experimental trials.

2.3. Materials

Task stimuli included paper-based model maps and onscreen referent maps. Model maps were presented on an A3 flipchart. Each map was positioned in the centre of a white A3 page. Onscreen referent maps including both correct and distractor maps were presented on a 13 inch Hewlett Packard touch-screen laptop in a 2 × 2 arrangement. Model maps measured 8 cm × 8 cm, 16 cm × 16 cm and 32 cm × 32 cm, for trials at a scaling factor of 1, 0.5 and 0.25 respectively. These scaling factors equated to trials in which the lengths of the referent maps were, the same size, one half the size, and one quarter the size of the model map, relative to the participant. All referent maps were 8 cm × 8 cm in size. The model and referent maps were positioned equidistantly from the

Table 2
Demographic features of the study sample.

Age Group	N	Percentage of Males	Age in Years (mean ± sd)
5 Years	65	53.8	5.577 ± 0.266
6 Years	65	64.6	6.521 ± 0.289
7 Years	63	52.4	7.506 ± 0.314
8 Years	67	52.2	8.504 ± 0.262
9 Years	64	56.3	9.451 ± 0.272
10 Years	62	51.6	10.304 ± 0.365

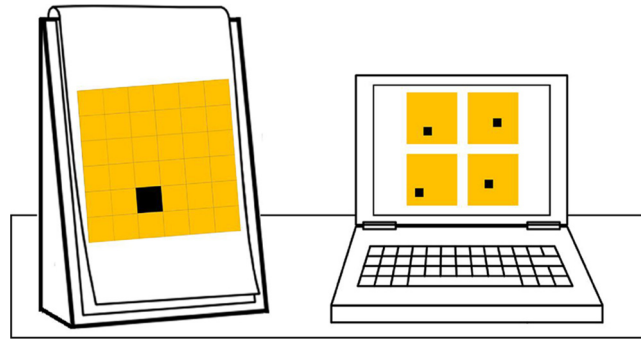


Fig. 1. Relative position of the model (left) and referent (right) maps relative to the participant, for the spatial scaling discrimination task.

participant. Consequently, the scaling factor in each trial was determined as the difference in the relative length of the referent and model maps with respect to the participant. All maps including both model and referent maps, were coloured yellow. Gridlines (for the model map only) and targets were presented in black ink.

The task included three blocks of six trials. Scaling factor varied by block. Within each block, the overall area of the maps, and by extension the scaling factor, did not change. However, the density of the grid on which targets were presented, and hence the size of the grid squares and visual acuity of the maps varied. As shown in Fig. 2, half of the trials in each block were presented using a 6 × 6 square grid (requiring gross-level acuity) while the remaining targets were presented using a 10 × 10 square grid (requiring fine-level acuity). The targets displayed on each map were methodically selected to ensure a balance of left and right side targets. No targets were selected in the outer columns or rows of each grid.

In order to counterbalance for any unintended effects of target position, or block order, five versions of the task were generated. First, for the initial set of targets generated, a second mirror-imaged target set was generated. This counter-balancing controlled for any potential left-right bias in target presentation. Second, to ensure that success on particular blocks could not be attributed to the specific targets used, presentation of target sets was counterbalanced between 0.5 and 0.25 scaling blocks. Overall, this created four versions of the task (Versions A-D). For these versions of the task, the order of block presentation was fixed and blocks were presented in order of increasing scaling factor (i.e. scaling factor was set at 1, 0.5 and 0.25 for Block A, B and C respectively). Finally, to confirm that the order of block presentation did not influence task performance, an additional version of the task, Version A2 was added. The targets included in Version A2 were identical to those in Version A. However, blocks B and C were presented in reverse order i.e. trials with a scaling factor of 0.25 were presented before trials with a scaling factor of 0.5. Approximately equal numbers of participants completed each task version.

For each trial, four onscreen referent maps were presented including 1 correct map (i.e. the scaled (or unscaled) correspondent of the model map) and 3 distractor maps. As shown in Fig. 3, the distractor maps displayed: a vertical distractor which displayed the target one row directly above or below the correct target (A); a horizontal distractor which displayed the target one column directly to the left or right of the correct target (B) and; a diagonal distractor in which the target was positioned at one of the 4 diagonal positions relative to the correct target (C). The onscreen position of the correct map relative to the three distractor maps was randomised across trials with the correct map appearing in each quadrant of the screen with equal frequency.

2.4. Analysis strategy

Statistical analyses were completed using IBM SPSS Statistics for windows (version 22). The use of parametric testing was determined by the outcomes of normality tests, the presence of outliers, and the relatively large sample size in this study. Unless otherwise reported, outcomes of parametric tests are reported. For analyses of variance (ANOVA) including task version, block order, gender or age group, where equal variances could not be assumed, the results for unequal variance are reported. Post-hoc Games-

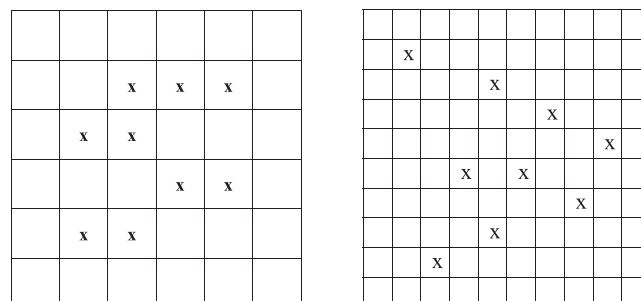


Fig. 2. Task Version A: Targets for trials requiring gross level acuity (left) and fine level acuity (right).

C	A	C
B	Correct target	B
C	A	C

Fig. 3. Position of distractor targets. A indicates a vertical distractor, B indicates a horizontal distractor, C indicates a diagonal distractor.

Howell or Tukey tests were used appropriately in cases where the assumption of homogeneity of variance was violated or met, respectively (Field, 2009). Performance accuracy, measured as percentage of correct trials, acted as the dependent variable in accuracy analysis. There was no missing performance accuracy data. However, thirty-two participants (Total N = 386) were excluded as their performance accuracy on unscaled trials was below chance (25%), suggesting that they were responding at random. Of those excluded, 15 participants were aged 5 years old, 11 participants were aged 6 years and 6 participants were aged 7 years.

Mean response times for correct trials (ms), acted as the dependent variable in response time analysis. In accordance with previous literature, any response times lower than 300 ms, or above 2.5 standard deviations from the median were treated as outliers and coded as missing, prior to the calculation of mean scores (Leys, Ley, Klein, Bernard, & Licata, 2013; Ratcliff & Tuerlinckx, 2002). As mean response times were calculated from correct trials only, it was not possible to calculate mean response times for participants who failed to accurately complete at least one trial out of three for each trial type. As such, we propose that the missing response time data in this study are not random but associated with performance accuracy (for further information see Table A1 in Appendix A). Hence, excluding participants with missing response time values for any trial type (as is the case in complete case analysis) causes bias, with significant under-representation of lower performing participants. Given this consideration, for this study, missing response time data were replaced with plausible values using multiple imputation. Multiple imputation is a statistical technique in which regression modelling is used to calculate plausible values to be used in place of missing data (Rubin, 1987). The use of multiple imputation in this study is supported by the availability of auxiliary variables including performance accuracy and age, that can be included in the multiple imputation model as causes of “missingness” (Collins, Schafer, & Kam, 2001). The number of imputations was calculated using the guidelines set by Graham, Olchowski, and Gilreath (2007) based on both the fraction of missing data (g) and the tolerance for power fall-off. In this study, the fraction of missing response time information is 0.118 and the tolerance for power fall-off was set at the lowest (most conservative) level, < 1%. Based on these parameters, it was determined that 20 imputations were required (Table 5, p212, Graham et al., 2007). Hence, for imputed data the pooled results from 20 imputations are reported.

3. Results

To ensure that performance was not attributable to the specific targets used or left-right bias, a one-way ANOVA comparing performance across task versions (versions A-D) was completed for both accuracy and response time. No significant effect of task version on performance accuracy, $F(3, 295) = .061$, $p = .980$, $\eta_p^2 = 0.001$, or response time, $F(3, 295) = 2.101$, $p = .102$, $\eta_p^2 = 0.021$, was found.

Secondly, to compare the effect of block order (the order of presentation of blocks of each scaling factor) on accuracy and response time, between subject t -tests (comparing Version A and Version A2) were completed. The results indicated no significant effect of block order on accuracy, $t(146) = 0.256$, $p = .798$, $d = 0.042$, or response time, $t(146) = 0.653$, $p = .514$, $d = 0.108$. As no significant effects of task version or block order were reported, these factors were not included in subsequent analyses.

3.1. Performance accuracy

A mixed ANOVA was completed with the between group factors of gender and age groups (5, 6, 7, 8, 9, and 10 years), and the within-group factors of scaling factor (1, 0.5 and 0.25) and acuity (gross and fine). As shown in Fig. 4, a significant main effect of scaling factor was found, $F(2, 684) = 42.074$, $p < .001$, $\eta_p^2 = 0.110$. Bonferroni corrected pairwise comparisons indicated higher accuracy for unscaled blocks compared to scaled blocks with a scaling factor of 0.5 ($p < .001$) or 0.25 ($p < .001$). No significant difference between scaled blocks was reported ($p > .05$). A significant main effect of acuity was also reported, $F(1, 342) = 281.376$, $p < .001$, $\eta_p^2 = 0.451$, with lower accuracy scores for trials requiring fine level acuity ($M = 49.377$, $SE = 0.912$) relative to gross level acuity ($M = 67.286$, $SE = 0.909$). A significant interaction was reported between scaling factor and visual acuity, $F(2, 684) = 10.711$, $p < .001$, $\eta_p^2 = 0.030$. To explore this interaction, two follow up, repeated measures one-way ANOVAs were completed for trials requiring fine level acuity and trials requiring gross level acuity respectively. For both trial types a significant effect of scaling factor was reported. However, the effect size for the effect of scaling factor, for trials requiring fine level acuity, $F(2, 706) = 42.324$, $p < .001$, $\eta_p^2 = 0.107$, was significantly larger than the effect size for trials requiring gross acuity, $F(2, 706) = 7.043$, $p < .001$, $\eta_p^2 = 0.020$. Across both trials types performance on the unscaled block was significantly higher than

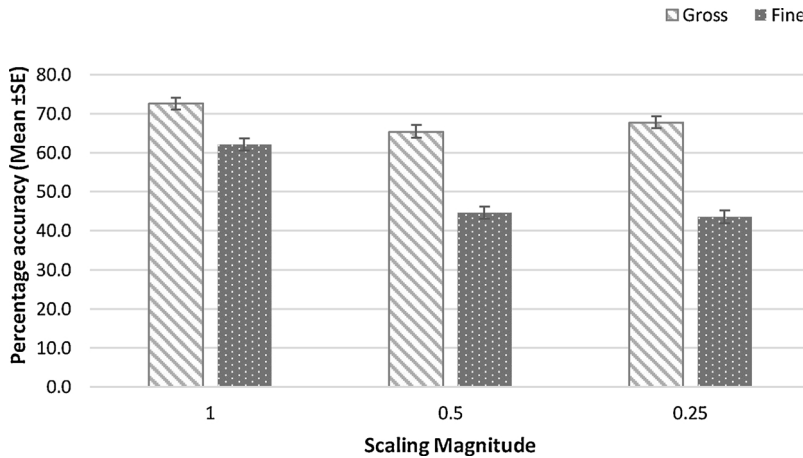


Fig. 4. Performance accuracy by scaling factor and acuity. Error bars indicate standard errors of the means.

performance on scaled blocks ($p < .05$ for both scaled blocks). For both fine and gross level acuity trials, no significant difference between scaled blocks was found ($p > .05$).

A significant main effect of age group on scaling accuracy was reported, $F(5, 342) = 30.718, p < .001, \eta_p^2 = 0.310$ (see Fig. 5). Tukey post-hoc tests indicated that performance accuracy increased with age. That is, 5-year-olds had significantly lower accuracy than 8–10 year olds only ($p < .001$). The performance of 5-year-olds was not significantly different to those aged 6 years ($p = .605$) or 7 years ($p = .157$). Similarly, 6-year-old children had significantly lower performance than all older children ($p < .001$) apart from those aged 7 years ($p = .966$). Accuracy scores at age 7 were also significantly lower than all older age groups ($p < .001$), while performance at 8 years was significantly lower than 10-year old performance only ($p = .026$). No other significant differences between 9 and 10-year-old’s performance were reported ($p > .05$). A significant interaction between age group and acuity was also reported, $F(5, 342) = 6.141, p < .001, \eta_p^2 = 0.082$.

As shown in Fig. 6, paired sample t -tests (with Bonferroni adjusted alpha levels of 0.008 [0.05/6] to account for multiple comparisons) indicated significantly higher performance accuracy for trials requiring gross compared to fine level acuity, for all age groups except for 5-year olds. It is interesting to note that the magnitude of the effect sizes reported differs by age group. 5-year old children show poor performance overall, with a relatively narrow performance gap between trials requiring gross and fine level acuity, $t(49) = 2.605, p = .012, d = 0.368$. As children get older, their accuracy for trials requiring gross level acuity improves prior to their accuracy for fine level trials. For examples see Fig. 6, for children aged 6 years, $t(53) = 5.478, p < .001, d = 0.754$ and children aged 7 years, $t(56) = 7.011, p < .001, d = 0.951$. Consequently, the gap in performance accuracy between trials with gross and fine level acuity grows. The greatest difference in performance across trial types is seen for children aged 8 years, $t(66) = 8.869, p < .001, d = 1.087$, and aged 9 years, $t(63) = 11.879, p < .001, d = 1.485$. It appears that this gap begins to close at 10 years where overall performance on both trial types is relatively strong, $t(61) = 7.032, p < .001, d = 0.904$. Furthermore, these findings are not attributable to the presence of floor effects in younger participants (both 5 and 6-year olds had above chance performance for trials requiring fine level acuity) or ceiling effects in the older age group (the average performance accuracy of all groups was less than or equal to 71.00%). No significant interaction was found for age group and scaling factor, $F(10,684) = 1.319, p = .218$,

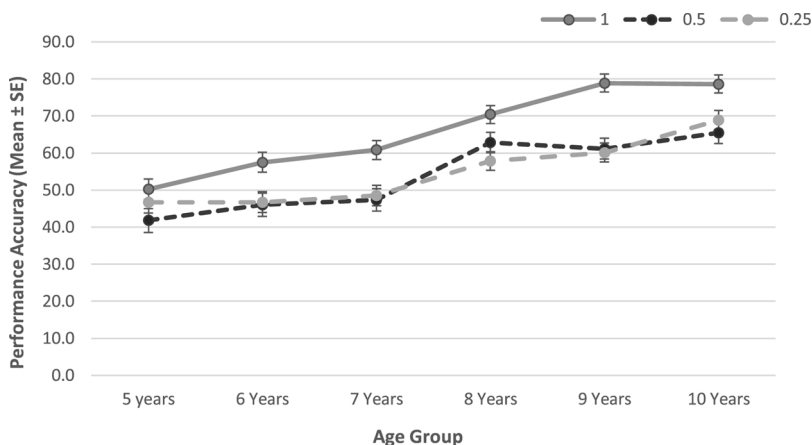


Fig. 5. Performance accuracy by age group and scaling factor. Error bars indicate standard errors of the means.

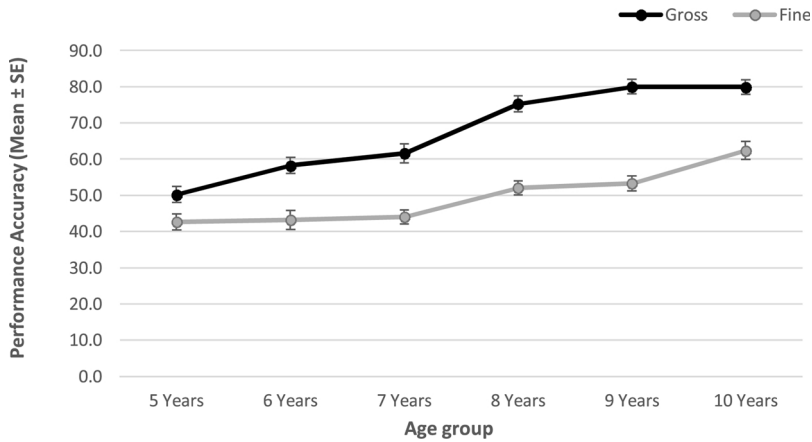


Fig. 6. Performance accuracy by age group and acuity. Error bars indicate standard errors of the means.

$\eta_p^2 = 0.019$. For gender, the overall performance of boys ($M = 60.521$, $SE = 0.980$) was significantly higher than that of girls ($M = 56.142$, $SE = 1.102$), $F(1,342) = 8.810$, $p = .003$, $\eta_p^2 = 0.025$. However, there were no significant interactions between gender and any other variables, including age ($p > .05$, $\eta_p^2 < 0.025$ for all).

3.2. Response time

Mean response time was calculated from correct trials for each trial type (i.e. all trials of the same scaling factor and acuity). Unless otherwise stated, the results reported are based on the multiple imputation data set. More information on results for complete case (where any participant with incomplete data was excluded) and multiple imputation data sets can be found in Table B1 in Appendix B. A mixed ANOVA was completed with the between-group measures of gender and age group (5, 6, 7, 8, 9, and 10 years), and within-group measures of scaling factor (1, 0.5 and 0.25) and acuity (gross and fine). No significant main effect of scaling factor, $F(2, 684) = 1.717$, $p = .260$, $\eta_p^2 = 0.005$, acuity $F(1, 342) = 0.353$, $p = .625$, $\eta_p^2 = 0.001$, or age group on response time was reported, $F(5, 342) = 0.930$, $p = .477$, $\eta_p^2 = 0.014$. A significant interaction was reported between scaling factor and acuity, $F(2, 684) = 5.255$, $p = .025$, $\eta_p^2 = 0.015$. Two follow up, repeated measures one-way ANOVAs indicated a significant effect of scale for gross level trials, $F(2, 706) = 4.676$, $p = .017$, $\eta_p^2 = 0.013$, but not fine level trials, $F(2, 708) = 1.198$, $p = .478$, $\eta_p^2 = 0.003$. For gross level trials, participants were significantly faster for trials at a scaling factor of 0.25 compared to unscaled trials ($p < .001$). No other significant differences between trials of different scaling factors were reported ($p > .05$). The results also indicated a significant main effect of gender, $F(1, 342) = 5.734$, $p = .047$, $\eta_p^2 = .016$. Overall, the average response time for females (5329.75 ± 154.022) was shorter than that of males (5784.884 ± 150.271).

3.3. Error patterns

As shown in Fig. 7, chi squared analysis was used to investigate differences in the relative proportions of Vertical (V), Horizontal (H), and Diagonal (D) type errors.

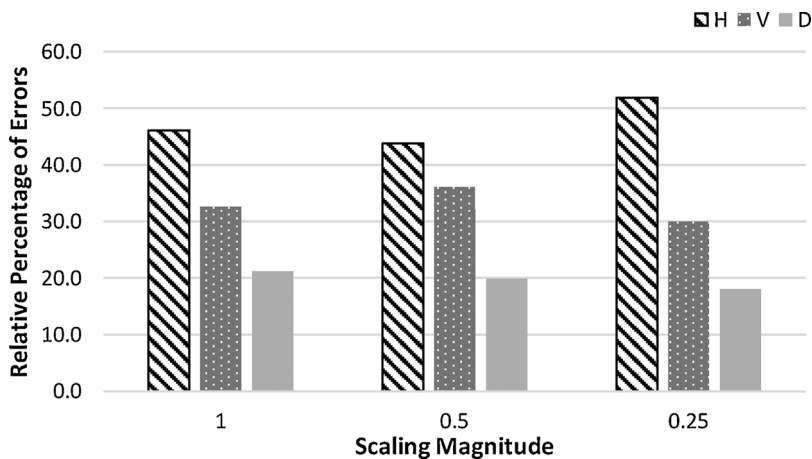


Fig. 7. Relative percentages of V, H and D type errors for trials at each level of scaling factor.

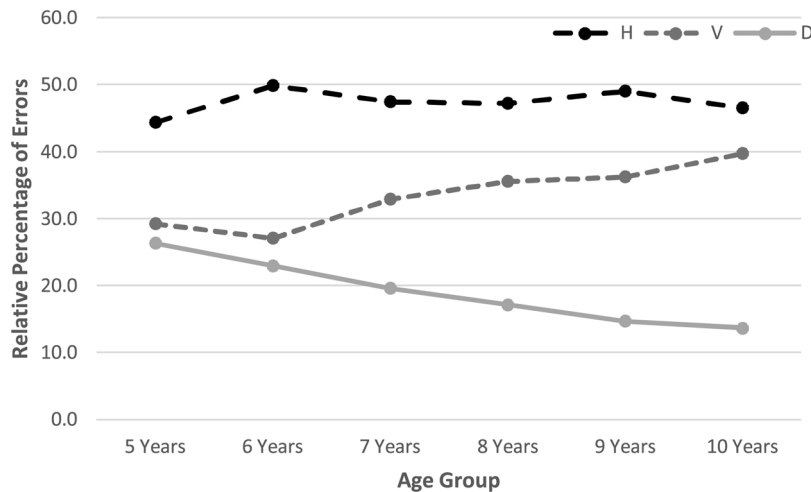


Fig. 8. Relative percentages of V, H and D type errors across different age groups.

(H) and Diagonal (D) type errors. For all scaling factors, significantly higher proportions of H errors relative to V and D errors were reported: unscaled trials: $\chi^2(2, N = 693) = 64.944, p < .001$; scaling factor of 0.5: $\chi^2(2, N = 948) = 84.994, p < .001$; scaling factor of 0.25: $\chi^2(2, N = 941) = 167.007, p < .001$. Fig. 8 highlights differences in the relative frequencies of V, H and D errors across age groups. Chi squared analysis for each age group indicated significantly more H errors compared to V and D errors for all age groups ($p < .001$ for all). Furthermore, V errors occurred with significantly higher frequency compared to D errors for all participants aged 8–10 years ($p < .001$ for all). However, there was no significant difference in the frequency of V and D errors for participants aged 5 years, $\chi^2(1, N = 268) = 0.731, p = .392$, and participants aged 6 years, $\chi^2(1, N = 240) = 1.667, p = .197$. For all analysis, N values indicate the number of trials.

4. Discussion

The primary aim of this study was to provide a developmental profile of spatial scaling in children between the ages of 5 and 10 years of age. Overall, significant age-based differences in scaling accuracy were reported. The results indicated performance gains in spatial scaling between 5 and 8-years-of-age, after which no significant improvements in task accuracy were found. These findings contrast with results from localisation paradigms, where children's accuracy on scaling tasks reaches a plateau by age 6 (Frick & Newcombe, 2012; Möhring et al., 2014). Conversely, children's scaling skills as measured using this discrimination paradigm continue to develop until 8-years-of-age. These contrasting results suggest that the scaling skills for placement, as required in localisation tasks, differ from those of discrimination tasks, enabling younger children to effectively complete tasks of this type (Frick & Newcombe, 2012; Möhring et al., 2014). This may be attributable to the increased scaling precision required for discrimination tasks, in addition to other domain general demands that may be needed for discrimination paradigms, such as working memory or inhibition. The findings of this study also add to previous literature on gender differences in spatial performance in childhood. Significant differences in both accuracy and response time were reported such that females had significantly faster but less accurate performance than males. This may suggest that females are applying scaling strategies less effectively than males. However, these findings should be interpreted in the context of the small effect sizes reported. These results are consistent with previous studies on spatial thinking in which non-significant or small effect sizes for significant gender differences between males and females were found (Alyman & Peters, 1993; Halpern et al., 2007; Lachance & Mazzocco, 2006; LeFevre et al., 2010; Manger & Eikeland, 1998; Neuburger, Jansen, Heil, & Quaiser-Pohl, 2011).

The secondary aim of this study was to explore the cognitive strategies used by children in the completion of discrimination type scaling tasks. As previously outlined, specific patterns of performance accuracy and response times in scaling tasks, have been associated with different cognitive strategies including absolute, relative and mental transformation strategies (see Table 1). As this study included a small number of trials at each scaling factor, the findings pertaining to response time should be interpreted cautiously and seen as complementary to the key findings on performance accuracy. Nonetheless, the findings reported in this study indicated no significant increase in response time with increasing scaling factor. Indeed for gross level trials, response times were significantly shorter for trials at a scaling factor of 0.25 compared to unscaled trials.

The results reported can be interpreted in the context of the aforementioned cognitive strategies for spatial scaling. Firstly, despite evidence that children use mental transformation strategies for spatial scaling in localisation tasks (for example see Möhring et al. (2014)), the use of mental transformation strategies is not supported in this study. This is particularly interesting given that mental transformation strategies have also been reported for discrimination tasks in adults (Möhring et al., 2016). However, the results reported in this study show neither reduced performance accuracy nor increased response times with increased scaling demands, which would be anticipated for mental transformation strategy use. In contrast, as scaling demands increase, relative scaling

strategies are associated with unchanged accuracy and response times, while absolute strategies are expected to generate reductions in performance accuracy only. The unusual pattern of results reported in this study, and mirrored in previous work by Frick and Newcombe (2012), is not entirely consistent with either of these models. Despite showing reduced performance for scaled relative to unscaled trials, no reduction in accuracy with increasing scaling factor is observed. As such the use of absolute strategies in the completion of this task is also deemed unlikely. Although the pattern of results reported in this study is not a perfect fit to the relative scaling model, both mental transformation and absolute strategies provide poor explanations for these findings. Furthermore, despite findings from adult populations where the physical size of the maps used was not found to significantly influence scaling performance (Möhrling et al., 2016), future research could investigate whether the physical size of the maps used influences performance in discrimination scaling tasks in children.

Overall, the results support the use of relative strategies for discrimination tasks of spatial scaling in children. As such, participants are proposed to encode relative distances to solve scaling problems, for example by encoding that the target is one third of the way between the two sides of the grid. These findings contrast with those for discrimination paradigms in adults, where mental transformation strategies are reported (Möhrling et al., 2016). These findings can be viewed in the context of other spatial domains such as mental rotation, for which there is evidence of variation in the strategies used for task completion by different individuals at different developmental stages (Geiser, Lehmann, & Eid, 2008; Glück, Machat, Jirasko, & Rollett, 2002; Janssen & Geiser, 2012). Consequently, it is unsurprising that scaling tasks with differing experimental paradigms may lead to the recruitment of differing scaling strategies. Perhaps therefore, it would not be safe to assume that all children at all ages deploy the same strategies in the completion of scaling tasks. Future studies should explore the conditions under which individuals might be encouraged to use specific cognitive scaling strategies, and whether particular features of task design promote the use of different strategies. For example, in this study, the inclusion of four maps may have encouraged the use of the relative strategy. Participants may have first encoded the target location, using relative distance information, across a single dimension (e.g. the horizontal axis), therefore enabling them to immediately discount some of distractors, before then relatively encoding the remaining maps on the other axis. This is cognitively less demanding than completing mental transformations on four individual maps. Furthermore, the presence of grid lines on the model map may have encouraged the use of relative distance information (i.e. the target is three units from the left of the map).

The discrimination task used in this study allowed for the categorisation of incorrect trials into discrete error sub-categories. For all scaling factors and age groups, higher proportions of H errors compared to V and D errors were reported. Furthermore, this pattern remained stable across scaling factors. These findings indicate interesting overall differences in individual's mapping abilities on the horizontal and vertical axis. This suggests that scaling demands do not induce specific negative effects for mapping on the horizontal or vertical axis respectively. The uniform error patterns may also suggest that participants use similar cognitive strategies to complete both scaled and unscaled trials. The use of relative scaling strategies would fit with this pattern. These findings beg the question as to why individuals appear to be more accurate in distinguishing D and V errors from the correct target, compared to H errors. For D errors, there is less shared contact area with the correct target. As such D errors are further away from the correct target and are understandably the easiest error type to distinguish from the correct target. However, shared contact with the target cannot explain the differences reported in the frequencies of H and V type errors. Both of these error types have the same degree of contact with, and are identical distances from the correct target.

Alternatively, differences in the frequency of H and V type errors may be attributable to the horizontal-vertical illusion (Oppel, 1855). This is the illusion that a vertical line appears longer than a horizontal line of the same length. This illusion leads to over-estimations in vertical segments relative to horizontal ones (Mamassian & de Montalembert, 2010). One explanation for this phenomenon is that vertical lines may be perceived as “receding into the third dimension” leading to perceptual errors in estimating spatial distance (p. 60, Girgus & Coren, 1975). In the context of this study, over-estimating vertical distances would lead participants to perceive V errors as further away from, and thus less likely to be confused with, the correct answer. While studies assessing the horizontal-vertical illusion typically include lines or dots displayed simultaneously on a single display (McGraw & Whitaker, 1999), the findings of the current study suggest that the horizontal-vertical illusion may extend to mapping tasks. Beyond this study, the implications of the horizontal-vertical illusion in spatial mapping tasks is largely unknown. Future studies could explore differences in horizontal mapping and vertical mapping using both localisation and discrimination paradigms. Alternatively, the observed differences in the frequency in H and V type errors may be attributable to the horizontal layout of the maps and referent spaces used in this task. As participants were required to transfer their attention horizontally from the target map to the referent spaces, focus may have been inadvertently directed to the horizontal axis. Future research could compare performance on this task when maps and referent spaces are presented using a vertical layout (i.e. with the target map presented above or below the referent spaces).

The use of a discrimination task in this study offers novel insights into the cognitive processes used by participants in the completion of spatial scaling. As previously outlined, performance on scaling tasks appears to be influenced by features of task design. For example, visual acuity was shown to influence performance in this study. Variation in acuity increased the suitability of this task for a wider age range of children, such that the inclusion of fine-level acuity trials allowed for performance variation for the oldest children, whilst the gross-level acuity trials avoided floor effects in performance for the younger children. The results show that performance accuracy for trials requiring gross level acuity stabilised at 7-years-of-age, in contrast to trials requiring fine level acuity, where performance accuracy stabilised at 8-years-of-age. Future research should further investigate the role of visual acuity in scaling success. As this is the first study to investigate scaling in children using a discrimination paradigm, the exact features of task design that best enable children to complete discrimination based scaling tasks, are largely unknown. Given that this task is the first to

require discrimination between four scaled spaces, and uses comparison between digital and paper-based formats, response times are high relative to other scaling tasks. Furthermore, it might be argued that the long-lasting comparison process required between four alternatives may be masking increases in response time across scaling factors. However, given that comparison between four alternatives is required for all trials, we would expect this to elevate response times across all scaling factors uniformly and not to selectively interfere with any effects of scaling or acuity. Overall the findings from this study, in particular findings on age-based performance differences, should also be viewed in light of the high levels of individual variation in task performance across children. However, taken together, the specific features of task design used in this study, have led to the generation of a measure suitable for assessing individual differences in spatial scaling abilities in older children (without ceiling effects) and younger children (without floor effects).

In this study, through the use of an age-appropriate discrimination task, it was shown that children achieve performance gains in spatial scaling until 8-years-of-age, some two years older than is typically seen for localisation paradigms. These contrasting results suggest that the spatial skills required for different scaling paradigms (localisation v's discrimination tasks) may vary. Perhaps, spatial scaling through development may best be understood by combining findings from localisation, discrimination and other paradigms. For example, future work with children of this age could investigate further unexplored paradigms such as the use of scaled maps for navigation or the reconstruction of maps from scaled models. Furthermore, the patterns of errors seen in this study suggest that mapping between differing spaces may be influenced by the horizontal-vertical illusion. Better understanding of this illusion may offer a novel way of teaching spatial mapping and improving scaling skills in children. By achieving a better understanding of the development of spatial scaling skills, children can be encouraged to engage in age appropriate spatial scaling tasks aimed at improving their scaling skills.

Given that other aspects of spatial cognition are predictive of success in Science, Technology, Engineering and Mathematics domains (STEM) domains in adults, spatial scaling may also play an important role in STEM achievement (Casey, Nuttall, & Pezaris, 2001; Gunderson, Ramirez, Beilock, & Levine, 2012; Taylor & Hutton, (2013); Wai, Lubinski, & Benbow, 2009). There are many practical applications of spatial scaling in the classroom including map reading, achieving proportionality in drawing, or classroom activities in science for example, such as relating larger scaled diagrams (e.g. on a whiteboard) to smaller printed diagrams. Given their relevance in the science and maths classroom there is a need to better understand the cognitive underpinnings and development of scaling abilities. Through the use of an age-appropriate discrimination paradigm, this study offers novel insights into children's spatial scaling abilities, highlighting the continuing development of scaling abilities in children up to 8 years of age, and the presence of substantial individual differences in spatial scaling at all developmental ages. These findings suggest room for improvement in children's spatial scaling skills, and highlight the potential of scaling interventions in the early primary school years. Such interventions may infer both direct benefits on spatial thinking and indirect advantages for STEM achievement more generally.

Acknowledgements

Funding for this research was provided by The Bloomsbury Colleges Ph.D. Scholarship Programme and the Economic and Social Research Council (ESRC), and was supported by the National Centre for Curriculum and Assessment (NCCA) Ireland.

Appendix A

Table A1

Percentage of participants with response time data for each trial type (trial types include all combinations of scaling factor (1, 0.5 and 0.25) and acuity (gross and fine).

Age Group	N	Scaling Factor					
		1		0.5		0.25	
		Gross	Fine	Gross	Fine	Gross	Fine
Age 5	50	90.00	82.00	84.00	70.00	92.00	76.00
Age 6	54	92.59	94.44	94.44	74.07	98.15	72.22
Age 7	57	94.74	89.47	91.23	75.44	96.49	77.19
Age 8	67	97.01	94.03	98.51	86.57	91.04	77.61
Age 9	64	100.00	96.88	98.44	76.56	98.44	76.56
Age 10	62	100.00	96.77	95.16	91.94	98.39	83.87
Total	354	96.05	92.66	94.07	79.66	95.76	77.40

Appendix B

Table B1

Summary of results for complete case and multiple imputation data sets.

Variables	Data Set	N	Df(n)	Df(d)	F	p	η_p^2
Acuity	Complete Case	184	1	172	2.022	.157	0.012
	Imputed	354	1	342	0.353	.625	0.001
Scaling factor	Complete Case	184	2	344	2.438	.102	0.014
	Imputed	354	2	684	1.717	.260	0.005
Scaling factor * Acuity	Complete Case	184	2	344	3.285	.053	0.019
	Imputed	354	2	684	5.255	.025	0.015
Age	Complete Case	184	5	172	1.573	.170	0.044
	Imputed	354	5	342	0.930	.447	0.014
Gender	Complete Case	184	1	172	0.224	.637	0.001
	Imputed	354	1	342	5.734	.047	0.016

Note. For complete case analysis, the results pertaining to age group should be interpreted with caution as there were only 16 5-year-old participants in this group, for whom response time data were available for all trial types.

References

- Alyman, C., & Peters, M. (1993). Performance of male and female children, adolescents and adults on spatial tasks that involve everyday objects and settings. *Canadian Journal of Experimental Psychology = Revue Canadienne De Psychologie Experimentale*, 47(4), 730–747.
- Bushnell, E., McKenzie, E., Lawrence, A., & Connell, S. (1995). The spatial coding strategies of one-year-old infants in a locomotor search task. *Child Development*, 66(4), 937–958. <http://dx.doi.org/10.1111/j.1467-8624.1995.tb00914.x>.
- Casey, B., Nuttall, R., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 32(1), 28–57. <http://dx.doi.org/10.2307/749620>.
- Collins, L., Schafer, J., & Kam, M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods*, 6(4), 330–351.
- DeLoache, J. (1987). Rapid change in the symbolic functioning of very young children. *Science*, 238(4833), 1556–1557. <http://dx.doi.org/10.1126/science.2446392>.
- DeLoache, J. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development*, 4(2), 121–139. [http://dx.doi.org/10.1016/0885-2014\(89\)90012-9](http://dx.doi.org/10.1016/0885-2014(89)90012-9).
- Field, A. (2009). *Discovering statistics using SPSS*. SAGE Publications Ltd.
- Frick, A., & Newcombe, N. (2012). Getting the big picture: development of spatial scaling abilities. *Cognitive Development*, 27(3), 270–282. <http://dx.doi.org/10.1016/j.cogdev.2012.05.004>.
- Frick, A., Möhring, W., & Newcombe, N. (2014). Picturing perspectives: development of perspective-taking abilities in 4- to 8-year-olds. *Frontiers in Psychology*, 5(386), 1–7. <http://dx.doi.org/10.3389/fpsyg.2014.00386>.
- Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence*, 36(6), 556–563. <http://dx.doi.org/10.1016/j.intell.2007.12.003>.
- Girgus, J., & Coren, S. (1975). Depth cues and constancy scaling in the horizontal-vertical illusion: The bisection error. *Canadian Journal of Psychology*, 29(1), 59–65.
- Glück, J., Machat, R., Jirasko, M., & Rollett, B. (2002). Training-related changes in solution strategy in a spatial test. *Learning and Individual Differences*, 13(1), 1–22. [http://dx.doi.org/10.1016/S1041-6080\(01\)00042-5](http://dx.doi.org/10.1016/S1041-6080(01)00042-5).
- Graham, J., Olchowski, A., & Gilreath, T. (2007). How many imputations are really needed? Some practical clarifications of multiple imputation theory. *Prevention Science*, 8(3), 206–213. <http://dx.doi.org/10.1007/s11121-007-0070-9>.
- Gunderson, E., Ramirez, G., Beilock, S., & Levine, S. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, 48(5), 1229–1241. <http://dx.doi.org/10.1037/a0027433>.
- Halpern, D., Benbow, C., Geary, D., Gur, R., Hyde, J., & Gernsbacher, M. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8(1), 1–51. <http://dx.doi.org/10.1111/j.1529-1006.2007.00032.x>.
- Huttenlocher, J., Newcombe, N., & Sandberg, E. (1994). The coding of spatial location in young children. *Cognitive Psychology*, 27, 115–147. <http://dx.doi.org/10.1006/cogp.1994.1014>.
- Huttenlocher, J., Newcombe, N., & Vasilyeva, M. (1999). Spatial scaling in young children. *Psychological Science*, 10(5), 393–398. <http://dx.doi.org/10.1111/1467-9280.00175>.
- Janssen, A., & Geiser, C. (2012). Cross-cultural differences in spatial abilities and solution strategies—An investigation in Cambodia and Germany. *Journal of Cross-Cultural Psychology*, 43(4), 533–557. <http://dx.doi.org/10.1177/0022022111399646>.
- Lachance, J., & Mazzocco, M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences*, 16(3), 195–216. <http://dx.doi.org/10.1016/j.lindif.2005.12.001>.
- LeFevre, J., Fast, L., Skwarchuk, S., Smith-Chant, B., Bisanz, J., Kamawar, D., et al. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, 81(6), 1753–1767. <http://dx.doi.org/10.1111/j.1467-8624.2010.01508.x>.
- Leys, C., Ley, C., Klein, O., Bernard, P., & Licata, L. (2013). Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology*, 49(4), 764–766. <http://dx.doi.org/10.1016/j.jesp.2013.03.013>.
- Liben, L., & Downs, R. (1993). Understanding person space map relations: Cartographic and developmental perspectives. *Developmental Psychology*, 29(4), 739–752. <http://dx.doi.org/10.1037/0012-1649.29.4.739>.
- Möhring, W., Newcombe, N., & Frick, A. (2014). Zooming in on spatial scaling: Preschool children and adults use mental transformations to scale spaces. *Developmental Psychology*, 50(5), 1614–1619. <http://dx.doi.org/10.1037/a0035905>.
- Möhring, W., Newcombe, N., & Frick, A. (2015). The relation between spatial thinking and proportional reasoning in preschoolers. *Journal of Experimental Child Psychology*, 132, 213–220. <http://dx.doi.org/10.1016/j.jecp.2015.01.005>.
- Möhring, W., Newcombe, N., & Frick, A. (2016). Using mental transformation strategies for spatial scaling: Evidence from a discrimination task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 42(9), 1473–1479. <http://dx.doi.org/10.1037/xlm0000240> No Pagination Specified.
- Mamassian, P., & de Montalembert, M. (2010). A simple model of the vertical–horizontal illusion. *Vision Research*, 50(10), 956–962. <http://dx.doi.org/10.1016/j.visres.2010.03.005>.
- Manger, T., & Eikeland, O. (1998). The effects of spatial visualization and students' sex on mathematical achievement. *British Journal of Psychology*, 89(1), 17–25.

- <http://dx.doi.org/10.1111/j.2044-8295.1998.tb02670.x>.
- McGraw, P., & Whitaker, D. (1999). Perceptual distortions in the neural representation of visual space. *Experimental Brain Research*, 125(2), 122–128. <http://dx.doi.org/10.1007/s002210050667>.
- Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2011). Gender differences in pre-adolescents' mental-rotation performance: Do they depend on grade and stimulus type? *Personality and Individual Differences*, 50(8), 1238–1242. <http://dx.doi.org/10.1016/j.paid.2011.02.017>.
- Newcombe, N., Huttenlocher, J., & Learnmonth, A. (1999). Infants' coding of location in continuous space. *Infant Behavior and Development*, 22(4), 483–510. [http://dx.doi.org/10.1016/S0163-6383\(00\)00011-4](http://dx.doi.org/10.1016/S0163-6383(00)00011-4).
- Newcombe, N., Sluzenski, J., & Huttenlocher, J. (2005). Preexisting knowledge versus on-line learning. *Psychological Science*, 16(3), 222–227. <http://dx.doi.org/10.1111/j.0956-7976.2005.00807.x>.
- Newcombe, N., Uttal, D., & Sauter, M. (2013). Spatial development. In P. Zelazo (Ed.). *Oxford handbook of developmental psychology: Vol. 1. Body and mind* (pp. 564–590). New York, NY: Oxford University Press.
- Oppel, J. (1855). *Über geometrisch-optische Täuschungen* [Translation: About geometric-optical illusions]. *Jahresbericht Des Frankfurter Vereins, 1854–1855*, 37–47.
- Piaget, J., & Inhelder, B. (1948). *The child's conception of space*. Norton: New York.
- Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review*, 9(3), 438–481. <http://dx.doi.org/10.3758/BF03196302>.
- Rubin, D. (1987). *Frontmatter. Multiple imputation for nonresponse in surveys*. John Wiley & Sons Inc. <http://dx.doi.org/10.1002/9780470316696.fmatter> [(pp. i–xxix).
- Shusterman, A., Ah Lee, S., & Spelke, E. (2008). Young children's spontaneous use of geometry in maps. *Developmental Science*, 11(2), F1–F7. <http://dx.doi.org/10.1111/j.1467-7687.2007.00670.x>.
- Taylor, H., & Hutton, A. (2013). Think3d!: training spatial thinking fundamental to STEM education. *Cognition and Instruction*, 31(4), 434–455. <http://dx.doi.org/10.1080/07370008.2013.828727>.
- Uttal, D. (1996). Angles and distances: Children's and adults' reconstruction and scaling of spatial configurations. *Child Development*, 67(6), 2763–2779.
- Vasilyeva, M., & Huttenlocher, J. (2004). Early development of scaling ability. *Developmental Psychology*, 40(5), 682–690. <http://dx.doi.org/10.1037/0012-1649.40.5.682>.
- Wai, J., Lubinski, D., & Benbow, C. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <http://dx.doi.org/10.1037/a0016127>.