Space for Mathematics: Spatial cognition as a contributor to the development of mathematics skills in children

Katie Gilligan

A thesis submitted for the degree

of Doctor of Philosophy

University College London

2019

I, Katie Gilligan confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Signature: …………………………………………………

Abstract

There is evidence to suggest that associations exist between spatial skills and mathematics in pre-school and adult populations. However, relatively few studies explore these associations in primary school aged children. The experimental studies presented in this thesis investigated the developmental relations between spatial and mathematical skills in children aged 5 to 10 years, including the transfer of spatial training gains to mathematics. Associations between spatial thinking and mathematics were observed longitudinally and cross-sectionally. Secondary data analysis of the Millennium Cohort Study, a longitudinal study of children in the United Kingdom, indicated that spatial performance at 5 years was a significant longitudinal predictor of mathematics at 7 years. Spatial skills explained 15% of the variation in mathematics achievement at 7 years even after controlling for gender, socio-economic status and language skills (N = 12099). Findings from a cross-sectional study of children aged 6 to 10 years found that spatial skills explained 7% to 13% of the variation across three mathematics performance measures (standardised mathematics, approximate number system, and number line estimation skills) (N = 155). Some relations reported between spatial and mathematical skills were sub-domain specific. While spatial scaling was a significant predictor of all mathematics outcomes, disembedding was associated with standardised mathematics performance only. Certain spatial-mathematical relations were also sensitive to developmental age. Mental rotation had a greater influence on mathematics for younger compared to older children. These insights on the selectivity and developmental sensitivity of spatial-mathematical relations were used to design an intervention study, which targeted mental rotation and spatial scaling skills. In this study, spatial training led to gains in the spatial skill trained (near transfer), transfer of gains to un-trained spatial domains (intermediate transfer), and transfer of gains to mathematical domains (far transfer). It was concluded that spatial skills have a causal role in mathematics outcomes in childhood.

Impact statement

Significant associations were found between spatial thinking and mathematics skills in children aged 5 to 10 years. These spatial-mathematical relations were supported by both longitudinal and cross-sectional evidence. Furthermore, training spatial thinking led to gains in the spatial skill trained (near transfer), transfer of gains to untrained spatial domains (intermediate transfer), and transfer of gains to mathematical domains (far transfer). The implications of these findings are far reaching.

The findings of this thesis fine-tune the relationship between different sub-domains of spatial thinking and different mathematical skills, demonstrating that spatial-mathematical relations show specificity to certain spatial and mathematical sub-domains, and sensitivity to developmental age. They add to the theoretical debate on the causal relationship between spatial and mathematics skills, outlining a causal effect of spatial thinking on mathematics outcomes that was previously only inferred based on correlational evidence. The findings also contribute to the current understanding of transfer of cognitive training gains to untrained domains. It is proposed that the choice of cognitive training be determined by an understanding of the underlying cognitive mechanisms of training targets. The training gains reported in this research highlight the importance of choosing task and age sensitive targets for cognitive training.

The evidence presented in this thesis strongly advocates for the *spatialisation* of primary school mathematics curricula such that children are taught how to read diagrams and graphs, encouraged to sketch and draw, exposed to spatial language, and given hands on opportunities to manipulate and explore with 3D materials. Given the ease with which they can be delivered, the findings from this thesis highlight the potential of instructional videos (explicit instruction) as a practical tool for *spatialising* the classroom. The introduction of spatial training, and the use of spatial tools in mathematics classrooms, are proposed as novel ways of improving mathematics performance in primary school children. Beyond individual gains, improving spatial and mathematics skills may lead to national improvements on international assessments of mathematics, in which children from the UK typically perform less favourably on space and shape related domains, compared to other aspects of mathematics (Greany, Barnes, Mostafa, Pensiero, & Swensson, 2016).

Outside the classroom, improving mathematics outcomes may have a wider economic impact. Improving mathematics attainment in the early school years may lead to related improvements in the quality of Science, Technology, Engineering, and Mathematics (STEM) graduates, with consequent implications for the STEM industry. In recent years, many employers have reported STEM personnel shortages and difficulties recruiting suitably qualified STEM graduates (National Audit Office UK, 2018). Given the continuous expansion of the STEM industry, improving STEM skills has become a national priority. The findings presented in this thesis suggest that targeting spatial skills offers a promising avenue to tackle this challenge.

Dissemination

Peer Reviewed Publications

Data from Chapter 4 are under review for publication as a Journal Article in Developmental Science:

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (submitted). First demonstration of effective spatial training for near-transfer to spatial performance and far-transfer to a range of mathematics skills at 8 years. *Developmental Science Special Issue.*

Data from Chapter 3 are published as a Journal Article in Developmental Science:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2018). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science.* https://doi: 10.1111/desc.12786

Collaboration from this thesis using data from Chapter 3 is published as a Journal Article in the British Journal of Educational Psychology.

Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology.* https://doi.org/10.1111/bjep.12211

Extended research using data from Chapter 3 is published as a Journal Article in Cognitive Development:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2018). The use of discrimination scaling tasks: a novel perspective on the development of spatial scaling. *Cognitive Development*. https://doi.org/10.1016/j.cogdev.2018.04.001

Data from Chapter 2 are published as a Journal Article in the Journal of Experimental Child Psychology:

Gilligan, K. A., Flouri, E., & Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology*, 163, 107–125. https://doi.org/10.1016/j.jecp.2017.04.016

Conference Proceedings and Seminars

Data from Chapter 4 were presented at the British Psychological Society, Developmental Section Annual Conference, Liverpool, UK:

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2018). The effectiveness of spatial training videos to improve spatial and mathematical thinking in primary school children. Oral presentation.

Extended data from Chapter 3 were presented at Mathematics Education Centre's 3rd annual symposium: The symbol grounding problem, Loughborough University, UK:

Gilligan K. A., Hodgkiss, A., Patel, P. C., & Farran, E. K. (2018). The role of spatial language in the primary school mathematics classroom. Poster presentation.

Data from Chapter 4 were presented at the European Association for Research on Learning and Instruction (EARLI) Sig 4 Conference, London, UK:

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2018). Brain space: Improving spatial thinking with instructional videos. Poster presentation.

Data from Chapter 4 were presented to the Developmental Neurocognition Lab, Birkbeck University, UK:

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2018). Evidence of near and far transfer: Using instructional videos of spatial thinking to improve spatial and mathematics skills in children. Oral presentation.

Data from Chapter 4 were presented at the British Psychological Society, Developmental Section Annual Conference, Stratford Upon Avon, UK:

Gilligan, K. A., Thomas, M. S. C., & Farran, E. K. (2017). Brain space: Using instructional videos of spatial thinking to improve spatial and mathematics skills. Oral presentation.

Data from Chapter 3 were presented at the 39th Annual Meeting of the Cognitive Science Society, London, UK:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). The role of spatial skills in mathematics cognition: Evidence from children aged 5-10 years. Poster presentation.

Data from Chapters 2 and 3 were presented at the 20th European Society for Cognitive Psychology (ESCOP) Conference, Potsdam, Germany:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). Spatial thinking and mathematics: a developmental perspective. Poster presentation.

Data from Chapters 2 and 3 were presented at the Doctoral Education Summer Conference, UCL Institute of Education, UCL, UK:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). The role of spatial thinking as a predictor of mathematics performance in primary school students. Oral presentation.

Data from Chapters 2 and 3 were presented at the Royal Society meeting on the origins of numerical abilities, London, UK:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). Finding Space for Maths: An Investigation of the Associations between Spatial and Maths Skills in the Primary School Years. Poster presentation.

Data from Chapters 2 and 3 were presented at the Centre for Educational Neuroscience Seminar Series, Birkbeck University, UK:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). Exploring the developmental relations between spatial cognition and mathematics in primary school. Oral presentation.

Data from Chapters 2 and 3 were presented at the British Psychological Society, Developmental Section Annual Conference, Belfast, UK:

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2017). Exploring associations between spatial cognition and mathematics: A developmental perspective. Poster presentation.

To my mother Lorraine,

for inspiring me with your kindness, strength and resilience

and to my father Paul,

for showing me the value of dedication, determination and hard work.

Acknowledgements

I would like to thank the Bloomsbury Scholarship Programme and the National Centre for Curriculum and Assessment, for generously funding my doctoral research. Very special thanks are due to the 500+ children who took part in this research and to all the teachers who welcomed me into their classrooms. Thank you for taking an interest in my research. I had so much fun working with you all. This thesis would not have been possible without you.

I owe a huge debt of thanks to my two supervisors, Prof. Emily Farran and Prof. Michael Thomas. I could not have asked for better mentors. Thank you for your time, patience, and guidance. I would like to particularly thank Emily for going above and beyond all my expectations of a PhD supervisor, and for making this PhD a wonderful experience. Thank you for your constant reassurance, and for giving me both guidance and freedom when I needed them most. You have been an inspirational role model. I am very excited to continue working with you. I would also like to thank Prof. Eirini Flouri for all of her help and advice during my time at UCL.

Throughout my doctorate, I was lucky enough to be surrounded by other fantastic PhD students, who I now call good friends. Thank you to Su, Nelly, Lizzie, Annie, and all other members of the CoGDeV lab and the CEN PhD group. I especially want to thank Alex Hodgkiss who has made this PhD infinitely more enjoyable. I could not ask for a better collaborator or friend. Thank you for occasionally pandering to my desires for cinema trips in the middle of the day.

This thesis would not have been possible without the love of my friends and family. I have been lucky enough to have wonderful friends who have helped and encouraged me through this thesis. I particularly want to thank Alan, Sam, Rachel, and Darina for providing motivation and distraction in equal measure. To my grandfathers, Tony Davis and Brian Gilligan who sadly did not see the completion of this thesis, thank you for always encouraging my love of learning. To my parents, Lorraine and Paul, thank you for your love and understanding during this PhD. Thank you for always enthusiastically supporting my goals and helping me to achieve them. To my sister

Aisling, thank you for being my best friend, for listening to endless PHD rants, for teaching me how to use commas, and most importantly, for always helping me to keep things in perspective with your witty humour and endless kindness.

To my fiancé Ciarán Lee. Quite simply this thesis would not have been possible without your love. Thank you for endless cups of Barry’s Tea, for always having secret stashes of chocolate buttons, and for inspiring me on countless occasions with your own passion for research. Thank you for coming on this adventure with me and for filling my life with happiness.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANOVA	Analysis of Variance
ANS	Approximate Number System
BAS II	British Ability Scales II
BPVS	British Picture Vocabulary Scale
CEBR	Centre for Economics and Business Research
CMAQ	Child Math Anxiety Questionnaire
CMTT	Child Mental Transformation Test
CEFT	Children's Embedded Figures Test
CBI	Confederation of British Industry's
CFA	Confirmatory Factor Analysis
EEG	Electroencephalogram
EFA	Exploratory Factor Analysis
FMRI	Functional Magnetic Resonance Imaging
L1	Level 1
L2	Level 2
R^2_{LIN}	Linear Model
R^2_{LOG}	Logarithmic Model
MANOVA	Multivariate Analysis of Variance
MANCOVA	Multivariate Analysis of Covariance
M4YC	Math for Young Children
MCS	Millennium Cohort Study
NFER	National Foundation for Educational Research
NHS	National Health Service
NP	Number Estimation
OFCOM	Office of Communications
OECD	Organisation for Economic Co-Operation and Development
PAE	Percentage Absolute Error
PiM	Progress in Mathematics
PISA	Programme for International Student Assessment
PN	Position Estimation

STEM	Science, Technology, Engineering and Mathematics
SES	Socio-Economic Status
SNARC	Spatial Numerical Association of Response Codes
SFON	Spontaneous Focus on Number
TIMSS	Trends in International Mathematics and Science Study
TOSA	Test of Spatial Assembly
TOL	Tolerance
VSWM	Visuo-Spatial Working Memory
WIAT	Wechsler Individual Achievement Test
WM	Working Memory

Chapter 1 Literature review and introduction to thesis

1.1 Introduction and Rationale

> "Our world is a world that exists in space, and a world without space is literally inconceivable. Given this basic truth, it is clear that living in the world requires spatial functioning of some kind." (Newcombe & Shipley, 2015)

Spatial thinking has diverse and wide-ranging applications in everyday life, from navigation which allows individuals to move around their environment, to tool use and the manipulation of objects (Newcombe, 2018). Spatial representations are required in the use of gesture, maps, diagrams, spatial language and mental images (Newcombe, 2018). Recent studies suggest that spatial skills also play an important role in Science, Technology, Engineering and Mathematics (STEM) learning (e.g., Wai, Lubinski, & Benbow, 2009). In both childhood and adolescence, spatial skills have been identified as significant longitudinal predictors of STEM outcomes (e.g., D. I. Miller & Halpern, 2013; Verdine et al., 2014). Based on these associations, spatial ability training has been proposed as a novel means of improving both spatial and STEM skills. The overarching aim of this thesis is to explore the role of spatial thinking for one important aspect of STEM achievement, mathematics performance.

Despite its importance, there is still no single well-accepted typology of spatial thinking, and consequently there are gaps in the current understanding of how and when specific spatial skills develop. Given the range of tasks for which spatial skills are necessary, there is a need to better understand when different spatial skills develop, and why individual differences in spatial thinking are observed. In addition, initial attempts to elicit transfer of gains from spatial training to mathematics in children have rendered mixed results (Cheng & Mix, 2014; Hawes, Moss, Caswell, & Poliszczuk, 2015). These mixed findings suggest that transfer of spatial training gains is specific to certain spatial and certain mathematics tasks, and that the success of training studies is developmentally (age) selective. This will be discussed further in section 1.4.

This thesis is motivated from three perspectives. First, from a theoretical standpoint there is a lack of research directed at fine-tuning the relations between spatial and mathematics skills through development. As will be further outlined in the literature review, given the proposed multi-dimensionality of spatial and mathematical cognition, it is somewhat surprising that studies to date typically focus on individual spatial sub-domains and individual aspects of mathematics. There is a need to explore the developmental relations between different spatial and mathematical sub-domains, the possible underlying causal mechanisms that underpin these relations, and developmental variations in these spatial-mathematical relationships. This foundational research is needed to further develop this field and to enable the selection of age-appropriate training targets for future interventions.

Second, from an educational perspective, integrating spatial thinking into STEM classrooms may offer a novel way of improving students' academic outcomes. Students from England typically perform poorly on mathematical space and shape domains in international assessments (Greany, Barnes, Mostafa, Pensiero, & Swensson, 2016; Jerrim & Shure, 2016). Furthermore, spatial skills are largely absent from science and mathematics curricula at both primary and secondary level education in the UK (UK Department for Education, 2013). Fine-tuning the role of spatial skills for STEM learning across development and integrating spatial thinking into STEM lessons may help to improve STEM achievement.

Third, improving STEM success is a particularly pertinent economic issue. STEM-related industries contribute over 400 billion pounds to the UK economy per year (Berressem, 2011; Centre for Economics and Business Research [CEBR], 2015) yet over 39% of firms in need of STEM employees have reported difficulties recruiting suitably qualified candidates (Confederation of British Industry's [CBI], 2013). If not addressed, it is predicted that a shortfall in the STEM workforce will cost the UK economy over 27 billion pounds per annum by 2020 (Engineering UK, 2018). Identifying the role of spatial skills for STEM learning, and developing ways of improving STEM outcomes through spatial thinking, may improve the quality of STEM graduates with knock-on improvements for the STEM industry.

If effective, spatial training interventions could offer a promising alternative to traditional attempts at improving STEM achievement, which could in turn confer both educational and economic benefits. Findings from spatial training interventions could also provide theoretical insights into the spatial-STEM relationship. The first step in designing effective interventions must be to establish a better understanding of the developmental trajectories of spatial and mathematics skills through childhood, and to elucidate the developmental relations of these constructs.

1.2 The development of spatial thinking

Spatial cognition, as described by Hart and Moore (1973), is ”the knowledge and internal or cognitive representation of the structure, entities, and relations of space; in other words, the internalised reflection and reconstruction of space in thought” (p. 248). This section outlines the current understanding of the typology and development of spatial thinking. Spatial thinking can be explored from a number of inter-related avenues. In section 1.2.1, three main theories of spatial development are outlined, i.e., Piagetian, Vygotskian and Nativist theories of development. These theories provide a framework for understanding the structure and development of spatial thought within the broader context of innate developmental starting points, and environmental and social influences. In section 1.2.2, different typologies of spatial thinking are outlined and reviewed. In this thesis the Uttal *et al.* (2013) typology of spatial thinking is used. This is also supported by Newcombe and Shipley (2015). Theoretical, neurological and behavioural evidence is presented to support the selection of this model. In section 1.2.3, current behavioural evidence on the development of spatial ability is reviewed in the context of the Uttal *et al.* (2013) typology of spatial thinking.

1.2.1 Theoretical perspectives on spatial development

Three theories have historically dominated the literature on the normative development of spatial skills. These theories reflect Piagetian, Vygotskian and Nativist perspectives respectively (Newcombe & Huttenlocher, 2003). Piaget’s interactionist theory proposes that infants are born with no knowledge of space, object permanency or occupation of space by matter (Piaget, 1951; 1952; 1954). Piaget argued that infants learn proficient spatial skills through interactions with their physical environment, and proposed that children continue to accomplish spatial milestones such as topological, projective and euclidian thinking until 9 to 10 years (Piaget & Inhelder, 1948). As outlined by Piaget, topological thinking uses concepts of proximity, order, separation and enclosure. Projective thinking develops when children begin to incorporate perspective into their understanding of spatial relationships, allowing them to perceive objects in relation to each other, accounting for vertical and horizontal relationships. Euclidian thinking involves the introduction of measurement concepts such as the length of lines or number of lines, which adds relative proportion into spatial thinking. Critics have highlighted several weaknesses in accepting Piaget's interactionist approach to spatial development, including evidence that many spatial skills and competencies develop before 10 years (Frick, Möhring, & Newcombe, 2014; Frick & Wang, 2014). Piagetian theory also fails to account for individual differences and error-making in mature spatial performance in adults (Möhring, Newcombe, & Frick, 2016). Despite its shortcomings, Piagetian theory highlights the important roles of both environmental interaction and biological maturation in the acquisition of spatial skills.

In contrast, Vygotskian theories state that spatial competencies are acquired through social interaction, language, and the social environment (Newcombe & Huttenlocher, 2003). Children are proposed to develop spatial skills through "guided participation" involving a teacher or model who has higher levels of expertise (Rogoff, 1990). The cultural transmission of symbolic systems from teachers to students is also thought to enhance spatial development (Gauvain, 1993;1995; Hutchins, 1995). While a role for cultural and social influence on childhood development seems likely, critics of this approach argue that Vygotskian models overestimate the influences of adult instruction and cultural contribution, characterising children as passive entities in their own spatial development (Newcombe & Huttenlocher, 2003).

Nativist theories propose that individuals are born with an innate spatial ability (Spelke & Newport, 1998). These theories propose that the development of spatial skills occurs through enrichment of innate neonatal starting points, or biological maturation in specific brain regions (Diamond, 1991). In support of this theory, there is evidence that children may have core modules (biological correlates) for object representation and geometric relations (Baillargeon, Spelke, & Wasserman, 1985; Kellman & Spelke, 1983; Spelke & Kinzler, 2007). Object representation is the ability to perceive and represent objects based on their spatio-temporal features (Spelke & Kinzler, 2007). An understanding of geometric relations reflects an ability to perceive the geometry of an environment, including relations such as the distance and angle between objects in a layout (Spelke & Kinzler, 2007). However, there is also conflicting evidence that does not support the existence of core modules of spatial thinking. The core module approach fails to: explain the roles of visual learning and manual exploration on spatial development, account for our ability to learn to navigate, justify why there is differing performance on spatial tasks on account of task design features, and explain training effects in reorienting experiments that are proposed to recruit the innate geometric module (Johnson, 2009; Needham, 2009; Twyman & Newcombe, 2010). There is reason to believe that innate core modules (biological correlates) are essential as cognitive starting points for spatial thinking. However, Nativist theories often under acknowledge the importance of environmental input and experience in spatial development (Carey, 1991; Newcombe, Uttal, & Sauter, 2013).

Review and criticism of these theoretical perspectives has led to the emergence of the *adaptive combination theory*, an alternative, neoconstructivist approach to understanding spatial development (Newcombe & Huttenlocher, 2000; 2006; Newcombe et al., 2013). This theory encompasses and combines the strengths of Piagetian, Vygotskian and Nativist perspectives (Newcombe & Huttenlocher, 2000). The *adaptive combination theory* supports Piaget's interactionist approach while also placing a greater importance on early cognitive starting points and the influences of cultural and social factors. This theory acknowledges that individuals are born with certain spatial abilities or may acquire these skills in the first few months of life. Environmental interaction at both physical and social levels, enables the growth and development of these skills over the first decade of a child's life (Newcombe et al., 2013). Hence, individual differences in spatial performance may be attributable to a range of biological, cognitive, genetic and/or environmental factors. Environmental factors are likely to include both universally available and variable aspects of a child's environment. For example, experiences such as interaction with solid objects and experience gained through movement, are available to most children (Newcombe & Huttenlocher, 2003). In contrast, other factors may be differentially experienced by some, but not all children, in their natural environments. That is, experiences of spatial language, building block play, map use and gesture, are likely to be beneficial to spatial development, and vary substantially across children (Newcombe & Huttenlocher, 2003).

Adaptive combination theory is supported by behavioural findings that, while normative spatial development appears to follow a somewhat universal trajectory, there are also substantial individual differences in spatial performance at all stages of development (Mix et al., 2016). This reflects the importance of a) establishing developmental trajectories of spatial thinking and b) understanding the environmental inputs that influence and modify spatial abilities across childhood. This thesis is framed using the *adaptive combination theory* of spatial development.

1.2.2 Typology of spatial thinking

Spatial cognition was first distinguished from general intelligence in the 1930s when unitary intelligence models were rendered inadequate and spatial cognition was recognised as a distinct contributor to variance in intelligence tests (Eysenck, 1939; Thurstone, 1938). Since this time, research on spatial cognition has been complicated by variations in both the spatial terminology and spatial typologies used. Attempts at defining a typology for spatial thinking have been approached from both psychometric and theoretical perspectives which has led to the emergence of many contrasting typologies (Linn & Petersen, 1985). Factor analysis studies throughout the 1950s and 60s used psychometric approaches to sub-classify spatial cognition into a series of spatial sub-components (Guilford & Lacey, 1947; Voyer, Voyer, & Bryden, 1995; Zimmerman, 1954). However, the sub-divisions generated were highly unstable, with overlap between spatial sub-categories and large differences in categories based on the inclusion or exclusion of certain spatial tasks (Carroll, 1993; Höffler, 2010; Lohman, 1988). As shown in Table 1.1, factor analysis studies led to the emergence of many contrasting models of spatial cognition, each with different definitions and classifications of spatial skills (Hegarty & Waller, 2004). Additional limitations of using factor analysis studies to establish the underlying structure of spatial thinking include the assumptions that, all participants employ similar cognitive strategies in spatial task completion, participants will continue to use the same cognitive strategy throughout completion of a given spatial task, and in a given study all sub-components of spatial ability have been represented by cognitive tasks (Hegarty & Waller, 2004). As outlined by Newcombe (2018) navigation has typically been omitted from models of spatial thinking generated by factor analysis, as traditionally navigation was a very difficult construct to assess.

Table 1.1

Spatial ability factors generated using factor analysis approaches, adapted from Hegarty and Waller (2004)

Study	**Factors Identified**	**Tests cited as typical markers for each factor**
Michael, Guilford, Fruchter, & Zimmerman, 1957	1. Spatial Visualization	-Paper Folding, Form Board
	2. Spatial Relations and Orientation	-Cube Comparisons Test, Guildford-Zimmerman Spatial Orientation, Card Rotations
	3. Kinesthetic Imagery	-Hands test
McGee, 1979	1. Spatial Visualization	-Paper Folding
	2. Spatial Orientation	-Cube Comparisons, Guildford-Zimmerman Spatial Orientation
Lohman, 1988	1. Spatial Visualization	-Paper Folding, Form Board, Cube Comparisons
	2. Spatial Relations	-Card Rotations
	3. Spatial Orientation	-Guilford-Zimmerman Spatial Orientation,
Carroll, 1993	1. Spatial Visualization	-Paper Folding, Form Board, Cube Comparisons, Guildford-Zimmerman Spatial Orientation
	2. Spatial Relations	-Card Rotations
	3. Closure Speed	-Snowy Pictures
	4. Flexibility of Closure	-Hidden Figures
	5. Perceptual Speed	-Identical Pictures
	6. Visual Memory	-Silverman-Eals visual memory task

Alternatively, typologies of spatial cognition can be derived using iterative, theoretical approaches, categorising spatial tasks based on the cognitive or perceptual processes required to complete them (Uttal et al., 2013). One such typology is Uttal *et al.'s* (2013) theoretical, top-down model of spatial skills (see also Newcombe and Shipley [2015]). As shown in Figure 1.1*,* this model is built on two fundamental theoretical distinctions. The first is between intrinsic and extrinsic representations and the second is between static and dynamic representations. Intrinsic representations are those pertaining to the size and orientation of an object, its parts and their relationships. In contrast, extrinsic representations relate to the location of an object, the relationship between objects, and the relationship between objects and their reference frames. Dynamic representations require movement such as bending, moving, folding, scaling or rotation, whilst static representations do not. By combining these two fundamental distinctions, Uttal *et al.* (2013) propose a two-by-two classification of spatial skills with four distinct sub-domains: intrinsic-static, intrinsic-dynamic, extrinsic-static and extrinsic-dynamic.

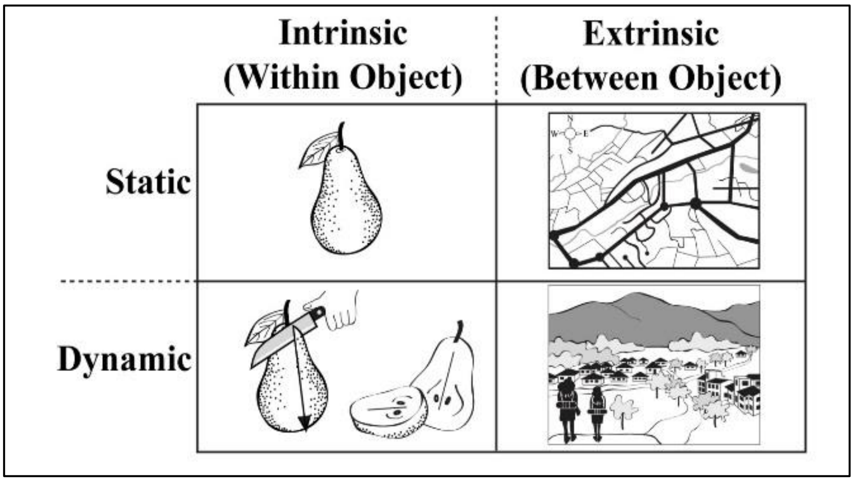

Figure 1.1. Uttal *et al.'s* (2013) two-by-two classification of spatial skills, taken from Newcombe (2018).

There is convincing theoretical, neurological and behavioural evidence to support the Uttal *et al.* (2013) model of spatial thinking (Chatterjee, 2008; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Palmer, 1978; Talmy, 2000). Although Uttal *et al.'s* (2013) model has been designed from a top-down perspective, its findings are complimentary to, and facilitate the incorporation of previous models. As shown in Table 1.2, early categorisations of spatial skills based on factor analysis studies or early theoretical models, can be mapped onto Uttal *et al.'s* (2013) spatial sub-domains. For example, Linn and Peterson (1985) outlined three theoretically driven categories of spatial thinking: spatial perception, mental rotation and spatial visualisation. As shown in Table 1.2, spatial perception and mental rotation fall within Uttal *et al.'s* (2013) extrinsic-static and intrinsic-dynamic sub-domains respectively, while spatial visualisation tasks, which are diverse in their nature, require both intrinsic-static and intrinsic-dynamic spatial skills. More extensive descriptions of the tasks listed in Table 1.2 can be found in Appendix A.

The Uttal *et al.* (2013) model also aligns with spatial models that are based on the evolutionary origins of spatial skills. From an evolutionary perspective, Newcombe (2018) proposed three kinds of spatial cognition with separate functions: navigation, tool use, and spatialisation. Tool use falls into the category of intrinsic relations, e.g., mental rotation could be used for correctly positioning stones to build a wall. Navigation can be described as an extrinsic spatial task, e.g., perspective taking to judge the field of vision of a predator. Newcombe (2018) additionally describes spatialisation as higher order tasks that use abstract spatial representations such as spatial language, gesture and sketches. These representations can be used to assist in the completion of other spatial tasks.

Neurological evidence for Uttal *et al.'s* (2013) model stems from functional Magnetic Resonance Imaging (fMRI) and Electroencephalogram (EEG) studies that highlight localisation, and processing differences between "what" (intrinsic) and "where" (extrinsic) information in the brain (Chatterjee, 2008). Differences in neural performance patterns measured using EEG have also been reported between intrinsic (mental rotation) and extrinsic (perspective taking) tasks (Christoforou, Hatzipanayioti & Avraamides, 2018). Behavioural studies also support the Uttal *et al.* (2013) model. Significant differences have been reported between object visualisers and spatial visualisers who excel at intrinsic-static and intrinsic-dynamic spatial tasks respectively (Kozhevnikov, Hegarty, & Mayer, 2002; Kozhevnikov, Kosslyn, & Shephard, 2005). Furthermore, recent confirmatory factor analysis (CFA) found evidence for a distinction between intrinsic and extrinsic spatial skills at 6 and 9 years (Mix, Hambrick, Satyam, Burgoyne, & Levine, 2018). Specifically, mental rotation, block design, figure copying and visuo-spatial working memory (VSWM) loaded onto one factor (intrinsic) and map reading, perspective taking and proportional reasoning loaded onto another factor (extrinsic). Taken together, there is convincing evidence to support the use of Uttal *et al.'s* (2013) model.

Table 1.2

Mapping of spatial categories from previous models onto the Uttal et al. (2013) model of spatial skills, adapted from Uttal et al. (2013).

Uttal et al. sub-domain (2013)	**Description**	**Examples of measures**	**Linn & Petersen (1985)**	**Carroll (1993)**
Intrinsic and static	Perceiving objects, paths, or spatial configurations amid distracting background information	Embedded Figures tasks, flexibility of closure, mazes	Spatial visualization	Visuospatial perceptual speed
Intrinsic and dynamic	Piecing together objects into more complex configurations, visualizing and mentally transforming objects, often from 2-D to 3-D, or vice versa. Rotating 2-D or 3-D objects	Form Board, Block Design, Paper Folding, Mental Cutting, Mental Rotations Test, Cube Comparison, Perdue Spatial Visualization Test, Card Rotation Test	Spatial visualization, mental rotation	Spatial visualization, spatial relations/speeded rotation
Extrinsic and static	Understanding abstract spatial principles, such as horizontal invariance or verticality	Water-Level, Water Clock, Plumb-Line, Cross-Bar, Rod and Frame Test	Spatial perception	Not included
Extrinsic and dynamic	Visualizing an environment in its entirety from a different position	Piaget's Three Mountains Task, Guildford-Zimmerman spatial orientation	Not included	Not included

The use of Uttal *et al.'s* (2013) model should be viewed in the context of its limitations. As is the case for other top-down models, it is unclear the degree to which the proposed spatial sub-divisions reflect the true cognitive (latent) structure of spatial thinking (Burgess, 2006; Mix et al., 2016). Furthermore, the use of this model is sometimes complicated by the fact that some spatial activities, including spatial tasks in the classroom, may recruit a number of Uttal *et al.'s* (2013) spatial sub-domains in combination (Okamoto, Kotsopoulos, McGarvey, & Hallowell, 2015). For example, some tasks require a series of steps such as choosing the correct size card and folding it to match a sample. A child would be required to use extrinsic-static spatial skills to scale between the various pieces of card to select the correct one. Intrinsic-dynamic spatial skills would then be required to rotate and re-orient the card and fold it correctly (Hawes, Tepylo, & Moss, 2015).

Of note, this thesis investigates spatial thinking in small-scale spaces only. Different scale spaces are defined by their perceptual and motor requirements (Broadbent, 2014). Spatial thinking in large-scale spaces is that which requires movement and observations from a number of vantage points (Broadbent, 2014; Kuipers, 1978;1982). In contrast, spatial thinking in small-scale spaces has no requirements for movement or for changing location. Although spatial thinking across small and large spaces may share processing requirements (Hegarty et al., 2006), there is also evidence of processing differences across differently sized spaces (Tversky, Morrison, Franklin, & Bryant, 1999; Zacks, Mires, Tversky, & Hazeltine, 2000). As such, this thesis investigates spatial skills in the context of small-scale spaces only, where movement and multiple vantage points are not required. Spatial navigation or spatial processing in large-scale spaces is beyond the remit of this thesis.

1.2.3 Current behavioural literature on spatial development

The existing literature on the development of spatial skills in childhood from 5 to 10 years can be reviewed in the context of Uttal *et al.'s* (2013) classification of spatial thinking.

1.2.3.1 The development of intrinsic-static spatial skills

Intrinsic-static spatial thinking requires coding of spatial features of objects, including their size and the arrangement of their parts, e.g., identifying objects as members of categories (Newcombe & Shipley, 2015). Intrinsic-static thinking is also required for carving shapes into parts (Newcombe & Shipley, 2015). The Children's Embedded Figures Task (CEFT) is the most commonly used measure of intrinsic-static spatial thinking and few other spatial tasks assess performance on this spatial sub-domain. The CEFT requires identification of the spatial configuration of one object against a distracting background (Ekstrom, French, Harman, & Dermen, 1976; Okamoto et al., 2015; Witkin & Goodenough, 1981; Witkin, Otman, Raskin, & Karp, 1971). Children have the ability to complete pre-school versions of the CEFT by 3 years, and performance on the pre-school version of this task continues to improve from 3 to 5 years (Busch, Watson, Brinkley, Howard, & Nelson, 1993).

Intrinsic-static spatial skills measured using disembedding tasks like the CEFT, show developmental progression through the primary school years (Witkin et al., 1971). Between 6 and 11 years, performance on the CEFT improves significantly with age, with significant differences in performance between all consecutive age groups (Amador-Campos & Kirchner-Nebot, 1997). However, other studies suggest that the developmental differences in CEFT performance may be subtler, with smaller between-age group effects (Guisande, Fernanda Páramo, Tinajero Vacas, & Almeida, 2007). Furthermore, notwithstanding developmental differences in CEFT performance, individuals also show substantial individual variation in disembedding skills, which continues into adulthood (Jia, Zhang, & Li, 2014). More research using new tasks is needed to better understand the development of intrinsic-static spatial skills.

1.2.3.2 The development of intrinsic-dynamic spatial skills

The majority of studies that have investigated spatial development, focus on intrinsic-dynamic spatial skills. Intrinsic-dynamic spatial skills include mental transformations like mentally rotating, folding, bending or breaking objects (Newcombe & Shipley, 2015). A considerable amount of research has focused on mental rotation, the ability to imagine rotations of an object in 2D or 3D space (Frick, Ferrara, & Newcombe, 2013). Early precursors of successful mental rotation have been reported at 16 months. Frick and Wang (2014) presented children with a toy on a turntable that was covered and rotated by 90 degrees. The turntable was uncovered to reveal that the toy had moved with the turntable (probable outcome) or had remained in the original location (improbable outcome). At 16 months, but not at 14 months, children demonstrated prolonged eye gazes for improbable outcomes. This suggests that children at 16 months expected the toy to turn with the turntable and were capable of anticipating rotational outcomes (Frick & Wang, 2014). However, children at 14 months did not show these precursors to successful mental rotation. Object fitting tasks are also useful for measuring precursors of mental rotation ability. Örnkloo & von Hofsten (2007) presented children with a box with a hole at the top (the shape of this varied by trial) and a series of objects at a perpendicular orientation to the hole (ensuring that rotation was required to fit the object into the hole). The authors recorded how children handled the objects before attempting to place them into the hole. At 14 months, children ignored the orientation of the objects when attempting to place them into the hole, at 18 months children appeared to realise that the objects required rotation but completed inaccurate rotations, while at 22 months children were largely accurate in rotating objects and fitting them into the hole (Ornkloo & von Hofsten, 2007). Successful precursors to mental rotation have also been reported in other studies. For example, at 25 months, children can rotate shapes (cylindrical oblong and square shaped oblong) to place them on top of 2D outlines, and to stack them in towers (Shutts, Ornkloo, von Hofsten, Keen, & Spelke, 2009).

In contrast to the tasks above, results from studies using more typical mental rotation paradigms with more complex shapes, and imagined rotations, show that children find mental rotation tasks challenging until approximately 5 years. For these tasks, several studies report above chance accuracy on mental rotation tasks from 5 years only (Broadbent, 2014; Dean & Harvey, 1979; Dean & Sherzer, 1982; Frick et al., 2013; Frick, Hansen, & Newcombe, 2013; Marmor, 1975; 1977; Okamoto-Barth & Call, 2008). Like adults, children show increased reaction times and increased error rates for trials at higher degrees of rotation (e.g., Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990). Some studies report above chance accuracy on mental rotation tasks at 4 years. For example, Estes (1998) reported a mean performance accuracy of 60% at 4 years on a computerised rotation-matching task, where chance performance was 50%. This was compared to 83% at 6 years (Estes, 1998). While above chance accuracy was reported at 4 years, performance remained relatively low. In another study in which children were asked to decide whether two images, including one rotated image, were matching or not, Marmor (1977) reported that 75% of children at 4 years demonstrated an adult-like linear increase in response time with increasing degrees of rotation. However, a subsequent study using a similar experimental paradigm failed to replicate these findings (Dean & Harvey, 1979). Recent findings from Frick e*t al.* (2013) also report above chance accuracy on mental rotation in some, but not all, children at 4 years (approximately 40% of children). This percentage increases to 95% at 5 years. In other mental rotation studies, even after task modifications such as providing manual or observational experience, children at 4 years do not perform above chance (Frick et al., 2013).

Similar findings have been reported for other intrinsic-dynamic spatial tasks. Performance on the Child Mental Transformation Task (CMTT) which requires mentally moving objects along diagonal lines and mentally rotating objects, shows significant age-based improvements between 4 and 7 years (Levine, Huttenlocher, Taylor, & Langrock, 1999). For mental folding, which requires imagining what an object will look like after it has been folded, there is evidence that by 5 years the majority of children demonstrate above chance performance which improves with age until it plateaus at 7 to 8 years (Harris, Newcombe, & Hirsh-Pasek, 2013). Overall, the findings indicate that precursors of successful intrinsic-dynamic spatial skills are evident in infancy. Although above chance accuracy on intrinsic-dynamic spatial tasks is reported in some studies at 4 years, in the majority of studies it appears that children are not capable of achieving above chance accuracy on intrinsic-dynamic tasks until the age of 5 years. These intrinsic-dynamic spatial skills continue to develop until at least 7 to 8 years.

1.2.3.3 The development of extrinsic-static spatial skills

Extrinsic-static spatial tasks require mapping of an object's location in relation to a reference system (Okamoto et al., 2015). Historically, horizontal and vertical invariance tasks were used to assess extrinsic-static spatial skills. For example, the Rod and Frame Test examines the ability to accurately code horizontal and vertical dimensions of a rod as defined by gravity, while ignoring the reference of a tilted frame (Newcombe & Shipley, 2015). Performance on this task gradually improves with age from 4 years until adulthood (Bagust, Docherty, Haynes, Telford, & Isableu, 2013; Haywood, Teeple, Givens, & Patterson, 1977; Witkin, Goodenough, & Karp, 1967). More recently, spatial scaling tasks have been introduced to measure extrinsic-static spatial thinking (Newcombe & Shipley, 2015). The prerequisite skills required for spatial scaling, including symbolic correspondence and metric encoding, emerge in early childhood (Huttenlocher, Newcombe & Sandberg, 1994; Newcombe, Sluzenski, & Huttenlocher, 2005; Vasilyeva and Huttenlocher, 2004). Comprehension of symbolic correspondence, or the correspondence between a model and a reference space has been reported in children as young as 3 years (DeLoache, 1987; DeLoache, 1989). At this age, children recognise that features on a map or model represent features in the real world. Metric encoding, or the ability to encode distances metrically, has been reported in infants as young as 5 months, with some infants demonstrating sensitivity to distance differences of just 20cm (Newcombe, Huttenlocher, & Learmonth, 1999; Newcombe et al., 2005). Bushnell, McKenzie, Lawrence and Connell (1995) demonstrated metric encoding in infants at 12 months showing that they 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.

Beyond these prerequisite skills, there is evidence that the ability to successfully map encoded distances between different sized spaces, i.e., spatial scaling, develops significantly between 3 and 5 years. For example, 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, at which time children's accuracy levels are broadly comparable to adult scores. No significant difference in performance between children at 5 and 6 years was reported. In a computer-based study Möhring, Newcombe and Frick (2014) demonstrated improvements in spatial scaling across different scaling factors between 4 and 5 years. Similar results have also been reported in studies using more natural environments. For example, Vasilyeva and Huttenlocher (2004) reported that 90% of children at 5 years could successfully place objects on a rectangular rug using a two-dimensional map. In comparison, at 4 years only 60% of children 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 objects are presented along a single dimension. Huttenlocher, Newcombe and Vasilyeva (1999) reported accurate spatial scaling for most children at 3 and 4 years when tested using a scaling paradigm with a single dimension; the horizontal axis. Similarly, at 4 years children can successfully use a map to locate one of three target bins, positioned along a single spatial dimension (Shusterman, Ah Lee, & Spelke, 2008). In studies requiring scaling in two dimensions, at 4 years, children find it difficult to correctly place an object in a target location within a room, based on locations learnt from a corresponding map (Uttal, 1996). Indeed, in certain spatial paradigms children up to 10 years find task completion difficult (Libens & Downs, 1993).

In summary, children as young as 3 years demonstrate symbolic correspondence and metric encoding, prerequisite skills for spatial scaling. Above chance accuracy on scaling tasks in one and two dimensions is typically evident from 3 and 5 years respectively. However, task performance is influenced by the number of spatial dimensions used and the task features of a given measure. There is limited information on spatial scaling performance in children older than 5 years.

1.2.3.4 The development of extrinsic-dynamic spatial skills

Extrinsic-dynamic spatial tasks require an understanding of the changing relations between two or more objects, or between the observer and other objects (Okamoto et al., 2015). These tasks are based on the fact that all objects have a location that can be coded using either an object-based reference frame (allocentric) or using a body-relative reference frame (egocentric) (Newcombe & Shipley, 2015). Perspective taking and other extrinsic-dynamic spatial tasks require an ability to use an allocentric reference frame, to represent a viewpoint that differs from one's own (Frick, Möhring, & Newcombe, 2014b) or to imagine observer movements (Newcombe & Frick, 2010). Perspective taking tasks are often used to measure extrinsic-dynamic spatial skills. Piaget and Inhelder (1956) first measured perspective taking with the Three Mountains Task and reported that perspective taking skills do not develop until 9 to 10 years. However, several studies have since contradicted these findings. The paragraphs below detail that the precursors to perspective taking have been reported from 2 years while more sophisticated forms of perspective taking are evident from approximately 6 years (Frick et al., 2014b).

Perspective taking is proposed to develop in two stages Level 1 (L1) and Level 2 (L2) (Flavell, Everett, Croft, & Flavell, 1981; Masangkay et al., 1974). During L1, children demonstrate precursors to successful perspective taking and understand that different standpoints give rise to different views, i.e., individuals with L1 knowledge understand that another person can see something different to themselves (Flavell, et al., 1981; Masangkay et al., 1974). L1 skills have been reported in children at 24 months (Moll & Tomasello, 2007; Sodian, Thoermer, & Metz, 2007). However, at this developmental stage, children find it difficult to imagine exactly what can be seen from a contrasting view point and can only do so in certain environmental conditions (Newcombe & Huttenlocher, 1992; 2003). During L1, children often demonstrate egocentric representations that are based on their own perspective (Newcombe & Huttenlocher, 1992). L2 capabilities in perspective taking develop at approximately 6 years, at which time children are capable of imagining a scene from an alternate perspective, i.e., using an allocentric reference frame to encode the location of an object relative to other objects (Frick et al., 2014b; Masangkay et al., 1974; Pillow & Flavell, 1986).

Perspective taking skills continue to develop and improve through childhood, with particular increases in L2 abilities between 7 and 8 years (Flavell et al., 1981; Frick et al., 2014b; Masangkay et al., 1974; Pillow & Flavell, 1986). There is evidence that it is not until 8 years that individuals fully develop the ability to integrate egocentric and allocentric reference frames and use them to successfully navigate within their surroundings (Nardini, Jones, Bedford, & Braddick, 2008). It is noteworthy that successful perspective taking at both L1 and L2 appears to be dependent on task design features including the complexity of the task, the number of objects involved, and the presence of conflicting frames of reference (Newcombe & Huttenlocher, 1992). For example, a study by Frick *et al.* (2014b) reported that increasing the number of objects used in perspective taking tasks led to reductions in task performance.

Taken together, there is evidence suggesting that L1 perspective taking skills are present from 24 months while L2 perspective taking skills develop at approximately 6 years. L2 skills continue to develop throughout childhood with particular increases in L2 abilities between 7 and 8 years (Frick et al., 2014b; Salatas & Flavell, 1976). However, there is limited evidence exploring perspective taking abilities after this age.

1.2.3.5 Summary of the development of spatial skills

Traditional Piagetian theories suggest that the skills required for the completion of spatial tasks are not evident until 10 to 11 years when children no longer hold topological views of spatial concepts (Piaget & Inhelder, 1948). In contrast to this, the literature highlighted here suggests that children show early precursors to successful spatial thinking in infancy, with marked improvements in spatial task performance between 5 and 8 years. The literature suggests that there may be subtle differences in the early developmental profiles of different spatial sub-domains. There is evidence that children demonstrate intrinsic-static spatial skills at 3 years and intrinsic-dynamic spatial skills at 4 years. For extrinsic-static spatial skills, there is above chance accuracy at approximately 5 years, while extrinsic-dynamic spatial abilities appear to emerge slightly later at approximately 6 years. Across all spatial sub-domains, there is evidence that task performance is dependent on features of task design. Unfortunately, few studies explore spatial development in later childhood, and no one study includes multiple measures of spatial thinking at consecutive developmental stages. Comparing spatial development across different sub-domains and across different studies is hindered by the different populations and testing paradigms used, thus the comparative findings outlined here should be interpreted with caution. This concern is further addressed in section 3.3.

Importantly, although this section highlighted developmental differences in spatial thinking, substantial individual differences in spatial performance are also reported across all of Uttal *et al.'s* (2013) spatial sub-domain (e.g., Liben and Downs, 1993, Newcombe & Frick, 2010). While children's spatial skills improve as they get older, factors such as environmental, biological and cultural influences may explain the large disparities in spatial performance between children of the same age. The roles of both development and individual differences in performance must be considered in any discussion of spatial thinking. These findings are consistent with the *adaptive combination theory* of spatial development. Beyond biological starting points, the findings emphasise the role of experience in the development of spatial skills. They highlight the fact that differences in personal experiences may lead to different spatial outcomes.

As outlined further in section 1.4, this thesis focuses on the important role that spatial thinking may play in educational and applied settings such as in the development of STEM skills. However, spatial thinking also has a practical significance in everyday life. The vast implications of spatial cognition are outlined by Newcombe (2018):

> "Any kind of action in a spatial world is in some sense spatial functioning, and hence can sensibly be called spatial cognition" (Newcombe, 2018, p.2)

Spatial thinking is required when driving a car, packing items into a suitcase, finding the freezer aisle at the supermarket, wrapping a gift, and playing tennis, among other examples. Spatial cognition has been identified as a sub-component of intelligence and has been reported as a distinct factor (beyond verbal and mathematical domains) in many factor analysis studies (Newcombe et al., 2013). From an evolutionary perspective, the development of advanced spatial cognition has facilitated humans in tool use, a skill that is largely unique to humans and sets us aside from other primates (Okamoto-Barth & Call, 2008). For these reasons, understanding the development of spatial thinking and factors that predict individual variation in spatial thinking, has significant practical and theoretical implications. Chapters 2 and 3 of this thesis, provide insights into the development of spatial thinking in the context of the main aim of this thesis; exploring spatial-mathematical relations. Subsequent sections outline the development of mathematical skills and explore the role of spatial thinking for mathematics outcomes.

1.3 The development of mathematical thinking

This section reviews current literature on the structure and development of mathematical thinking. As described for spatial ability, the development of mathematical thinking is explored in the context of three main theoretical perspectives; Piagetian, Vygotskian and Nativist theories (section 1.3.1). Framed in the context of these theoretical perspectives, one proposed typology of mathematical thinking, the von Aster and Shalev (2007) typology of mathematical thinking is described. Behavioural evidence on the development of mathematical skills is also outlined (section 1.3.2). This section also highlights the role of other factors in predicting mathematics outcomes including general cognitive abilities, language skills and socio-demographic factors (section 1.3.3).

Although mathematics is often taught as a single subject in schools, the domain of mathematics contains several different components and success in mathematics requires a range of skills and competencies. Distinctions between different strands of mathematics in the classroom are derived from similarities in content, and not from cognitive principles (Mix et al., 2016). Modern mathematics curricula have developed over time, shaped by economic and social influences (Newcombe & Huttenlocher, 2003). Despite cultural differences, mathematics programmes across countries, are often very similar, that is, they are based around mathematical competencies (e.g., factual knowledge of mathematical concepts, problem solving using mathematical concepts, applying mathematical concepts in novel contexts) and content areas (e.g., number and measurement, space and geometry, algebra). However, as suggested by Davis, Drefs, and Francis (2015) it is unclear whether the focus of school-based mathematics curricula today reflects the day-to-day mathematics skills required for 21^{st} century life. For example, given the steep rise of careers in STEM domains it is noteworthy that spatial thinking is typically absent from modern mathematics curricula (Davis et al., 2015). As outlined in further detail in section 1.4.1, spatial thinking has been associated with success in STEM performance, particularly mathematics performance, in several studies.

Understanding numbers is pivotal to developing more advanced skills in mathematics. However, the terms “numerical cognition” and “mathematical cognition” are often incorrectly used interchangeably. Although they are related, there is an important distinction between these terms and numerical cognition is an important prerequisite for other aspects of mathematical cognition. Numerical cognition relates to the acquisition and development of quantitative skills. In contrast, mathematical cognition relates to a wider, more comprehensive range of mathematics skills, beyond number, to other content strands such as algebra, geometry, statistics and, trigonometry, among others (Butterworth, 1999). Understanding numerosities and developing quantitative skills, are pivotal to the development of other mathematical competencies (Träff, 2013). Identifying the developmental trajectory of quantitative skills is particularly important in the study of both numerical cognition, and mathematics cognition more generally. As outlined in the next section, the first step in identifying how and why numerical (and mathematical) skills develop, is to understand the main theoretical perspectives of childhood development (Piagetian, Vygotskian and Nativist perspectives) and how they relate to mathematics.

1.3.1 Theoretical perspectives on the development of numerical cognition

As highlighted above, the development of numerical cognition will now be discussed within the context of Piagetian, Vygotskian and Nativist perspectives (Newcombe & Huttenlocher, 2003). Piaget proposed that children cannot reason quantitatively until approximately 11 years, and that quantitative abilities develop through interaction with the environment (Piaget, 1941). This includes experiences such as a stimulating home numeracy environment and early learning about numbers (Cankaya & LeFevre, 2016; Skwarchuk, Vandermaas-Peeler, & LeFevre, 2016). However, the Piagetian perspective is weakened by studies that have demonstrated quantitative abilities in infants and pre-school children younger than 11 years (Gelman & Gaillistel, 1978; Starkey & Cooper, 1980; Starkey, Spelke, & Gelman, 1990; Wynn, 1992).

Alternatively, the Nativist approach posits that individuals are born with an innate quantitative ability (Starkey, 1992). This is supported by findings that infants as young as 1 week show sensitivity to changes in numerosities (Antell & Keating, 1983). Similar findings have been reported in EEG studies and habituation studies in older infants (Izard, Sann, Spelke, & Streri, 2009; Starkey & Cooper, 1980; Xu & Spelke, 2000). Habituation studies are based on the idea that a stimulus loses novelty when it is presented repeatedly, and an individual will eventually stop responding to it (dishabituation). If an individual perceives a novel stimulus, a response is elicited showing that the individual can tell the difference between the habituated and the novel stimuli (Phelps, 2011). These habituation studies in infants suggest that infants have a concept of quantity, i.e., they perceive a difference when they are presented with different numbers of items. However, the findings of other studies suggest that infant competencies regarding quantity may be less advanced than previously believed. In one habituation study, Clearfield and Mix (2001) found that children at 6 months responded to differences in amount (area or length) and not to differences in number. Mix, Levine, and Huttenlocher (1997) found that children at 7 months are not capable of accurately detecting numerical correspondence between sounds and visual displays. Furthermore, children at 4 years, but not at 3 years, can solve nonverbal addition and subtraction questions (Levine, Jordan, & Huttenlocher, 1992). Thus, the behavioural findings supporting the Nativist approach should be interpreted with caution. The Nativist approach is also supported by biological studies highlighting a role of genetics on mathematics performance (Plomin & Kovas, 2005). However, while genetics may play a role in cognition, this does not dictate that cognitive skills cannot be shaped by experience (Fisher, 2006). Research to date indicates that while some quantitative abilities may show innate qualities, this does not preclude them from development through experience. The role of experience in development should not be underestimated. As highlighted by Newcombe (2017)

> "strong starting points are not mature ending points" (Newcombe, 2017, p.51)

Vygotskian theories highlight the role of social factors on quantitative development including cultural influences on quantitative skills (Miller & Stigler, 1987; Saxe et al., 1987). In comparison to spatial cognition, there is a large amount of research investigating the environmental factors thought to explain individual differences in quantitative skills, and mathematical skills more generally. The importance of cultural transmission in the development of quantitative skills is highlighted by the central role of cultural environment in the teaching of number symbols, i.e., the concept that the quantity 5 is linked to the written word five, the digit 5, and the verbal word five (Dehaene, 1997). The acquisition of symbolic number understanding is driven by cultural experiences of schooling and education. Further support for Vygotskian theories comes from evidence of inter-country, and indeed inter-school differences in mathematics performance (Cowan, 2015). For example, in comparison to children from the UK, superior performance is often reported on international mathematics assessments for children from East Asian countries (Greany et al., 2016). Among other reasons, this may be attributable to the fact that, in comparison to children from the UK, East Asian children learn number skills at a younger age, they spend more time learning mathematics outside school, and their parents are less likely to believe that mathematics skills are determined by biological factors (Cowan & Saxton, 2010).

Within countries, there are also social influences on mathematics performance that are reflected in between-school variations in mathematics achievement (Cowan &

Donlan, 2010; Goldhaber, Liddle, Theobald, & Walch, 2010; Gutman & Feinstein, 2008). For example, higher levels of mathematics anxiety in teachers has been associated with reduced student achievement in mathematics (Beilock, Gunderson, Ramirez, & Levine, 2010). Higher motivation and higher self-belief in mathematical ability has been reported in low ability students whose teachers reported a flexible (not fixed) mindset towards mathematical learning (Rattan, Good, & Dweck, 2012). Collaborative work between groups of students has been shown to improve mathematics achievement (Desforges & Cockburn, 1987), while no significant benefit of ability grouping in mathematics has been demonstrated (Ireson, Hallam, & Hurley, 2005). However, Vygotskian theories cannot account for evidence of quantitative abilities in very early childhood in children who have yet to encounter formal education. Vygotskian perspectives are further weakened by evidence of quantitative skills in non-human primates (Flombaum, Junge, & Hauser, 2005; Xu & Spelke, 2000).

This section has outlined evidence supporting and critiquing Piagetian, Vygotskian and Nativist approaches to numerical development. There is convincing evidence supporting each of these approaches. However, there is no reason to assume that the approaches are mutually exclusive. Therefore, in this thesis, it is proposed that quantitative development can best be understood by combining these three theoretical perspectives, similarly to the adaptive combination approach outlined for spatial development in section 1.2.1. This combined approach recognises that individuals have some innate capacity (skill) to represent number, and that these number skills develop and improve with experience. Furthermore, in contrast to spatial cognition which has largely been forgotten in the primary school classroom (Davis et al., 2015), the cultural role of schooling and education may be particularly influential for quantitative and mathematical development. Outlined in further detail in the next section, the von Aster and Shalev (2007) typology of numerical cognition was selected for use in this thesis, as it fits with the proposed theoretical approach to numerical development. Von Aster and Shalev's (2007) model proposes that the building blocks for numerical cognition may originate from innate, biological starting points (Nativist approach). However, environmental interaction (Piagetian approach) and cultural influence (Vygotskian approach) are required for the development of proficient mathematical competencies.

1.3.2 Typology and development of numerical cognition

This study adopts von Aster and Shalev's (2007) typology of numerical cognition which posits that individuals are equipped with an innate, core system for representing numbers; the approximate number system (ANS). The ANS stores approximate representations of numerical magnitude in the brain without symbols (Cordes, Gelman, Gallistel, & Whalen, 2001; Dehaene, 2011; Feigenson, Dehaene, & Spelke, 2004). These representations are believed to be stored on a mental number line (Dehaene, Bossini, & Giraux, 1993; de Hevia, Vallar, & Girelli, 2006; Le Corre & Carey, 2007). Evidence for an ANS includes findings that very young infants are capable of discriminating, representing, and remembering large numbers of items (Feigenson et al., 2004; Libertus & Brannon, 2010; Lipton & Spelke, 2003; Xu & Spelke, 2000). For ANS tasks, where participants are asked to determine the more numerous of two dot arrays, there is typically a distance effect in performance. Participants respond more accurately and faster, when the numerical distance separating two numbers is relatively large, e.g., 7 (3 vs. 10,) than when it is small, e.g., 2 (3 vs. 5) (Buckley & Gillman, 1974; Dehaene, Dupoux, & Mehler, 1990; Moyer & Landauer, 1967). These distance effects have been reported for infants, children and adults, in both behavioural and imaging studies (Butterworth, 2005; Butterworth & Varma, 2013; Girelli, Lucangeli, & Butterworth, 2000; Pinel, Dehaene, Rivière, & LeBihan, 2001; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). Individuals who have larger numerical distance effects are proposed to have less distinct representations of numerical magnitude (Holloway & Ansari, 2008). Although the ANS can be described as an innate system, this does not preclude it from change and development with experience. There is evidence that for ANS tasks, performance improves through childhood, adolescence and early adulthood until approximately 30 years (e.g., Halberda, Ly, Wilmer, Naiman, & Germine, 2012; Halberda & Feigenson, 2008; Piazza et al., 2010; Purpura & Simms, 2018). This suggests that while basic ANS abilities are innate, this system undergoes development with experience.

The symbolic number system is the way in which symbolic numerals are represented in the brain (Mussolin, Nys, Content, & Leybaert, 2014). The von Aster and Shalev (2007) model states that the core number system (the ANS) provides a foundation from which the symbolic number system develops (the ANS Mapping Account of symbolic number development). These two systems in combination, provide a platform upon which more complex mathematics abilities are established (Barth, La Mont, Lipton, & Spelke, 2005; Butterworth, 1999; Feigenson et al., 2004; Piazza, 2010). There are two tasks commonly used to measure the symbolic number system. First, symbolic number representations can be measured using symbolic comparison tasks in which participants must compare the size of two symbolic numbers (De Smedt, Noël, Gilmore, & Ansari, 2013; Gilmore, McCarthy, & Spelke, 2007). Many studies have demonstrated that performance on symbolic magnitude comparison tasks improves with age (Matejko & Ansari, 2016; Moore & Ashcraft, 2015; Vanbinst, Ceulemans, Peters, Ghesquière, & De Smedt, 2018; Xenidou-Dervou, Molenaar, Ansari, van der Schoot, & van Lieshout, 2017).

Second, symbolic number skills can be measured using symbolic number line estimation tasks (Geary, Hoard, Byrd-Craven, Nugent, & Numtee, 2007; LeFevre et al., 2010; Siegler & Opfer, 2003). For symbolic number line estimation tasks participants are required to estimate the location of symbolic numbers on a number line (the start and end points of which are clearly indicated, e.g., 0-10, 0-100) (Siegler & Opfer, 2003). Performance on tasks of this type is typically measured in two ways, first as percentage absolute error (PAE) which indicates the difference between a participant's estimate and the actual position of a number on a number line, relative to the length of the line (Siegler & Booth, 2004). Second, curve estimation compares the fit of linear models (R^2_{LIN}) and logarithmic models (R^2_{LOG}) to participants' performance using correlations between participants' estimates and the actual positions of numbers on the number line (Siegler & Opfer, 2003). Logarithmic performance patterns suggest that individuals represent smaller numbers in a spaced-out manner at the lower end of the number line, while the positions of larger numbers are condensed at the top of the line. Hence performance decreases with increasing numerical magnitude. In contrast, linear performance patterns suggest that participants spread numbers evenly along the number line. Hence, participants perform similarly regardless of numerical magnitude (Simms, Clayton, Cragg, Gilmore, & Johnson, 2016). The proximity of R^2_{LIN} scores and R^2_{LOG} scores to the value 1, indicates the degree to which a participant's estimates reflect a linear or logarithmic pattern respectively. Demonstration of linear performance patterns on number line estimation tasks reflects a more accurate representation of symbolic number. Thus, comparing whether participants' estimates are more reflective of R^2_{LIN} or R^2_{LOG} representations is one way of measuring number line estimation performance.

Regardless of the metric used, there is consistency across studies such that performance on symbolic number line estimation improves with development (e.g., Ashcraft & Moore, 2012; Friso-van den Bos et al., 2015; Moore & Ashcraft, 2015; Praet & Desoete, 2014; Sasanguie, Göbel, Moll, Smets, & Reynvoet, 2013). For example, Siegler, Thompson, and Opfer (2009) reported a transition from logarithmic to linear representations for 0-100 number line estimation tasks between kindergarten and Grade 2, and for 0-1000 number line estimation tasks between Grades 2 and 4.

The exact process by which the ANS might give rise to the symbolic number system is unclear. As outlined above, one proposal, the ANS Mapping Account, suggests that the ANS is the foundation onto which symbolic representations are subsequently mapped, giving rise to a logarithmic, and eventually linear, representation of symbolic numbers (Feigenson et al., 2004). Children are proposed to learn symbols and number words through rote-counting and map these onto the ANS (Ansari, 2008; Halberda & Feigenson, 2008; Mundy & Gilmore, 2009; Siegler & Booth, 2004; von Aster & Shalev, 2007). Following the development of symbolic number abilities, the ANS may become a checking system for symbolic arithmetic. An alternative view, the Dual Representation View, proposes that symbolic numbers are processed and acquire meaning in a fundamentally different way from non-symbolic (ANS) representations (e.g., Carey, 2004; 2009; Lyons, Ansari, & Beilock, 2012; Rips, Bloomfield, & Asmuth, 2008; Wiese, 2007). It is suggested that learning number words and symbols leads to new "exact" numerical representations with exact ordinal content information. These representations may subsequently map with ANS representations causing increased ANS precision (Mussolin et al., 2014; Pica, Lemer, Izard, & Dehaene, 2004; Verguts & Fias, 2008). Under this proposal non-symbolic foundations (the ANS) do not act as the starting point for the development of symbolic numbers (Piazza et al., 2010; Piazza, Pica, Izard, Spelke, & Dehaene, 2013).

The eventual development of the symbolic number system enables individuals to represent large numbers exactly (Carey, 2004; Dehaene, 2011; Le Corre & Carey, 2007; von Aster & Shalev, 2007). The age at which the symbolic number system develops is dependent on environmental exposure to symbolic language and symbols. Increases in symbolic task performance are seen throughout the primary school years (Mundy & Gilmore, 2009; Sekuler & Mierkiewicz, 1977; Vanbinst, Ceulemans, Peters, Ghesquière, & De Smedt, 2018). This is mirrored by differences in brain activation patterns for symbolic number tasks for individuals of differing ages (Butterworth & Varma, 2013). Areas of the prefrontal cortex are typically activated during the completion of symbolic number tasks in children, while the intraparietal sulcus is activated when adults complete similar tasks (Ansari, Garcia, Lucas, Hamon, & Dhital, 2005). These changes may reflect a shift from processing and problem solving during symbolic number tasks in childhood, to automatic memory retrieval, symbol processing and magnitude processing in the completion of symbolic number tasks in adulthood (Butterworth & Varma, 2013). Like the ANS, these developmental differences highlight the capacity for change in the symbolic number system with increasing age.

In combination, it is proposed that basic number abilities including ANS skills and symbolic number skills act as a platform for the development of more complex mathematical skills such as multi-digit calculation, word problems, algebra, measurement and data handling skills (Feigenson, Libertus, & Halberda, 2013; Träff, 2013). This is supported by evidence that basic number abilities are longitudinal predictors of later mathematics achievement. Significant concurrent and longitudinal associations have been reported between symbolic number representations (using both symbolic number comparison and number line estimation tasks) and mathematics outcomes in several studies (De Smedt et al., 2013; Friso-van den Bos et al., 2015; Muldoon, Towse, Simms, Perra, & Menzies, 2013; Price & Fuchs, 2016; Sasanguie et al., 2013; Schneider et al., 2017; Schneider et al., 2018; Xenidou-Dervou et al., 2017). Intervention studies have also shown that training in symbolic number skills leads to gains in other mathematical domains (Honoré & Noël, 2016; Obersteiner, Reiss, & Ufer, 2013; Van Herwegen, Costa, Nicholson, & Donlan, 2018). As stated by DeSmedt *et al.* (2013) this suggests a causal relationship between symbolic number skills and school- relevant mathematical competencies.

As outlined in a review by De Smedt *et al.* (2013) there is mixed evidence on the existence of associations between the ANS and mathematics achievement. Several meta-analyses have found significant correlations between the ANS and mathematics performance (Chen & Li, 2014; Fazio, Bailey, Thompson, & Siegler, 2014; Halberda, Mazzocco, & Feigenson, 2008; Mazzocco, Feigenson, & Halberda, 2011; Schneider et al., 2017). However, these results are not uncontested and in other studies no associations between the ANS and mathematics outcomes have been found (Holloway & Ansari, 2008; Lyons, Price, Vaessen, Blomert, & Ansari, 2014). One proposed explanation of these conflicting findings is that the relationship between the ANS and mathematics performance is sensitive to variations in participant ages, the ANS and mathematics tasks used, and stimuli employed (De Smedt et al., 2013; Schneider et al., 2017).

Although they are not adopted in this thesis, other slightly adapted, similar models have also been proposed for the development of numerical cognition. For example, in contrast to an ANS, Butterworth (1999; 2010) suggests that quantities are represented exactly in the ANS and not as approximate representations. This model, the Numerosity Coding Theory, suggests that acquisition of the symbolic number system is not required for representing and manipulating exact numerosities (Butterworth, 1999; 2010). Others have argued that an ANS does not exist at all, but, that infants reason about number by opening “object files” for each new object seen (Hauser & Carey, 1998; Simon, 1997). In this model, nonverbal calculation is thought to develop through the maintenance and manipulation of mental models of objects (Huttenlocher et al., 1994). Despite their differences, models explaining the development of numerical cognition show similarities in that they each demonstrate developmental transitions from limited representations to exact knowledge of quantities and number (Newcombe & Huttenlocher, 2003). Furthermore, for each of these models, developmental transitions are dependent on both biological starting points and environmental influences including the cultural environment (Newcombe & Huttenlocher, 2003).

The von Aster and Shalev (2007) typology provides one model with which mathematical development can be explored. In this section, evidence was presented supporting the use of this typology of mathematical thinking. Evidence was presented, supporting the idea that the ANS and symbolic number skills provide a platform for the development of more complex mathematical skills. For example, there is evidence that both ANS and symbolic number skills are predictors of later mathematics achievement. Evidence was also presented that number skills improve and develop with experience. However, as outlined in the next section, beyond developmental differences, a range of other factors, have also been proposed to influence mathematical performance.

1.3.3 Predictors of individual variation in mathematics skills

Although there is evidence that mathematical skills develop with age, there is also substantial variation in children's individual mathematical abilities within age groups (e.g., Friso-van den Bos et al., 2015). Cockcroft (1982) suggested that there might be as much as a seven-year-difference in mathematics skills in children at 11 years. Understanding the causes of individual variability in mathematics is pivotal to finding ways of improving children's mathematical outcomes. Beyond genetics, this variation may be attributable to cognitive, demographic or dispositional factors.

Mathematics is a multi-dimensional skill that requires several general cognitive abilities other than numerical skills alone. Beyond the role of spatial abilities which will be discussed in the next section, success in mathematics has been associated with higher scores in measures of general cognitive ability (von Aster & Shalev, 2007), working memory (Alloway & Alloway, 2010; Andersson, 2006; Bull, Espy & Wiese, 2008; Passolunghi, Mammarella & Altoe, 2008; Raghubar, Barnes, & Hecht, 2010), executive functioning (Cragg & Gilmore, 2014; Cragg, Keeble, Richardson, Roome, &

Gilmore, 2017; Fuchs et al., 2010; Purpura, Schmitt, & Ganley, 2017; Verdine et al., 2014), reasoning (Nunes et al., 2007), processing speed (Berg, 2008), and attention (Merrell & Tymms, 2001). Early language skills including expressive and receptive language have also been associated with success in mathematics (Cowan, 2015). Findings from Le Fevre *et al.* (2010) show that linguistic measures are a reliable early predictor of achievement in mathematics, while Moll, Snowling, Göbel, and Hulme (2015) reported that individuals with language difficulties or reading problems also demonstrate poor performance in mathematical achievement tests. However, while language might play a distinct role in mathematical development, correlations between numeracy and literacy achievement may also reflect the presence of an underlying general intelligence or "g" factor (Alloway & Alloway, 2010; Mayes, Calhoun, Bixler, & Zimmerman, 2009).

Differences in mathematical performance have also been associated with social and demographic factors including socio-economic status (SES) (Byrnes & Wasik, 2009), gender (Halpern et al., 2007) and ethnicity (Sonnenschein & Galindo, 2015). Children from low SES backgrounds typically perform less favourably on mathematical measures when compared to their higher SES counterparts (Byrnes & Wasik, 2009; Oakes, 2005; Sirin, 2005). These differences continue into adolescence with lower SES schools having lower mathematics achievement than higher SES schools (McConney & Perry, 2010). Based on the findings of a meta-analysis, Banerjee (2016) outlined several underlying factors that may explain the reduced performance of individuals from lower SES backgrounds. The lack of a positive environment, negative attitudes towards school and learning, and a lack of support from teachers and schools were all outlined as possible factors (Banerjee, 2016).

Evidence for gender differences in mathematics achievement is less well supported, and many studies argue against gender differences in this domain. In a meta-analysis of gender differences in mathematics in adolescence, Lindberg, Hyde, Petersen and Linn (2010) found no significant gender difference in mathematics performance, and, on average, the effect sizes reported were small ($.05 < d < .07$). Similarly, in childhood populations, Hyde *et al.* (2008) reported that the average effect size for gender differences in standardised mathematics performance was small (based on over 7 million US students aged 7 to 16 years) ($d < 0.1$). Where significant differences were reported, the differences often favoured females. In another large study of 1391 participants, no significant gender differences were reported by Hutchinson, Lyons and Ansari (2017) between 6 and 13 years, on a range of numerical tasks. However, the trend of a female advantage in mathematics is mirrored in other studies (Robinson & Lubienski, 2011). Overall, recent research contradicts historical views of a male advantage in mathematics and suggests that where male performance in mathematics is higher, the size of performance differences is small. Possible explanations for gender differences in mathematics include suggestions that they are attributable to variations in interests, neurological, or cognitive outcomes, which are in turn shaped by biological, genetic and environmental influences (Halpern et al., 2007; Penner & Paret, 2008).

Finally, there is relatively less information on differences in mathematics performance across ethnic groups. In a US-based study by Hall, Davis, Bolen and Chia (2010), scores for mathematical-concepts and mathematical-computation were lower for Black, compared to White students (10 and 13 years). In younger children from the US, significant differences in mathematics performance were also reported at 4 years (Sonnenschein & Galindo, 2015). These differences favoured White compared to Black and Latino groups. One proposed explanation is that these differences are attributable to variations in the early home numeracy environment across ethnic groups (Brooks-Gunn & Markman, 2005). This may reflect cultural distinctions, socioeconomic differences or a mismatch between the culture of parents and the school system that their child is a part of (Sonnenschein & Galindo, 2015). In the UK, evidence on mathematics performance differences across ethnic groups is limited. Statistics from the UK Department for Education (2017) show that on average 75% of students aged 10 to 11 years met the expected standard of mathematics by the end of Key Stage 2, however, this percentage differed across ethnic groups (Chinese 92%; Black Caribbean 67%; White British 62%; White Irish 80%; Indian 86%) (UK Department for Education, 2017). Given that these results are based on one specific age group, there is a need to investigate the role of ethnicity in mathematics outcomes across different developmental ages.

To summarise, there is convincing evidence that cognitive, demographic and dispositional factors play a role in the development of mathematical skills. While the primary focus of this thesis is to delineate the relationship between spatial thinking and mathematics performance, the findings need to be considered in the context of other known predictors of mathematics achievement. This is discussed in the next section, with reference to spatial cognition.

1.4 Spatial ability and success in mathematics

1.4.1 Evidence for spatial- mathematical relations

1.4.1.1 Adult and adolescent studies

Spatial ability has been identified as a reliable predictor of STEM outcomes in many large-scale longitudinal studies (*N* > 500), following both normative and intellectually gifted populations through adolescence and adulthood (Shea, Lubinski, & Benbow, 2001; Wai et al., 2009). Talent Search participants are young people from the United States who qualify for special educational programmes due to high performance on college entrance exams at a young age (Wai et al., 2009). Even after controlling for quantitative and verbal skills, longitudinal studies of Talent Search participants have reported significant correlations between high spatial ability scores (intrinsic-dynamic spatial skills) at 13 years and later STEM outcomes (Shea et al., 2001). The STEM outcomes measured included: a preference for mathematics as a high school subject at 18 years, achievement of undergraduate and graduate degrees in STEM measured at 23 years, and future careers in STEM domains relative to careers in the humanities measured at 33 years (Shea et al., 2001). Similar findings have been reported in studies of non-gifted students. It has been reported that those who pursue STEM careers and complete STEM degrees at both undergraduate and masters level have higher spatial ability scores at 13 years (Wai et al., 2009). The spatial ability measure used in these studies was a composite of performance across a range of spatial tasks, predominantly targeting intrinsic-dynamic spatial skills.

This pattern of associations between spatial thinking and STEM outcomes in adults is mirrored in cross-sectional studies. Spatial ability has been implicated as an important factor in undergraduate science success, medicine, dentistry, and engineering performance (Hegarty, 2014; Hegarty, Keehner, Cohen, Montello, & Lippa, 2007; Hegarty, Keehner, Khooshabeh, & Montello, 2009; Uttal, Miller, & Newcombe, 2013). Mental rotation skills (intrinsic-dynamic spatial skills) have been associated with undergraduate students' abilities to translate organic chemistry diagrams (Stull, Hegarty, Dixon, & Stieff, 2012). For physics, spatial visualisation skills (intrinsic-dynamic sub-domain) are significantly correlated with mechanics problem solving (Kozhevnikov & Thornton, 2006) while for engineering, mental rotation skills (intrinsic-dynamic sub-domain) are significantly associated with an individual's efficiency in learning to use computer aided design software (Sorby & Baartmans, 2000).

More specifically for mathematics, mental rotation skills (intrinsic-dynamic sub-domain) has been associated with mathematical performance in adults using number line estimation and magnitude comparison tasks (Thompson, Nuerk, Moeller, & Cohen Kadosh, 2013). Similarly, in adolescents intrinsic-dynamic spatial skills are significantly correlated with mental arithmetic and problem solving at 15 to 16 years (Reuhkala, 2001), geometry performance at 13 years (Delgado & Prieto, 2004), and mathematical word problems at 12 years (Hegarty & Kozhevnikov, 1999). Neuroimaging findings suggest that these spatial-mathematical associations may be attributable to shared processing requirements for spatial and mathematical tasks. There is evidence that overlapping circuits in the parietal lobe are activated in the completion of both number, and spatial tasks (Cutini, Scarpa, Scatturin, Dell'Acqua, & Zorzi, 2014; Hubbard, Piazza, Pinel, & Dehaene, 2005; Winter, Matlock, Shaki, & Fischer, 2015).

Overall, the evidence from longitudinal, cross-sectional and neuroimaging studies of adults and adolescents supports the existence of associations between spatial thinking and STEM domains; in particular the mathematics domain. However, as outlined in sections 1.2 and 1.3, both spatial and mathematical skills undergo significant development in childhood (before 13 years). Therefore, it is important to establish whether spatial-mathematical relations are present at all stages of development, or whether they emerge when mature performance levels are reached. The next section reviews the evidence for spatial-mathematical associations in childhood populations.

1.4.1.2 Longitudinal studies in childhood populations

Longitudinal studies in childhood populations have also measured the associations between spatial skills and mathematics. These studies predominantly focus on the predictive role of pre-school spatial skills. Verdine *et al.* (2014) reported that spatial skills at 3 years, assessed using the Test of Spatial Assembly (TOSA), a measure of intrinsic-dynamic spatial ability, predicted a significant 27% of the variation in mathematical problem solving, measured using the Wechsler Individual Achievement Test (WIAT) at 4 years. Similarly, a preliminary report from Farmer *et al.* (2013) indicated that spatial performance on the TOSA at 3 years is significantly correlated with a combined mathematics measure, at 5 years. Wolfgang, Stannard, and Jones (2001) demonstrated that spatial play in the pre-school years, in particular adaptiveness and integration in block play, is associated with mathematics achievement at 12 years. However, these results should be interpreted cautiously as interpretation of free block play is subjective and subject to errors. Furthermore, block play does not exclusively measure spatial thinking as it is influenced by a range of cognitive skills including attention and executive functions (Wolfgang et al., 2001). In a study of primary school children, Gunderson, Ramirez, Beilock, and Levine (2012) reported that performance on the Thurstone Mental Rotation Task (intrinsic-dynamic sub-domain) at 7 years, predicted improvement in number line estimation 6 months later. Gunderson *et al.* (2012) extended these results to show that performance on the CMTT, also a measure of intrinsic-dynamic spatial skills, at 5 years was predictive of approximate symbolic calculation at 8 years. These results were found to be mediated by number line estimation scores at 6 years.

As seen for adult studies, a majority of longitudinal studies that have explored spatial-mathematical associations in children, measure spatial skills in the intrinsic-dynamic spatial sub-domain. However, there is also some evidence that these associations hold for other spatial sub-domains and mathematics. This suggests that the association between spatial ability and mathematics competence is wide-ranging.

Performance on a spatial relations task, which required input from both intrinsic-static and intrinsic-dynamic spatial sub-domains, was found at 3 years to be a significant predictor of arithmetic at 10 years (Zhang et al., 2014). Similarly, a composite measure of spatial skills, assessing both intrinsic-static and intrinsic-dynamic sub-domains at 7 years, significantly predicted mathematics achievement at 10 years (Carr et al., 2017). Casey *et al.* (2015) reported that spatial skills in girls, assessed using a composite measure generated from block design (intrinsic-dynamic spatial sub-domain) and mental transformation tasks (intrinsic-static and intrinsic-dynamic spatial sub-domains), at 7 years were a significant predictor of mathematics reasoning at 11 years. Longitudinal studies of primary school students have also reported correlations between visuospatial skills, including visual perception and motor integration at 6 years, and mathematics achievement at 9 years. However, these findings were confounded by the visual and motor demands of the tasks used (Lachance & Mazzocco, 2006; Mazzocco & Myers, 2003).

Overall, there is evidence that spatial abilities in the pre-school years, particularly intrinsic-dynamic spatial skills, are associated with later mathematics performance. In older children there is evidence that general spatial abilities in the primary school years are associated with later mathematics outcomes at 9 to 11 years. However, because most studies of primary school aged children use spatial composite scores, it is unclear which spatial sub-domains drive associations between spatial and mathematical performance in middle childhood.

1.4.1.3 Cross-sectional studies in childhood populations

Further insights into spatial-mathematical relations can be obtained from cross-sectional studies in primary school populations (from 5 to 10 years). Significant correlations have been reported between mental rotation (an intrinsic-dynamic spatial skill) and both calculation and arithmetic in children aged 6 to 8 years (Cheng & Mix, 2014; Hawes et al., 2015). For other intrinsic spatial tasks including disembedding (an intrinsic-static spatial skill) and spatial visualisation (an intrinsic-dynamic spatial skill), performance has been associated with a range of mathematics achievement measures at 10 and 11.5 years respectively ($.37 < r < .42$) (Tosto et al., 2014). Performance on mental rotation and disembedding tasks (intrinsic-static and intrinsic-dynamic sub-domains) and VSWM, was also identified as a significant predictor of standardised mathematics achievement (measured using the WIAT) at 8 to 10 years (Simms et al., 2016). In contrast, Carr, Steiner, Kyser, and Biddlecomb (2008) reported no significant association between mental rotation (an intrinsic-dynamic spatial skill) and standardised mathematics performance at 7 years.

Mix *et al.* (2016; 2017) have completed the most extensive cross-sectional research to date on spatial and mathematical thinking in the primary school years. In both initial exploratory factor analysis (EFA) (2016) and follow-up CFA (2017) studies, Mix *et al.* found that, although spatial and mathematics tasks are highly correlated, they form distinct factors (Mix et al., 2016; 2017). By comparing children of differing ages on the same spatial and mathematics tasks, Mix *et al.* (2016; 2017) have provided important preliminary evidence that there are distinct relations between individual spatial sub-domains and specific aspects of mathematics performance, and that these relations vary with age. More specifically, mental rotation (an intrinsic-dynamic spatial skill) was a significant predictor of mathematics (a general mathematics factor derived from performance on a range of mathematics measures) at 6 years only, while VSWM was a significant predictor at 11 years only. VSWM was measured using a spatial location memory task. No spatial predictors were identified for mathematics at 9 years. These findings suggest that associations between spatial thinking and mathematics in the primary school years may not be limited to the intrinsic-dynamic spatial domain. However, of note, some cross-factor loadings were not replicated across both the EFA and CFA studies. These inconsistencies suggest that there is instability of cross-factor loadings across different populations, which weakens the generalisability of the results. Thus, the findings should be interpreted cautiously (Mix et al., 2016; 2017).

In summary, current literature supports the organisation of spatial and mathematics domains as two distinct factors, with some cross-factor loadings. Cross-sectional studies provide evidence that different sub-domains of spatial thinking and different aspects of mathematics are differentially associated. That is, not all spatial and mathematics skills are associated to the same degree. Furthermore, there is evidence that the relationship between spatial and mathematical skills changes with development. Associations between some spatial and mathematics skills are present at specific developmental stages only. However, no known study investigates the role of different spatial sub-domains for mathematics, at different developmental ages in primary school.

These findings across both longitudinal and cross-sectional studies in children highlight a need to further elucidate the specificity of spatial-mathematical relationships across different tasks and skills. In particular, there is limited research on spatial-mathematical relations across the primary school years. Elucidating these relations in primary school children is important as there is evidence that the relationship between spatial skills and mathematics is sensitive to developmental age.

1.4.2 Explaining associations between spatial and mathematics skills

There is a need to move beyond the question of “whether” to “why” significant correlations are repeatedly reported between mathematical and spatial constructs. Using evidence from longitudinal studies, Bailey (2017) presented a convincing argument for a causal effect of spatial skills on mathematics in the pre-school years. However, these findings are not definitive and cannot easily be translated to older children. Understanding the causal relationship and underlying explanations for spatial-mathematical associations, is key to integrating spatial learning into the mathematics classroom and to developing successful classroom interventions (Clements & Sarama, 2004; Hawes, Tepylo & Moss, 2015; Mix & Cheng, 2012).

Findings on spatial-mathematical relations do not support a simple linear coupling between spatial and mathematical cognition (Fias & Bonato, 2018). Instead, it has been proposed that several different explanations underpin spatial-mathematical associations, depending on the mathematical and spatial sub-domains assessed (Fias & Bonato, 2018). Historically the Mental Number Line, or the idea that numbers are represented spatially in the brain, was proposed to explain observed associations between spatial and mathematical constructs (Barsalou, 2008; Lakoff & Núñez, 2000). The Spatial-Numerical Association of Response Codes (SNARC) effect, thought to reflect the presence of the Mental Number Line, has been demonstrated in several studies. For example, individuals are faster to respond to small numbers with their left hand and larger numbers with their right hand, suggesting that small numbers are spatially represented to the left and larger numbers are represented to the right in the brain (Dehaene et al., 1993). Findings from arithmetic-based studies show that individuals typically overestimate addition results (right-side-of-space bias) and underestimate subtraction results (left-side-of-space bias) (Fischer & Shaki, 2014; Werner & Raab, 2014). In a similar way arithmetic performance is also influenced by the spatial presentation of equations and numbers (Fisher, Borchert, & Bassok, 2011; Landy & Goldstone, 2007; McNeil & Alibali, 2004). However, accepting the Mental Number Line as the driver of all spatial-mathematical relations is inconsistent with the differential associations observed between certain spatial and mathematical sub-domains, reported by Mix *et al.* (2016; 2017) among others. Instead, it has been suggested that not all associations between spatial and mathematical tasks can be explained in the same way, and a range of other explanations have subsequently been proposed as theoretical accounts for specific spatial-mathematical relations.

First, it has been proposed that extrinsic-static spatial tasks, particularly spatial scaling tasks, rely on intensive quantification skills, or proportional reasoning (Newcombe, Möhring, & Frick, 2018). Magnitude can be encoded using two different quantification systems, an extensive system (using absolute amounts) or an intensive system (using proportions or ratios). Accurate spatial scaling between two different sized spaces requires the intensive coding strategy, with proportional mapping of relative, not absolute, distances. In mathematics, similar proportional mapping between extensive discrete representations of numbers to continuous intensive representations, is required for number line estimation and reasoning about formal fractions (Möhring, Newcombe, Levine, & Frick, 2016; Rouder & Geary, 2014). Theoretically, ANS tasks may also require proportional reasoning to facilitate ordinal comparisons, while performance on some geometry, area and distance tasks also rely on proportional and not absolute judgements (Barth & Paladino, 2011; Dehaene, Piazza, Pinel, & Cohen, 2003). In support of this, at 4 to 5 years significant correlations between spatial scaling performance and proportional reasoning performance (identification of the strength of flavour of different combinations of cherry juice and water) have been reported (Möhring, Newcombe, & Frick, 2015). Taken together, extrinsic-static spatial task performance is expected to correlate with mathematics tasks that rely on intensive quantity processing or proportional reasoning.

Second, for intrinsic-dynamic and extrinsic-dynamic spatial tasks, active processing including mental visualisation and manipulation of objects in space is thought to be required for successful task completion (Lourenco, Cheung, & Aulet, 2018; Mix et al., 2016). It is postulated that the generation of mental models allows individuals to visualise not only individual components of problems but also the relations between parts (Lourenco et al., 2018). Theoretically, in mathematics, individuals may use mental visualisations to represent and solve complex mathematical word problems, e.g., by keeping terms together and structuring order of operations tasks, or to represent and organise complex mathematical relationships such as multi-digit numbers (Huttenlocher, Jordan, & Levine, 1994; Laski et al., 2013; Thompson et al., 2013). Mental visualisations may also be used to ground abstract concepts. For example, in missing term problems of the format 4 + __ = 5, individuals may use visualisations of blocks or other concrete objects to balance the equation presented (Lourenco et al., 2018). Dynamic spatial tasks are thus expected to correlate with mathematical tasks requiring the mental manipulation or organisation of numbers.

Third, intrinsic-static spatial tasks are reliant on form perception; the ability to distinguish shapes from a more complex background or to break more complex pictures into parts (Mix et al., 2016). Form perception is theoretically useful for spatial tasks such as map reading and figure drawing (Newcombe & Shipley, 2015), and for mathematics tasks such as distinguishing symbols such as + and × symbols, interpreting charts and graphs, and accurately completing multistep calculations which require an understanding of the spatial relations between symbols (Landy & Goldstone, 2007; 2010; Mix et al., 2016). As such, intrinsic-static spatial skills are predicted to relate to mathematics tasks that require identification and use of symbols or visual aids.

Proposing theoretical explanations for associations between spatial and mathematical skills is further complicated by the role of developmental age. As outlined in the previous section, the relationship between spatial skills and mathematics appears to show sensitivity to developmental age. One explanation for this is that the role of spatial skills as predictors of mathematics may be greater for novel tasks compared to practiced, automatic mathematics skills (Ackerman, 1988; Uttal & Cohen, 2012; Young, Levine, & Mix, 2018). Spatial skills may provide scaffolding when students are faced with new mathematics material but may play a reduced role as mathematics skills become increasingly automatic or familiar (Mix et al., 2016). Alternatively, age-based differences in spatial-mathematical associations may be due to variations in the mathematical content that children are exposed to across school years (Mix et al., 2016). For example, in the early school years spatial scaling may be associated with number line estimation skills, but scaling performance is unlikely to correlate with performance on complex equations, to which children are exposed to in later school years.

In this section, it has been outlined that not all associations between spatial and mathematical tasks can be explained in the same way and a range of explanations have been proposed as theoretical accounts for specific spatial-mathematical relations. However, it is noteworthy that these accounts are based on theoretical predictions and, to date, there is limited evidence exploring the specificity of spatial-mathematical associations in primary school children in the context of these mechanistic accounts. Furthermore, this list of possible underlying mechanisms put forward to explain spatial-mathematical associations is not exhaustive and there may be additional explanations for other spatial-mathematical relations that have yet to be identified. Further research is needed to corroborate and refine the proposed explanations for spatial-mathematical relations, which considers the specificity and developmental sensitivity of these associations.

1.5 Cognitive training

Moving beyond associational studies, cognitive training offers a method of exploring the direction of causality between different cognitive skills. The current literature on the malleability of spatial thinking and the evidence that spatial training can foster improvements in mathematics is important to consider and critically assess. In addition to theoretical implications, there are significant educational implications of identifying effective training paradigms that render gains across spatial and mathematical skills. Such training could have direct benefits to student learning in the mathematics classroom.

1.5.1 Training spatial skills

There is a large body of evidence supporting the malleability of spatial thinking through intervention (Uttal et al., 2013). Baenninger and Newcombe (1989) were the first to classify spatial training paradigms into two types: direct and indirect training. Direct training paradigms involve task-specific spatial training, with training being provided on test items relevant to the spatial skill or range of spatial measures being assessed. For example, Lizarraga and Ganuza (2003) reported gains in mental rotation performance after 12 weeks of training with mental rotation practice worksheets and experience manipulating cubes (d = 0.788), compared to a control group who received no intervention. One limitation of direct training is that it is difficult to distinguish training gains from practice effects. For indirect spatial training, it is proposed that participants' exposure to spatially rich experiences increases their subsequent spatial task performance. However, indirect training paradigms do not include experiences that are directly related to a specific spatial task (Baenninger & Newcombe, 1989). In one example of indirect spatial training, Blüchel, Lehmann, Kellner, and Jansen (2013) reported significant gains in spatial (mental rotation) performance at 9 years following a two-week motor training programme, that was not targeted towards improving spatial thinking skills. The programme involved training in motor skills such as catching, juggling and bouncing balls, but did not have a direct spatial training component. An extended conceptualisation of spatial training was proposed by Uttal *et al.* (2013) who expanded Baenninger and Newcombe's (1989) description of indirect training to include two distinct forms: video game training in which training is delivered using video games and course-based training in which participants are exposed to a semester long spatially relevant course (Uttal et al., 2013). Like Baenninger and Newcombe (1989), Uttal *et al.* (2013) also recognised direct training as spatial task training involving task practice, strategic instruction or computer-based lessons.

The malleability of spatial thinking has been well summarised in two meta-analyses (Baenninger & Newcome, 1989; Uttal et al., 2013). First, across 26 studies of spatial training, significantly larger effect sizes were reported for groups who received direct spatial training, compared to control groups who completed no training (Baenninger & Newcombe, 1989). In contrast, no significant spatial gains were reported for groups receiving indirect spatial training compared to controls. However, these findings should be interpreted in the context of the small number of studies that investigated indirect training (n = 2 studies). More recently, Uttal *et al.* (2013) completed a second extensive meta-analysis investigating the malleability of spatial thinking (N = 217 studies). An effect size of almost one half a standard deviation was reported for spatial training compared to control conditions (*Hedges G* = 0.47) (Uttal et al., 2013). Unlike Baenninger and Newcombe (1989) no differences in effect size were reported for different types of spatial training. Uttal *et al.* (2013) also explored the durability and transferability of spatial training effects. Some studies administer post-testing immediately following training while others wait days, weeks, or even months until post-testing (Uttal et al., 2013). However, Uttal *et al.* (2013) found no significant difference in the size of training gains reported, based on the timing of post testing. This suggests that spatial gains achieved through training are durable. Uttal *et al.* (2013) also investigated gains in novel task performance following training, i.e., gains in tasks/skills that had not been trained. Gains with an effect size of 0.48 (*Hedges G*) were reported for novel tasks after spatial training. This is convincing evidence that spatial training transfers to other untrained skills.

Of the 217 studies included in Uttal *et al.'s* (2013) meta-analysis, 26% were completed with children. The average effect size for studies of children under 13 years was 0.61 (*Hedges G*), higher than the effect size for all older age groups. As proposed by Heckman and Masterov (2007) this may be because cognition is particularly malleable in childhood. Of the child-based studies reviewed by Uttal *et al.* (2013), 66% included direct spatial task training. A measure of intrinsic-dynamic spatial skills was included in 68% of studies; an intrinsic-static spatial task was used in 28% of studies, while only 7.5% and 11% included an extrinsic-static or extrinsic-dynamic spatial task respectively. These findings suggest that spatial thinking is malleable, and that spatial training can lead to large gains in spatial performance. However, there is a clear bias in the training methods typically employed in spatial training studies involving children, with a disproportionate emphasis on intrinsic-dynamic spatial skills.

1.5.2 Evidence of transfer of spatial training gains to mathematics

Despite the malleability of spatial skills and the known associations between spatial ability and mathematics competence, only two known studies have investigated transfer of spatial training gains to mathematics outcomes in children, using spatial training in which there is no mathematical component in the training paradigm. The findings of these studies are inconsistent. Cheng and Mix (2014) reported significant gains in mathematical calculation following a single 40-minute mental rotation training session (intrinsic-dynamic spatial skill) in children aged 6 to 8 years. Gains were specific to missing term arithmetic problems, e.g., 4 + __ = 9, and no similar improvements were reported for children in the control condition who completed crossword puzzles. In a similar study also using mental rotation training in children aged 6-8 years, Hawes *et al*. (2015) failed to replicate these findings. Here, participants completed 15 sessions of computerised mental rotation training (intervention) or literacy training (control) respectively. Despite improvements in mental rotation and mental transformation (an untrained spatial skill) for the intervention group, Hawes *et al.* (2015) did not report improvements in mathematics measured using nonverbal arithmetic and missing term arithmetic problems.

The inconsistencies between the findings reported in these two studies may be explained by several factors. First, individual training was delivered in the Cheng and Mix (2012) paradigm, while Hawes *et al.* (2015) administered group training in a classroom setting. Gains reported by Cheng and Mix (2012) may therefore be attributable to the motivational benefits of one-to-one interaction with a researcher. Without the direct supervision of a researcher, it is unclear to what degree participant motivation and engagement with training may have influenced outcomes in the Hawes *et al.* (2015) paradigm. Second, the timing of post-testing differed between the studies. Cheng and Mix (2012) delivered post-testing immediately following training, while Hawes *et al.* (2015) administered post-testing one week after training. There is no guarantee that the gains reported by Cheng and Mix (2012) are durable, and instead they may reflect a priming effect. Thus, the timing of post-testing may have influenced the results of Cheng and Mix (2012). Third, the training modes differed somewhat between the studies. Cheng and Mix (2012) provided participants with physical manipulatives (shapes) and instructed participants to move the shapes provided, to check their answers. In contrast, Hawes *et al.* (2015) provided participants with feedback on the accuracy of their responses, but no explanation was provided to explain accuracy. The possible explanations for differences in the outcomes of the two training studies in this domain are explored further in Chapter 4. However, in short, there is mixed evidence on the effectiveness of spatial training for mathematics. There is a need for future research on the features of spatial training that may promote mathematical gains.

Further insight can be gained from studies that integrate spatial skills into mathematical training and instruction. Hawes, Moss, Caswell, Naqvi, and MacKinnon (2017) reported significant gains in both spatial and mathematical outcomes following a 32-week classroom-based intervention in which spatial visualisation activities were integrated into mathematics, geometry-based lessons (Math for Young Children [M4YC] project). Based on a sample of children aged 4 to 7 years, spatial gains were found in mental rotation (η_p^2 = .16), spatial language (η_p^2 = .16), and visuospatial reasoning (η_p^2 = .19). Gains were also reported in symbolic number processing (η_p^2 = .10) but not non-symbolic comparison (η_p^2 = .03) or number knowledge (η_p^2 = .01) compared to controls. Cohen (1988) defined η_p^2 values of 0.01, 0.06, and 0.14, as small, medium, and large effect sizes respectively. The authors suggest that the mathematical gains may reflect shared processing requirements for spatial and symbolic number tasks. Alternatively, they may be accounted for by improved spatial representation of number after training, or improvements in executive functions. Improved executive functions have been associated with higher mathematics achievement in previous studies (e.g., Hawes, Moss, Caswell, Naqvi, &

MacKinnon, 2017). Also based on the M4YC project, Bruce and Hawes (2015) reported significant gains in mental rotation at 6 to 8 years, following a teacher-led intervention of geometry and spatial based activities. Of note, this study did not include a control group. In older children aged 10 to 12 years, similar findings were reported in a classroom-based study where children completed 2 hours of training per week, over a 10-week period (Lowrie, Logan, & Ramful, 2017). Training was delivered by teachers and included lessons that focussed on developing spatial constructs, in addition to lessons that integrated spatial thinking into problem-solving tasks. Significant gains were reported in spatial visualisation (d = 0.65), mental rotation (d = 0.43), and geometry-based mathematics items (d = 0.34). However, no gains in spatial orientation or number-based mathematics items were found. This study did not include a control group; therefore, it is unclear to what degree the performance gains reported were attributable to practice effects. Overall, these studies highlight the effectiveness of incorporating spatial thinking into mathematics lessons as a means of improving mathematics outcomes.

In many studies that integrate spatial thinking into mathematics lessons, teachers are permitted to customise and adapt the proposed lessons, tailoring them to their classrooms (Hawes et al., 2017; Lowrie et al., 2017). Thus, not all participating children are exposed to identical training paradigms, and while unlikely, it is possible that adaptations made by teachers to their lessons may contribute to the performance gains reported. Another limitation of integrating spatial thinking and mathematics in training paradigms is that studies of this type cannot offer insight into the underlying causal relationship between spatial and mathematical constructs. In these studies, it is not possible to disentangle the impact of the spatial and mathematical aspects of training respectively. Additionally, given that the training materials (lessons) require a range of skills and processes, it is also not possible to elucidate which of these mechanisms has contributed to the gains reported. Finally, many classroom-based studies investigating spatial and mathematical training do not include a control group (Bruce & Hawes, 2015; Lowrie et al., 2017). This limits the inferences that can be made as any gains reported following training might be due to practice effects. Despite these limitations, from an educational perspective, these studies offer useful tools for mathematical teaching. They also offer valuable insight (both practical and theoretical) into the impact of embedding spatial thinking into classroom-based mathematics activities.

1.5.3 Insights into cognitive training from other cognitive domains

Failure to find transfer of spatial training gains to mathematics, an untrained cognitive domain, may be due to poor selection of training tasks. As reported in section 1.5.2, there is mixed evidence for transfer of training gains from spatial domains to mathematics skills (Cheng & Mix, 2014; Hawes et al., 2015). Within the broader cognitive training literature, similar mixed findings have been reported for transfer of training gains in other untrained domains, e.g., working memory (WM). WM is the ability to store information (verbal or visuo-spatial) for short amounts of time and to manipulate this information (Baddeley & Hitch, 1974). Like spatial thinking the malleability of WM has been demonstrated in many studies with improved performance on WM tasks after behavioural training (*Hedges G* = .31 for verbal WM training; *Hedges G* = .28 for VSWM training)(Melby-Lervåg, Redick, & Hulme, 2016). Significant correlations have also been reported between WM and mathematics outcomes across a range of studies (Friso-van den Bos, van der Ven, Kroesbergen, & van Luit, 2013; Fuchs et al., 2010). Therefore, similarly to spatial thinking, it has been proposed that WM training may lead to transfer of gains to mathematics outcomes. Although WM is a cognitive ability that has been targeted extensively in training studies, there is very little evidence of transfer of WM training gains to distantly related tasks, such as general cognitive abilities or academic outcomes (Melby-Lervåg & Hulme, 2013; Melby-Lervåg et al., 2016; Schwaighofer, Fischer, & Bühner, 2015). Obtaining transfer of training gains to skills beyond those that have been targeted by intervention is not easily achieved (Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015). The findings from WM studies highlight the apparent selectivity of training and raise the question as to why one might expect to see transfer of training gains between two seemingly distinct cognitive skills. In this thesis, it is proposed that the success of cognitive training is contingent on an understanding of the underlying cognitive mechanisms between training targets and transfer domains. It is proposed that the decision to complete training studies should be supported not only by correlations between training targets and transfer domains, but also some understanding of how, and why these domains might be associated.

In summary this section has provided convincing evidence that spatial thinking, particularly intrinsic-dynamic spatial thinking, is malleable. However, from the current findings in this domain, no clear conclusion can be drawn regarding the transfer of spatial training gains to mathematics in childhood populations. Furthermore, the findings from cognitive training studies in other domains such as WM suggest that transfer of cognitive training gains to untrained domains is difficult to achieve. In particular, correlations between measures may be an insufficient basis for establishing training paradigms. This is explored further in Chapter 4.

1.6 Conclusions and thesis directions

The use of spatial training to improve mathematics is promising because spatial thinking is malleable, and leads to gains in spatial task performance (Bruce & Hawes, 2015; Taylor & Hutton, 2013; Uttal et al., 2013). While there is convincing evidence from correlational studies that spatial and mathematical skills are associated in pre-school and adult populations, the findings in the primary school years are less established (e.g., Verdine et al., 2014; Wai et al., 2009). Furthermore, few studies have employed spatial training in an effort to show transfer of spatial training gains to mathematics, and the results of these training studies are variable. Some studies report a positive impact of spatial training on mathematics (Cheng & Mix, 2014; Hawes et al., 2018; Lowrie et al., 2017) while others report no transfer of spatial training gains to mathematics (Hawes et al., 2015). Beyond features of task and study design, such as those outlined in section 1.5.2, the inconsistencies in the results may be attributable to the fact that spatial and mathematical thinking are often treated as unitary constructs. However, as outlined in sections 1.2 and 1.3, both spatial and mathematical cognition are complex cognitive domains.

This thesis explores spatial thinking in the context of Uttal *et al.'s* (2013) theoretical classification of spatial skills. This classification has four spatial sub-domains (intrinsic-static, intrinsic-dynamic, extrinsic-static and extrinsic-dynamic) through which the development of spatial thinking and its role for mathematics in childhood are explored. Sub-domains of mathematical thinking are considered using von Aster and Shalev's (2007) model of numerical cognition. This model proposes an innate, core system for representing numbers, the ANS. Through development, the ANS becomes integrated with a symbolic number system providing a platform for more complex mathematics abilities such as multi-digit calculation, word problem solving, algebra, measurement and data handling skills (Butterworth, 1999; Feigenson et al., 2004; Piazza, 2010; Träff, 2013).

This thesis includes three inter-related experimental studies, outlined in Chapters 2, 3 and 4, each of which presents specific research questions and employs a distinct methodological approach. The thesis focuses on exploring these important questions in children aged 5 to 10 years. The findings from these studies are drawn together in Chapter 5 to form conclusions and identify future research directions. The study presented in Chapter 2 explores the longitudinal and concurrent relationships between spatial and mathematical skills in children aged 5 and 7 years, controlling for socio-demographic factors and language skills. The study involves secondary data analysis of 12,099 children who participated in the Millennium Cohort Study (MCS). It expands on previous findings by using a large-scale, longitudinal sample of primary school children, a population that have been largely omitted from research on the associations between spatial ability and mathematics achievement. In this chapter, the differential associations between spatial and mathematical skills for children of different genders and those in different SES groups are also explored.

Building on this, in Chapter 3 the developmental relations between spatial and mathematics skills across 5 consecutive age groups in the primary school years (6, 7, 8, 9 and 10 years) are explored. Using a cross-sectional approach, this study compares performance across Uttal *et al.'s* (2013) four spatial sub-domains and each of von Aster and Shalev's (2007) mathematical sub-domains, including classroom-based mathematics skills (*N* = 155). It provides important insights into the specificity of associations between spatial and mathematical skills, acknowledging that both spatial and mathematical thinking are multi-dimensional constructs. Importantly it provides evidence that spatial-mathematical associations are age-dependent, and highlights age and task relevant targets for spatial training.

In Chapter 4, the efficacy of explicit (instructional videos) and implicit (practice with feedback) methods of training spatial skills at 8 years are compared, and the transfer of spatial training gains to other spatial and mathematical domains are investigated. Informed by the longitudinal and cross-sectional findings reported in Chapters 2 and 3, the study outlined in this chapter uses an intervention-based design including pre-testing, training and post-testing (N = 250). The outcomes provide insights into the malleability of spatial thinking, and the causal relationship between different sub-domains of spatial and mathematical thinking.

Together, the three components of this thesis provide important evidence for the complex relationship between spatial skills and mathematics, the specificity of spatial-mathematical relations across sub-domains, the age dependency of spatial-mathematical relations, and the efficacy of spatial skills training as a novel means of improving mathematics performance. The implications and importance of these findings and the areas for further research are discussed in Chapter 5.

Chapter 2 The longitudinal contribution of spatial ability to mathematics achievement in the early primary school years

2.1 Introduction

As outlined in Chapter 1 there is evidence that spatial and mathematical skills are associated in pre-school aged children. For example, findings from Verdine *et al.* (2014) showed that spatial skills at 3 years, significantly predicted mathematical problem solving at 4 years, assessed using the TOSA, an intrinsic-dynamic spatial measure. Beyond the pre-school years, relatively few studies explore spatial-mathematical relations in middle childhood, and those that do, present somewhat mixed results. In this chapter the role of spatial skills at the age at which children first enter formal schooling, an age group of children that are largely absent from previous literature, is explored. In the UK, children begin formal education at approximately 5 years. At this age, they are presented with a range of novel mathematical content. Previous studies have suggested that improved spatial thinking may assist in learning novel information (Ackerman, 1988; Uttal & Cohen, 2012). Therefore, it stands to reason that individual differences in spatial thinking may influence learning of novel mathematical content of the curriculum. Hence, in this chapter it is proposed that individual differences in spatial skills on entry to primary school have an important predictive role in supporting mathematical success in subsequent years.

More specifically, the study presented in this chapter explores the role of intrinsic-dynamic spatial skills as predictors of mathematics ability in the early primary school years. The use of an intrinsic-dynamic spatial task is useful, given the strong association of this spatial sub-domain with mathematics observed in studies with older children and adults (for example, Reuhkala, 2001; Thompson et al., 2013; Wai et al., 2009). Findings from correlational studies in children suggest that intrinsic-dynamic spatial skills specifically, have particular associations with mathematics at 6 years but not at 9 years (Mix et al., 2016). To date, the most convincing evidence that spatial skills may play a predictive, beneficial role in mathematics outcomes during the early years of schooling comes from Gunderson *et al.* (2012) who also measured intrinsic-dynamic spatial performance. In their study, Gundersen *et al.* (2012) demonstrated that intrinsic-dynamic spatial skills at 5 years are predictive of approximate calculation performance at 8 years. The authors suggest that these benefits are mediated by number line estimation skills. However, a major limitation to this study is that approximate calculation skills are one small sub-component of mathematics performance. To enhance the generalisability of the results reported and to enable application to successful classroom intervention, there is a need to explore the role of spatial thinking on more comprehensive measures of mathematics skills beyond calculation alone.

The finding that spatial and mathematical skills are associated in middle childhood is also limited by the fact that few studies control for other known predictors of mathematics achievement, including language skills. As outlined in Chapter 1, it is not yet known whether there is a direct relationship between spatial and mathematical skills or whether these associations might be attributable to the overlapping language demands of the tasks used, or by an underlying intelligence (IQ) factor (Alloway & Alloway, 2010; Mayes et al., 2009). Previous studies have demonstrated significant associations between mathematics and language skills (LeFevre et al., 2010; Moll et al., 2015). Therefore, there is a need to control for language ability when exploring the role of other mathematical predictors. By comparing models that include or exclude shared variance with language skills respectively, the unique and shared variance in mathematics performance that is attributable to spatial skills can be established.

In addition, as outlined in Chapter 1, there is evidence that socio-demographic factors influence mathematics outcomes. There are inconsistent findings on whether mathematical performance differs across genders, with no reliable evidence for a male or female advantage (Halpern et al., 2007; Hyde et al., 2008). Furthermore, where gender effects are reported, the size of performance differences is often small (Hyde et al., 2008). To better understand gender effects in mathematics outcomes, there is a need to use large scale studies with representative populations. There is more convincing evidence that mathematical performance differs across different SES groups (Byrnes & Wasik, 2009). Yet, no known studies on spatial-mathematical relations generalise their findings to investigate whether socio-demographic factors mediate the observed associations between spatial and mathematical skills. For example, by exploring whether spatial-mathematical relations are stronger for children from high compared to low income groups, or for males versus females. Furthermore, no known studies explore differences in spatial and mathematics skills across ethnic groups, or the role of ethnicity in mediating spatial-mathematical relations.

This is the first study to investigate both the concurrent and longitudinal relationships between intrinsic-dynamic spatial skills and mathematics in the early primary school years (5 to 7 years). While most studies to date focus on specific sub-components of mathematics such as arithmetic or calculation, this study explores associations between spatial skills and mathematics achievement more generally. There are benefits of exploring the role of spatial skills for mathematics from both a holistic perspective and in the context of individual mathematical sub-domains. This study explores the value of intrinsic-dynamic spatial skills as a predictor of mathematics achievement, measured using a standardised mathematics measure that is proposed to reflect the range of skills and competencies required in the mathematics classroom. As such, the findings of this study have practical importance for influencing mathematical achievement in the classroom. Using data from the MCS, the study presented in this chapter explores associations between spatial skills and mathematics in the early primary school years using a large-scale, general population longitudinal sample. It investigates changes in intrinsic-dynamic spatial skills over time and identifies the contribution of spatial skills at 5 and 7 years to achievement in mathematics at 7 years. Importantly, it extends previous research by exploring the role of spatial skills for mathematics performance while accounting for the roles of other known and possible predictors of mathematics performance, i.e., gender, SES, ethnicity and language skills. In short, this study identifies reliable associations between a specific spatial skill and mathematics achievement at early primary school ages which, if significant, could enable the effective design of targeted age-based mathematics interventions, the outcomes of which may have both educational and economic implications.

2.2 Materials and Methods

2.2.1 The Millennium Cohort Study (MCS)

The MCS is a longitudinal population-based study of children born in the United Kingdom between 2000 and 2002. Participants of the MCS were sampled using a stratified, clustered design, ensuring adequate representation of disadvantaged and ethnic minority groups and over-representation of children living in the smaller UK countries including Scotland, Northern Ireland and Wales. To date, the MCS has collected 6 waves of data during which the children in the study were approximately 9 months, and 3, 5, 7, 11 and 14 years respectively. The MCS uses questionnaires, interviews, and a range of cognitive assessments with cohort members, their families and teachers to collect information on a wide range of variables including; cognitive development; child and parental physical and mental health; income and poverty; parenting; ethnicity and schooling among others.

The current study focuses on the Millennium Cohort during Waves 3 and 4, for which suitable measures of spatial ability are available. Wave 3 was completed between February 2006 and January 2007 when the study participants (*N* = 15,460) were approximately 5 years. Wave 4 was completed between January 2008 and February 2009 when the participants (*N* = 14,043) were approximately 7 years. The Centre for Longitudinal Studies, who manage the MCS, attained ethical approval for Wave 3 of the MCS from the London Multi-Centre Research Ethics Committee of the National Health Service (NHS), while ethical approval for Wave 4 of the study was obtained from the Northern and Yorkshire Research Ethics Committee of the NHS. No additional ethical approval was required for this study. The data used in this study was open access. It was accessed and downloaded by registering with the UK Data Service. For more details see https://www.ukdataservice.ac.uk.

2.2.2 Participants

Power analysis, based on the largest possible regression model (20 predictors), indicated that to achieve power of 0.8, with a small effect size of (f^2 = 0.02), 1064 participants were required. The initial study sample included the eldest cohort child from each MCS family (N = 19,244). The inclusion of a single participant from each family ensured that clustering effects did not occur. Participants with missing data on any of the cognitive or educational measures chosen for this study (see below) were subsequently excluded from the sample rendering a sample size of 12,537 participants. Participants who did not indicate that they spoke English only or mostly English at home were excluded from this study to remove variance created by differences in language comprehension (438 participants excluded). The final sample size for this study was 12,099 participants. Thus the desired power was achieved. The Organisation for Economic Co-Operation and Development (OECD) equivalised income scores at Wave 4 were used as a measure of SES in this study. OECD equivalised income scores convert reported household income into a modified scale based on the number and age of all members of the family (Hansen & Joshi, 2008). Any missing income data were replaced and each observation was weighted to reflect the original sampling probability and attrition (Hansen & Joshi, 2008). The final income distribution was divided, generating five equal-sized quintiles.

The demographics of the final sample compared to those of the excluded sample are shown in Table 2.1. The excluded sample includes all participants present in the original MCS sample who were excluded from this analysis. The demographics shown in Table 2.1 are based on unweighted data. Data were unweighted as some of the excluded sample were not present at Wave 4. Therefore, application of Wave 4 weights accounting for sampling design, non-response and attrition was not suitable for this group. Hence, Wave 4 weights were applied to neither the excluded nor the final samples. As shown, the selection criteria used to generate the final study sample led to small but significant differences in the ages of the samples at Wave 3 and Wave 4. Across both waves, the mean age for the excluded sample was higher than that of the included sample. Although there is a significant difference in the gender ratio between the samples, the table indicates that the final sample has a more balanced gender distribution, compared to the excluded sample. As expected, the percentage of participants in all non-white ethnic groups was reduced in the final sample leading to a 13.4% increase in the percentage of white participants in the study compared to the relative percentage of white participants in the excluded sample. This is most likely explained by the English language exclusion criteria. The excluded sample also has significantly higher proportions of participants in the lowest and second income quintiles. This may be explained by higher rates of non-response and attrition in the lower income groups. In comparison, the final sample includes approximately even percentages of participants in each income-based quintile, with a slight under-representation of the lower income groups. The final results should be viewed in light of the slight under-representation of participants in lower income families, and the slight over-representation of white participants relative to all other ethnic groups.

Table 2.1

Demographic characteristics of the final study sample compared to participants excluded from analysis (unweighted data)

		Excluded Sample		Final Sample		Test
		N	*% total*	*N*	*% total*	*Pearson's χ^2*
Gender						
	Male	3818	53.4	6079	50.2	18.33 ***
	Female	3327	46.6	6020	49.8	
Ethnic group						
	White	5220	73.1	10463	86.5	578.57***
	Mixed	265	3.7	324	2.7	
	Indian	237	3.3	259	2.1	
	Pakistani & Bangladeshi	800	11.2	534	4.4	
	Black or Black British	384	5.4	340	2.8	
	Other Ethnic group	177	2.5	122	1.0	
	Missing	62	.9	57	0.5	
OECD Equivalised Income Quintiles						
	Lowest	589	8.2	2267	18.7	344.35***
	Second quintile	468	6.6	2394	19.8	
	Third quintile	295	4.1	2502	20.7	
	Fourth quintile	224	3.1	2475	20.5	
	Highest quintile	173	2.4	2450	20.2	
	Missing	5396	75.4	11	0.1	
		Mean	**SD**	**Mean**	**SD**	**T (D)**
Age Wave 3 (years)						
	Male	5.23[a]	.25	5.22	0.25	3.55 (.068)***
	Female	5.23[a]	.26	5.21	0.24	
Age Wave 4 (years)						
	Male	7.30[a]	.30	7.23	0.25	10.90 (.296)***
	Female	7.30[a]	.28	7.22	0.25	

Note. * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$, [a] For the excluded sample, ages at Waves 3 and Wave 4 are based on a sample size of 3146 and 1745 participants respectively. This reduction in sample size is due to the large number of participants in the initial sample who did not participate in Wave 3 and/or Wave 4.

2.2.3 Measures

As shown in Table 2.2, all participants completed a series of cognitive measures across Wave 3 and Wave 4 of the MCS. This included a subset of items from a standardised test of mathematics for 7 year olds (National Foundation for Educational Research [NFER], 2004) in addition to a selection of measures taken from the British Ability Scales II (BAS II), a standardised test battery that measures cognitive ability (Elliott, Smith, & Mc Cullock, 1996). For all test measures, age-based standardised test scores converted to z-scores, are reported.

Table 2.2

Cognitive measures included in the MCS Waves 3 and 4

Test Measure	Wave 3	Wave 4
BAS II-Pattern Construction	✓	✓
BAS II-Naming Vocabulary	✓	
BAS II-Word Reading		✓
NFER-Progress in Maths		✓

Note. BAS II = British Ability Scales II*;* NFER = National Foundation for Educational Research.

2.2.3.1 Mathematics skills

A shortened version of the National Foundation for Educational Research Progress in Maths (NFER PiM) test for 7 year olds was administered at Wave 4 as a measure of mathematics (NFER, 2004). The NFER PiM is an assessment of mathematics ability and includes a wide assortment of items on all aspects of the National Mathematics Curricula including questions on numbers, shapes, measurement and data handling. Age-based standardised scores were based on 6-month age intervals and were calculated based on the full-length NFER PiM test normed in 2004.

2.2.3.2 Spatial skills

This study used the Pattern Construction subscale of the BAS II as a measure of spatial ability (BAS II; Elliott et al., 1996; Hill, 2005). This nonverbal reasoning task is modelled on Kohs' traditional Block Design Test (Kohs, 1919). The task requires participants to copy a stimulus pattern using a set of blocks. The block faces are either all yellow, all black, or half-yellow, half-black. Participants must re-create a stimulus pattern by rotating, re-arranging and joining the blocks. The task falls within the intrinsic-dynamic sub-domain of spatial cognition as described by Uttal *et al.* (2013). In easier trials, the stimulus pattern is presented using 3-D blocks. Harder trials use 2-D picture representations of the stimulus pattern. Task success is measured as accuracy in block orientation and positioning, and response time. Age-based standardised scores were calculated based on 3-month age intervals (BAS II; Elliott et al., 1996; Hill, 2005).

2.2.3.3 Control variables

Additional sub-tests of the BAS II included as covariates in analyses were the Naming Vocabulary subscale (Wave 3) which measures expressive verbal ability and the Word Reading subscale (Wave 4) which measures educational knowledge of reading. In the Naming Vocabulary scale children are shown a series of pictures and are asked to name each of them. In the Word Reading scale, children are shown words on cards and are asked to read them aloud. Age-based standardised scores for these measures were based on 3-month age intervals. Due to the age difference of participants at different waves of the MCS, different language measures were included at Wave 3 and Wave 4. No single language measure was available for both waves.

2.2.4 Analysis strategy

Missing OECD equivalised income values, which accounted for 0.1% of cases, and missing ethnicity values, which accounted for 0.5% of cases, were calculated using the multiple imputation function in SPSS. MCS weights to account for the original stratified, clustered design of the MCS sample and sample attrition and non-response, were applied to all analyses unless otherwise stated. All sample sizes reported are based on unweighted data. Initial descriptive statistics were completed to provide an overview of overall performance patterns across tasks. T-tests and analysis of variance (ANOVA) were used to investigate the main effects of gender and SES (income groups) on task performance for all test measures including both language and spatial-based cognitive tasks, and mathematics achievement. Where equal variances could not be assumed, the results for unequal variance were reported. Post-hoc Games-Howell or Hochberg's GT2 tests were used appropriately in cases where the assumption of homogeneity of variance was violated or met respectively. A correlation matrix was completed to investigate the relative associations between performance measures and to inform subsequent general linear models. *T*-statistics were used to compare correlation coefficients (Field, 2013).

2.2.4.1 Regression models

To explore the role of spatial skills as a predictor of mathematics achievement, general linear models in SPSS were used. General linear models allow for the use of MCS weights to account for sample design, attrition and non-response. The use of age-adjusted z-scores for all cognitive task measures and age allowed for meaningful comparison of unstandardised *b* values within models. Although age-based standardised scores were used throughout, these scores were based on 3-month (BAS II) or 6-month (NFER PiM) age intervals and did not account for age-based variability within these age brackets. Hence, exact age at Wave 4 was included as a predictor in all models. While this extra adjustment for age is a more conservative approach, comparable results were found when models were run which did not include age as a predictor.

Model 1 was the most conservative and investigated the additional variation in mathematics explained by spatial skills, above that explained by demographic and language measures. As outlined in section 2.3.4.1, Model 1 presented the influence of spatial skills on mathematics, controlling for other variables including gender, SES (income-based quintiles), ethnicity, age and language skills (Naming Vocabulary and Word Reading at Wave 3 and 4 respectively). In this model, spatial task performance (performance on the Pattern Construction Task) at Wave 3 and 4 was considered simultaneously. Model 1 also explored the role of gender and SES as moderators in the relationship between spatial and mathematics skills by adding interaction terms for gender*spatial skills and SES*spatial skills at Wave 4. These interaction terms were included due to the identification of gender and SES (income-based quintiles) differences in spatial task performance in the bivariate analysis (further details in section 2.3.2).

Model 2, presented in section 2.3.4.2, was a less conservative model and investigated shared variation between spatial and language skills as predictors of mathematics. This model explored the role of spatial skills as a predictor of mathematics when controlling for demographic factors only. Language skills were only included after spatial skills in this model.

Model 3 sought to determine the longitudinal contribution of spatial and language skills at Wave 3 as predictors of mathematics achievement at Wave 4. No Wave 4 measures were included as predictors in Model 3. In the two previous models it is likely that the longitudinal value of Wave 3 measures in predicting mathematics achievement was underestimated due to shared variance between Wave 3 and Wave 4 spatial and language measures respectively. Hence, Model 3 was included to explore the longitudinal contributions of cognitive skills to mathematics achievement in the absence of concurrent predictors.

Model 4 investigated the role of concurrent predictors of mathematics achievement at Wave 4. Model 4, presented in section 2.3.4.4, investigated the role of spatial and language measures at Wave 4 as concurrent predictors of mathematics achievement at Wave 4. No Wave 3 measures were included as predictors in Model 4. The inclusion of Models 3 and 4 allowed for the comparison of concurrent and longitudinal predictors of mathematics respectively. To allow for more meaningful comparison of Model 3 and Model 4, the order of inclusion of variables in Model 4 was identical to Model 3.

In summary comparison between Models 1 and 2 provided the range of potential variance in mathematics achievement explained by spatial skills (in the presence and absence of other predictors of mathematics). Comparison across Model 3 and Model

4 outlined the roles of longitudinal and concurrent predictors (both spatial and language) of mathematical achievement.

2.3 Results

2.3.1 Overall task performance

Descriptive statistics for each of the cognitive and academic measures used in this study are shown in Table 2.3. While these results are specifically based on the sample included in this study, they are comparable to those describing the performance of the total MCS sample at Waves 3 and Wave 4 (Hansen, Jones, & Budge, 2010; Hansen & Joshi, 2008).

Table 2.3

Descriptive statistics for task performance across Waves 3 and 4 (z-scores, unweighted data)

Wave	Test Measure	N	Max	Min	Mean	SD
Three						
	BAS II- Pattern Construction	12,099	2.95	-3.12	0	1.00
	BAS II- Naming Vocabulary	12,099	2.34	-3.25	0	1.00
Four						
	BAS II- Pattern Construction	12,099	2.44	-3.05	0	1.00
	BAS II- Word Reading	12,099	1.86	-3.16	0	1.00
	NFER PiM	12,099	2.42	-1.88	0	1.00

Note. BAS II = British Ability Scales II*;* NFER PiM= National Foundation for Educational Research Progress in Mathematics

2.3.2 Performance differences based on gender and SES

2.3.2.1 Gender differences

Independent t-tests were carried out to identify differences in task performance based on gender. As shown in

Table 2.4, the results indicated that there was a significant difference in performance between males and females for all tasks. The mean score for females exceeded that for males on all tasks with the exception of mathematics performance where male scores were above those of females. These results should be viewed in light of the relatively small effect sizes reported for all t-tests. Cohen described values of *d* below 0.2 as small effects (Cohen, 1988; 1992). Hence, the magnitude of Cohen's *d* observed in Table 2.4, ranging from 0.053 to 0.177, suggests that the reported differences in performance of males and females on academic and cognitive measures are relatively small.

Table 2.4

Gender differences in cognitive and mathematics task performance (z-scores, weighted data).

Test Measure	**Gender**				**Statistics**	
	Male (n = 6079)		*Female (n = 6020)*		*Test statistic*	*Effect size*
	Mean	SD	Mean	SD	*T* value	Cohen's *D*
Wave 3						
BAS II- Pattern Construction	-.09	1.04	.09	0.94	-9.81**	0.177
BAS II- Naming Vocabulary	-.01	1.01	.05	0.95	-3.18**	0.057
Wave 4						
BAS II- Pattern Construction	-.04	1.04	.02	0.96	-3.09**	0.058
BAS II- Word Reading	-.05	1.05	.10	0.92	-8.60**	0.154
NFER PiM	.01	1.04	-.04	0.95	2.97**	0.053

Note. * indicates $p < .05$, ** indicates $p < .01$, all n's are based on unweighted data. BAS II = British Ability Scales II; NFER PiM = National Foundation for Educational Research Progress in Mathematics

2.3.2.2 SES differences

One-way ANOVAs with SES as a between participant factor (5 levels: 5 equal-sized income quintiles) demonstrated significant differences in cognitive and mathematics performance across income levels. As shown in Figure 2.1, significant differences in performance across income groups were reported for all tasks as follows: Word Reading (Wave 4), F (4, 12320) = 268.18, $p < .001$, $\eta_p^2 = .075$; Pattern Construction (Wave 4), F (4, 12320) = 146.05, $p < .001$, $\eta_p^2 = .050$; NFER PiM (Wave 4), F (4, 12320) = 197.93, $p < .001$, $\eta_p^2 = .058$; Naming Vocabulary (Wave 3), F (4, 12320) = 291.96, $p < .001$, $\eta_p^2 = .096$; and Pattern Construction (Wave 3), F (4, 12320) = 120.28, $p < .001$, $\eta_p^2 = .036$). Post-hoc tests revealed significant differences between all SES groups (p's $< .010$). However, the effect sizes (η_p^2) reported can be classified as small (Cohen, 1988).

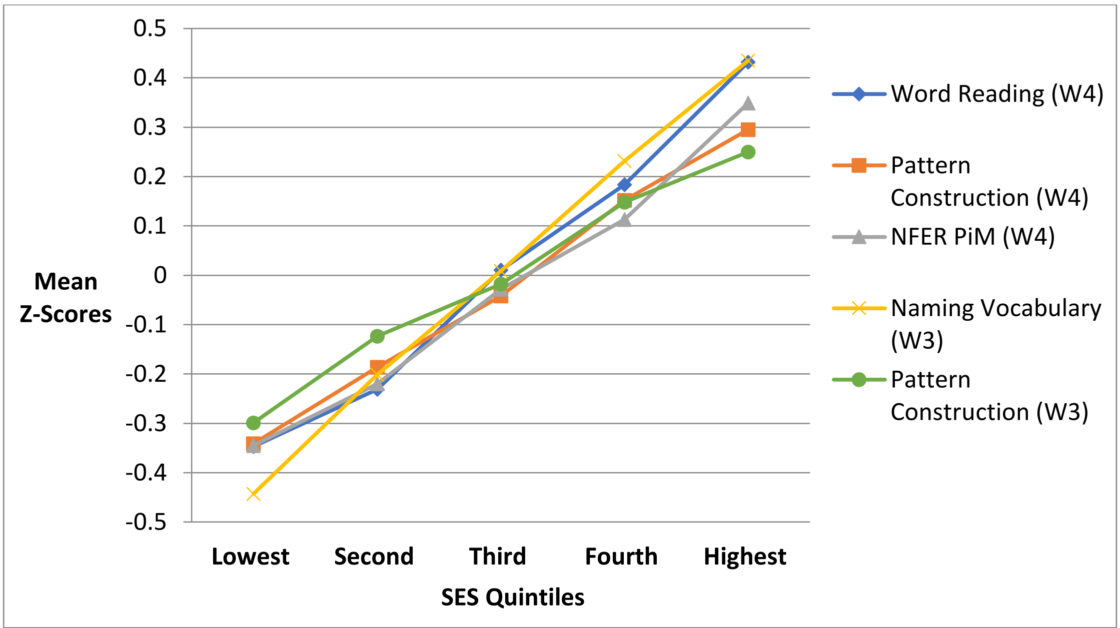

Figure 2.1. Cognitive and mathematics task performance across SES groups. *Note.* SES groups are income quintiles (z-scores, weighted data)

2.3.3 Associations between mathematics and cognitive measures

Bivariate correlations between mathematics performance scores at 7 years (NFER PiM) and all cognitive measures are shown in Table 2.5. There were medium to strong correlations between mathematics and all cognitive measures. As defined by Cohen (1988) correlations between .3 and .5 can be defined as having a medium to large effect. Word Reading at Wave 4 had a larger correlation with NFER PiM scores, r (12099) = .53, p < .001, followed by Pattern Construction scores at both Wave 4, r (12099) = .48, p < .001, and Wave 3, r (12099) = .43, p < .001. There were significant differences (p < .001) between these correlations (measured using t-statistics).

Table 2.5

Correlations between mathematics and cognitive measures (z-scores, unweighted data)

		Wave 4		**Wave 3**	
	Measure	BAS II-Pattern Construction	BAS II-Word Reading	BAS II-Naming Vocabulary	BAS II-Pattern Construction
	NFER PiM	.48	.53	.39	.43
Wave 4	BAS II- Pattern Construction		.33	.32	.56
	BAS II- Word Reading			.37	.35
Wave 3	BAS II- Naming Vocabulary				.33

Note. All correlations were significant at the p < .001 level, unweighted N = 12,099. BAS II = British Ability Scales II*;* NFER PiM = National Foundation for Educational Research Progress in Mathematics

2.3.4 Regression analyses

2.3.4.1 Model 1

The results of all models are summarised in Table 2.6. Model 1 sought to determine the contribution of spatial ability to the variation in mathematics achievement while controlling for other known or possible predictors of mathematics ability including language skills, gender, age, ethnicity and SES. This model is the most conservative. Word Reading at Wave 4 and Naming Vocabulary at Wave 3 were both included as language measures, accounting for language skills across two time points. Spatial measures included Pattern Construction scores at both Wave 3 and Wave 4. As the correlations between language and mathematics scores at Wave 4 were greater than those between spatial skills and mathematics performance (see Table 2.5), language measures were added to the model before spatial measures.

Overall, the model accounted for 42.4% of the variation in mathematics scores at 7 years. The demographic measures entered in step 1 including gender, age at Wave 4, ethnicity and SES accounted for 7.3% of the variation, adjusted R^2 = .073, F (11, 11667) = 85.13, p < .001, while the language measures added in step 2 accounted for 26.3% of the variation, adjusted R^2 = .336, F (13, 11665) = 456.54, p < .001. The spatial measures entered in step 3 accounted for an additional 8.8% of the variation, even after accounting for all other predictors, adjusted R^2 = .424, F (15, 11663) = 575.01, p < .001. No significant interactions between gender and spatial skills, or SES and spatial skills were reported in step 4 (p > .05 for both). All variables, with the exception of ethnic group, were significant predictors in the final model. The b values, t-statistics and effect sizes indicated that Word Reading, β = .35, t (11663) = 42.67, p < .001, η_p^2= .135, and Pattern Construction, β = .26, t (11663) = 13.50, p < .001, η_p^2 = .015, at Wave 4 had the most significant impact on predicting mathematics achievement at Wave 4.

2.3.4.2 Model 2

Model 2 explored the role of spatial skills as a predictor of mathematics when controlling for demographic factors only. As seen in Model 1, the demographic measures accounted for 7.3% of the variation in mathematics. Spatial scores at Waves 3 and 4 were entered simultaneously in step 2 and accounted for 22.6% of the variation, adjusted R^2 = .299, F (13, 11665) = 384.30, p < .001. The language measures entered in step 3 explained an additional 19.8% of the variation, adjusted R^2 = .424,

F (15, 11663) = 575.01, p < .001. All spatial and language measures were significant predictors in this model (p's < .001, η_p^2's > .017).

2.3.4.3 Model 3

Model 3 explored the variation in mathematics achievement predicted by cognitive measures at Wave 3 only. Overall, the model accounted for 27.7% of the variation in mathematics scores at 7 years, with demographic measures accounting for 7.3% of this variation. Based on the magnitude of correlations between Wave 3 measures and mathematics achievement (see Table 2.5), spatial scores were added to the model before language scores. The spatial measure accounted for 15.4% of the variation in mathematics, adjusted R^2 = .227, F (12, 11666) = 287.00, p < .001, while the language measure accounted for an additional 5.0% of the variation, adjusted R^2 = .277, F (13, 11665) = 344.47, p < .001. The b values, t- statistics and effect sizes suggest that Pattern Construction makes the most significant impact on predicting mathematics achievement, β = .34, t (11665) = 39.89, p < .001, η_p^2 = .120, followed by Naming Vocabulary, β = .26, t (11665) = 28.26, p < .001, η_p^2 = .064.

2.3.4.4 Model 4

Finally, Model 4 explored the variation in mathematics achievement predicted by cognitive measures at Wave 4 only. The final model accounted for 40.1% of the variation in mathematics scores at 7 years, with demographic measures accounting for 7.3% of this variation. Spatial skills were entered in step 2 accounting for 19.2% of the variation in mathematics, adjusted R^2 = .266, F (12, 11666) = 352.83, p < .001, while the language measure added in step 3 accounted for an additional 13.5% of the variation, adjusted R^2 = .401, F (13, 11665) = 601.43, p < .001. Word Reading made the most significant impact on predicting mathematics achievement, β = .41, t (11665) = 51.29, p < .001, η_p^2 = .184, followed by Pattern Construction scores, β = .33, t (11665) = 28.26, p < .001, η_p^2 = .133.

2.3.4.5 Additional Information

For all models, the assumptions of normality were met. Outliers were defined as any individuals falling outside three standard deviations of the mean for at least one of the continuous variables in a given model. In Models 1 and 2, 396 cases were identified as outliers (3.3% of the sample). In Model 3, 289 cases (2.4% of the sample) and in Model 4, 141 cases (1.2% of the sample) were identified as outliers. All outliers were included as they account for very small proportions of the sample population and do not significantly influence the findings reported. In addition, there was no justifiable reason to exclude these cases as it is likely that they reflect natural variation in the population.

Table 2.6

General linear models predicting mathematics achievement at 7 years (weighted data)

Model 1		***B***	**SE**	***t***	***p***	***Partial η²***	***F***	***df***	***p***	***Adj. R²***	***Δ Adj. R²***
Step 1											
SES (income quintiles) [a]	Lowest	-0.11	0.02	-4.40	< .001	.002	85.13	11, 11667	< .001	0.073	
	Second	-0.11	0.02	-4.47	< .001	.002					
	Third	-0.07	0.02	-2.86	.004	.001					
	Fourth	-0.10	0.02	-4.25	< .001	.002					
Ethnicity [b]	White	0.02	0.07	0.26	.794	0					
	Mixed	0.04	0.08	0.55	.583	0					
	Indian	-0.01	0.09	-0.10	.918	0					
	Pakistani, Bangladeshi	-0.05	0.08	-0.62	.536	0					
	Black, Black British	-0.10	0.08	-1.27	.204	0					
Gender	Male	0.14	0.01	10.12	< .001	.009					
Age		-0.05	0.01	-6.81	< .001	.004					

Model 1 cont.		*B*	SE	*t*	*p*	*Partial* η^2	*F*	*df*	*p*	*Adj.* R^2	Δ *Adj.* R^2
Step 2											
Word Reading (W4)		0.35	0.01	42.67	< .001	.135	456.54	13, 11665	< .001	0.336	0.263
Naming Vocabulary (W3)		0.12	0.01	14.77	< .001	.18					
Step 3											
Pattern Construction (W4)		0.25	0.02	13.50	< .001	.015	575.01	15, 11663	< .001	0.424	0.088
Pattern Construction (W3)		0.13	0.01	14.33	< .001	.017					
Step 4											
Gender* Pattern Construction (W4)		-0.01	0.01	-0.35	.727	0	431.63	20, 11658	< .001	0.424	0
SES* Pattern Construction [a] (W4)	Lowest	-0.01	0.02	-0.38	.701	0					
	Second	0.01	0.02	0.53	.600	0					
	Third	-0.03	0.02	-1.38	.169	0					
	Fourth	0.02	0.02	0.92	.358	0					

Model 2	***B***	**SE**	***t***	***p***	***Partial η^2***	***F***	***df***	***P***	***Adj. R^2***	***Δ Adj. R^2***
Step 1										
As seen for model 1[c]						85.13	11, 11667	< .001	0.073	
Step 2										
Pattern Construction (W4)	0.25	0.01	28.12	< .001	.063	384.30	13, 11665	< .001	0.299	0.226
Pattern Construction (W3)	0.13	0.01	14.32	< .001	.017					
Step 3										
Word Reading (W4)	0.12	0.01	14.79	< .001	.018	575.01	15,11663	< .001	0.424	0.198
Naming Vocabulary (W3)	0.35	0.01	42.66	< .001	.135					
Model 3										
Step 1										
As seen for model 1[c]						85.13	11, 11667	< .001	0.073	
Step 2										
Pattern Construction (W3)	0.34	0.01	39.89	< .001	.120	287.00	12, 11666	< .001	0.227	0.154
Step 3										
Naming Vocabulary (W3)	0.26	0.01	28.26	< .001	.064	344.47	13, 11665	< .001	0.277	0.050

Model 4	B	SE	t	p	Partial η^2	F	df	p	Adj. R^2	Δ Adj. R^2
Step 1										
As seen for model 1[c]	0.41	0.01	51.29	< .001	.184	85.13	11, 11667	< .001	0.073	
Step 2										
Pattern Construction (W4)	0.33	0.01	28.26	< .001	.133	352.83	12, 11666	< .001	0.266	0.192
Step 3										
Word Reading (W4)	0.41	0.01	51.29	< .001	.184	601.43	13, 11665	< .001	0.401	0.135

Note. [a]The reference category is highest SES quintile, [b]The reference category is other ethnic group, [c]The parameter estimates for the demographic measures entered in step 1 varied very subtly for each of Models 1-4, due to differences in the predictors included in each of the models. The exact parameter estimates for step 1 of each model are available on request. W3 = Wave 3; W4 = Wave 4

2.4 Discussion

Intrinsic-dynamic spatial skills explained a significant proportion of the variance in mathematics achievement in the early primary school years, above that explained by other demographic factors, or language skills alone. Based on a sample of over 12,000 participants, these findings add substantial support for both a concurrent and longitudinal role of spatial skills for general mathematics achievement. They also extend previous findings by assessing mathematics using a more comprehensive measure of mathematics than calculation skills alone (Gundersen et al., 2012). The results of this study also extend previous longitudinal findings in pre-school populations and older children (Casey et al., 2015; Verdine et al., 2014; Zhang et al., 2014) to children in the early primary school years. This study demonstrates that spatial skills at 5 years explain a unique proportion of the variance in mathematics achievement at 7 years, in middle childhood. Owing to the design of the study, it was also possible to investigate shared variation between spatial and language measures. By comparing models that include and exclude language skills, the true proportion of variation in mathematics explained by spatial skills could be estimated. This value is predicted to fall between the more conservative 8.8% result and the more liberal 22.6% result, generated by models that either include or exclude shared variance with language skills respectively.

Further analyses highlighted the individual and unique contributions of Wave 3 measures at 5 years and Wave 4 measures at 7 years to the variation in mathematics outcomes at 7 years (Wave 4). In both models, spatial skills explained a substantial proportion (over 15%) of the variation in mathematics performance at 7 years. The findings of this study are particularly applicable to the classroom, as this study included a classroom-based, standardised measure of mathematics (the NFER PiM), for the first time. This test includes items on a range of mathematical skills required by children in the UK classroom including questions on numbers, shapes, measurement and data handling (NFER, 2004), thus increasing the generalisability of these findings to real-world contexts. It is also interesting to note that the profile of associations between spatial versus language predictors and mathematics achievement at Wave 4, contrasts with that seen for Wave 3. Based on the observed *b* values, *t*-statistics and effect sizes, language at 7 years is a stronger predictor of mathematics when compared to spatial skills. In contrast, at 5 years, spatial skills are a stronger predictor of subsequent mathematics achievement at 7 years, when compared to language skills. Although this pattern of findings may be due to the different language measures used in the two waves, it may also suggest that while language skills are stronger concurrent predictors of mathematics, spatial skills are stronger longitudinal predictors of mathematics achievement. While spatial skills do have an important concurrent role in mathematics performance, these findings highlight particular longitudinal connections between spatial skills and mathematics performance between 5 and 7 years, in the context of language measures. Previous findings show that spatial skills may be more important for novel mathematics tasks compared to practiced, automatic mathematics skills (Ackerman, 1988; Uttal & Cohen, 2012). At 5 years, children in the UK begin formal schooling and thus are faced with large amounts of new mathematics material. The findings of this study are consistent with the notion that children with strong spatial skills at 5 years are better able to learn novel mathematical concepts, which in turn impacts their later mathematics performance. This finding is particularly interesting as it highlights a particular, positive role for early spatial skill training for later mathematics achievement.

Another notable finding was the difference in performance on the Pattern Construction task between Wave 3 and Wave 4. While this may reflect the test-retest reliability of the Pattern Construction task, previous test-retest correlations of .88 have been reported for this measure (Elliott, Smith, & McCulloch, 1997). Another explanation for these differences is that the sample in this study may have differed from the standardisation sample for the Pattern Construction Task at Wave 3 or Wave 4. Alternatively, performance differences seen in Pattern Construction scores across waves may reflect the malleability of spatial skills in middle childhood. As the spatial scores calculated account for age, the findings suggest that other environmental factors or experiences, aside from age-dependent developmental change alone, may influence spatial development between 5 and 7 years. These factors may include developmental strategy change or environmental factors such as early schooling experiences, exposure to technology or gaming (Office of Communications [Ofcom], 2015; Spence & Feng, 2010). Identifying these factors could improve understanding of individual differences in spatial skills.

The study demonstrated that both gender and income were significantly associated with differences in task performance across all measures investigated. In line with other studies such as Byrnes and Wasik (2009), the findings show that children from higher SES backgrounds outperformed their lower SES counterparts consistently across all tasks. Gender differences were also reported such that females outperformed males in all test measures except for mathematics achievement where male performance was above that of females. It is important to recognise that the effect sizes of these findings were very small, suggesting that although gender differences in performance may exist, the size of these differences may be negligible. Nonetheless, the findings do highlight a slight female advantage in spatial task completion. This contrasts with previous studies in which males (in the pre-school and primary school years) have been reported to outperform females on a range of spatial measures (e.g., Carr et al., 2008; Casey et al., 2008; Casey, Pezaris, & Nuttall, 1992; Johnson & Meade, 1987; Levine et al., 1999; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005). Thus, the findings add to a growing body of literature challenging the existence of a significant male advantage in spatial cognition in young children (Alyman & Peters, 1993; Halpern et al., 2007; Lachance & Mazzocco, 2006; LeFevre et al., 2010; Manger & Eikeland, 1998; Neuburger, Jansen, Heil, & Quaiser-Pohl, 2011).

Beyond main effects of gender and SES on task performance, the results do not suggest differential relations between spatial and mathematics skills for children of different genders or those in different SES groups. No significant interactions were reported between gender and spatial thinking, nor SES and spatial thinking, in predicting mathematics outcomes. Given the size of the sample tested in this study, these findings offer convincing support that spatial and mathematics skills are associated similarly across different demographic groups, and that targeting future training studies to distinct SES groups or to males or females specifically is not necessary.

2.4.1 Strengths and limitations

An important strength of the study was the use of large-scale, general population data, which ensured the generalisability of the findings to children in the UK. The nature of the sampling protocol employed in the MCS enhances the generalisability of the results reported, due to the inclusion of participants from a range of SES backgrounds. However, the use of secondary data to answer novel research questions is dependent on the availability of suitable test measures. While the MCS dataset provides an excellent resource for the examination of the relationship between intrinsic-dynamic spatial skills and mathematics achievement at 7 years, these findings cannot be generalised beyond the intrinsic-dynamic sub-domain to other spatial sub-domains. Another limitation of using the MCS dataset was the lack of a mathematics achievement measure at Wave 3. Without mathematics achievement scores at Wave 3, it was not possible to run a cross-lagged panel correlation to assess whether early mathematics abilities are predictive of later spatial skills, as well as whether earlier spatial skills influence later mathematics outcomes. Similarly, it was not possible to measure what cognitive skills predicted mathematics gains over time. The results reported here are also limited to children in the UK school system. Owing to differences in school environments cross-culturally, further research is needed to establish whether these findings have international generalisability.

In support of the results reported in this chapter, previous findings indicate that intrinsic-dynamic spatial tasks may be particularly useful to mathematics as they require the accurate completion of mental transformations. For example, it has been proposed that intrinsic-dynamic spatial skills can also be applied in the completion of mathematics tasks including measurement tasks, lines of symmetry tasks, and equations that are presented in atypical formats (Bruce & Hawes, 2015; Mix & Cheng, 2012). Strong intrinsic-dynamic spatial skills may be useful for certain mathematics tasks of this type as they may allow children to cognitively manipulate aspects of a given task, for example, by folding shapes or re-arranging the order of equations. While associations between other sub-domains of spatial thinking and mathematics are less well understood, there is some indication that different spatial sub-domains may be particularly important for different mathematics tasks at different developmental ages (Mix et al., 2016). For example, extrinsic tasks such as spatial scaling may be particularly important for the ordinal comparison of numbers (Mix, Prather, Smith, & Stockton, 2014) and the use of a mental number line (Dehaene et al., 1993). In Chapter 3, the findings reported here are extended beyond intrinsic-dynamic skills, to explore associations between mathematics and intrinsic-static, extrinsic-static and extrinsic-dynamic spatial skills. Similarly, while this study focused on associations between spatial and mathematics skills at 5 and 7 years only, in Chapter 3 these results are extended by testing associations between spatial and mathematical thinking across development in primary school children aged 6 to 10 years.

2.4.2 Conclusion

In this chapter, significant associations between intrinsic-dynamic spatial skills and mathematics achievement are reported, such that spatial task performance at both 5 and 7 years can explain a significant proportion of variation in mathematics scores at 7 years, above that described by socio-demographic or language measures. This suggests the potential of training early intrinsic-dynamic spatial skills as a novel method of improving classroom-based mathematics achievement. The use of this type of training is explored in Chapter 4.

Chapter 3 The developmental relations between spatial cognition and mathematics in primary school children

3.1 Introduction

Building on the findings outlined in Chapter 2 and acknowledging the neurological, behavioural and linguistic evidence that spatial thinking is not a unitary construct (see section 1.2.2), this study sought to measure developmental and individual differences in spatial thinking across each of Uttal *et al.*'s (2013) spatial categories. As previously outlined, these categories are founded on two dimensions, distinguishing skills as being intrinsic or extrinsic along one dimension, and as being static or dynamic along the other. In the study outlined in this chapter, a carefully selected task was used to examine each of Uttal et *al.'s* (2013) spatial sub-domains: intrinsic-static, intrinsic-dynamic, extrinsic-static and extrinsic-dynamic sub-domains. The role of each individual spatial sub-domain in explaining mathematics outcomes was explored.

Despite a bias towards studies investigating the role of intrinsic-dynamic spatial skills for mathematics, there is some evidence from studies of older children that other spatial sub-domains may impact on mathematics outcomes. As outlined in section 1.4, there is cross-sectional evidence that intrinsic-static spatial skills are correlated with mathematics performance at 10 and 11 years ($.37 < r < .42$) (Markey, 2010; Tosto et al., 2014). Intrinsic-static spatial skills at both 3 and 7 years are also significant longitudinal predictors of mathematics at 10 years ($.31 < r < .49$) (Carr et al., 2017; Casey et al., 2015; Zhang et al., 2014). These findings suggest that associations between spatial and mathematics skills in the primary school years are not limited to the intrinsic-dynamic spatial domain. However, there is a need for more detailed investigation to elucidate whether spatial-mathematical associations are consistent across all spatial sub-domains, at all ages. Further refining the findings in this field would facilitate a better understanding of not just *if*, but *why* significant correlations are often reported between spatial and mathematical constructs.

As outlined in section 1.4.2, findings on spatial-mathematical associations do not suggest a simple linear coupling between spatial and mathematical cognition (Fias &

Bonato, 2018). Consistent with the multi-dimensionality of both spatial and mathematical thinking, it has been proposed that some spatial skills may contribute to some mathematics skills and not others, and some spatial skills may not have a role in mathematics performance (Fias & Bonato, 2018). While the Mental Number Line was historically proposed to explain all observed associations between spatial and mathematical constructs (Barsalou, 2008; Lakoff & Núñez, 2000), this model does not fit with evidence that there are differential associations observed between specific spatial and mathematical sub-domains (Mix et al., 2016; 2017). Therefore, it has been proposed that not all spatial-mathematical associations can be explained in the same way, and as outlined in section 1.4.2, a range of theoretical explanations have been proposed for specific spatial-mathematical relations.

First, it has been proposed that extrinsic-static spatial task performance may rely on intensive quantification skills (proportional reasoning) and is thus expected to correlate with mathematics tasks that may also require proportional reasoning, e.g., number line estimation and approximate number comparisons (Newcombe, Levine, & Mix, 2015; Newcombe, et al., 2018; Rouder & Geary, 2014). Second, active processing including mental visualisation and manipulation of objects has been proposed as a requirement for intrinsic-dynamic and extrinsic-dynamic spatial tasks (Lourenco et al., 2018; Mix et al., 2016). Consequently, performance on these spatial sub-domains are expected to correlate with mathematics tasks requiring the mental manipulation or organisation of numbers, e.g., to ground abstract concepts in complex mathematical word problems, to complete missing term problems, or to solve multidigit calculations (Lourenco et al., 2018). Third, form perception is theoretically useful for intrinsic-static spatial tasks when distinguishing shapes from more complex backgrounds (Newcombe & Shipley, 2015). Intrinsic-static spatial tasks are therefore expected to correlate with mathematics activities that require the use of symbols or charts (Landy & Goldstone, 2007; 2010; Mix and et al., 2016). Based on these theoretical explanations for specific spatial-mathematical relations (proportional reasoning, mental visualisation and form perception), the a priori prediction for this study is that certain spatial sub-domains will be differentially associated with different mathematics outcomes.

As outlined in 1.2.3.5, there is evidence that spatial performance across each of Uttal *et al.*'s (2013) spatial sub-domains improves with developmental age (Newcombe et al., 2013). Comparison of spatial performance across studies suggests that there may be subtle differences in the developmental profiles of different spatial sub-domains. For example, success on intrinsic spatial tasks has been reported at a younger age (3 to 4 years) than extrinsic spatial tasks (5 to 6 years) (Frick et al., 2013; Frick et al., 2014a; Frick & Newcombe, 2012). However, no one study includes multiple measures of spatial thinking at consecutive developmental stages. Therefore, comparative findings should be interpreted cautiously as comparing spatial development across different sub-domains and across different studies, is hindered by the varying populations and testing paradigms used. This gap in the literature is addressed in this study. The development of, and associations between, different aspects of spatial thinking across 5 consecutive age groups in the primary school years (6, 7, 8, 9 and 10 years) are investigated for the first time.

Developmental differences in spatial-mathematical relations are also investigated in this chapter. It is hypothesised that some spatial-mathematical relations are stronger at specific developmental ages. Recent findings from Mix *et al.* (2016; 2017) provided a first-step to this understanding by showing age specific spatial-mathematical relations, such that intrinsic-dynamic spatial skills were significant predictors of mathematics at 6 years only, while VSWM was a significant predictor at 11 years only. This may reflect a developmental transition in the spatial skills that are important for mathematics. As described in Chapter 2, the role of spatial skills may be greater for novel mathematics tasks compared to automatic mathematics skills (Ackerman, 1988; Uttal & Cohen, 2012; Young et al., 2018). Alternatively, as the mathematical content that children are exposed to varies across school years, and spatial-mathematical associations are proposed to be specific to certain spatial tasks and mathematical content, this may lead to developmental variation in observed spatial-mathematical associations (Mix et al., 2016). The developmental relations between spatial and mathematics skills across consecutive age groups in middle childhood are explored in this chapter.

The selection of mathematics measures for inclusion in this study was driven by von Aster and Shalev's (2007) model of numerical cognition (further details in 1.3.2). This model proposes that individuals have an innate ANS to measure approximate representations of numerical magnitude in the brain (Cordes et al., 2001; Feigenson et al., 2004). Through development individuals are proposed to acquire a symbolic number system to represent symbolic numerals (Le Corre & Carey, 2007). While the exact process, by which the ANS might give rise to the symbolic number system is unknown, these systems are proposed to act as a platform for the development of other mathematical skills such as multi-digit calculation, word problem solving, algebra, measurement and data handling skills (Barth et al., 2005; Butterworth, 1999; Feigenson et al., 2004; Piazza, 2010; Träff, 2013). The study presented in this chapter investigates the role of spatial skills for mathematics across each of von Aster and Shalev's (2007) components of numerical thinking. This study includes a measure of both ANS and symbolic number skills, in addition to a standardised mathematics measure that identifies more complex mathematical skills including multi-digit calculation, missing term problems, fractions, etc. The inclusion of a standardised mathematics measure, reflective of the range of skills and competencies that are required in the mathematics classroom, also increases the practical implications of the findings. More specifically, the NFER PiM was chosen as this test is specifically designed to reflect the UK mathematics curriculum. Additionally, age-appropriate forms of the test and age-based standardised scores (based on a UK-based population) were available for each of the age groups included in this study (NFER, 2004). The investigation of the relations between spatial and mathematical skills outlined in this chapter also controls for other known predictors of mathematics performance including gender (Halpern et al., 2007) and language skills (LeFevre et al., 2010; Moll et al., 2015).

The study presented in this chapter has three aims. The first aim is to provide a developmental profile of spatial thinking in consecutive age groups from 6 to 10 years. The inclusion of consecutive age groups in this study provides strong acuity of this developmental change. Previous studies highlight preliminary evidence of subtle differences in developmental profiles of spatial thinking across spatial sub-domains

(Frick et al., 2013; Frick et al., 2014a). The inclusion of a range of spatial measures in this study, allows direct comparison of the developmental profiles of each of the sub-domains in the Uttal *et al.* (2013) typology. The second aim is to compare the roles of different spatial sub-domains in explaining mathematics performance, controlling for gender and language skills. Based on the aforementioned theoretical explanations for specific spatial-mathematical relations (proportional reasoning, mental visualisation and form perception), the a priori prediction for this study is that different spatial sub-domains will be differentially associated with mathematics outcomes. The third aim is to explore differences in spatial-mathematical relations from the ages of 6 to 10 years. It is hypothesised that some spatial-mathematical associations are age-dependent. There is evidence for a developmental transition in the spatial skills that are important for mathematics, which is proposed to occur in middle childhood (Mix et al., 2016; 2017). This study aims to refine the timing of this developmental transition.

3.2 Materials and Methods

3.2.1 Participants

The sample size for this study was determined using GPower. Power analysis was based on the largest possible regression model which included three control variables (age, vocabulary scores and gender), four spatial predictors and four age*spatial task interaction terms (see section 3.4.2). Based on the study presented in Chapter 2 which also explored the role of spatial thinking as a predictor of mathematics, a medium to large effect size was expected (f^2 = .217). To achieve power of 0.8, it was calculated that 78 participants were required. This study included 155 children across five age groups. Participants were drawn from a culturally diverse, London-based school with a 19% eligibility for free school meals which is slightly above the national average of 11% (UK Department for Education, 2017b). The age and gender of the participants in the study are shown in Table 3.1.

Table 3.1

Demographic features of the study sample

Age group	Sample size	% Male	Age years (mean ± SD)
6 years	30	53.3	6.0 ± 0.34
7 years	31	41.9	7.0 ± 0.29
8 years	32	56.3	8.0 ± 0.28
9 years	31	45.2	9.0 ± 0.33
10 years	31	51.6	10.0 ± 0.33

3.2.2 Spatial skills assessed, and measures used

3.2.2.1 Disembedding (intrinsic-static sub-domain)

To assess disembedding the CEFT was used. The CEFT is a measure of intrinsic-static spatial ability and measures the ability to dis-embed information from a larger context (Witkin et al., 1971). The task was delivered in accordance with the administration guidelines (Witkin et al., 1971). Participants were required to locate a target shape embedded within a more complex, meaningful picture. The task was presented as two blocks in a fixed order. Within each block, participants were introduced to a reference target shape (house and tent shape for Blocks A and B respectively). For each block, participants first completed 4 discrimination trials during which they were required to identify the target shape from a selection of other similar shapes. Discrimination trials were repeated until participants correctly answered two items in succession. Following this, participants completed two practice trials (Block A) or a single practice trial (Block B) in which they were required to locate the target shape within a more complex picture and to outline the target shape with their finger (see Figure 3.1). Performance feedback was given for practice trials. Participants repeated each practice trial until successfully locating the target shape. Practice trials were followed by 11 and 14 experimental trials, for Block A and Block B respectively. As for practice trials, participants were required to locate the target shape within more complex pictures. No feedback was given for experimental trials. As per the guidelines, for the first three experimental trials in each block, the reference shape was visible to the participant. From the fourth trial, the reference target shape was hidden from view. Only participants failing all trials in Block A did not progress to Block B. The task finished when participants failed five consecutive trials within Block B. Performance was measured as percentage accuracy (Min: 0%; Max: 100%). This was based on the maximum possible score (i.e., 28) and not the total number of trials completed by the participant.

Figure 3.1. Example stimulus from the CEFT

3.2.2.2 Mental rotation (intrinsic-dynamic sub-domain)

Mental rotation skills were measured using The Mental Rotation Task, a computerised measure of intrinsic-dynamic spatial ability. The protocol and stimuli were modified from Broadbent, Farran and Tolmie (2014). In each trial participants were asked to identify which of two monkey images located above a horizontal line, matched the target monkey image below the line. As shown in Figure 3.2, the images above the line included a mirror image of the target image, and a version of the target image rotated by a fixed degree from the target image. Participants completed four practice trials at 0° followed by 36 experimental trials. Only participants achieving at least 50% in the practice trials were deemed to understand the task instructions and continued to complete the experimental trials. Experimental trials were randomly presented and included equal numbers of clockwise and anti-clockwise rotations at 45°, 90° and 135° (eight trials for each degree of rotation), eight trials at 180° and four trials at 0°. Prior to analysis, performance scores on clockwise and anti-clockwise trials for each degree of rotation was combined (i.e., all 90° and -90° trials were collapsed). Participants used labelled keys on the left and right of the computer keyboard to respond. Percentage accuracy was recorded (Min: 0%; Max: 100%).

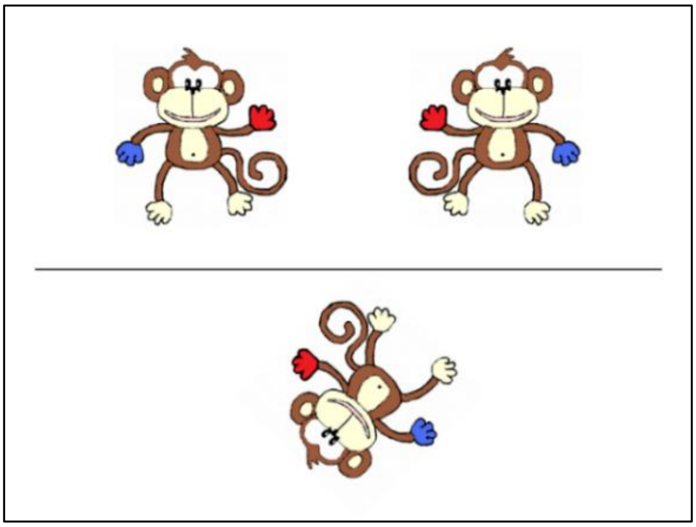

Figure 3.2. Sample item from the Mental Rotation Task (135° anti-clockwise trial)

3.2.2.3 Spatial scaling (extrinsic-static sub-domain)

A novel spatial scaling task was specifically designed as an extrinsic-static spatial task for use in this study. Further details pertaining to the design of this task have been published in Gilligan, Hodgkiss, Thomas, and Farran (2018). In this task, participants were required to use a printed "Pirate map" with a target to identify a corresponding onscreen referent map from four options (one correct and three distractor maps). Model maps were either the same size as the onscreen referent maps or were scaled-up versions of the referent maps (see Figure 3.3). In each trial, participants were encouraged to respond as quickly and accurately as possible by manually pressing one of the maps on the touch screen laptop 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 two practice trials with a scaling factor of 1 (both the model and referent maps measured 8cm x 8cm). Feedback was given for practice trials. If incorrect, participants were asked to repeat the trial until the correct answer was selected. Only participants correctly answering at least one of the two practice items on their first attempt continued to participate in the experimental blocks. Between each block, the task instructions were repeated. Participants received no feedback on their performance during experimental trials.

Scaling factor varied by experimental block and was determined as the difference in the relative size of the referent and model maps with respect to the participant. Scaling factor was set at 1, 0.5 and 0.25, i.e., model maps measured 8cm x 8cm, 16cm x 16cm and 32cm x 32cm, for trials at a scaling factor of 1, 0.5 and 0.25 respectively. Referent maps measured 8cm x 8cm in every trial. 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. Blocks were presented in order of increasing scaling factor. 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 was varied. This led to a corresponding difference in the visual acuity of the maps. As shown in Figure 3.4, half of the trials in each block were presented using a 6 x 6 square grid (requiring gross-level acuity) while the remaining targets were presented using a 10 x 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.

As outlined, for each trial four onscreen referent maps were presented including one correct map (i.e., the scaled (or unscaled) correspondent of the model map) and three distractor maps. As shown in Figure 3.5, 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 four 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. Performance on the task was measured as percentage accuracy (Min: 0%; Max: 100%).

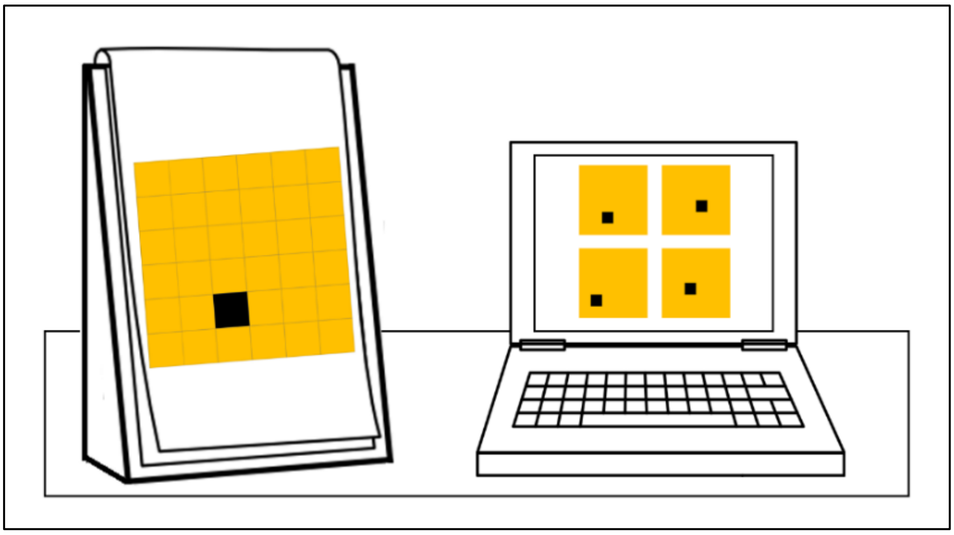

Figure 3.3. Relative position of model (left) and referent (right) maps relative to the participant, in the Spatial Scaling Task

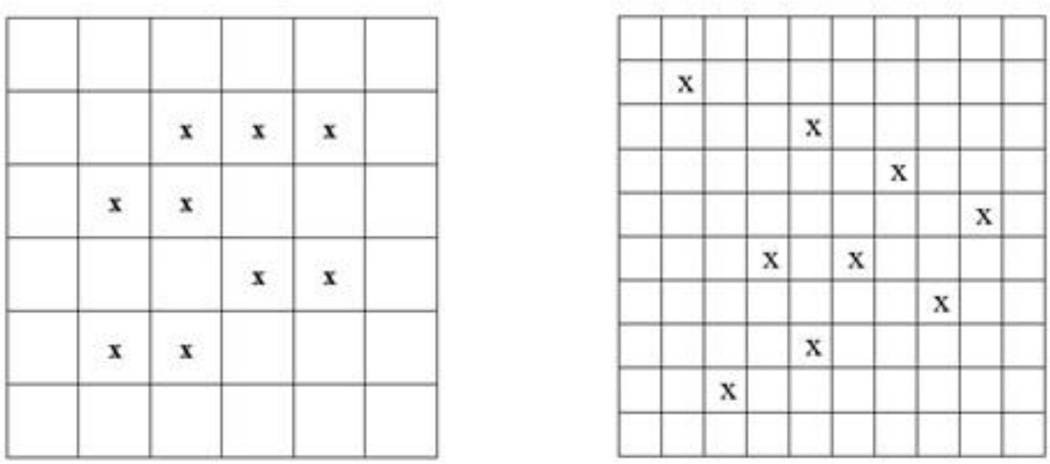

Figure 3.4. Sample spatial scaling targets for trials requiring gross level acuity (left) and fine level acuity (right)

C	A	C
B	Correct target	B
C	A	C

Figure 3.5. Position of distractor targets in the Spatial Scaling Task. *Note.* A indicates a vertical distractor, B indicates a horizontal distractor, C indicates a diagonal distractor

3.2.2.4 Perspective taking (extrinsic-dynamic sub-domain)

The Perspective Taking Task was included as a measure of extrinsic-dynamic spatial thinking and was taken from Frick *et al.* (2014b). Participants were required to identify which of four photographs had been taken from the perspective of a photographer, based on a 3-D or pictorial representation of the photographer in an arrangement. Participants completed four practice trials with real, 3-D objects and play-mobil characters holding cameras (to denote photographers). For each practice trial, participants were shown a photograph and were asked to identify which, if either, of the play-mobil photographers had taken the photograph. Participants confirmed their answers by standing up and looking over the shoulders of the photographers. Feedback was given for practice trials. If unsuccessful, participants were given sufficient attempts to correctly complete each practice trial. Participants were required to successfully answer all practice trials before moving to the computer-based experimental trials. In each of 18 experimental trials, participants were presented with a stimulus picture including a photographer and several objects in a spatial arrangement (see Figure 3.6). Participants were asked to select which of four photographs best represented the photograph that the photographer in the stimulus picture had taken from where they were standing. Complexity was introduced by increasing the number of objects in the stimulus picture (one, two or four objects). Trials also differed in the angular difference between the participant and the photographer. Participants completed equal numbers of trials in which they were positioned at 0°, 90° and 180° from the photographer respectively. Participants completed two trials for each complexity and angle combination. Trial order was fixed such that the angular difference changed between adjacent trials. The character acting as a photographer and the objects (colour, shape, relative position) were also changed between trials. Percentage accuracy was recorded (Min: 0%; Max: 100%).

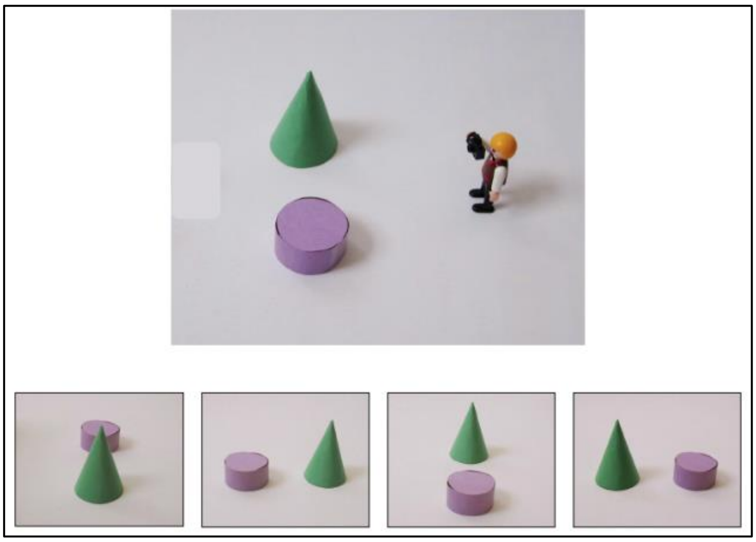

Figure 3.6. Sample trial from the Perspective Taking Task (2 items at 90°)

3.2.3 Mathematics ability measures

3.2.3.1 NFER Progress in Mathematics

The NFER PiM was administered as a measure of standardised mathematics performance. As outlined in Chapter 2, the NFER PiM test is an assessment of mathematics achievement designed to address the National Mathematics Curriculum in England, Wales and Northern Ireland (NFER, 2004). The test series includes items assessing number, algebra, data handling, shape, space and measures. Age-appropriate NFER PiM tests were administered to each age group of participants as per the test guidelines (NFER, 2004). Age-based standardised scores with a mean of 100 and a standard deviation of 15, were used in all analyses (Min: 69; Max: 141).

3.2.3.2 Approximate Number System Task

A dot comparison task was used to measure ANS skills in this study. This computerised task was taken from Gilmore, Attridge, De Smedt, and Inglis (2014). In each trial participants were required to compare and identify the more numerous of two dot arrays (see Figure 3.7). Each set of dot arrays was presented for 1500ms (or until a key press) and was followed by a fixation dot. Participants used labelled keys on the left and right of the computer keyboard to respond. Only participants who achieved at least 50% on the practice trials (eight trials) continued to the 64 randomly presented experimental trials. The quantity of dots in each comparison array ranged from 5 to 22. The ratio between the dots in each array varied between 0.5, 0.6, 0.7 and 0.8, with approximately equal numbers of trials assessing each of these ratios. This ratio effect is characteristic of performance on ANS tasks, and reduced performance is typically observed as the ratio between item sets approaches 1. For example, participants are expected to have higher performance when comparing 5 to 10 dots (a ratio of 0.5) than when comparing 5 to 6 dots (a ratio of approximately 0.8) (Barth, Kanwisher, & Spelke, 2003; Gilmore et al., 2014). The colour of the more numerous array (red or blue), in addition to the size and the density of dot presentation, were counterbalanced between trials. Task performance was measured as percentage accuracy (Min: 0%; Max: 100%).

It is noteworthy that performance on ANS tasks can be measured using several different metrics including performance accuracy, Weber fractions and the numerical ratio effect (NRE) for accuracy or reaction time (Inglis & Gilmore, 2014). Measuring ANS performance using the Weber fraction (w) assumes that when an individual is presented with an array of n dots, they form a representation of the dots that follows a normal distribution (with mean n and standard deviation w) (Inglis & Gilmore, 2014). However, there is evidence that the use of the Weber fraction leads to highly skewed distributions and that this metric has low test-retest reliability (Inglis & Gilmore, 2014). Additionally, this metric is highly sensitive to context and differs with task and stimulus properties (DeWind & Brannon, 2016). Furthermore, there is evidence that the Weber Fraction is highly correlated with performance accuracy on ANS measures, which poses the question as to what additional information the

Weber fraction provides, beyond performance accuracy scores. For the NRE, scores are calculated as the slope of the line created by plotting an individual's accuracy against the ratio of dots being compared (or alternatively plotting response times against the ratio of dots being compared) (Gilmore, Attridge & Inglis, 2011). However, there is also evidence that the NRE has poor test-retest reliability and that this outcome does not correlate with either accuracy or Weber fraction measures of ANS performance (Inglis & Gilmore, 2014). Taken together, and as advocated in several other papers (e.g., Inglis & Gilmore 2014; Guillaume & Van Rinsveld, 2018), performance accuracy was used as the outcome measure in this study.

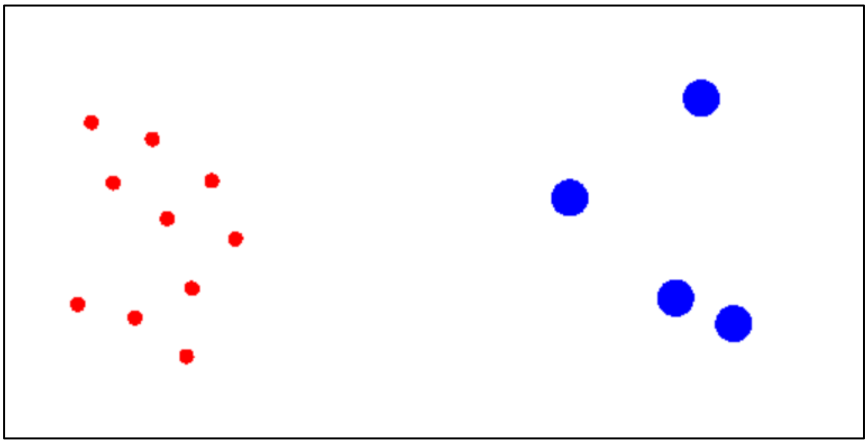

Figure 3.7. Sample dot arrays from the ANS Task

3.2.3.3 Number line estimation

The paper-based Number-Line Estimation Task used to assess symbolic numerical representation in this study was adapted from Siegler and Opfer (2003). Two trial types were included, number estimation (NP) and position estimation (PN) trials. As shown in Figure 3.8, for NP trials, participants were presented with a target number and were asked to estimate its location on a number line by drawing a straight line (hatch mark) through the number line at their selected location. As shown in Figure 3.8, for PN trials participants were presented with a vertical hatch mark on a number line and were asked to estimate what number was represented by the mark. This task was comprised of three blocks. Within each block participants completed two practice trials (one NP and one PN) followed by eight experimental trials (equal numbers of NP and PN trials presented alternately). Performance on NP and PN trials were collapsed across blocks. Blocks differed in the number line range presented. As per the Siegler and Opfer (2003) method, the number line in Block B ranged from 0-100 (numbers included 2, 3, 6, 18, 20, 24, 42, 50, 67, 71) and the number line in Block C ranged from 0-1000 (numbers included 2, 6, 18, 24, 71, 230, 250, 390, 500, 810). In this study, Block A with a range of 0-10 was added to reduce floor effects in younger children who may be less familiar with larger numbers (numbers included 1, 2, 3, 4, 5, 6, 7, 8, 9, 10).

Trial order was fixed and increased in difficulty. Participants began with Block A, followed by Block B and Block C. The numbers included in each block were chosen to enhance the identification of children's use of logarithmic and linear models and to minimize the impact of content knowledge (e.g., 50 is one half of 100). Similarly to other studies, there was over-sampling of numbers below 20 (Friso-van den Bos et al., 2015; Laski & Siegler, 2007). As outlined in section 1.3.2 performance was measured using PAE scores (Min: 0%; Max: 100%) and curve estimation (R^2_{LIN} scores; Min: 0; Max: 1). As the results were broadly similar for these measures, only R^2_{LIN} scores are reported in this chapter. Similar patterns of performance, with smaller effects, were found for PAE scores (see Appendix B). Participants were given the opportunity to complete all blocks. However, the 0-10 block was considered an age specific measure, and was analysed, at 6 and 7 years only. For each block where a participant's mean PAE scores for the practice trials were greater than 15%, or where participants failed to answer at least 80% of items, they were excluded from analysis for this block. For the 0-1000 block, only four participants at 6 years were eligible for inclusion, hence this age group was excluded from analysis.

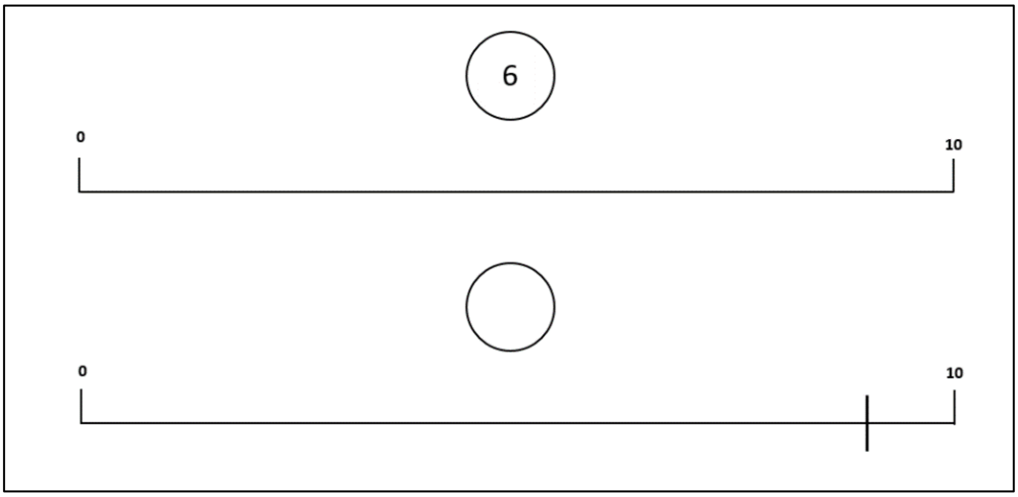

Figure 3.8. Sample items from the Number Line Estimation Task. Number to Position trials are shown above and Position to Number trials are shown below

3.2.4 Receptive Vocabulary Measure

The British Picture Vocabulary Scale (III) (BPVS) was administered as a measure of receptive vocabulary (Dunn, Dunn, Styles, & Sewell, 2009). Given that vocabulary is highly correlated with IQ (Sattler, 1988), the BPVS-III also acted as an estimate of general IQ. As per the administration guidelines, participants were asked to select which of four coloured pictures, best illustrated the meaning of a given word. Raw scores were used in analysis (Min: 0; Max: 168).

3.2.5 Procedure

Prior to the commencement of this study, ethical approval was granted by the UCL, IOE Department of Psychology and Human Development. A Disclosure and Barring Service Clearance Certificate for any researchers working on data collection was attained through UCL. With school permission, opt-out consent forms were sent to all parents/guardians. Prior to taking part, all participants were given an age-appropriate verbal description of the study and were informed that they could withdraw from the study at any time. The confidentiality and anonymity of their identifying information and task scores were emphasised. Participants were offered the opportunity to ask any questions and each individual participant was asked for verbal consent to participate.

The tasks used in this study were part of a larger test battery investigating associations between spatial thinking, mathematics and science achievement. For completeness, the full set of tasks comprising the procedure is described, but those not relevant to this thesis are not discussed in detail. Each participant completed a battery of mathematics, science, spatial and language measures, across four test sessions. In each session, mathematics tasks were completed prior to spatial tasks, to avoid possible mathematics improvements due to spatial training effects (Cheng & Mix, 2014). Beyond this stipulation, the order of task presentation within each session was randomised with equal numbers of participants completing each task order. During Session 1, a one-hour classroom-based session, the NFER PiM Test and the Number-Line Task (for children aged 8, 9 and 10 years only) were completed. Session 2, a 35-minute session, was completed in the school's computer suite in groups of 8 children, supervised by a minimum of two researchers. For computerised tasks, Hewlett Packard computers with a screen size of 17 inches were used. Children completed mathematics measures (the ANS Task, the Child Math Anxiety Questionnaire [CMAQ] [not discussed here] and the Number-Line Task [children aged 7 and younger]) and spatial measures (the Mental Rotation Task and a folding task [not discussed here]). Equal numbers of children completed each of the task orders shown in Table 3.2.

Table 3.2

Task orders for session 2

Order A	**Order B**	**Order C**	**Order D**
Number Line	Number Line	Number Line	Number Line
CMAQ	ANS	CMAQ	ANS
ANS	CMAQ	ANS	CMAQ
Rotation	Folding	Folding	Rotation
Folding	Rotation	Rotation	Folding

In session 3, participants were tested individually in a quiet room using a 13-inch Hewlett Packard touch-screen laptop. This session lasted 45 minutes. One mathematical task, a Spontaneous Focus on Number (SFON) Task [not discussed here] was completed at the beginning of the session. This was followed by spatial tasks (the Perspective Taking Task, the CEFT and the Scaling Task) and language tasks (the BPVS and a spatial language task [not discussed here]). Task order was randomised between participants such that each spatial and language task was completed in each position of the test battery with equal frequency, and the order of task presentation was not fixed. Each participant was presented with one of the five task orders shown in Table 3.3. For older students, a science assessment was also completed in a fourth session. For more details on the relations between spatial ability and science performance see Hodgkiss, Gilligan, Tolmie, Thomas, and Farran (2018). All computer-based measures were designed and implemented using the programme Open Sesame.

Table 3.3

Task orders for Session 3

Position	Order A	Order B	Order C	Order D	Order E
1	SFON	SFON	SFON	SFON	SFON
2	Scaling	Embedded figures	Perspective taking	Spatial language	BPVS
3	Perspective taking	Spatial language	BPVS	Scaling	Embedded figures
4	Spatial language	BPVS	Embedded figures	Perspective taking	Scaling
5	Embedded figures	Scaling	Spatial language	BPVS	Perspective taking
6	BPVS	Perspective taking	Scaling	Embedded figures	Spatial Language

3.2.6 Data analysis

The results of this study are presented in two parts. In Part A, descriptive statistics are presented including information on above chance performance on individual tasks, and the influence of features of task design on task performance.

Developmental differences and gender differences in spatial and mathematics task performance are also investigated. In Part B, the relations between spatial and mathematical thinking are explored.

Due to school absences and technical errors, 10 participants had missing scores for a single task in the battery (the proportion of missing data was 0.9%). Missing data was distributed as follows: the CEFT (one participant); the Perspective Taking Task (two participants); the NFER PiM Test (two participants); the ANS Task (five participants); the Number Line Task (one participant) and; the BPVS (two participants). For all measures except for the Number Line Estimation Task, all participants successfully completed the practice trials and proceeded to the experimental trials. Failure to complete a sufficient number of trials, or to achieve less than 15% error in the practice trials, led to the exclusion of 24, 19 and 17 participants on the 0-10, 0-100 and 0-1000 blocks of the Number Line Estimation Task respectively. However, as discussed further in section 3.5.4, there are no well recognised methods for determining floor or ceiling performance on number line estimation tasks. Therefore, the exclusion criterion used to measure whether participants understood the aims of the Number Line Estimation Task in this study, may have been overly conservative. The cut-off score (< 15% error) applied may have led to the exclusion of lower performing participants from the sample, i.e., participants who understood the task aims but had poor performance. The results should be interpreted in light of this limitation.

Due to missing data for some tasks, the desired participant numbers were not achieved for all models. Post-hoc power analysis was completed to determine the achieved power for each model. Except for the 0-10 Number Line Estimation Task (Model 3), all models achieved a power level greater than .919, which is above the suggested power level of 0.8 (Cohen, 1988). The results for the 0-10 Number Line Estimation Task should be interpreted cautiously due to the relatively low power of this model (0.754) (see Table 3.4).

For all measures, performance across age groups was explored graphically. For measures in which a ceiling (or floor effect) was suspected, one sample t-tests were completed against ceiling (or floor) performance. However, no significant floor or ceiling effects were found. Parametric analysis was completed as tests of normality indicated that all measures were broadly normal, and for most measures (except for the Number Line Estimation Task) there was no significant skewness or kurtosis. The accepted range was defined as ± 2.56 (Field, 2013). Furthermore, the overall sample size (N = 155), and the sample sizes in each age group ($n \geq 30$ for all) were sufficiently large for the Central Limit Theorem to apply (Field, 2013). For overall performance, across all age groups, there were no outliers for any measures. For performance split by age groups there was a relative absence of outliers, except for two high performers on the NFER PiM test (one male and one female aged 9 years), one low performer on the BPVS (one male aged 6 years), six low performers on 0-10 Number Line Estimation block (three males aged 6 years), five low performers on the 0-100 Number Line Estimation block (two females and one male aged 6 years, one male aged 7 years and one female aged 9 years), two low performers on the 0-1000 Number Line Estimation block (one female aged 9 years). All outliers were retained as they were deemed to reflect normal variation in the population.

Table 3.4

Post-hoc power analysis for regression models

Model	N	Effect size f^2	Power
Model 1	155	.292	.999
Follow up: Younger age group	61	.528	.997
Follow up: Older age group	94	.275	.988
Model 2	155	.173	.992
Model 3	48	.252	.754
Model 4	136	.125	.919
Model 5	108	.408	.999
Follow up: Younger age group	83	.327	.986
Follow up: Older age group	NA	NA	NA

3.3 Results Part A: Descriptive statistics

All results reported in Part A are based on complete case data only. Any participant with missing data for a given task was excluded from analysis for that task.

3.3.1 Gender differences

Gender differences in spatial and mathematics performance were investigated using Bonferroni adjusted t-tests to account for multiple comparisons (alpha levels of .004 [.05/13]). Where Levene's test was violated, the results for unequal variances were reported. As shown in Table 3.5 , there were no significant gender differences for any of the spatial measures or the BPVS ($p > .05$). For unadjusted p-values, significant differences favouring males were reported for both the 0-100 ($p = .025$, $d = 0.383$) and the 0-1000 ($p = .007$, $d = 0.518$) block of the Number Line Estimation Task. These differences were not significant when the results were adjusted for multiple comparisons (alpha level = .004). However, to ensure that the influence of gender was not overlooked, gender was included as a control variable in subsequent regression analysis for the 0-100 and 0-1000 blocks of the Number Line Estimation Task.

Table 3.5

Gender differences in performance on spatial, mathematics and language measures (unadjusted p values)

Test Measure (*n*-males, *n*-females)	Gender				Statistics		
	Male		Female		Test statistic	Sig	Effect size
	Mean	*SD*	*Mean*	*SD*	*T*	*P*	*D*
Spatial Measures							
Disembedding (77,77)	47.48	18.43	42.65	17.60	1.67	.097	0.268
Mental Rotation (77,78)	72.65	17.80	70.86	20.58	0.84	.401	0.135
Spatial Scaling (77,78)	57.07	20.10	52.64	20.22	1.37	.173	0.220
Perspective Taking (76,77)	56.81	20.47	58.28	20.84	0.43	.671	0.068
Mathematics Measures							
NFER PiM Standard Score (76,77)	97.57	14.75	97.19	15.28	0.16	.875	0.025
ANS Task (74,76)	60.97	13.39	61.98	14.25	0.45	.650	0.073
No. Line 10 R^2_{LIN} (20,28)	.88	.16	0.89	0.11	0.35	.725	0.072
No. Line 100 R^2_{LIN} (66,70)	.89	.15	0.82	0.20	2.27	.025	0.383
No. Line 1000 R^2_{LIN} (50,58)	.87	.17	0.74	0.31	2.64	.007	0.518
Language measure							
BPVS Raw Score (74,76)	95.85	21.91	95.91	22.09	0.02	.987	0.003

Note. NFER PiM = National Foundation for Educational Research Progress in Mathematics; ANS = Approximate Number System; No. Line = Number Line; R^2_{LIN} = Linear response patterns; BPVS = British Picture Vocabulary Scale

3.3.2 Spatial task performance

3.3.2.1 Disembedding (intrinsic-static sub-domain)

To explore differences in disembedding skills across age groups, a one-way ANOVA was completed with age as a between participant variable (5 levels: 6, 7, 8, 9, 10 years). For this, and all other ANOVA analyses described in this section, where Mauchly's Test of Sphericity was violated, Huynh-Feldt corrected values were reported. Age had a statistically significant effect on disembedding ability, F (4, 149) = 15.51, p < .001, η_p^2 = .294. Tukey post-hoc comparisons indicated that while performance did not differ significantly between 6 and 7 years (p = .682), these younger age groups had significantly lower performance than all older ages (*p's* < .002). No significant differences in performance between children aged 8 to 10 years were found (*p's* > .577). To allow comparison of performance on different spatial measures across age groups, performance on all spatial tasks including disembedding is summarised and displayed in Figure 3.12.

3.3.2.2 Mental rotation (intrinsic-dynamic sub-domain)

One sample t-tests were used to explore whether performance was above chance at each degree of rotation (0°, 45°, 90°, 135°and 180°) for each age group (6, 7, 8, 9, 10 years). Chance was set at 50% as there were two possible response options in each trial. At 6 years, performance was not significantly above chance on trials at 135°, t (29) = - .98, p = .337, d = - 0.178, or at 180°, t (29) = -1.10, p = .281, d = -0.200. At 7 years, performance was not above chance on trials at 180°, t (30) = .43, p = .667, d = 0.078. For all other degrees of rotation, above chance performance was reported at 6 and 7 years (*p's* < .003, *d's* > 0.580). This suggests that children aged 6 and 7 years understood the task aims as they could complete trials at smaller degrees of rotation. For those aged 8, 9 and 10 years, above chance performance was reported for all degrees of rotation (*p's* < .001, *d's* > .0.950). For more details see Appendix C.

To investigate the effect of age and degree of rotation on task performance, a two-way ANOVA was completed with age group as a between participant variable (5 levels: 6, 7, 8, 9, 10 years) and degree of rotation as a within participant variable (5 levels: 0°, 45°, 90°, 135°,180°). Age had a significant effect on performance accuracy, F (4, 150) = 16.64, p < .001, η_p^2 = .307. Tukey post-hoc comparisons indicated that performance at 6 years was significantly lower than all older age groups including those aged 7 years (*p's* < .035). At 7 years, performance was lower than all older age groups including those aged 8 years (*p's* < .019). No significant differences in performance between children aged 8 to 10 years were found (*p's* > .949). Mental rotation performance across age groups is displayed in Figure 3.12.

A significant main effect of degree of rotation was reported, F (4,600) = 47.96, p < .001, η_p^2 = .242. As shown in Figure 3.9, this was best explained using a linear contrast, such that performance decreased with increasing degree of rotation, F (1,150) = 121.27, p < .001, η_p^2 = .447. This was also supported by Bonferroni corrected pairwise comparisons (p < .008 between all degrees of rotation). A significant interaction between degree of rotation and age group was also reported, F (16,600) = 2.34, p = .004, η_p^2 = .059. Follow-up one-way repeated measures ANOVA's, for each age group, found a significant effect of degree of rotation, (*p's* < .002) that was best described by a linear contrast (*p's* < .005). The interaction between degree of rotation and age group was driven by differences in the effect sizes reported for different age groups. The degree of rotation effect was largest at 6 years (η_p^2 = .686) and 7 years (η_p^2 = .519). The effect sizes were smaller for those aged 8 (η_p^2 = .350), 9 (η_p^2 = .233) and 10 years (η_p^2 = .380). Overall these performance patterns are in line with other studies of mental rotation, such that there is reduced performance for trials at higher degrees of rotation (Kosslyn et al., 1990).

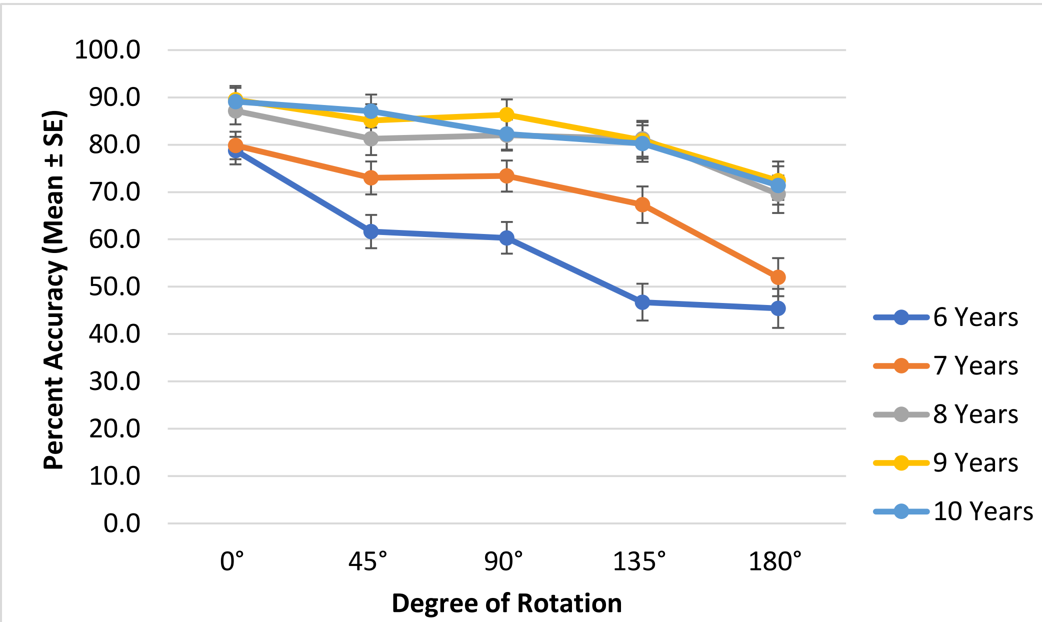

Figure 3.9. Performance on the Mental Rotation Task across different degrees of rotation and different age groups

3.3.2.3 Spatial scaling (extrinsic-static sub-domain)

One sample t-tests were used to investigate above chance performance at each scaling factor (1, 0.5, 0.25) for each age group (6, 7, 8, 9, 10 years). As each trial included four possible response options, chance was set at 25%. Above chance performance on the Spatial Scaling Task was reported for all scaling factors, for all age groups (*p's* < .005, *d's* > 0.557).

A 3-way ANOVA was completed with scaling factor (3 levels: 1, 0.5, 0.25) and visual acuity (2 levels: gross, fine) as within participant factors, and age group (5 levels: 6, 7, 8, 9, 10 years) as a between participant factor. There was a significant effect of age group on performance, F (4, 150) = 17.07, p < .001, η_p^2 = .313. Tukey post-hoc tests indicated no significant differences in performance between any consecutive age groups. At 6 years, performance was significantly lower than those at 8, 9 and 10 years (p < .001), but not at 7 years (p = .298). At 7 years performance was lower than at 9 and 10 years (p < .001) but not 8 years (p = .105). There was a marginally significant difference in performance between children at 8 and 10 years, favouring the older group (p = .054). No significant differences in performance between children aged 8 and 9 years (p = .396) or between children aged 9 and 10 years were found (p = .874). Differences in performance on the Spatial Scaling Task across age groups are displayed in Figure 3.12.

There was also a significant main effect of scaling factor, F (2, 300) = 15.80, p < .001, η_p^2= 0.950. Bonferroni corrected pairwise comparisons indicated significantly higher performance for unscaled relative to scaled trials (p < .001 for both a scaling factor of 0.5 and 0.25). No significant difference between trials at a scaling factor 0.5 and 0.25 was reported (p = 1.00). A significant main effect of visual acuity was also found, F (1, 150) = 146.99, p < .001, η_p^2 = 0.495, with lower accuracy for trials requiring fine level acuity relative to gross level acuity. There was a significant interaction between scaling factor and visual acuity, F (2, 300) = 11.52, p < .001, η_p^2= 0.071. Two follow-up repeated measures one-way ANOVAs were completed for trials requiring fine level acuity and trials requiring gross level acuity respectively. As shown in Figure 3.10, for gross level acuity, no significant effect of scaling factor was found, F (2, 308) = .20, p

= .821, η_p^2= .001. A significant effect of scaling factor was reported for fine level acuity, *F* (2, 308) = 24.18, *p* < .001, η_p^2= .136. There was significantly higher performance on unscaled trials relative to trials at a scaling factor of 0.25 (*p* < .001) and a scaling factor of 0.5 (*p* < .001). There was no difference in performance between trials at a scaling factor of 0.25 and 0.5 (*p* = 1.00). No significant interactions with age were reported for scaling factor or visual acuity (*p's* > .117, η_p^2*'s* < .048).

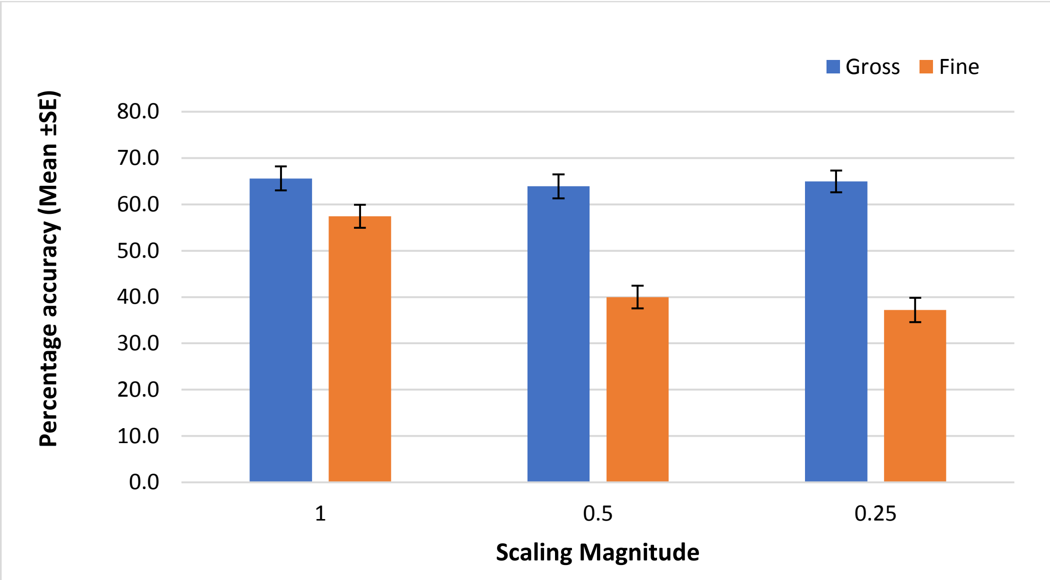

Figure 3.10. Performance accuracy on the Spatial Scaling Task across trials at different scaling factors and different levels of acuity

3.3.2.4 Perspective taking (extrinsic-dynamic sub-domain)

One sample t-tests were used to investigate whether participants in different age groups (6, 7, 8, 9, 10 years) performed above chance on trials at different angles (0°, 90°, 180°). As each trial in this task had four possible response options, chance was set at 25%. All age groups performed above chance for 0° trials (*p's* < .001, *d's* > 2.723). This suggests that participants of all ages understood the task instructions. For 90° trials performance was below chance at 6 years, *t* (28) = - 0.97, *p* = .339, *d* = - 0.181, and 7 years, *t* (30) = 1.21, *p* = .236, *d* = 0.217. Above chance performance was reported for all older age groups (*p's* < .001, *d's* > 0.599). A similar pattern was reported for 180° trials with above chance performance for older age groups (*p's* <

.001, *d's* > 0.688) but not for those at 6 years, *t* (28) = 2.00, *p* = .06, *d* = 0.371, or 7 years, *t* (30) = 1.46, *p* = .154, *d* = 0.263.

To investigate the effects of age, angle and complexity on perspective taking performance, a 3-way ANOVA was completed with angle (3 levels: 0°,90°, 180°) and complexity (3 levels: 1,2,4 objects) as within-participant factors and age group (5 levels: 6, 7, 8, 9, 10 years) as a between-participant factor. A significant effect of age group was reported, *F* (4, 148) = 12.17, *p* < .001, η_p^2 = .248. Games-Howell post-hoc tests indicated that there were no significant performance differences between any consecutive age groups (*p's* > .05). Performance at 6 years was lower than at 8 years (*p* = .008), 9 years (*p* < .001) and 10 years (*p* < .001). However, there was no significant difference in performance at 6 and 7 years (*p* = .698). At 7 years performance was significantly lower than at 9 and 10 years (*p* < .001 for both). There was no significant difference in performance at 7 and 8 years (*p* = .214). At 8 years, performance was not significantly different to performance at 9 years (*p* = .443) or 10 years (*p* = .061). Similarly, there was no significant difference in performance at 9 and 10 years (*p* = .905). These age-based differences in performance on the Perspective Taking Task are outlined in Figure 3.12.

Significant main effects of angle, *F* (2, 296) = 223.67, *p* < .001, η_p^2 = .602, and complexity, *F* (2, 296) = 18.80, *p* < .001, η_p^2 = .113, were found. As shown in Figure 3.11, Bonferroni corrected pairwise comparisons indicated that performance on 0° trials was significantly higher than performance on both 90° and 180° trials (*p* < .001). However, no significant difference in performance was seen for trials at 90° and 180° (*p* = 1.00). Pairwise comparisons also indicated a reduction in performance as the number of objects included in the task increased. As demonstrated in Figure 3.11, participants did significantly better on trials with only one object compared to trials with two or four objects (*p* < .001 for both). There was also higher performance for trials with two compared to four objects (*p* = .028).

No significant interactions between age, angle or complexity were reported (*p's* > .070, η_p^2*'s* < .049). The expected performance patterns were observed for this task, i.e., the patterns of performance are consistent with other studies of perspective taking where performance is lower for trials that are not presented at 0° and trials with greater numbers of objects (Frick et al., 2014b).

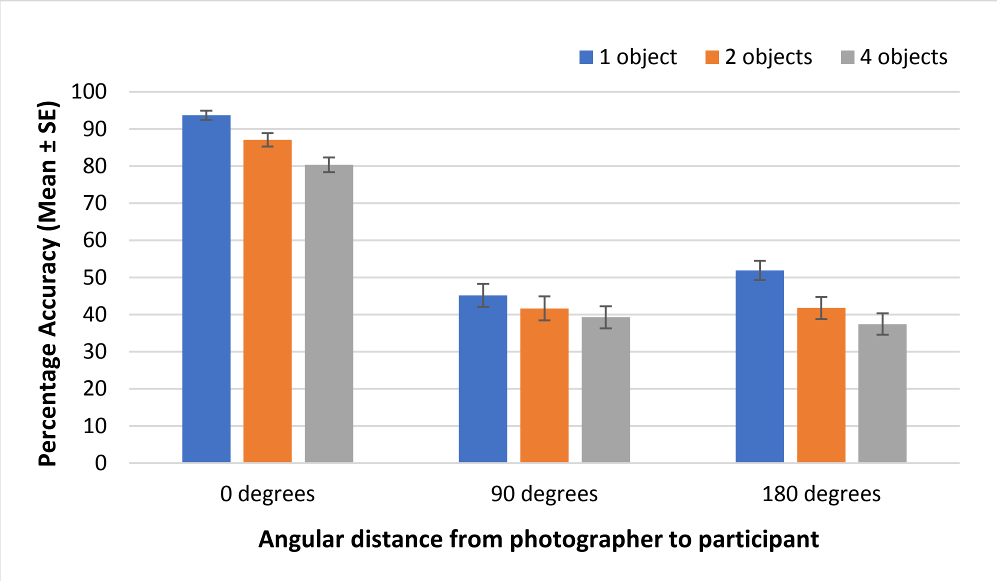

Figure 3.11. Performance accuracy on the Perspective Taking Task across different angle and complexity conditions

3.3.3 Summary of the development of spatial skills

To summarise, age-based differences were reported for all four of Uttal *et al.'s* (2013) spatial sub-domains. The post-hoc comparisons reported suggest slight differences in the developmental progression of different skills. As shown in Figure 3.12, performance on disembedding and mental rotation improved rapidly before 8 years. However, for scaling and perspective taking improvements were more gradual, with no significant differences in performance between consecutive age groups. This information is presented in table format in Appendix D.

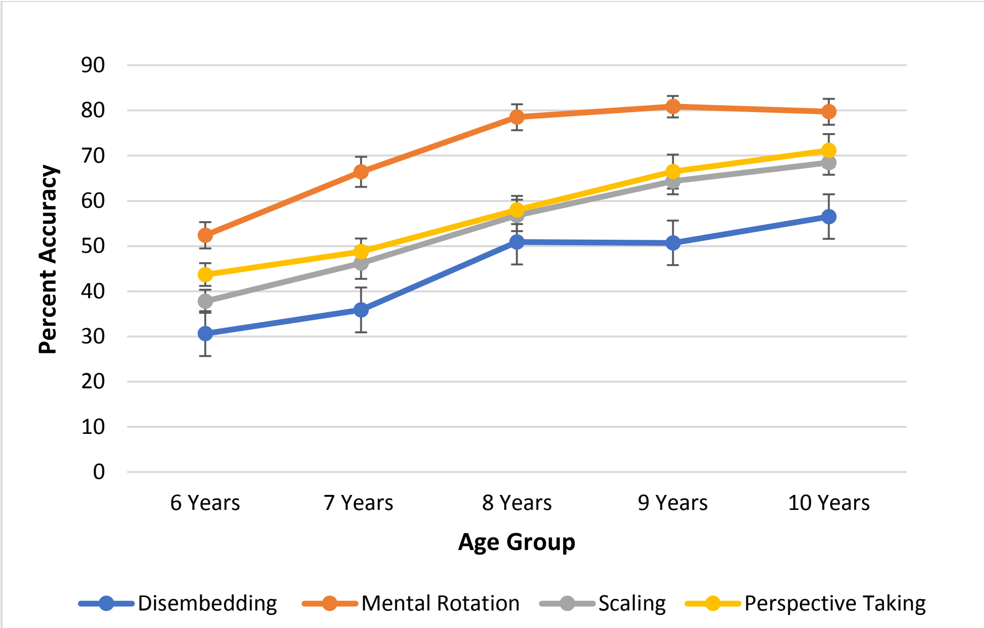

Figure 3.12. Spatial task performance across development

3.3.4 Mathematics performance

3.3.4.1 Mathematics achievement

For the NFER PiM test, each age group of children completed a different, age appropriate test. Performance was measured against standardised age-based scores where each participant's performance was compared to other children of their age (national averages). As each age group of children completed a different age appropriate task, comparing performance across age groups was not deemed suitable. Descriptive statistics for the NFER PiM are reported in Table 3.7.

3.3.4.2 ANS skills

One sample t-tests were used to explore whether participants in each age group (6, 7, 8, 9, 10 years) performed above chance. As each trial had two possible response options, chance was set at 50%. At 6 years, children did not perform significantly above chance on the ANS Task, t (28) = 1.82, p = .079, d = -0.345. Given that this might reflect lower ability rather than a weak understanding of the task aims, performance of this group on the ANS task was retained. For all other age groups performance on this task was significantly above chance; 7 years, t (30) = 3.72, p < .001, d = 0.668; 8

years, *t* (31) = 6.39, *p* < .001, *d* = 1.129; 9 years, *t* (30) = 7.79, *p* < .001, *d* = 1.397; 10 years, *t* (26) = 7.57, *p* < .001, *d* = 1.549.

To investigate the distance effect (ratio between the two dot arrays presented) and the effect of age group on performance, a two-way ANOVA was completed with distance (4 levels: 0.5, 0.6, 0.7, 0.8) as a within-participant factor and age group (5 levels: 6, 7, 8, 9, 10 years) as a between-participant factor. A significant effect of age group on ANS performance was reported, *F* (4, 145) = 14.28, *p* < .001, η_p^2 = .283. Games-Howell post-hoc tests indicated that performance at 6 years was significantly lower than performance at 7 years (*p* = .012), 8, 9 and 10 years (*p's* < .001). There was no significant difference in performance at 7 years and 8 years (*p* = .078). However, performance at 7 years was significantly lower than at 9 and 10 years (*p* < .001). No other significant differences in performance accuracy between at 8, 9 and 10 years were reported (*p's* > .733) (see Table 3.7).

A significant distance effect was found, *F* (3, 345) = 68.84, *p* < .001, η_p^2 = .322. Bonferroni comparisons indicated significant differences between all ratios (p < .001) except for ratios of 0.5 and 0.7 (p = .568). Distance effects are characteristic of tasks of this type (Buckley & Gillman, 1974; Dehaene et al., 1990; Moyer & Landauer, 1967). Thus, the expected performance patterns were observed for this task suggesting typical performance on the ANS task for the population of children included in this study.

3.3.4.3 Number line estimation

As outlined in section 3.2.3.3, all results reported in the chapter are based on R^2_{LIN} scores. Similar results were found when PAE was used as the outcome variable. These results are reported in Appendix B.

For each block of the Number Line Estimation Task, curve estimation was used to calculate linear (R^2_{LIN}) and logarithmic response patterns (R^2_{LOG}). For each block, the value of linear and logarithmic response patterns were compared for each individual. For all blocks, these simple comparisons indicated that a higher percentage of participants had estimates that were best described by described by a linear compared to a logarithmic function (i.e., the participant had a higher R^2_{LIN} score compared to R^2_{LOG} score) (see Table 3.6). Thus, linear estimates (R^2_{LIN} values) were used as the outcome variable in all subsequent analysis (Simms et al., 2016).

Table 3.6

Percentage of participants demonstrating linear estimates across different blocks of the number line task (sample size shown in brackets)

Range	6 Years	7 Years	8 Years	9 Years	10 Years	Total
0-10	82.61 *(23)*	84.0 *(25)*	NA	NA	NA	83.33 *(48)*
0-100	47.6 *(21)*	71.4 *(28)*	84.4 *(32)*	90.0 *(30)*	92.0 *(25)*	78.7 *(136)*
0-1000	NA	31.8 *(22)*	73.3 *(30)*	71.0 *(31)*	92.0 *(25)*	68.5 *(108)*

Age-based differences in performance were investigated for each block of the Number Line Estimation Task individually. The 0-10 block of the task was analysed at 6 and 7 years only. An independent t-test indicated significant differences in R^2_{LIN} estimates between 6 and 7 years, t (46) = .38, p = .709, d = 0.112. As described in Table 3.7, the differences were due to higher accuracy at 7 years.

As significant gender differences were reported for the 0-100 block of the Number Line Estimation Task, a two-way ANOVA was completed with age (5 levels: 6, 7, 8, 9, 10 years) and gender (2 levels: male, female) as between participant variables. As shown in Table 3.7, there was a significant effect of age group, F (4, 126) = 14.52, p < .001, η_p^2 = .315. This was explored using Games-Howell post-hoc tests. At 6 years, R^2_{LIN} estimates were significantly lower than at 8, 9 and 10 years (p < .001). No significant difference in performance between 6 and 7 years was found (p = .165). At 7 years R^2_{LIN} estimates were significantly lower than at 10 years (p < .001) but not at 8 years (p = .101) or 9 years (p = .067). No significant differences in performance were reported between 8, 9 and 10 years (*p's* > .599). No significant interaction between gender and age group was found, F (4, 126) = 0.47, p = .759, η_p^2 = .015.

R^2_{LIN} estimates for the 0-1000 block of the Number Line Estimation Task were only considered for participants aged 7 and older. As significant gender differences were reported for the 0-1000 block, a two-way ANOVA was completed with age (4 levels: 7, 8, 9, 10 years) and gender (2 levels: male, female) as between participant variables. There was a significant effect of age, F (3, 100) = 9.49, p < .001, η_p^2 = .222. Games Howell post-hoc tests indicated significantly lower scores at 7 years compared to 8 years (p = .026), 9 years (p = .015) and 10 years (p < .001). At 8 years performance was significantly lower than at 10 years (p = .046) No significant differences in performance were reported between 8 and 9 years (p = .991) or 9 and 10 years (p = .106) (see Table 3.7). There was a significant interaction between gender and age group, F (3, 100) = 3.18, p = .027, η_p^2 = .087. Follow up one-way ANOVAs investigating differences in performance across age groups were completed for males and females respectively. For males, no main effect of age was found, F (3, 46) = 1.23, p = .308, η_p^2 = .074. For females, there was a main effect of age, F (3, 54) = 10.57, p < .001, η_p^2 = .370. Games Howell post hoc tests indicated significant differences in performance between girls at 7 and 9 (p = .018), 7 and 10 (p < .001), and 8 and 10 years (p = .042).

3.3.5 Language performance

To explore age-based differences in performance on the BPVS, a one-way ANOVA was completed with age group as a between participant measure (5 levels: 6, 7, 8, 9, 10 years). A main effect of age group was found, F (4, 148) = 26.28, p < .001, η_p^2 = .415. As described in Table 3.7 the results showed improved performance with increasing age. Tukey post-hoc comparisons indicated significantly lower performance at 6 years compared to 8, 9 and 10 years (p < .001) but not compared to 7 years (p = .141). Performance at 7 years was significantly lower than at 9 and 10 years (p < .001) but not 8 years (p = .080). At 8 years there was significantly lower performance than at 10 years (p < .001) but not at 9 years (p = .184). No significant difference in performance between 9 and 10 years was found (p = .239).

Table 3.7

Descriptive statistics for mathematics and language task performance across age groups

Task	**Metric**	**6 Years**	**7 Years**	**8 Years**	**9 Years**	**10 Years**
ANS Task Accuracy	Mean ± SE	47.85 ± 1.14	56.55 ± 1.76	64.31 ± 2.24	69.05 ± 2.45	69.10 ± 2.22
	Max	57.81	78.69	89.06	89.06	92.19
	Min	34.38	43.75	43.75	40.63	45.31
No. Line 10 R^2_{LIN}	Mean ± SE	0.88 ± .03	0.89 ± .02			
	Max	0.99	0.99	NA	NA	NA
	Min	0.32	0.56			
No. Line 100 R^2_{LIN}	Mean ± SE	0.66 ± .04	0.79 ± .03	0.90 ± .03	0.91 ± .03	0.96 ± .01
	Max	0.95	0.97	1.00	1.00	1.00
	Min	0.23	0.34	0.39	0.30	0.72
No. Line 1000 R^2_{LIN}	Mean ± SE		0.57 ± .07	0.82 ± .04	0.83 ± .04	0.94 ± .02
	Max	NA	1.00	1.00	1.00	1.00
	Min		0.11	0.20	0.28	0.28
BPVS Standard Score	Mean ± SE	75.27 ± 2.76	85.45 ± 2.99	96.61 ± 2.66	106.16 ± 3.75	115.20 ± 2.94
	Max	102	129	126	139	147
	Min	42	35	61	64	73

Note. NFER PiM = National Foundation for Educational Research Progress in Mathematics; ANS = Approximate Number System; R^2_{LIN} = Linear response patterns; No. Line = Number Line; BPVS = British Picture Vocabulary Scale

3.4 Results Part B: Spatial-Mathematical Relations

As no individual participant was missing data for more than one task, and to optimise power, missing values for overall task performance were replaced by mean scores on that task for a participant's age group. To investigate the effect of mean replacement of missing data, all regression analyses were repeated using pairwise deletion (see section 3.4.2.7). Comparable results were reported.

3.4.1 Associations between task performance on different measures

Pearson correlations were used to investigate the relative associations between measures and to inform regression models. The results of bivariate correlations between all measures are outlined in Table 3.8. Significant correlations at the $p < .001$ level were reported between performance accuracy scores for all spatial measures. For mathematics measures, the NFER PiM test and the ANS Task were significantly correlated with all spatial measures and the BPVS ($p < .001$). The 0-100 and 0-1000 blocks of the Number Line Estimation Task were significantly correlated with the spatial measures and the BPVS, with the exception that the 0-1000 task was not correlated with mental rotation ($p = .080$). For the 0-10 block of the Number Line Estimation Task significant associations were found for spatial scaling ($p = .034$) and the 0-100 block of the Number Line Estimation Task ($p < .001$) only.

Table 3.8

Correlations between test measures

	Spatial Measures				**Mathematics Measures**				**BPVS**
	2	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1. Disembedding	.29***	.45***	.44***	.35***	.36***	.09	.47***	.43***	.38***
2. Mental Rotation	/	.46***	.39***	.33***	.44***	-.079	.33***	.17***	.49***
3. Spatial Scaling		/	.52***	.52***	.59***	.31*	.52***	.51***	.59***
4. Perspective Taking			/	.30***	.43***	-.01	.40***	.31***	.45***
5. NFER PiM				/	.37***	.10	.35***	.34***	.52***
6. ANS Task					/	.14	.40***	.25***	.46***
7. No. Line 10 R^2_{LIN} (*n* = 48)						/	.54***	.42	.09***
8. No. Line 100 R^2_{LIN} (*n* = 136)							/	.37***	.47***
9. No. Line 1000 R^2_{LIN} (*n* = 108)								/	.41***
10. BPVS									/

Note. * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. Unless otherwise stated $N = 155$ and percentage accuracy scores are reported. NFER PiM = National Foundation for Educational Research Progress in Mathematics; ANS = Approximate Number System; R^2_{LIN} = Linear response patterns; No. Line = Number Line; BPVS = British Picture Vocabulary Scale

3.4.2 Information on collinearity

Collinearity was assessed using Tolerance and VIF scores (Field, 2013). Collinearity statistics indicated appropriate Tolerance and VIF scores for all regression models, where a cut off of > 0.2 was used for Tolerance scores (Menard, 1995) and a cut off of < 10 was used for VIF scores (Myers, 1990) (see Table 3.9 and Table 3.10).

Table 3.9

Co-linearity analysis for each of the main regression models

Predictors	**Metric**	**Model 1**	**Model 2**	**Model 3**	**Model 4**	**Model 5**
Age (months)	Tol	0.42	0.43	/	0.47	0.54
	VIF	2.39	2.30	/	2.11	1.85
BPVS	Tol	0.49	0.49	/	0.55	0.62
	VIF	2.03	2.03	/	1.82	1.62
Gender	Tol	/	/	/	0.93	0.92
	VIF	/	/	/	1.07	1.09
Disembedding	Tol	0.66	0.66	0.80	0.67	0.53
	VIF	1.52	1.51	1.25	1.49	1.89
Mental Rotation	Tol	0.69	0.69	0.71	0.57	0.68
	VIF	1.45	1.45	1.41	1.75	1.47
Spatial Scaling	Tol	0.52	0.53	0.59	0.72	0.85
	VIF	1.91	1.89	1.69	1.39	1.18
Perspective Taking	Tol	0.58	0.63	0.80	0.54	0.58
	VIF	1.73	1.60	1.25	1.84	1.73
Age*Mental Rotation	Tol	0.87				/
	VIF	1.15				/
Age*Spatial Scaling	Tol					0.62
	VIF					1.61
Age* Disembedding	Tol					0.63
	VIF					1.59

Note. BPVS = British Picture Vocabulary Scale; Tol = Tolerance

Table 3.10

Co-linearity analysis for each of the follow-up regression models

Predictor	Metric	Model 1: Mental Rotation		Model 5: Scaling & Disembedding	
		Younger	*Older*	*Younger*	*Older*
Age (months)	Tol	0.746	0.76	0.688	/
	VIF	1.340	1.32	1.454	/
BPVS	Tol	0.745	0.71	0.763	/
	VIF	1.342	1.41	1.311	/
Gender	Tol	/	/	0.897	/
	VIF	/	/	1.115	/
Disembedding	Tol	0.773	0.91	0.754	/
	VIF	1.293	1.09	1.327	/
Mental Rotation	Tol	0.656	0.83	0.741	/
	VIF	1.525	1.21	1.349	/
Spatial Scaling	Tol	0.664	0.71	0.820	/
	VIF	1.507	1.42	1.219	/
Perspective Taking	Tol	0.828	0.64	0.629	/
	VIF	1.208	1.57	1.590	/

Note. BPVS, British Picture Vocabulary Scale, *Tol*, Tolerance

3.4.3 Identifying predictors of mathematics outcomes

Hierarchical regression models were completed for each mathematical outcome. These models investigated the proportion of mathematical variation explained by spatial skills after accounting for other known predictors of mathematical performance including language ability (the BPVS) and age. Gender was included as a control variable for mathematics tasks for which significant gender differences were reported (see section 3.3.1). All predictors were converted to z-scores prior to entry into the regression models. For all models, the control variables were added in step 1. In step 2, the spatial measures were entered together, as there was no strong evidence as to which skills might best predict different aspects of mathematical performance. In step 3 interaction terms between age and each spatial skill were added using forward stepwise entry. Only significant interactions were retained in the final models. The results reported in Table 3.11 to Table 3.15 reflect the regression statistics (*b, SE, ß, t* and *p)* for the final models (i.e., when all predictors had been entered). For all regression analyses, adjusted r^2 values are reported.

3.4.3.1 Model 1: Identifying predictors of standardised mathematics performance

Model 1 sought to determine the contribution of different spatial skills to the variation in standardised mathematics performance, as measured using the NFER PiM. As shown in Table 3.11, the final model accounted for 42.6% of the variation in mathematical achievement, adjusted R^2 = .282, F (3, 152) = 31.28, p < .001. In step 1, the control variables including age[1] and language ability were added to the model accounting for 28.2% of the variation. In step 2, the spatial measures were added to the model, uniquely predicting an additional 12.4% of the variation, Δ adjusted R^2 = .124, F (7, 148) = 18.58, p < .001. Finally, in step 3 interaction terms between each spatial skill and age were entered into the model. Only the interaction between mental rotation and age was retained. This accounted for an additional 2.0% of the variation in standardised mathematics performance, Δ adjusted R^2 = .020, F (8, 147) = 17.32, p < .001. Taken together, age, language ability, spatial scaling, disembedding and the interaction term between mental rotation and age, were all significant predictors of mathematics achievement in the final model.

The interaction was further explored graphically by plotting standardised mathematics scores against mental rotation scores for each age group (see Figure 3.13). The graph indicated a difference in the relationship between measures at 6 and 7 years compared to 8, 9 and 10 years. The sample was divided accordingly, and the regression analysis was re-run using younger (6 and 7 years; n = 60) and older groups (8, 9 and 10 years; n = 93) respectively. As shown in Table 3.11, the patterns reported for both age groups were broadly similar to the overall model, with spatial scaling

[1] Although year-group based standardised scores were used for the NFER PiM task, these scores were standardised across an entire academic year group. As such, exact age (in months) on day one of testing was also included as a predictor, to account for age-based variability within each year group

and disembedding identified as important predictors in both models. However, for younger participants mental rotation approached significance (p = .058) and the *ß* values were similar for mental rotation (*ß* = .20) compared to disembedding (*ß* = .22) and spatial scaling (*ß* = .27). This pattern was not present for the older group, where a non-significant *ß* value was reported for mental rotation (*ß* = *-.13)*.

3.4.3.2 Model 2: Identifying predictors of ANS performance

Model 2 investigated the role of spatial skills in explaining ANS performance. The final model explained 40.4% of the variation in ANS skills. As before, the control variables were entered in step 1 and explained 32% of ANS variation, adjusted R^2 = .320, F (3, 152) = 37.16, p < .001. The four spatial measures were added in step 2, accounting for an additional 8.4% of the variation, Δ adjusted R^2 = .084, F (7, 148) = 18.37, p < .001. Interaction terms between each spatial skill and age were entered in step 3. No interactions with age were retained in the final model. As shown in Table 3.12, spatial scaling and age were significant predictors in the final model.

3.4.3.3 Model 3: Identifying predictors of 0-10 number line estimation performance

In Model 3 the role of spatial skills as a predictor of R^2_{LIN} values on the 0-10 Number Line Estimation Task was explored. The control variables including gender were added in step 1, which led to a negative adjusted R^2 value (-3.6%). Hence, these variables were removed, and the regression was re-run. In the revised model, the spatial tasks were added to the model in step 1, explaining 12.6% of the variation, adjusted R^2 = .126, F (5, 43) = 2.70, p = .043. Interaction terms between each spatial skill and age were entered in step 3, however none were retained in the final model. The final model accounted for 12.6% of the variation. Spatial scaling and rotation were the only significant predictors (see Table 3.13).

3.4.3.4 Model 4: Identifying predictors of 0-100 number line estimation performance

Model 4 explored the role of spatial skills in explaining R^2_{LIN} performance on the 0-100 Number Line Estimation Task. The control variables were added in step 1 and accounted for 32.9% of the variation, adjusted R^2 = .329, F (4, 132) = 23.08, p < .001.

In step 2 spatial skills accounted for an additional 5.6% of the variation, Δ adjusted R^2 = .056, F (8, 128) = 13.05, p < .001. None of the interaction terms added in step 3 were retained in the model. As shown in Table 3.14, the final model accounted for 38.5% of the variation. Disembedding and spatial scaling were significant predictors in the final model.

3.4.3.5 Model 5: Identifying predictors of 0-1000 number line estimation performance

Model 5 explored the contribution of spatial skills to R^2_{LIN} scores on the 0-1000 Number Line Estimation Task. The control variables including gender added in step 1 explained 28.3% of the variance in task performance, adjusted R^2 = .283, F (4, 104) = 15.08, p < .001. The spatial skills added in step 2 accounted for an additional 8.6% of the variation, Δ adjusted R^2 = .086, F (8, 100) = 9.93, p < .001. In step 3 interaction terms between each spatial skill and age were added. The interaction between age and spatial scaling was retained an explained an additional 6.6% of the variation, Δ adjusted R^2 = .066, F (9, 99) = 11.32, p < .001. The interaction between age and disembedding was also retained, explaining 2.4% of the variation, Δ adjusted R^2 = .024, F (10, 98) = 11.09, p < .001. The final model outlined in Table 3.15 explained 45.9% of the variation on the 0-1000 block of the Number Line Estimation Task. Age, language ability, gender, spatial scaling, disembedding and the interaction terms (between spatial scaling and age, and disembedding and age) were significant predictors in the final model. The interactions were explored graphically (see Figure 3.13). For both spatial scaling and disembedding, the figure indicated a linear relationship with number line estimation performance at 7, 8 and 9 years. However, there was no linear relationship between these spatial skills and number line performance at 10 years. The graphs indicated that this might be due to ceiling performance on the 0-1000 block of the Number Line Estimation Task at 10 years. Alternatively, these differences may have been driven by differences in strategy use for the 0-1000 Number Line Estimation Task at 10 years. Regardless of their origins, given the different performance patterns at 10 years compared to all other age groups, it was not deemed appropriate to include all age groups in a single analysis. Therefore, the sample was divided into a younger group (7, 8 and 9 years; n = 83) and older group (10 years: *n* = 25). As shown in Table 3.15, spatial scaling and disembedding were significant predictors for the younger group. For the older group, the sample size was too small to complete regression analysis. Instead correlations were used to show that there was no significant association between spatial scaling (*r* = -.16) or disembedding (*r* = -.30) and 0-1000 number line estimation at 10 years. The limitations of this analysis are outlined in the discussion.

Table 3.11

Regression Model 1: Factors predicting standardised mathematics achievement (NFER PiM) (N = 155)

Model 1	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***Adj. R^2***	***Δ Adj.R^2***
Step 1										
Age (months)	-6.90	1.41	-0.46	-4.88	< .001	31.28	152	< .001	.282	
BPVS	7.32	1.30	0.49	5.62	< .001					
Step 2										
Disembedding	3.10	1.13	0.21	2.75	.007	18.58	148	< .001	.406	.124
Mental Rotation	0.25	1.10	0.02	0.22	.824					
Spatial Scaling	5.13	1.26	0.34	4.06	< .001					
Perspective Taking	0.77	1.20	0.05	0.64	.523					
Step 3										
Mental Rotation*Age	-2.26	0.92	-0.16	-2.45	.015	17.32	147	< .001	.426	.02

Follow Up: Younger Group	b	*SE*	*ß*	*t*	P	*F*	*df*	*p*	*Adj.* R^2	*Δ Adj.* R^2
Step 1										
Age (months)	2.40	3.36	0.07	0.71	.478	22.42	58	< .001	.417	
BPVS	7.29	1.83	0.38	3.99	< .001					
Step 2										
Disembedding	4.45	1.87	0.22	2.37	.021	15.40	54	< .001	.590	.173
Mental Rotation	3.07	1.59	0.20	1.93	.058					
Spatial Scaling	4.56	1.70	0.27	2.68	.010					
Perspective Taking	-1.61	1.77	-0.08	-0.91	.369					
Follow Up: Older Group	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***Adj.* R^2**	***Δ Adj.* R^2**
Step 1										
Age (months)	-5.47	2.41	-0.22	-2.26	.026	14.28	91	< .001	.222	
BPVS	7.19	1.72	0.41	4.19	< .001					
Step 2										
Disembedding	3.03	1.41	0.19	2.15	.034	9.78	87	< .001	.403	.181
Mental Rotation	-2.40	1.62	-0.13	-1.48	.142					
Spatial Scaling	5.19	1.72	0.30	3.01	.003					
Perspective Taking	2.08	1.59	0.14	1.31	.194					

Table 3.12

Regression Model 2: Factors predicting ANS performance (N = 155)

Model 2	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	R^2	ΔR^2
Step 1										
Age (months)	2.56	0.83	0.29	3.08	.002	37.16	152	< .001	.320	
BPVS	0.03	0.78	0.00	0.04	.969					
Step 2										
Disembedding	-0.09	0.68	-0.01	-0.13	.893	18.37	148	< .001	.404	.084
Mental Rotation	0.74	0.66	0.08	1.11	.267					
Spatial Scaling	3.11	0.76	0.35	4.12	< .001					
Perspective Taking	0.55	0.69	0.06	0.79	.429					

Table 3.13

Regression Model 3: Factors predicting R^2_{LIN} scores on the 0-10 Number Line Estimation Task (n = 48)

Model 3	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***R^2***
Step 1									
Disembedding	0.00	0.03	0.02	0.12	.902	2.70	43	.043	.126
Mental Rotation	-0.05	0.02	-0.36	-2.20	.033				
Spatial Scaling	0.08	0.03	0.55	3.11	.003				
Perspective Taking	-0.03	0.03	-0.20	-1.34	.188				

Table 3.14

Regression Model 4: Factors predicting R^2_{LIN} scores on the 0-100 Number Line Estimation Task (n = 136)

Model 4	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***R^2***	**Δ *R^2***
Step 1										
Age (months)	0.04	0.02	0.19	1.94	.054	23.08	132	< .001	.329	
BPVS	0.03	0.02	0.15	1.65	.101					
Gender	-0.05	0.03	-0.13	-1.93	.056					
Step 2										
Disembedding	0.03	0.02	0.19	2.29	.023	13.05	128	< .001	.385	.056
Mental Rotation	0.00	0.01	0.02	0.22	.825					
Spatial Scaling	0.04	0.02	0.23	2.52	.013					
Perspective Taking	0.00	0.02	0.01	0.17	.867					

Table 3.15

Regression Model 5: Factors predicting R^2_{LIN} scores on the 0-1000 Number Line Estimation Task (n = 108)

Model 5	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***R^2***	**Δ *R^2***
Step 1										
Age (months)	0.10	0.03	0.30	3.10	.002	15.08	104	< .001	.283	
BPVS	0.05	0.03	0.18	2.02	.046					
Gender	-0.08	0.04	-0.15	-2.08	.040					
Step 2										
Disembedding	0.07	0.03	0.25	2.59	.011	9.93	100	< .001	.369	.086
Mental Rotation	-0.02	0.02	-0.06	-0.77	.441					
Spatial Scaling	0.09	0.03	0.33	3.52	< .001					
Perspective Taking	0.01	0.02	0.04	0.50	.616					
Step 3										
Scaling*Age	-0.09	0.03	-0.27	-2.99	.004	11.32	99	< .001	.435	.066
Step 4										
Disembedding*Age	-0.06	0.03	-0.21	-2.31	.023	11.09	98	< .001	.459	.024

Follow Up: Younger Group	**b**	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	***Adj. R^2***	***Δ Adj.R^2***
Step 1										
Age (months)	0.05	0.05	0.11	1.04	.300	11.269	79	< .001	.273	
BPVS	0.05	0.03	0.17	1.72	.089					
Gender	-0.12	0.05	-0.22	-2.41	.018					
Step 2										
Disembedding	0.07	0.03	0.23	2.30	.024	8.712	75	< .001	.397	.124
Mental Rotation	0.00	0.03	-0.01	-0.10	.924					
Spatial Scaling	0.09	0.03	0.31	2.85	.006					
Perspective Taking	0.00	0.03	0.00	0.05	.961					

Note. B = unstandardized coefficient; SE = Standard Error; ß = standardised coefficient; NFER PiM = National Foundation for Educational Research Progress in Mathematics; ANS = Approximate Number System; R^2_{LIN} = Linear response patterns; BPVS = British Picture Vocabulary Scale

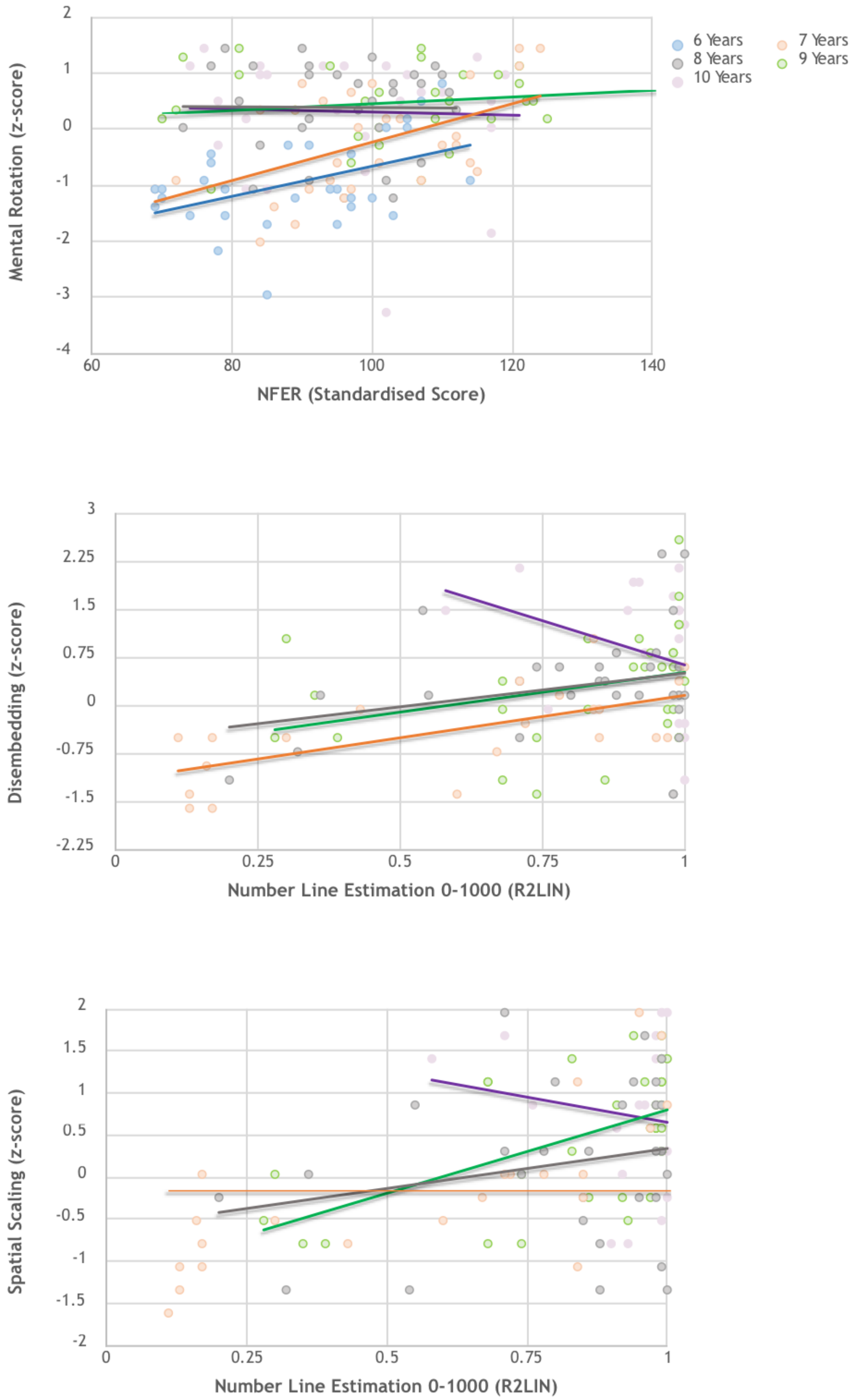

Figure 3.13. Significant interactions between age and spatial skills

3.4.3.6 Regression results using pairwise deletion

To ensure that mean replacement of missing data did not influence the results reported, each of the regression models reported above, was repeated using pairwise deletion. As shown in Table 3.16 the performance patterns when pairwise deletion was used were very similar to the patterns reported when mean replacement of missing data was used. For the 0-100 block of the Number Line Estimation Task, there was a slight difference in which predictors were significant in the final model, when pairwise deletion was used (BPVS, Scaling-age interaction) compared to mean replacement (gender). For all other models, the significant predictors were the same regardless of whether pairwise deletion or mean replacement was used.

Table 3.16

Comparison of outcomes of regression analyses based on pairwise deletion and mean replacement of missing data

	Pairwise Deletion			Mean Replacement		
	Adj R^2	***N***	***Significant predictors***	***Adj R^2***	***N***	***Significant predictors***
Model 1						
Step 1	.293			.282		
Step 2	.418			.406		
Step 3	.436	147	Age, BPVS, Scaling, Disembedding, Age*Rotation	.426	155	Age, BPVS, Scaling, Disembedding, Age*Rotation
Model 2						
Step 1	.332			.320		
Step 2	.422	144	Age, Scaling	.464	155	Age, Scaling
Model 3						
Step 1	.150	45	Rotation, Scaling	.126	48	Rotation, Scaling
Model 4						
Step 1	.354			.329		
Step 2	.416			.385		
Step 3	.434	132	Age, BPVS, Disembedding, Scaling, Scaling*age	NA	136	Age, gender, Disembedding, Scaling
Model 5						
Step 1	.300			.283		
Step 2	.401			.369		
Step 3	.486			.435		
Step 4	.509	103	Age, BPVS, Gender, Disembedding, Scaling, Scaling*age, Disembedding*age	.459	108	Age, BPVS, Gender, Disembedding, Scaling, Scaling*age, Disembedding*age

3.5 Discussion

In this study, spatial skills were identified as significant predictors of several mathematics outcomes, even after controlling for other known predictors of mathematics. The results highlight improvements in spatial task performance across development from 6 to 10 years. As discussed further below, developmental differences in spatial-mathematical relations are also evident such that some spatial tasks have a role for mathematics at all ages (spatial scaling and disembedding) while others have age specific effects (mental rotation). The study was completed with a population of children aged 6 to 10 years. The results reported provide the first known evidence on spatial-mathematical relations across consecutive age groups in primary school children.

3.5.1 Overview of findings

Addressing its first aim, this study provides developmental profiles for each of Uttal *et al.'s* (2013) spatial sub-domains from 6 to 10 years, showing that performance on all four spatial sub-domains improves with developmental age. Although not all between group comparisons were significant, for most tasks, other than mental rotation, there were increases in performance until 10 years. For mental rotation, performance plateaued at 8 years. These developmental patterns are consistent with findings from previous studies that explore the development of individual spatial sub-domains (Frick et al., 2013; Frick et al., 2014a). The current study also found subtle differences in the development of extrinsic compared to intrinsic spatial skills. For extrinsic spatial tasks (the Spatial Scaling Task and the Perspective Taking Task), there were no significant differences in performance between consecutive age groups. This suggests a gradual, steady increase in performance accuracy between 6 and 10 years that can best be observed by comparing children across a wide age range. In contrast, for the intrinsic measures (the CEFT and the Mental Rotation Task) there was significantly lower performance at 6 and 7 years compared to 8 years, with large gains in accuracy in the early primary school years and slower development thereafter. This study extends the current understanding of spatial development, as most previous studies are based on children under 8 years (Frick et al., 2013; 2014a).

These differing developmental patterns between intrinsic and extrinsic spatial skills are interesting for two reasons. First, they lend support to the intrinsic vs. extrinsic distinction in Uttal *et al.'s* (2013) model of spatial thinking. The results reported here suggest that these two spatial categories (intrinsic vs. extrinsic) may have differing developmental patterns, which may suggest that they are distinct constructs. This is supported by a recent CFA study by Mix *et al.* (2018) who also found stronger evidence for the intrinsic vs. extrinsic, compared to the static vs. dynamic distinction of spatial thinking. Second, as outlined further in the next section, some spatial-mathematical relations are age-dependent. A developmental transition in the role of intrinsic tasks, for mathematics, is proposed to occur at approximately 8 years. This rapid development of spatial thinking in the early primary school years may explain the age-dependent associations that are reported between some intrinsic spatial tasks and mathematics (both in this chapter and elsewhere, e.g., Mix et al., 2016). In short, before 8 years there appears to be substantial development of spatial thinking, particularly of intrinsic spatial skills. However, after 8 years, developmental improvement in spatial task performance is smaller and for some tasks such as the Mental Rotation Task, performance levels out.

Addressing the second aim, the findings reported indicate a significant role for spatial skills in predicting mathematical outcomes. For some spatial sub-domains, their role in predicting mathematics was consistent across age groups. Spatial skills explained 12.4% of general mathematics performance with disembedding (intrinsic-static sub-domain) and spatial scaling (extrinsic-static sub-domain) identified as significant predictors. For the ANS task, although spatial skills predicted 8.4% of the variation in performance, spatial scaling (extrinsic-static sub-domain) was the only significant spatial predictor. In contrast, spatial skills explained 12.6%, 5.6% and 8.6% of the variation on the 0-10, 0-100 and 0-1000 blocks of task respectively. Spatial scaling (extrinsic-static sub-domain) was a significant predictor for all three blocks of the Number Line Estimation Task. The study addressed its second aim, to provide evidence that different spatial sub-domains are differentially associated with mathematics outcomes.

Addressing its third aim, the findings of this study demonstrate age-dependent effects and indicate that for some spatial skills, their role in predicting mathematics changes through development. A role of mental rotation (intrinsic-dynamic sub-domain) in predicting standardised mathematics outcomes was found at 6 and 7 years only. Furthermore, mental rotation was a significant predictor of 0-10 number line estimation, which was completed at 6 and 7 years only. For the 0-100 and 0-1000 blocks of the Number Line Estimation Task, mental rotation was not a significant predictor for any age groups. These findings are consistent with Mix *et al.* (2016; 2017) and suggest a transition in the spatial skills that are important for mathematics, which occurs in middle childhood at approximately 7 to 8 years (Mix et al., 2016; 2017). Here, this transition is defined by a reduction in the role of mental rotation (intrinsic-dynamic spatial skills) for mathematics performance. As discussed further in section 3.5.2, successful performance on mental rotation tasks requires mental visualisation. Therefore, these performance patterns may reflect a reduction in the use of mental visualisation strategies in the completion of mathematics tasks at approximately 8 years.

For the 0-1000 Number Line Estimation Task (the most difficult of the three blocks of the Number Line Estimation Task) age-dependent performance patterns were also found. Static tasks including spatial scaling and disembedding were important predictors at 7[2], 8 and 9 years. No significant correlations were reported between these spatial skills and 0-1000 number line performance at 10 years. These findings may reflect another developmental shift in the role of spatial skills for mathematics performance. As suggested by Mix *et al.* (2016; 2017) at 10 years individuals may rely more heavily on verbal or VSWM strategies for mathematics performance, in place of spatial strategies. However, the correlations reported here at for children at 10 years should be interpreted with caution, as they do not control for other predictors of number line performance. Further research is required to confirm these results.

[2] Children at 6 years were not included in analysis of the 0-1000 Number Line Estimation Task

Taken together, these results support multi-dimensional models of spatial thinking (Buckley, Seery, & Canty, 2018). The four spatial predictors included in this study, measuring each of Uttal *et al.'s* (2013) four theoretically motivated spatial sub-domains, were found to have varying roles in explaining mathematics outcomes. As outlined in Chapter 2, previous studies of primary school children have typically explored associations between intrinsic-dynamic spatial tasks and mathematics. The results of this study highlight the importance of other spatial sub-domains in explaining mathematics outcomes, particularly spatial scaling (extrinsic-static sub-domain). The failure of some previous studies to find significant spatial-mathematical associations may reflect the limited spatial sub-domains assessed, or the age of the participants tested (Carr et al., 2008).

3.5.2 Mechanisms of spatial-mathematical associations

Spatial scaling was a significant predictor of all mathematics measures in this study. In line with Möhring *et al.* (2015) shared proportional reasoning requirements are highlighted here, as a likely underlying mechanism explaining these findings. For the Number Line Estimation Task, there is a clear role for proportional reasoning. For example, 28 can be positioned on a 0-100 number line with relatively high accuracy by dividing the line into 4 portions. For standardised mathematics performance, there are a range of mathematics topics that may require proportional reasoning such as reasoning about fractions or completing area and distance questions. For the ANS Task, proportional reasoning can be used to compare the ratios of the dot arrays presented. Importantly, the relations between spatial scaling and ANS performance reported in this study suggest that associations between scaling and mathematics are not caused by a symbolic number mechanism such as the Mental Number Line, as symbolic number representations are not required for dot comparison in the ANS Task. Taken together, these findings support the proposal that proportional reasoning is the underlying shared cognitive mechanism between spatial scaling and mathematics skills.

Disembedding was a significant predictor of both number line estimation and standardised mathematics performance. These associations may be attributable to shared form perception demands of these tasks. As outlined in section 1.4.2, for standardised mathematics, form perception is theoretically useful for distinguishing symbols and digits, interpreting charts, and completing multistep calculations (Mix and et al., 2016; Landy & Goldstone, 2007; 2010). For the Number Line Estimation Task and other mathematics tasks, form perception is required for the identification of numeric symbols and for interpreting and using visual diagrams.

Finally, mental rotation was a significant predictor of mathematics outcomes for younger participants only. For both standardised mathematics and the 0-10 block of the Number Line Estimation Task, mental rotation played an important role at 6 and 7 years. This suggests that there may be a developmental transition in the role of intrinsic-dynamic spatial skills for mathematics at approximately 7 to 8 years. Mental rotation is proposed to require active processing including mental visualisations (Lourenco et al., 2018; Mix et al., 2016). Hence, the findings reported here suggest that younger children may use mental models to visualise problems, including mathematics problems. For example, mental visualisations may be used to represent and organise complex word problems or mathematical relationships (Huttenlocher et al., 1994; Laski et al., 2013; Thompson et al., 2013). However, these results suggest that the use of mental visualisation strategies in mathematics is less common in older age groups. As children get older, they may learn new strategies for completing mathematics tasks, such as WM or verbal strategies, rendering mental visualisation strategies redundant. Older children may rely on mental visualisations less, as the mathematical problems that they are required to complete may not require them. Alternatively, mental visualisations may be more useful for novel mathematics tasks compared to automatic mathematics skills (Ackerman, 1988; Uttal & Cohen, 2012; Young et al., 2018). The tasks presented here may have been more novel for children in the younger age groups, thus they may have resorted to mental visualisation to solve them. For older children for whom the tasks were more familiar, other strategies such as memory strategies may have been used.

As outlined at the start of this chapter, the Perspective Taking Task was also hypothesised to recruit mental visualisations. However, this task was not a significant predictor of any of the mathematics outcomes in this study. These findings highlight an important distinction between different types of mental visualisations based on the frame of reference being transformed. This is supported by both behavioural and neuro-imaging evidence in adults. Hegarty *et al.,* (2006) found that object transformation ability and viewer/perspective transformation ability form two distinct spatial factors in adults. There is also evidence that these differing skills recruit distinct but overlapping neural systems (Broadbent et al., 2014). For example, there is evidence that visual perspective taking leads to greater activation of parieto-occipital areas compared to object-based rotation tasks (Zacks, Vettel, & Michelon, 2003). In the current study, it is proposed that different mental transformation abilities (object and viewer/perspective transformations) are differentially associated with mathematics in children. Object-based transformations such as those required for mental rotation and other intrinsic-dynamic spatial tasks are important for mathematics. However, allocentric viewer transformations (requiring imagined self-movement) as required for perspective taking and other extrinsic-dynamic spatial tasks are not, at least for the age-range measured. This is an important distinction, particularly for the design of training studies targeting mental visualisation skills.

The findings in this study provide evidence for the proposal that there are different explanations underpinning spatial-mathematical associations, depending on the mathematical and spatial sub-domains assessed (Fias & Bonato, 2018).

3.5.3 The role of control variables

This study highlights associations between language skills and mathematics performance. Accounting for spatial ability and the other control variables, vocabulary remained a significant predictor of standardised mathematics performance, and the most difficult 0-1000 Number Line Estimation Task. These results are consistent with previous findings that language skills are a significant longitudinal predictor of general mathematics achievement in the pre-school and primary school years, even after controlling for spatial ability (see Chapter 2 and LeFevre et al., [2010]). The results are also consistent with findings that for primary aged children language is a significant predictor of achievement in other STEM domains such as science, even after controlling for spatial thinking (Hodgkiss et al., 2018). Taken together the evidence suggests that language and spatial skills have distinct relations to mathematics, and STEM performance more broadly.

There were no significant differences in performance between males and females on any of the spatial tasks included in the study. Historically, other studies have reported a male advantage in spatial task performance in childhood (e.g., Carr et al., 2008; Casey et al., 2008). Like the findings of Chapter 2, the results of this study add to the literature arguing that the spatial performance of girls and boys is equivalent (e.g., Halpern et al., 2007; LeFevre et al., 2010). In the domain of mathematical cognition, a significant male advantage was found for the 0-100 (d = 0.383) and 0-1000 (d =0 .518) blocks of the Number Line Estimation Task. No gender differences were reported for the other mathematics tasks. This is consistent with previous studies that have also reported mixed findings on gender differences in mathematics skills. Some studies such as the study presented in Chapter 2 report evidence for gender differences in mathematics performance (Halpern et al., 2007; Penner & Paret, 2008) while others have found no significant gender bias for mathematics (Lindberg et al., 2010). The findings reported in this study highlight the task specific nature of mathematical performance differences. The findings suggest that the mathematics outcomes used across previous studies may account for the variable results reported.

3.5.4 Future directions and limitations

In summary, spatial skills were significant predictors of performance across all mathematics measures, explaining approximately 5 to 14% of the individual variation in performance. However, interpretation of the findings reported in this chapter must be weighed against the methodological limitations of the study, particularly limitations with the Number Line Estimation Task. Due to time constraints, performance scores for each block of the Number Line Task were each based on a relatively small number of trials. The results reported here would be strengthened by a replication study using a number line measure with a greater number of trials. Second, in this study, performance was collapsed across NP and PN trials of the Number Line Estimation Task. This assumes that similar cognitive processes are recruited for these item types. Although no significant differences in performance between NP and PN items were found in this study, future research is needed to investigate the impact of item type on number line estimation. Third, as outlined in section 3.4.2.5, for the 0-1000 block of the Number Line Estimation Task, ceiling performance may have been reached at 10 years. However, as outlined in section 3.2.6, one sample t-tests indicated that performance at 10 years was significantly below ceiling on this task. These findings suggest that one-sample t-tests against ceiling performance are not a good method of measuring ceiling performance on number line estimation tasks. New methods of establishing floor and ceiling performance on number line tasks are required. These methods should incorporate both R^2_{LIN} and PAE measurements. Determining cut-off points for floor and ceiling effects should be completed through the collaboration of experts in the field. It should be informed by establishing age-based standardised scores of number line estimation (both R_2^{LIN} and PAE performance) across number line ranges. This would allow determination of a. the individual variation on number line performance that is expected within age groups, b. based on the variances observed, what range number line tasks are suitable for different age groups of participants. Age-based standardised scores would allow the identification of outliers, e.g., participants scoring more than two standard deviations above (ceiling performance) or below (floor performance) the mean for their age on a given block of the number line task.

There was also insufficient power to complete the desired analysis for the 0-1000 block of the Number Line Estimation Task at 10 years. Hence, the results reported for the 0-1000 block of the Number Line Task at 10 years do not control for age or gender. Using a larger sample of children at 10 years, these findings should be replicated. Similarly, the results for 0-10 number line estimation were also slightly under powered and should be replicated. Despite the weaknesses outlined above, the findings in this study are strengthened by the fact that the patterns of performance reported are consistent with other studies of spatial-mathematical relations in children of different ages (Mix, 2016; 2017).

Although charting the development of spatial skills is not the main aim of this thesis, the results reported in this chapter provide insights into the development of different spatial sub-domains across middle childhood. However, these findings are not withstanding limitations. The inferences made based on post-hoc comparisons should be interpreted cautiously. Although significant developmental differences are reported, there is also substantial individual variation in spatial task performance with age-groups. The cross-sectional design used in this study does not allow for the comparison of individual performance patterns across time. The findings in this domain would be enhanced by longitudinal research following a single cohort across the ages of 6 to 10 years. Furthermore, the results reported for extrinsic-dynamic spatial skills may have be influenced by the relatively high cognitive load of the Perspective Taking Task used in this study. The complexity of the task instructions may have influenced performance for younger children. Furthermore, beyond spatial skills, successful performance on this task may require attention, inhibition, memory skills and switching between different levels of representation. Therefore, the spatial performance scores reported for this task may be heavily influenced by other cognitive abilities. The results reported for extrinsic-dynamic skills in this study would be strengthened by replication using other extrinsic-dynamic tasks such as navigation tasks.

Finally, the causal inferences that can be drawn from the results reported in this chapter are limited. Although the findings provide important insights on the specificity of associations between spatial and mathematical sub-domains, at specific ages, the direction of these associations are undefined. For example, it is unknown whether spatial skills influence mathematics outcomes, whether mathematical skills influence spatial thinking or whether there is a bidirectional relationship between the skills. Having established the associations between spatial and mathematical thinking, the next logical step is to explore the causal relationship between these variables using training. The causal relationship between spatial and mathematical thinking is addressed in Chapter 4. The mechanisms underpinning spatial-mathematical relations suggested in this chapter also provide a platform from which the training study in Chapter 4 is designed.

3.5.5 Conclusion

This study extends previous findings by comparing the role of Uttal *et al.'s* (2013) four spatial sub-domains in predicting mathematics outcomes. Overall, spatial skills explained 5 to 14% of the variation across three mathematics performance measures, beyond other known predictors of mathematics. Spatial scaling (extrinsic-static sub-domain) was a significant predictor of all mathematics outcomes, across all ages, highlighting its importance for mathematics in middle childhood. Other spatial sub-domains were differentially associated with mathematics in a task and age-dependent manner. For example, mental rotation (intrinsic-dynamic sub-domain) was a significant predictor of mathematics at 6 and 7 years only, which suggests that at approximately 8 years of age there is a transition period regarding the spatial skills that are important for mathematics. This study emphasises the importance of choosing theoretically motivated, task and age sensitive targets for spatial training, to elicit transfer of training gains. The effects of such training on both spatial and mathematics outcomes, are explored in Chapter 4.

Chapter 4 Effective spatial training for near-transfer to spatial performance and for far-transfer to a range of mathematics skills at 8 years

4.1 Introduction

The results reported in Chapter 3 suggest that training spatial thinking could confer benefits for both spatial and mathematics outcomes. There are mixed findings on the transfer of training gains (to untrained skills) in other cognitive domains such as WM (for a review see Melby-Lervåg et al., 2016). However, far transfer of training gains may be constrained by an understanding of the underlying cognitive mechanisms of training targets. It is proposed that the task and age-dependent explanations for spatial-mathematical associations outlined in Chapter 3, strengthen the likelihood of achieving far transfer of gains from spatial to mathematics domains in the training study outlined in this chapter. The study outlined in this chapter sought to investigate the impact of spatial training on the spatial skills targeted in training (near transfer), un-trained spatial skills (referred to here as intermediate transfer) and mathematics skills (far transfer).

4.1.1 Rationale for the study

The proposal that spatial training interventions can improve mathematical ability in children is supported by evidence that spatial ability is malleable, and that there are significant associations between spatial and mathematics skills in childhood. Spatial thinking is one aspect of cognition that appears to be particularly amenable to change through intervention (Baenninger & Newcombe, 1989; Uttal et al., 2013). Uttal *et al.* (2013) reported an effect size of almost one half a standard deviation for training studies that compared spatial training to control conditions (*Hedges G* = .47). The effect size increased to 0.61 (*Hedges G*) when the analysis was limited to studies of children under 13 years (n = 53 studies). This demonstrates particular malleability of spatial thinking in childhood. There is also convincing evidence that spatial and mathematical thinking are associated longitudinally in childhood. Spatial thinking at 3 years, measured using the TOSA, predicted 27% of the variation in mathematics problem solving at 5 years (Verdine et al., 2014). Similarly, as outlined in Chapter 2, Pattern Construction skills at 5 years explained 8.8% of the variation in mathematics performance at 7 years.

There is convincing evidence that spatial-mathematical relations are specific to certain spatial and mathematics tasks and that spatial-mathematical relations differ across development (Fias & Bonato, 2018). In Chapter 3, spatial scaling was reported to be the strongest spatial predictor of standardised mathematics performance at 6 to 10 years when compared to perspective taking, disembedding and mental rotation. Mental rotation had an age-dependent role at 6 to 8 years only. Similar age-dependent findings were reported by Mix *et al.* (2016; 2017) who found that mental rotation was a significant predictor of mathematics performance at 6 years but not at 9 or 11 years. Taken together, the selection of spatial sub-domains for training studies should reflect the facts that a) not all spatial skills are equally associated with all mathematics outcomes and b) spatial-mathematical associations are developmentally sensitive.

This study included participants aged approximately 8 years. As outlined above, there is evidence of significant spatial-mathematics relations at this age (see Chapter 3). Furthermore, as described in the next section, this age range overlapped with other spatial training studies that investigated transfer of gains to mathematics (Cheng & Mix, 2014; Hawes et al., 2015). Additionally, children of this age were deemed old enough for independent computer-based training.

4.1.2 Transfer of spatial training gains to mathematics

As outlined in section 1.4.2, few studies have investigated transfer of gains from spatial training (with no mathematical component) to mathematical skills. Significant gains have been reported in both mental rotation performance (near transfer) and mathematical calculation skills (far transfer) following 40-minutes of mental rotation training at 6 to 8 years (Cheng & Mix, 2014). For mathematical calculation, gains were found for missing term arithmetic problems only. In a similar mental rotation training study, Hawes *et al.* (2015) failed to replicate these findings and reported no far transfer of spatial training gains in children of the same age.

As outlined in Chapter 1, these differing results may be explained by several factors. First, individual and group training were delivered by Cheng and Mix (2014) and Hawes *et al.* (2015) respectively. Without the direct supervision of a researcher, reduced engagement with training may have contributed to the results of the Hawes *et al.* (2015) study. The role of motivational factors including participant engagement in training is explored further in section 4.1.4. Second, post-testing was delivered immediately following training by Cheng and Mix (2014), while Hawes *et al.* (2015) delivered post-testing one week after training. Thus, caution must be taken in assuming that the gains reported by Cheng and Mix (2014) are durable. Third, the training method differed between the two studies. Hawes *et al.* (2015) used implicit instruction. Points were awarded for correct trials, but no instructions were given to explain correct (or incorrect) answers. In contrast, Cheng and Mix (2014) used explicit instruction by giving participants physical manipulatives (mirroring those included in the onscreen trials) and instructing them to move the shapes to check their answers.

Differences in the training modes used in the above two studies reflect a broader distinction between explicit and implicit instruction types. Both explicit and implicit instruction fall into the broader category of direct training (i.e., they involve task specific training). In this study, implicit instruction is defined as instruction in which students are not aware of learning and use their experiences to construct an understanding. In contrast, for explicit instruction, the instructor plays a key role in explaining concepts to students and the student is aware of the skill or knowledge being taught. There is mixed evidence regarding the effectiveness of explicit and implicit instruction in learning more generally (Kirschner, Sweller, & Clark, 2006). However, no known spatial training studies compare the efficacy of implicit and explicit instruction. Most studies of children have demonstrated the effectiveness of spatial training using implicit training, for example, where participants complete task practice with feedback (Uttal et al., 2013). Instructional videos are one tool that can be used to deliver explicit instruction. There is evidence that viewing an instructional video of successful task completion can improve subsequent performance in number line estimation and spatial cross-sectioning in adults (Cohen & Hegarty, 2014; Gallagher-Mitchell, Simms, & Litchfield, 2018). The success of instructional videos may be attributable to observational learning (Castro-Alonso, Ayres, & Paas, 2014; Paas & Sweller, 2012). In particular, for spatial thinking, instructional videos may activate the mirror neuron system as individuals imagine movements (Rizzolatti & Sinigaglia, 2010; Tettamanti et al., 2005). From a practical perspective, instructional videos could offer a novel, practical method of introducing spatial thinking into the classroom. This study compared the efficacy of explicit and implicit spatial instruction for the first time.

4.1.3 The selection of training targets

The findings reported in Chapter 3 emphasise the importance of choosing theoretically motivated, task and age sensitive, targets for spatial training. Mental rotation and spatial scaling were targeted for training in this study. These skills have previously been associated with mathematics achievement in children aged 6 to 8 years. Furthermore, as investigated in Chapter 3, underlying cognitive mechanisms have been proposed that may explain associations between these spatial skills and mathematics outcomes (e.g., Mix et al., 2016; 2017).

In Chapter 3 spatial scaling was highlighted as a particularly useful target for spatial skill training as it was a significant predictor of mathematics across a range of outcomes ($.23 < \beta < .55$). Here we propose two reasons to explain these associations. First, there is a proposed underlying mechanism (proportional reasoning) linking certain mathematics tasks (e.g., number line estimation and ANS performance) to spatial scaling. Based on this proposal, there is no theoretical reason to predict that spatial scaling would be associated with all mathematics tasks, particularly those with no proportional reasoning requirements, e.g., multi-digit calculation. Second, in spatial scaling tasks participants are required to compare two differently scaled spaces (i.e., it is an extrinsic-static task). However, in the context of an individual object, scaling can also be viewed as an object transformation, i.e., expanding or contracting an object (Newcombe & Shipley, 2015). Object transformations like this are required in intrinsic-dynamic spatial tasks. In this way, spatial scaling tasks may elicit both proportional reasoning and mental visualisation, two underlying cognitive processes that are required for different mathematics tasks.

The results of Chapter 3 also highlight mental rotation and disembedding as potential spatial training targets for some but not all aspects of mathematics at certain ages. In this study, mental rotation was selected as a training target for two main reasons. First, there is a proposed underlying mechanism explaining associations between mental rotation and mathematics outcomes. Specifically, mental rotation is proposed to elicit active processing, including mental visualisation and manipulation of objects (Lourenco et al., 2018; Mix et al., 2016). Thus, mental rotation training may have benefits for mathematics tasks requiring the mental manipulation or organisation of numbers. The second reason mental rotation was selected as a training target in this study was so that meaningful comparisons could be made between this study and previous studies in this domain, all of which administered mental rotation training. Although for practical reasons it was not chosen as a training target in this study, future research could also explore the effects of training disembedding skills on mathematics outcomes.

4.1.4 Motivational factors in training studies

One original aspect of this study is that it controlled for motivational factors including engagement with, and expectations of, spatial training. These factors may explain the mixed successes reported in previous cognitive training studies (Green et al., in press; Strobach & Karbach, 2016). As outlined by Green *et al.* (in press) there is a lack of research into the role of expectation effects in driving gains in cognitive training, and how best to measure them. Also referred to as placebo effects, expectation effects occur when the expectation that a training programme (intervention) will work, induces gains, independent of training content (Green et al., in press). While studies assessing the placebo effect are common in medicine (e.g., Finniss, Kaptchuk, Miller, Benedetti, 2010), only a small number of training studies in the domain of cognitive psychology, explore the influence of expectation effects. For example, despite completing identical cognitive training, Foroughi, Monfort, Paczynski, McKnight, and Greenwood (2016) reported gains in an adult placebo group following recruitment using a suggestive flyer that eluded to gains following cognitive training, but no gains were reported in a control group who were recruited with a non-suggestive flyer. This suggests that gains were due to the suggestive recruitment method and not the training itself. Furthermore, adults who believe that intelligence is malleable have better academic and cognitive outcomes (Dweck, 2000), in addition to larger gains in intelligence tests following WM training (Jaeggi, Buschkuehl, Shah, & Jonides, 2014). Despite evidence for expectation effects in adults, no known studies explore expectation effects of cognitive training in child populations.

The neural underpinnings of expectation effects are unknown. Gains associated with expectation effects may lead to changes in brain plasticity or may merely improve test-taking (Green et al., in press). For studies aiming to design training paradigms that generate optimum gains for participants, harnessing the power of expectation effects may be a valuable mechanism for cognitive enhancement (Green et al., in press). However, from a mechanistic perspective, expectation may act as a confound in cognitive training studies (Foroughi et al., 2016). Even in studies with an active control group, there is no guarantee that expectations of training will be equivalent across groups. Completing blinded interventions in cognitive psychology is difficult and participants are often aware that they are in the active control or treatment group respectively. This may influence their expectations of training. Thus, failure to control for differences in expectations is perceived by some to be a fundamental design flaw in training studies (Boot, Simons, Stothart, and Stutts, 2013). By controlling for expectation effects, the causal inferences made in this cognitive training study are enhanced (Boot et al., 2013).

Engagement with training and compliance with training protocols is another factor that may influence the outcomes of cognitive training (Hawes et al., 2015). It has been shown that participants who persist with WM training are more likely to improve (Shah, Buschkuehl, Jaeggi, & Jonides, 2012), and those who show higher levels of engagement with WM training are more likely to exhibit training gains (Jaeggi et al., 2014). Both the intrinsic motivation of individuals and extrinsic motivational features of a given training paradigm influence task engagement (Jaeggi et al., 2014). Previous research on intrinsic motivation in classroom learning shows that "academic engaged time" or "time on task" is a significant predictor of children's academic outcomes (Berliner, 1979; Denham & Lieberman, 1980). For extrinsic motivation, design elements of game-based training such as displaying prizes, certificates or high scores on screen, can increase motivation to complete training and improve engagement (Holmes, Gathercole, & Dunning, 2009; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Katz, Jaeggi, Buschkuehl, Stegman, & Shah, 2014; Wang, Zhou, & Shah, 2014). Overall, differences in the degree to which participants engage in training, may influence the reported success of training paradigms. By measuring and controlling for participant engagement, the rigour of this study is substantially stronger. It was possible to determine the extent to which cognitive training gains are attributable to training over and above differences in participant engagement.

In short, there is evidence that expectations of, and engagement with, training may influence training outcomes, and that the inclusion of an active control group is insufficient as a control measure for these effects. The inclusion of an active control group as a control for motivational factors assumes that, both training and control conditions are equally engaging, and that participants are unaware of which treatment condition they are in. These assumptions weaken the conclusions of training studies. Controlling for expectation and engagement effects in this study strengthens the causal inferences made.

4.1.5 Causality and training studies

Most cognitive training studies are founded on reports of significant correlations between the skill being trained, and the skill to which transfer is expected. However, as outlined in section 1.4.3, despite strong correlations between cognitive and academic skills, far transfer of training gains from cognitive training such as WM training to academic outcomes is not always observed (e.g., Melby-Lervåg et al., 2016). To move from correlation studies to designing meaningful interventions there is a need to explore more deeply what a correlation between two factors might indicate. For example, significant positive correlations have been reported between performance on mathematical arithmetic tasks and mental rotation in pre-school children (e.g., Verdine et al., 2014). As outlined by Reichenbach (1956), these correlations may be explained by various causal models, the most basic of which are outlined overleaf :

a) Mental rotation performance is a *cause* of arithmetic performance

b) Arithmetic performance is a *cause* of mental rotation performance

c) A common *cause* exists between both arithmetic and mental rotation, e.g., attention or a genetic factor

Failure to find transfer of training gains from mental rotation training to arithmetic performance may be because the causal relationship between these factors is best explained by models b or c above. In this way, training studies offer insight into the causal relationships between cognitive factors, moving beyond correlational findings. Determining a direction of causality between cognitive skills is challenging. The current study provides some of the first evidence on the causal relationship between spatial skills (mental rotation and spatial scaling) and mathematics outcomes.

Although it is not the main focus of this study, it is also important to consider the role of development in associational studies of cognitive skills. If spatial skills have a causal role in arithmetic performance, developmental timing may also be a factor. Consider the correlations between arithmetic and mental rotation described above. On one hand, mental rotation may play a role in the execution of arithmetic tasks. For example, when presented with equations in non-prototypical formats, individuals may mentally rotate these equations to a more favourable orientation. If this is the case one would expect that mental rotation training would improve subsequent arithmetic performance. The impact of spatial training on the execution of mathematics skills is investigated in this study. On the other hand, significant correlations between mental rotation and arithmetic may reflect a role for mental rotation in the acquisition and learning of new arithmetic material (Mix et al., 2016). In this case, one would not expect that mental rotation training would lead to immediate gains in arithmetic performance, unless participants were asked to learn new arithmetic skills. The current study does not explore the effect of spatial training on the acquisition of new mathematics skills.

4.1.6 Current study

The study presented in this chapter compared explicit and implicit instruction methods for training spatial skills in children aged 8 years. It explored transfer of spatial training gains to other spatial and mathematical domains. Explicit instruction was delivered using instructional videos, which were specifically designed for use in this study. To identify the causal relationship between spatial and mathematical thinking, the spatial training intervention used in the study was not embedded within a mathematical context. The choice of spatial scaling and mental rotation as spatial training targets was supported by both theoretical and behavioural evidence. The effectiveness of the intervention was assessed in the context of near, intermediate and far transfer of gains, whilst also controlling for expectation and engagement effects.

4.2 Materials and Methods

4.2.1 Participants

The sample size for this study was determined using GPower. Based on the studies presented in Chapter 2 and Chapter 3 which also explored the roles of spatial thinking for mathematics, a medium effect size was expected (f = .25). The power analysis was based on the largest ANOVA in this study. This included two between participant variables, training mode (2 levels: explicit, implicit) and training type (mental rotation, spatial scaling, literacy), and one within participant variable, time (2 levels: pre-training, post-training). To achieve power of 0.8, power analysis indicated that a minimum of 158 participants were required. As the study design included data collection at two-time points, it was anticipated that there would be some participant drop-off between Time 1 and Time 2. Therefore, the sample size was increased to account for possible attrition of the sample. Participants were 250 children from six primary schools across London, UK. All participants were in Year 3 (M_{age} = 8.09 years, SD = .41 years). The proportion of males (48%) and females (52%) was approximately equal.

4.2.2 Study Design

The UCL, IOE Department of Psychology and Human Development granted ethical approval for this study. Upon receiving school permission, opt-out consent forms were sent to all parents/guardians requesting permission for children to take part in this study. Furthermore, prior to taking part, all participants were also given an age-appropriate verbal description of the study and were informed that they could withdraw from the study at any time. All researchers involved in data collection held a Disclosure and Barring Service Clearance Certificate.

The study used a randomised, controlled, pre-post training design. All participants completed an identical battery of tasks one week pre- training ± 1 day (Time 1), and immediately post-training (within 5 minutes) (Time 2). All tasks and training procedures were computer-based and were delivered using *Gorilla*, an online testing platform (www.gorilla.sc). Participants were randomly assigned to one of six training groups using the randomisation function on the *Gorilla* platform. The task battery included two spatial measures, assessing mental rotation and spatial scaling respectively. These measures were included as potential targets of near transfer (spatial tasks trained on) and of intermediate transfer (untrained spatial tasks). Three mathematics measures were included as potential targets for far transfer (missing term problems, a number line estimation task and a geometry task).

To assess the role of motivational factors, two participant engagement measures, a pre-training expectations of training measure and a post-training engagement with training measure, were also administered. The order of task presentation for pre- and post-testing was randomised across participants. Participants completed testing in their school IT suites in groups of 6 to 8 participants supervised by at least one (but usually two) researchers. Sessions 1 and 2 were 45 and 60 minutes respectively, with breaks. All task instructions were incorporated into the *Gorilla* platform and were presented to participants using earphones. Participants moved through the task battery at their own pace. Motivational screens were presented at fixed intervals to encourage participants. These screens were presented independently of performance.

4.2.3 Training Procedures

Training groups differed by training mode (explicit vs. implicit) and training type (mental rotation vs. spatial scaling vs. control). Participants in explicit training conditions received explicit instruction of how to complete the task presented. As shown in Table 4.1, approximately equal numbers of participants were allocated to each group. For both implicit and explicit instruction, training lasted between 3 and 4 minutes. For implicit instruction, the length of training was dependent on each participants' performance (i.e., the speed taken to complete the items). For some participants in the implicit instruction group, training lasted up to 6 minutes.

Table 4.1

Number of participants in each training group

Training Type	**Training Mode**		
	Explicit	Implicit	Total
Mental Rotation	44	42	86
Spatial Scaling	41	43	84
Control	41	39	80
Total	126	124	250

4.2.3.1 Explicit Instruction

Three of the training groups viewed instructional videos that provided explicit task instructions. Two groups watched videos with spatial content, while the control group watched a video on word reading. The videos were specifically designed for use in this study using Vyond (www.vyond.com). All non-training content was uniform across videos, e.g., the characters, storyline and narration. The videos can be accessed using the links provided below. Group 1 viewed the instructional mental rotation video. Participants in this group were given a description and viewed eight examples of mental rotation (see Figure 4.1 for a screenshot). For more details go to https://youtu.be/18iyRsvtGAQ. Group 2 viewed the instructional scaling video, in which a description of spatial scaling, and eight examples of spatial scaling were shown (see Figure 4.2 for a screenshot). For more details go to https://youtu.be/grhxFEqgz51. For Group 3, the control video was shown. Participants watched eight examples of word-picture matching, in which the onscreen characters selected the correct picture to match a given word (see Figure 4.3 for a screenshot). Participants allocated to the control group did not view any spatial instruction. For more details go to https://youtu.be/qDmgRR2RLyE.

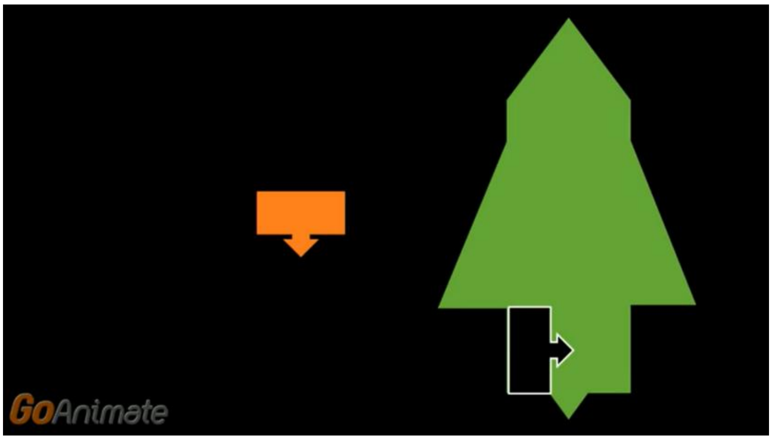

Figure 4.1. Screenshot taken from the instructional video of mental rotation (explicit instruction)

Figure 4.2. Screenshot taken from the instructional video of spatial scaling (explicit instruction)

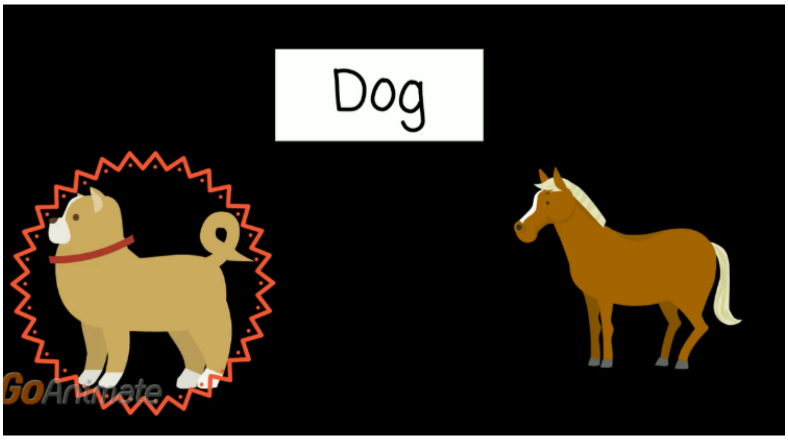

Figure 4.3. Screenshot taken from the control instructional video (explicit instruction)

4.2.3.2 Implicit Instruction

The three implicit training groups completed task practice with computer-based feedback. For each trial, participants were shown an onscreen tick or cross indicating the accuracy of their response. For incorrect trials, participants were given the opportunity to repeat the trial until they had selected the correct answer (all tasks had two possible response options). Participants were not given any explicit instruction on how to complete the trials. Participants moved to the next trial when the correct response was selected. For implicit training, two groups completed spatial tasks (the same tasks presented at pre and post testing), while the control group completed a word reading task. The number of trials included in implicit training was determined as the approximate number of trials that could be completed in the same length of time as the instructional videos described in 4.2.3.1. This was established through piloting. Group 4 completed implicit mental rotation training and were presented with 30 trials of the Mental Rotation Task on which they received feedback (further details of this task are outlined in 4.2.4.1). Group 5 completed implicit spatial scaling training comprising of 24 trials of the Spatial Scaling Task (further details of this task can be found in 4.2.4.2). Feedback was given for each trial. Group 6 completed implicit control training. These participants completed 30 trials of a word-picture matching task in which they were asked to match a word to one of two pictures by using labelled keys on the keyboard (see Figure 4.4). This was a reading task requiring minimal spatial skills. Feedback was provided.

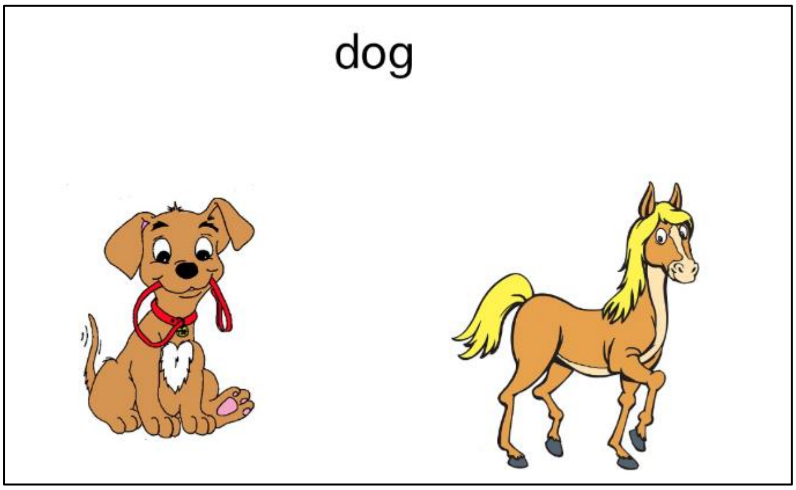

Figure 4.4. Sample trial from the control training task (implicit instruction)

4.2.4 Tasks and Measures

4.2.4.1 Mental Rotation Task

The Mental Rotation Task used in this study was similar to the task described in Chapter 3. In the current study, as this task was completed at least twice by each participant, modifications were made with the aim of improving participant's interest in the task. The monkey images used as stimuli in Chapter 3, were replaced with five other animal stimuli (dog, horse, zebra, elephant and lion) taken from Neuburger *et al.* (2011). All images covered an approximately equal surface area with equal numbers of animals facing the left and right side respectively.

In each trial of the Mental Rotation Task participants were required to identify which of two animal images located above a horizontal line matched the target image below the line. As shown in Figure 4.5, the images above the line included a mirror image of the target image, and a version of the target image rotated by a fixed degree from the target image. Participants used labelled keys on the computer keyboard to respond. Trials were separated by a fixation dot displayed for 500 milliseconds. Participants completed four practice trials at 0° where feedback was provided. For incorrect trials, participants were given the opportunity to answer the trial again.

Only participants achieving at least 50% in the practice trials, on their first attempt, continued to the 40 experimental trials. In practice, all participants achieved above 50% in the practice trials. The practice trials were followed by 40 experimental trials. No feedback was given for experimental trials at pre or post testing. The experimental trials included equal numbers of clockwise and anti-clockwise rotations at 45°, 90° and 135° (eight trials for each degree of rotation), and eight trials at 180° and 0°. For all analysis, performance on clockwise and anti-clockwise trials was collapsed (i.e., all 90° and - 90° trials were collapsed). The order of trial presentation was randomised for each participant. Trials were also counter balanced. Equal numbers of correct answers were presented on the left and right-hand side of the screen respectively. Each animal stimulus was presented at each degree of rotation with equal frequency. Percentage accuracy was recorded.

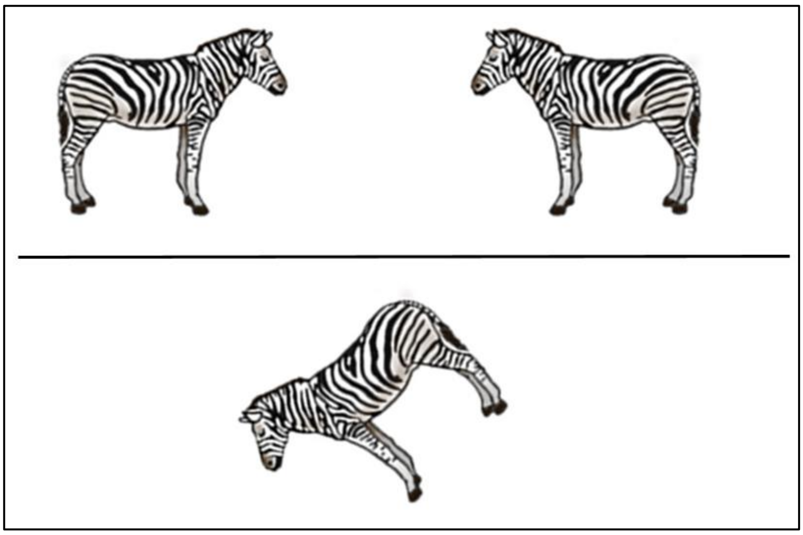

Figure 4.5. Sample item from the Mental Rotation Task (45° anti-clockwise trial)

4.2.4.2 Spatial Scaling Task

The Spatial Scaling Task designed for use in Chapter 3 was not suitable for use in this study as it was not entirely computer-based. Thus, a computer-based measure was required. The Spatial Scaling Task used in this study was modified from Möhring *et al*. (2016). In each trial participants were shown two 1D images of a circular space (a farmer's field) containing a target (an egg). Participants were asked to identify whether the eggs in the two fields were in the same position or in different positions (see Figure 4.6). For half of the trials, the targets were presented in the same position in both fields (match trials). For the remaining trials, the position of the target in one field was adjusted by 2cm (to the left or right) relative to the second field (mismatch trials). Participants responded using labelled keys on the computer keyboard. All trials were separated by a fixation dot displayed for 500 milliseconds. Participants completed six practice trials during which feedback was given and no time limit was imposed. The practice trials included 1 match and 1 mismatch trial at a scaling factor of 1, 0.625 and 0.375 respectively. Only participants achieving at least 50% in the practice trials continued to the experimental trials. In practice all participants achieved over 50% accuracy in the practice trials. The practice trials were followed by 72 randomly presented experimental trials. For pre and post testing no feedback was given for experimental trials. In line with the original protocol from Möhring *et al*. (2016), each trial was displayed for 5 seconds. Experimental trials differed by the location of the target on the horizontal axis, and by scaling factor. Six different target positions were included, a modification from the original study where 15 positions were used. Scaling factor was manipulated by keeping the size of one space constant while expanding the size of the second. In this way six scaling factors were included (1, 0.875, 0.75, 0.625, 0.5, and 0.375). Performance was measured as percentage accuracy.

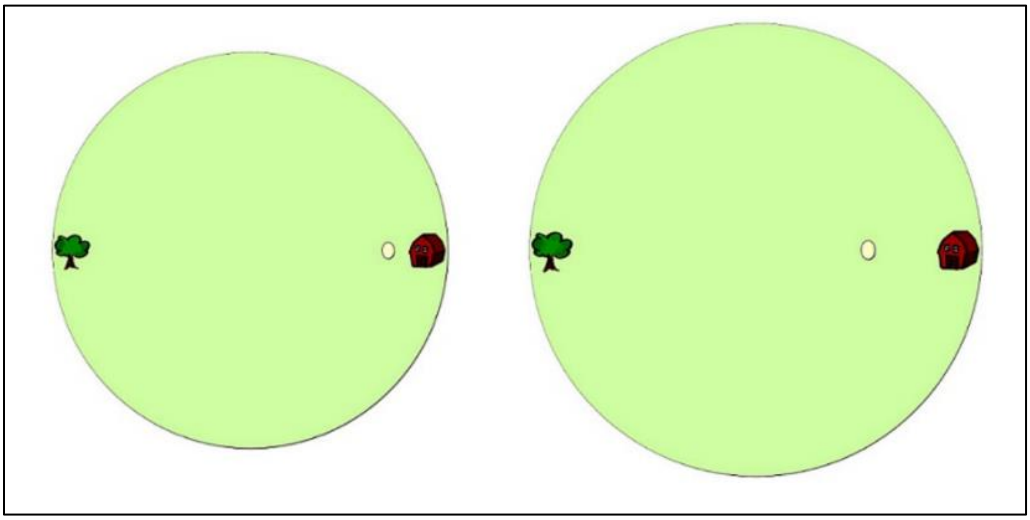

Figure 4.6. Sample mismatch trial at a scaling factor of 0.875 from the Spatial Scaling Task, taken from Möhring *et al.* (2016).

4.2.4.3 Missing Term Problems

The Missing Term Problems included in this study were modified from Hawes *et al.* (2015). For each item participants were required to complete the missing number(s) in a simple mathematical equation (see Figure 4.7). This task included 2 practice items where the solutions were shown after participants submitted an answer. Following this, 21 test items were displayed. No solutions were shown for these items. Test items included the original 18 items from Hawes *et al.* (2015) and three additional, low-difficulty items that were added to the task after piloting to alleviate floor effects. Items were presented in order of increasing difficulty and a time limit of 25 seconds was allocated to each test item. Approximately equal numbers of addition vs. subtraction items, and single vs. multi-digit numbers were included. The position of the missing box was also balanced across items. Performance on this task was measured as percentage correct.

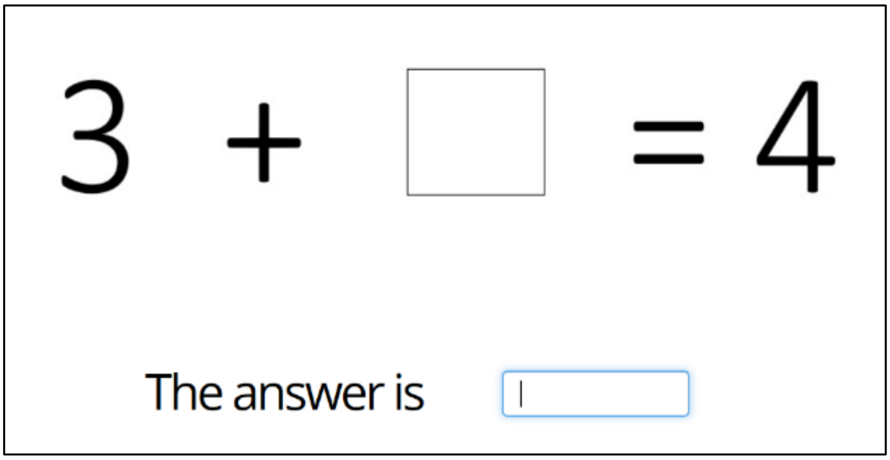

Figure 4.7. Sample Missing Term Problem

4.2.4.4 Number Line Estimation Task

Similarly to Chapter 3, a number line estimation task was used to measure symbolic numerical representations. The method of this task was adapted from Chapter 3, in order to address some of the limitations outlined in section 3.5.4. As this study had a relatively narrower age range of participants, compared to the sample of participants in Chapter 3, a 0-100 range number line was deemed suitable for all participants. Using a 0-100 scale, neither floor nor ceiling effects in performance were expected for children at 8 years. One limitation of the number line protocol used in Chapter 3 was the small number of trials administered for each block of the task. To address this limitation, the Number Line Estimation Task in this study included 30 trials on a number line ranging from 0-100. A second limitation of the protocol used in Chapter 3 was the use of both NP and PN type items. To reduce any possible confounding effects of item type (NP or PN), and in line with other studies that measure number line estimation in children (e.g., Simms et al., 2016), all trials included in the Number Line Estimation Task in this study were NP items.

As shown in Figure 4.8, for each item, participants were presented with a target number and were asked to estimate its location on a number line by using the mouse cursor to click on the number line at their selected location. For practice items (n = 2), solutions (50, 20) were shown onscreen after participants attempted an answer. No solutions were given for experimental items (n = 30). The target numbers included in the experimental items of the task (2, 6, 7, 13, 16, 19, 24, 27, 28, 35, 37, 38, 42, 46, 49, 54, 58, 59, 61, 63, 67, 71, 74, 79, 82, 83, 86, 91,92, 95) were taken from Gallagher-Mitchell, Romero-Rivas, Rodriguez-Cuadrado, and Dackermann (2017) . The order of experimental items was randomised. Performance was measured using PAE scores and curve estimation (see section 1.2.2).

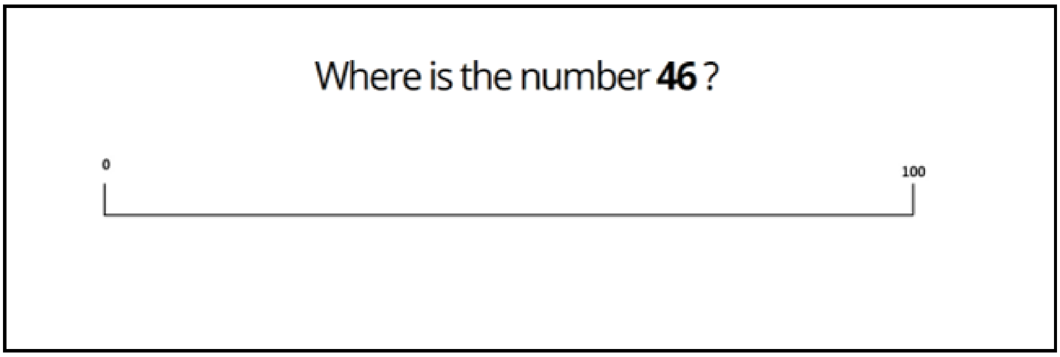

Figure 4.8. Sample item from the Number Line Estimation Task

4.2.4.5 Geometry Task

A geometry task was designed for use in this study. It was based on the statutory geometry learning requirements for Year 2 students in the UK (UK Department for

Education, 2013). The task included two item types, Shape Items and Symmetry Items. For Geometry Shape Items, participants were shown an image of a shape and were asked to select the correct number of sides (or faces) on the shape from four possible response options (see Figure 4.9). Participants completed a single practice item using a 2-D shape on which they were given feedback. All participants successfully completed this item. Geometry Shape Items differed in the dimensionality of the images shown and included six 2-D shapes and six 3-D shapes. Performance was measured as percentage accuracy collapsed across all items.

For each Geometry Symmetry Item, a target shape was displayed on screen and participants were asked to select which of four possible response options was the mirror image of the target shape (see Figure 4.10). Participants completed a single practice trial in which they received feedback. All participants successfully completed this item and continued to ten, randomly presented experimental items. For each item, the distractor images included a match error, a shape error and a symmetry error (see Figure 4.10). For match errors, the distractor was identical in both shape and position to the target shape (a). For shape errors, the distractor was in the correct position, however the shape was not a mirror of the target image, but another similar shape (b). Finally, for symmetry errors the distractor was the correct shape, but was in an incorrect position (c). Performance accuracy was recorded.

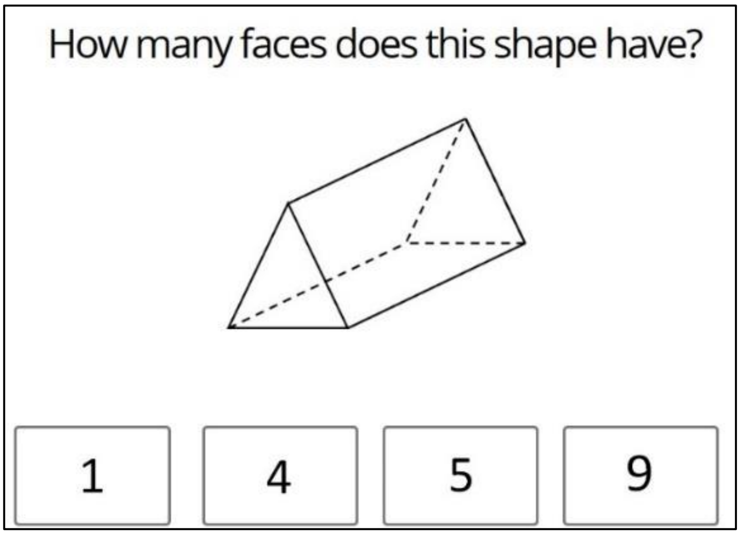

Figure 4.9. Sample Geometry Shape Item

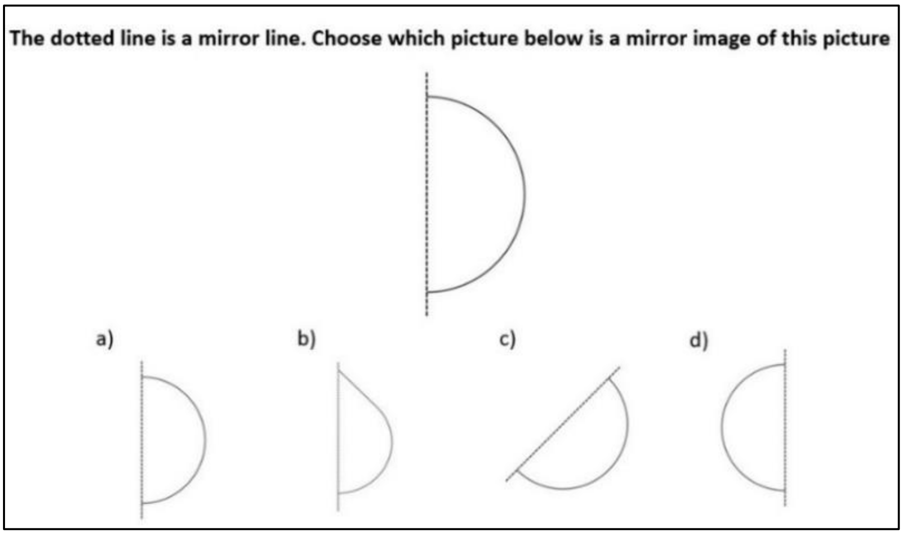

Figure 4.10. Sample Geometry Symmetry Item. *Note.* a) match error, b) shape error, c) symmetry error, d) correct answer

4.2.4.6 Expectations of the effectiveness of training

Prior to the delivery of training, all participants were asked a single question, measuring their expectations of the effectiveness of training, "We are going to be playing some games. How much do you think the games will help you with your maths?". The question was displayed alongside an onscreen scale (see Figure 4.11). Participants responded by selecting a point on the scale using the mouse cursor. Participant's responses were coded as 1-12 based on the onscreen position selected. A score of 1 was allocated for responses that indicated low expectations of training while a score of 12 was allocated for responses that indicated high expectations of training.

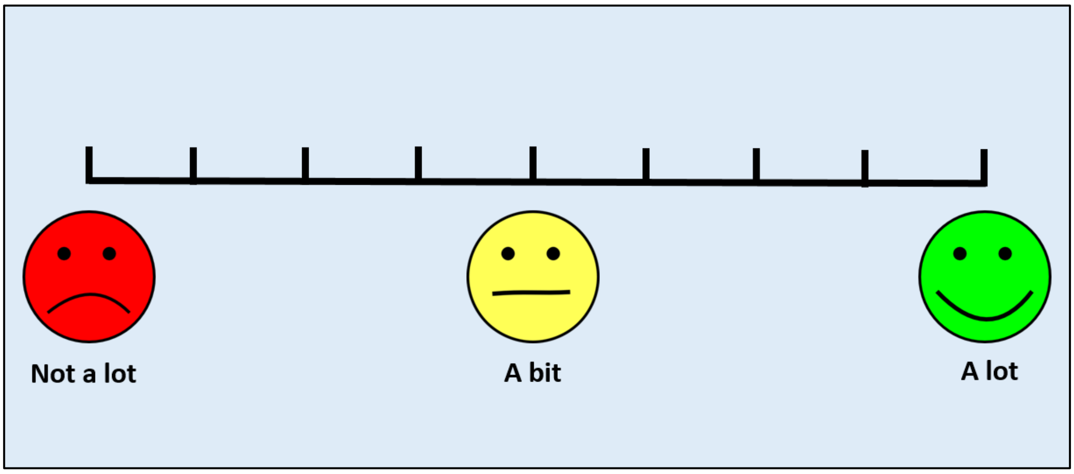

Figure 4.11. Response scale for measuring expectations of the effectiveness of training

4.2.4.7 Participant Engagement Questionnaire

A participant engagement questionnaire was delivered following training to assess participant's enjoyment of, and engagement with, the training that they had received. This questionnaire was specifically designed for use in this study. As shown in Table 4.2, the questionnaire included four questions, the phrasing of which varied slightly based on the type of training delivered. Each question was presented alongside an onscreen scale (see Figure 4.12). Participants responded to each question by selecting a point on the scale using the mouse cursor. Participant's responses were coded as 1-12 based on the onscreen position selected. A score of 1 was allocated for responses that indicated low engagement while a score of 12 was allocated for responses that indicated high engagement. Participants were awarded an overall engagement score, an average of their scores across all four questions (items were reverse coded where necessary).

Table 4.2

Items included in the Participant Engagement Questionnaire

Item	Explicit Training	Implicit Training
1	How much did you enjoy the video?	How much did you enjoy the game?
2	How exciting was the video?	How exciting was the game?
3	How easy was it to understand the video?	How easy was it to understand the game?
4	How much effort did it take to watch the video?	How much effort did it take to play the game?

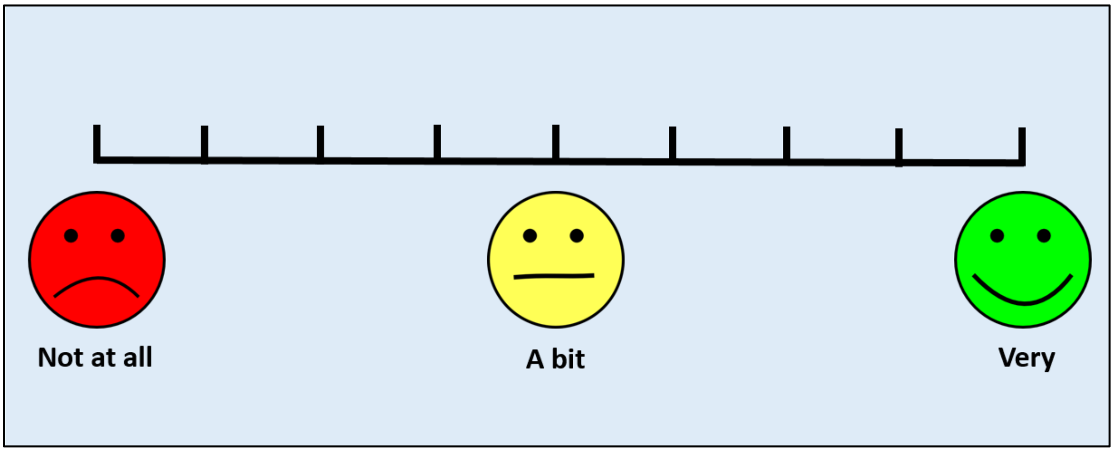

Figure 4.12. Sample scale from the Participant Engagement Questionnaire

4.2.5 Data treatment

The a priori power analysis for this study was based on a medium effect. In practice, the results of this study had small to medium effect sizes. Post-hoc power analysis indicated that the majority of the analyses achieved a power level higher than 0.8, the suggested level for adequately powered studies (Cohen, 1988). The only exception to this was the Number Line Estimation Task, for which a power level of 0.68 was reported. The results of this study should be interpreted in light of the low power reported for this task.

Due to technical errors, school disruptions and absences, data for a single task was lost for nine participants at Time 1 and for 15 participants at Time 2. For task analysis at Time 1, missing data were replaced using mean replacement. For analysis of training effects, there was no replacement of missing data. Participants were excluded from analysis for any tasks for which their data were missing.

For all measures, performance across age groups was explored graphically. For measures in which a ceiling (or floor effect) was suspected, one sample t-tests were completed against ceiling (or floor) performance. No significant floor or ceiling effects were found. Tests of normality indicated that most measures had broadly normal distributions. The main exception to this was performance on the Number Line Estimation Task, which was skewed. Boxplots were used to investigate outliers. At Time 1, there were two outliers for the Spatial Scaling Task (one high performer in the implicit spatial scaling training group and one low performer in the implicit control group). There were also outliers for number line estimation performance. These are likely attributable to the skewed levels of performance on this task. All outliers were deemed to reflect normal variation in the population and were retained. As all groups were large enough for the central limit theorem to apply (*n's* > 30) (Field, 2013), parametric analyses were used.

4.3 Results

4.3.1 Performance at Time 1

4.3.1.1 Overall task performance at Time 1

Unless otherwise stated all Time 1 analysis was based on the 250 participants in the overall sample, i.e., analysis at Time 1 is collapsed across training types and training modes. At Time 1, mean scores were used to replace missing data. Although the participant engagement measure was completed during session 2, this measure was completed prior to training. Thus, it can be considered a pre-training measure and it is included in Time 1 analysis. Descriptive information for performance on each of the tasks included in this study is shown in Table 4.3.

Table 4.3

Descriptive statistics at Time 1

Measure	Descriptive Statistics				
	Mean	**SE**	**SD**	**Min**	**Max**
Mental Rotation	59.00	0.99	15.64	25.00	100
Spatial Scaling	54.00	0.54	8.54	23.61	79.17
Missing Term Problems	56.42	1.56	24.68	0.00	100
Number Line R^2_{LIN}	0.93	0.01	0.08	0.63	1.00
Number Line PAE	0.10	0.01	0.06	0.03	0.30
Geometry Shape Items	63.73	1.05	16.54	16.67	100
Geometry Symmetry Items	54.36	2.08	32.94	0	100
Expectations (mean rating 0-12)	9.47	0.23	3.64	0	12.00

Note. Unless otherwise stated all results reported are percentage accuracy scores

4.3.1.2 Spatial task performance at Time 1

One sample t-tests were used to investigate above chance performance for both the Mental Rotation Task and the Spatial Scaling Task. For both tasks, each trial included two possible response options, therefore chance was set at 50%. For the Mental Rotation Task, participants performed above chance on 0° trials, t (249) = 19.34, p < .001, d = 1.223; 45° trials, t (249) = 15.06, p < .001, d = 0.952; 90° trials, t (249) = 9.77, p < .001, d = 0.618; and 135° trials, t (249) = 2.74, p = .012, d = 0.174. For trials at 180° performance was not above chance, t (249) = -.03, p = .975, d = - 0.002. For the Spatial Scaling Task above chance performance was reported for all scaling factors. This included: a scaling factor of 1, t (249) = 4.20, p < .001, d = 0.266; a scaling factor of 0.875, t (249) = 5.31, p < .001, d = 0.336; a scaling factor of 0.75, t (249) = 4.26, p < .001, d = 0.310; a scaling factor of 0.625, t (249) = 5.24, p < .001, d = 0.332; a scaling factor of 0.5, t (249) = 4.20, p < .001, d = 0.266, and; a scaling factor of 0.375, t (249) = 3.20, p < .001, d = 0.202. As participants performed above chance at most degrees of rotation, and at all scaling factors, this suggests that they understood the aims of the tasks.

One-way ANOVA analyses were used to investigate the effects of degree of rotation (5 levels: 0°, 45°, 90°, 135°, 180°) and scaling factor (6 levels: 1, 0.875, 0.75, 0.625, 0.5, 0.375) on mental rotation and spatial scaling performance respectively. For the Mental Rotation Task, the results indicated a significant main effect of degree of rotation, F (4, 996) = 87.578, p < .001, η_p^2 = .260. As shown in Figure 4.13, these differences were best explained by a linear contrast. Performance decreased with increasing degree of rotation, F (1,249) = 180.51, p < .001, η_p^2 = .420. This performance pattern was also supported by Bonferroni pairwise comparisons. Significant differences in performance were reported between each degree of rotation (*p's* < .001) except for 135° and 180° (p = .050). In contrast, as shown in Figure 4.14, no significant main effect of scaling factor was reported for the Spatial Scaling Task, F (5, 1245) = .747, p = .589, η_p^2 = .003. These patterns of performance are similar to those reported in previous studies for mental rotation and spatial scaling respectively (Broadbent, 2014; Frick & Newcombe, 2012).

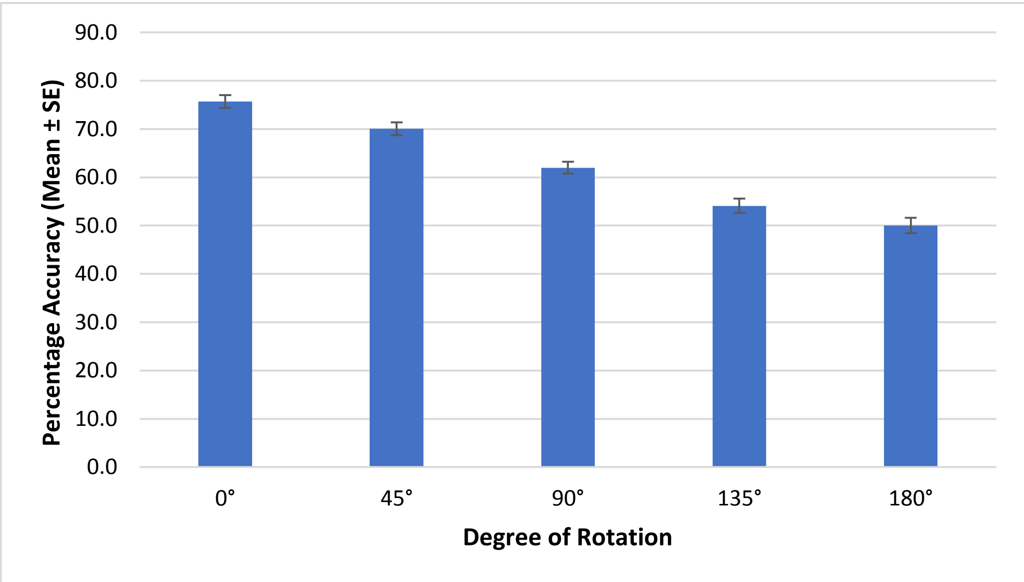

Figure 4.13. Performance on the Mental Rotation Task at Time 1 across different degrees of rotation

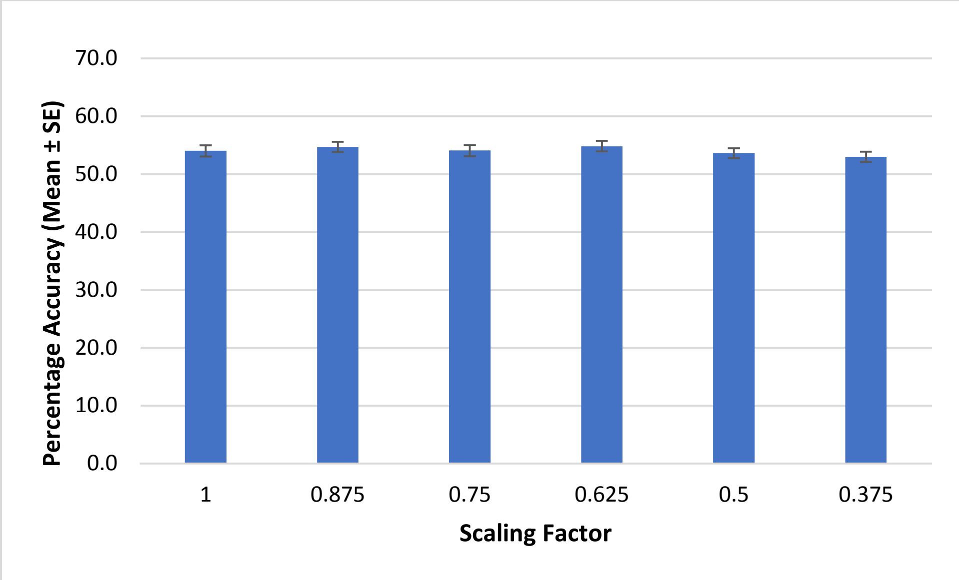

Figure 4.14. Spatial Scaling performance at Time 1 across different scaling factors

4.3.1.3 Mathematics task performance at Time 1

Responses on Missing Term Problems and the Number Line Estimation Task were open ended. As such, it was not possible to calculate above chance performance for these measures. Instead, for Missing Term Problems, participants who did not score higher than 10% at Time 1, were not deemed to understand the task aims and were excluded (*n* = 14). For the Number Line Estimation Task participants who didn't attempt at least 75% of items, or participants with a mean PAE score higher than 15% for practice items were also excluded (*n* = 0). For this task, the value of linear and logarithmic response patterns were compared for each individual. These simple comparisons demonstrated that 74% of participants had estimates that were best described by a linear compared to a logarithmic function (i.e., the participant had a higher R^2_{LIN} compared to R^2_{LOG} score). Therefore, linear estimates (R^2_{LIN}) were used as the outcome measure in all subsequent analysis (Simms et al., 2016).

For the Geometry Task, a paired samples t-test was used to investigate the effect of item type on task performance (2 levels: Shape Items and Symmetry Items). There was a significant difference in performance between Geometry Shape Items (63.73 ± 1.05) and Geometry Symmetry Items (54.36 ± 2.08), t (1,249) = 4.34, p < .001, d = 0.295. Based on this difference, Symmetry Items and Shape Items were considered separately in all subsequent analysis. One sample t-tests were used to explore above chance performance on each item type. As both item types included four possible response options, chance was set at 25%. Participants performed above chance on both Shape Items, t (249) = 37.04, p < .001, d = 2.342, and Symmetry Items, t (249) = 14.09, p < .001, d = 0.891.

4.3.2 Gender differences in task performance at Time 1

Independent T-tests (controlling for multiple comparisons [0.05/8 = 0.006]) were used to explore gender differences in task performance at Time 1. Where homogeneity of variance could not be assumed, the results for unequal variances were reported. As shown in Table 4.4, males had significantly lower PAE on the Number Line Estimation Task compared to females, t (148) = 3.15, p = .002, d = 0.401. No other significant gender differences were reported (*p's* > .05, *d's* < .261). Thus, gender was included as a control variable when investigating the effects of training on the Number Line Estimation Task only.

Table 4.4

Gender differences in task performance at Time 1

Test Measure	**Gender**				**Statistics**	
	Male (n = 121)		Female (n = 129)		Test statistic	Effect size
	Mean	**SD**	**Mean**	**SD**	***T***	***D***
Mental Rotation	60.38	16.05	57.76	15.19	0.74	0.09
Spatial Scaling	54.76	7.53	53.28	9.36	1.37	0.17
Missing Term Problems	59.71	24.57	53.34	24.47	2.05	0.26
No. Line Estimation R^2_{LIN}	.09	.05	.11	.06	3.15	0.40
No. Line Estimation PAE	.94	.07	.92	.08	1.44*	0.18
Geometry Shape Items	62.81	15.59	64.60	17.39	0.85	0.11
Geometry Symmetry Items	53.55	33.83	55.12	32.19	0.37	0.05
Expectations	9.13	3.86	9.79	3.39	1.45	0.18

Note. * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. R^2_{LIN} = linear response; PAE = Percentage Absolute Error; No. Line = Number Line

4.3.2.1 Differences in task performance across training groups at Time 1

To confirm that there were no significant performance differences between groups at Time 1, a two-way ANOVA was completed for each task. Training mode (2 levels: explicit vs. implicit) and training type (3 levels: mental rotation vs. spatial scaling vs. literacy) were included as between participant variables.

No significant differences in task performance across training types were reported for: the Mental Rotation Task, F (2, 244) = 2.43, p = .090, η_p^2 = .020; the Spatial Scaling Task, F (2, 244) = 1.77, p = .173, η_p^2 = .014; Missing Term Problems, F (2, 244) = 2.32, p = .100, η_p^2 = .019; PAE scores on the Number Line Estimation Task, F (2, 244) = 0.01, p = .920, η_p^2 = .000; R^2_{LIN} scores on the Number Line Estimation Task, F (2, 244) = 0.01, p = .991, η_p^2 = .000; Geometry Shape Items, F (2, 244) = 0.376, p = .687, η_p^2 = .003, or; Geometry Symmetry Items, F (2, 244) = 0.34, p = .709, η_p^2 = .003.

Similarly, no significant effect of training mode was reported for: the Mental Rotation Task, *F* (1, 244) = 0.02, *p* = .890, η_p^2 = .000; the Spatial Scaling Task, *F* (1, 244) = 1.07, *p* = .303, η_p^2 = .004; Missing Term Problems, *F* (1, 244) = 0.68, *p* = .410, η_p^2 = .003, PAE scores on the Number Line Estimation Task, *F* (1, 244) = 0.49, *p* = .613, η_p^2 = .004; R^2_{LIN} scores on the Number Line Estimation Task, *F* (1, 244) = 0.48, *p* = .490, η_p^2 = .002; Geometry Shape Items, *F* (1, 244) = 1.45, *p* = .230, η_p^2 = .006 or; Geometry Symmetry Items, *F* (1, 244) = 4.05, *p* = .060, η_p^2 = .015.

To assess differences in expectations of training across groups, a two-way ANOVA was also completed with training mode (2 levels: explicit vs. implicit) and training type (3 levels: mental rotation vs. spatial scaling vs. literacy) as between participant variables. There were no differences in expectations of training across training modes, *F* (1, 244) = 3.25, *p* = .072, η_p^2 = .013, or training types, *F* (2, 244) = 0.27, *p* = .763, η_p^2 = .002.

4.3.2.2 Associations between measures at Time 1

Pearson bivariate correlations were completed between all Time 1 measures. This allowed for the investigation of whether the observed associations between spatial and mathematics skills, that have been demonstrated in previous studies (e.g., Mix et al., 2016) and which form the rationale for the training paradigm used in this study, were present. Significant correlations were reported between all tasks, except for performance on Geometry Shape Items which was not correlated with accuracy on the Mental Rotation Task, *r* (249) = .09, *p* = .147 (Table 4.5). Expectations of the effectiveness of training were not correlated with any behavioural measures. For the Number Line Estimation Task, the correlations reported between the two number line outcome measures (PAE and R^2_{LIN} scores), and all other tasks, were similar. However, it was hypothesised that spatial scaling training would lead to improved proportional reasoning skills, which would subsequently reduce PAE scores, i.e., enabling participants to position estimates more accurately. Scaling training was not predicted to influence participants' understanding of symbolic number, i.e., R^2_{LIN} scores. Therefore, results based on PAE scores are reported in this chapter. However, patterns of performance for R^2_{LIN} scores were broadly similar (see Appendix E).

Table 4.5

Correlations between tasks at Time 1

	Spatial Tasks		**Mathematics Tasks**					**Expectations**
	1	2	3	4	5	6	7	8
1. Mental Rotation	/	.28***	.29***	- .21***	.25***	.09	.23***	.06
2. Spatial Scaling		/	.35***	-.30***	.33***	.16*	.26***	.04
3. Missing Term Problems			/	-.49***	.53***	.30***	.42***	-.02
4. No. Line PAE				/	-.83***	-.25***	-.33***	.01
5. No. Line R^2_{LIN}					/	.22***	.31***	-.02
6. Geometry Shape Items						/	.18***	.01
7. Geometry Symmetry Items							/	-.03
8. Expectations								/

Note. * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. R^2_{LIN} = linear response; PAE = Percentage Absolute error; No. Line = Number Line

4.3.3 Performance at Time 2

4.3.3.1 Near and intermediate transfer of gains

In the investigation of training effects, there was no mean replacement of data. Participants were excluded from analysis for any tasks for which they were missing data at Time 1 (n = 9) or Time 2 (n = 15). Participants scoring higher than 95% on a given task at Time 1, were deemed to have reached ceiling level performance on the task and were excluded from training analysis for that task only. This included two participants for the Mental Rotation Task, nine participants for the Missing Term Problems, ten participants for the Geometry Shape Items and 18 participants for the Geometry Symmetry Items.

Multivariate Analysis of Variance (MANOVA) tests were used to investigate training effects across near, intermediate and far transfer measures. Time was included as a within participant variable (2 levels: pre-training, post-training). Training mode (2 levels: explicit, implicit) and training type (3 levels: mental rotation, spatial scaling, literacy) were included as between participant variables. Where sphericity could not be assumed, Greenhouse- Geisser values were reported. Significant interactions were explored with paired t-tests. It has been argued that the power of training studies can be increased by analysing results using ANCOVA tests with post-training scores as the dependent variable and pre-training scores as a covariate (Van Breukelen, 2006). Therefore, the analysis reported in this section were repeated using ANCOVA with time one scores as a covariate. Comparable results were reported (see Appendix F).

4.3.3.1.1 Mental Rotation

For the Mental Rotation Task, there was a significant main effect of time. There was significantly higher performance at Time 2, F (1, 237) = 21.87, p < .001, η_p^2 = .084. This finding was best explored within the context of the significant interaction found between time and training type, F (2, 237) = 6.88, p < .001, η_p^2 = .055. As shown in Figure 4.15, t-tests indicated a significant improvement in performance accuracy following mental rotation training, t (83) = 5.49, p < .001, d = 0.581 (near transfer) and spatial scaling training, t (79) = 2.30, p = .024, d = 0.263 (intermediate transfer).

No significant improvement in performance accuracy was reported following control training, t (78) = 0.21, p = .837, d = 0.019. No other main effects or interactions with time were reported (p's > .05, η_p^2's < .005).

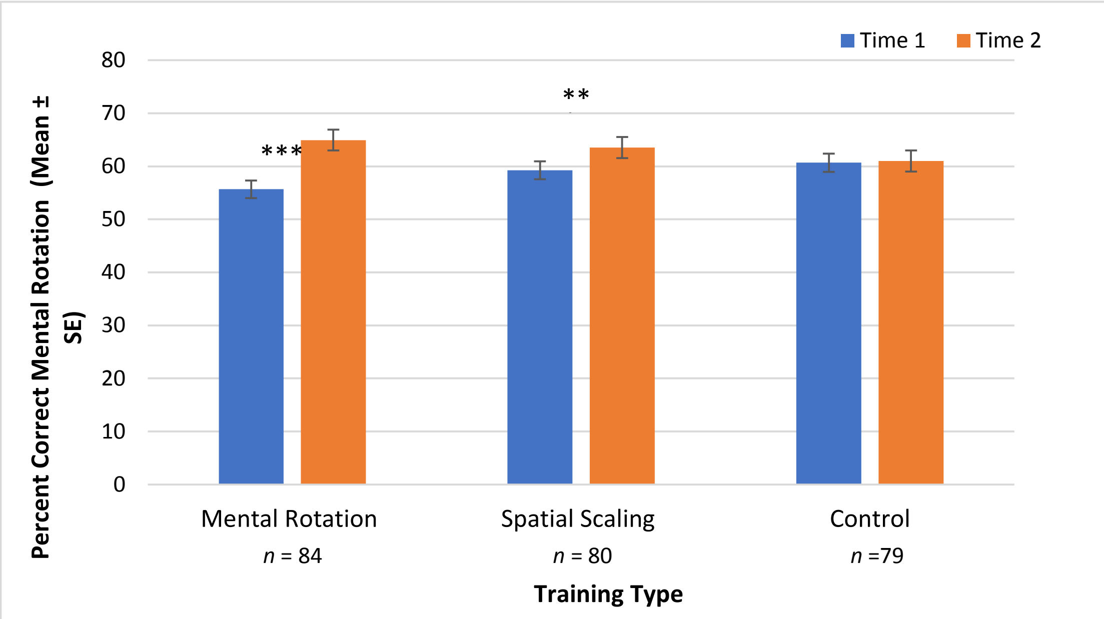

Figure 4.15. Mental Rotation accuracy at Time 1 and Time 2 for different training types. *Note.* * indicates p < .05, ** indicates p < .01, *** indicates p < .001. SE = Standard Error

4.3.3.1.2 Spatial Scaling

A significant main effect of training type was found, with higher overall performance for the spatial scaling training group compared to the other training groups, F (2, 232) = 8.28, p < .001, η_p^2 = .067. This was best explored in the context of the significant interaction between time and training type, F (2, 232) = 6.25, p = .002, η_p^2 = .051 (see Figure 4.16). T-tests indicated significant performance gains following spatial scaling training only, t (76) = 3.99, p < .001, d = 0.450 (near transfer). No significant gains were reported following mental rotation training, t (80) = 0.04, p = .972, d = 0.004, or control training, t (79) = 0.70, p = .485, d = 0.088. There were no other main effects or significant interactions with time (p's > .05, η_p^2's < .005).

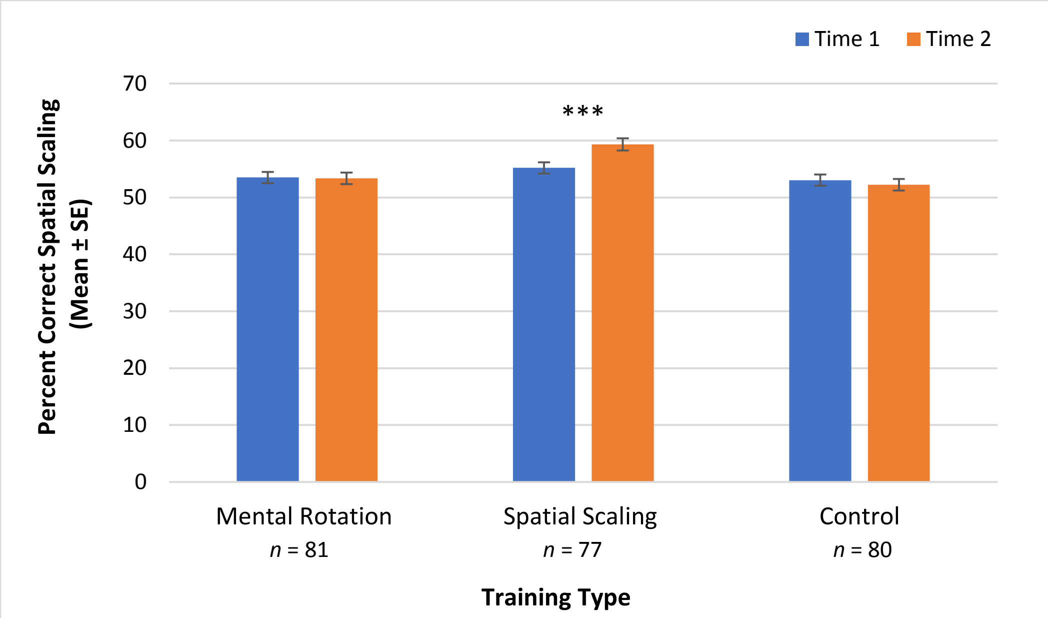

Figure 4.16. Spatial scaling accuracy at Time 1 and Time 2 for different training types. *Note.* * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. SE = Standard Error

4.3.3.2 Far transfer of gains

4.3.3.2.1 Missing Term Problems

A significant interaction between time and training type was found, F (2, 209) = 4.58, p = .011, η_p^2 = .042 (see Figure 4.17). T-tests indicated a significant improvement in accuracy following mental rotation training only, t (69) = 2.73, p < .008, d = 0.241 (far transfer). No significant improvements were reported following spatial scaling training, t (74) = 1.30, p = .197, d = 0.117, or control training, t (69) = 0.73, p = .466, d = 0.067. There were no other significant main effects or interactions with time (p's > .05, η_p^2's < .009).

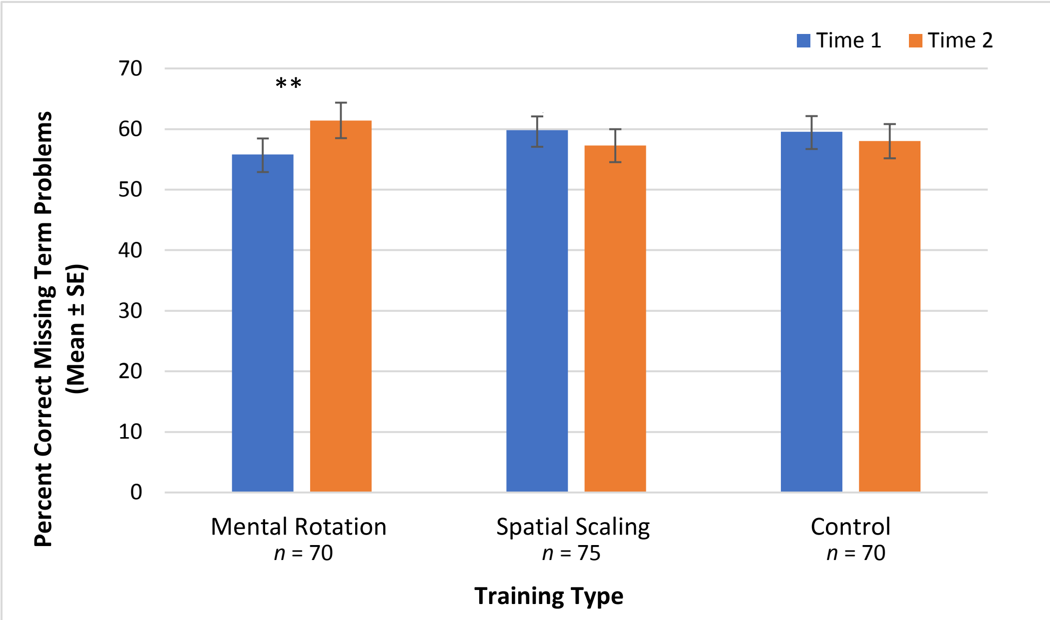

Figure 4.17. Percentage Correct on Missing Term Problems at Time 1 and Time 2 for different training types. *Note.* * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. SE = Standard Error

4.3.3.2.2 Number Line Estimation

As a significant gender effect was reported for PAE scores at Time 1, gender was included as a between participant variable. However, no significant main effect or interactions with gender were reported for this task (p's > .05, η_p^2's < .014). Hence, gender was removed, and the analysis was repeated. A significant main effect of time was reported, F (1,237) = 5.86, p = .016, η_p^2 = .024. This finding was best explored within the context of the interaction between time and training type. As shown in Figure 4.18, there was a significant interaction between time and training type, F (2, 237) = 6.05, p = .002, η_p^2 = .054. T-tests indicated a significant reduction in error following spatial scaling training, t (79) = 2.12, p = .037, d = 0.172 (far transfer). No significant difference in performance was found following mental rotation training, t (82) = 1.91, p = .060, d = 0.222. However, a significant increase in error was reported following control training, t (79) = 3.01, p = .003, d = 0.360. No other main effects or significant interactions with time were reported (p's > .05, η_p^2's < .005). As outlined in section 4.3.2.2, similar analysis was completed for R^2_{LIN} performance. The patterns of performance across time and training type were comparable to PAE scores (see Appendix E).

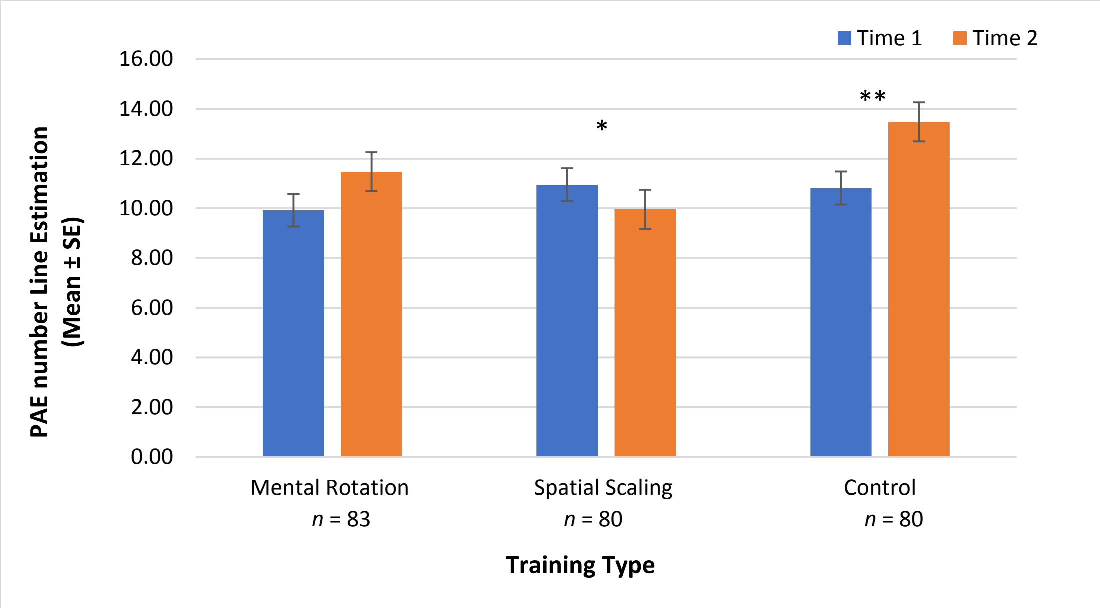

Figure 4.18. PAE on the Number Line Estimation Task at Time 1 and Time 2 for different training types. *Note.* * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. SE = Standard Error

4.3.3.2.3 Geometry Task

For Geometry Shape Items there were main effects of time, F (1, 219) = 12.93, $p <$ .001, η_p^2 = .056, training mode, F (1, 219) = 6.39, p = .012, η_p^2 = .028, and training type, F (2, 219) = 3.25, p = .041, η_p^2 = .029. These main effects were best explored in the context of the interactions reported below. There was a significant interaction between time and training type for Geometry Shape Items, F (2, 219) = 3.82, p = .022, η_p^2 = .034 (see Figure 4.19). T-tests indicated significant gains in performance accuracy following mental rotation training, t (75) = 2.93, p = .004, d = 0.308 (far transfer), and spatial scaling training, t (75) = 3.70 $p < .001$, d = 0.314 (far transfer). There were no significant gains following control training, t (72) = 0.21, p = .833, d = 0.024. There was also a significant interaction between time and training mode for Geometry Shape Items, F (1, 219) = 5.95, p = .016, η_p^2 = .026 (see Figure 4.20). There was a significant improvement in performance following implicit training, t (104) = 4.41, $p < .001$, d = 0.351, but not explicit training, t (116) = 0.85, p = .395, d = 0.069.

No significant three-way interaction between time, training mode and training type was reported, F (2, 219) = 1.60, p = .204, η_p^2 = .014. For Geometry Symmetry Items, all groups had improved performance between Time 1 and Time 2, F (1, 213) = 40.30, p < .001, η_p^2 = .159. However, there were no other main effects or significant interactions with time (p's > .05, η_p^2's < .013).

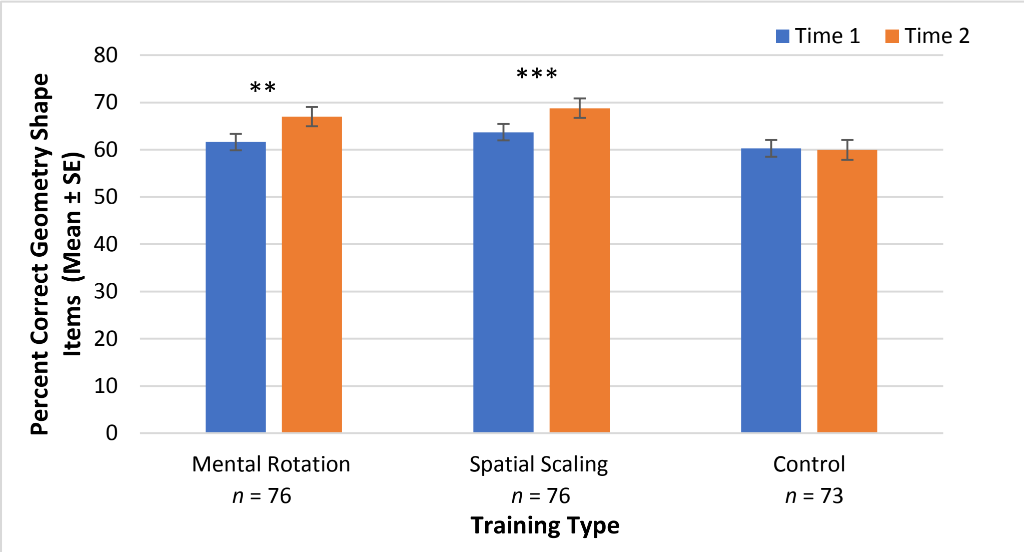

Figure 4.19. Accuracy on Geometry Shape Items at Time 1 and Time 2 for different training types. *Note.* * indicates p < .05, ** indicates p < .01, *** indicates p < .001.

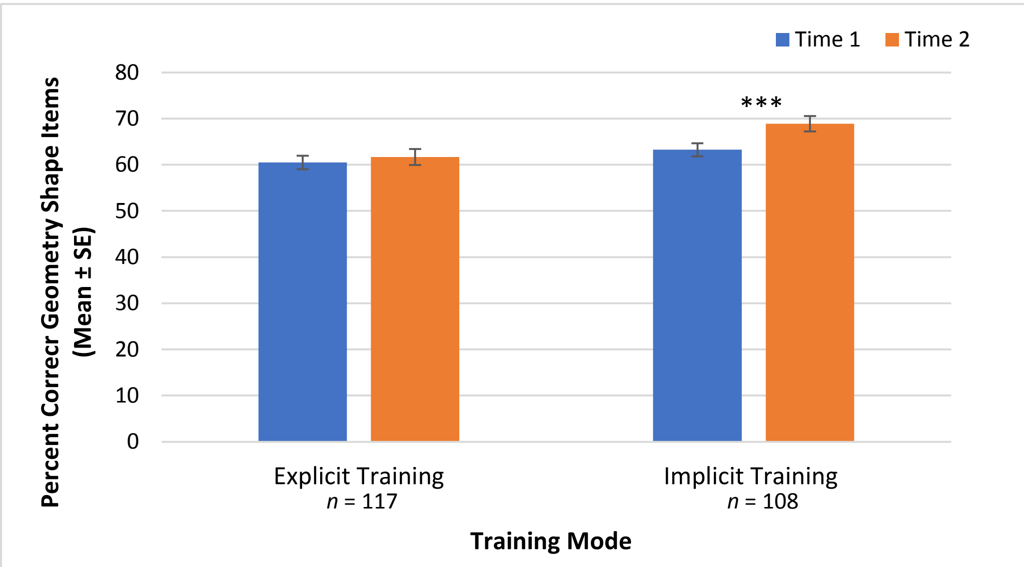

Figure 4.20. Accuracy on Geometry Shape Items at Time 1 and Time 2 for different training modes. *Note.* * indicates p < .05, ** indicates p < .01, *** indicates p < .001

4.3.3.3 Motivational factors

4.3.3.3.1 Expectations of training

A MANOVA was completed with training mode (2 levels: explicit, implicit) and training type (3 levels: mental rotation, spatial scaling, literacy) as between participant variables and expectations of training as the dependent variable. There were no significant differences in self-reported expectations of training for participants in different training mode conditions, F (1, 244) = 3.25, p = .072, η_p^2 = .013, or training type conditions, F (2, 244) = 0.27, p = .763, η_p^2 = .002. Multivariate analysis of covariance (MANCOVA) was used to explore whether individual participant gains on each outcome measure were predicted by expectations of training. A separate MANCOVA was completed for each training type (mental rotation, spatial scaling and control) and each training mode (explicit and implicit). Time was included as a between participant variable and expectation score was included as a covariate. There were no significant interactions between participant expectations of training and time for any of the training types (p's > .05, η_p^2's < .033) or any of the training modes (p's > .05, η_p^2's < .012).

4.3.3.3.2 Participant engagement with training

A MANOVA was completed with training mode (2 levels: explicit, implicit) and training type (3 levels: mental rotation, spatial scaling, literacy) as between participant variables and self-reported engagement levels as the dependent variable. There was a significant difference in engagement across training types, F (2, 244) = 3.37, p = .036, η_p^2 = .027. Bonferroni pairwise comparisons indicated significantly higher engagement levels following control training compared to spatial scaling training (p = .034). There was no main effect of training mode on engagement, F (1, 244) = 1.81, p = .180, η_p^2 = .007. There was a significant interaction between training type and training mode on engagement, F (2, 244) = 3.30, p = .039, η_p^2 = .026. For explicit training there were no differences in engagement across training types, F (2, 123) = 0.56, p = .573, η_p^2 = .009. For implicit training there was an effect of training type, F (2, 121) = 5.42, p = .006, η_p^2 = .082. As highlighted in Figure 4.21, post-hoc Bonferroni tests indicated significantly higher engagement following control training compared to spatial scaling training (p = .004).

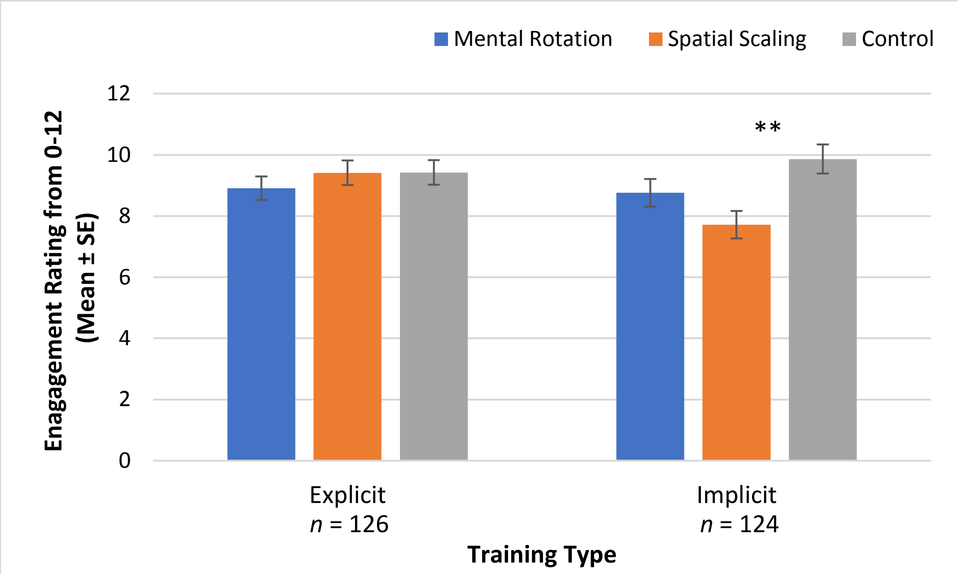

Figure 4.21. Self-reported levels of engagement following training across training modes and training types. *Note.* * indicates $p < .05$, ** indicates $p < .01$, *** indicates $p < .001$. SE = Standard Error

4.4 Discussion

The results of this study support previous correlational findings on spatial-mathematical relations and provide insight into the causal relationships between different aspects of spatial and mathematical thinking. It was demonstrated that training mental rotation skills and, for the first time, training spatial scaling skills, led to gains in spatial and mathematical thinking at 8 years. These gains were present following explicit and implicit instruction. Spatial training gains had near, intermediate, and far transfer effects. Spatial thinking is therefore one cognitive domain in which transfer of cognitive training gains is possible. The gains reported reflect the importance of choosing developmentally sensitive, theoretically motivated training targets.

4.4.1 Near, intermediate and far transfer of gains

Near transfer: Mental rotation and spatial scaling training led to significant gains in mental rotation (*d* = 0.581) and spatial scaling (*d* = 0.450) respectively. These findings are consistent with previous evidence that spatial skills are malleable in children (Uttal et al., 2013). Previous studies typically investigated the malleability of mental rotation or other spatial tasks that elicit mental visualisation (Uttal et al., 2013). However, this is the first known study to show the malleability of spatial scaling in children.

Intermediate transfer: Significant gains in mental rotation were reported following spatial scaling training (*d* = 0.263) providing evidence of intermediate transfer of spatial scaling training to an untrained spatial task. These findings are consistent with those of Uttal *et al.* (2012) who found that spatial training transferred to other untrained spatial tasks (*Hedges G* = .48). However, Uttal *et al.* (2013) reported that intermediate transfer was not evident in all studies and was more likely to occur where longer training sessions were included. The short training sessions used in this study (3-5 minutes) may explain why no intermediate transfer was reported following mental rotation training. As outlined in 4.1.3, one reason for transfer from spatial scaling training to mental rotation performance may be that spatial scaling training elicited mental visualisation, which is also required for mental rotation tasks.

Far transfer: Participants who completed mental rotation training had significant accuracy gains on Missing Term Problems (*d* = 0.241). These findings of far transfer of gains are consistent with the findings of Cheng and Mix (2014) who demonstrated that explicit mental rotation training led to gains in performance accuracy on a similar task. As outlined in Chapter 3, Cheng and Mix (2014) propose that these findings are due to the fact that children solve arithmetic problems of this type by mentally rotating the terms presented, i.e., by restructuring the equation in a more prototypical format. For example, 4 + __ = 9, can be mentally rotated to generate the equation __= 9 – 4. However, this mental manipulation would require a relatively advanced understanding of calculation rules, i.e., that a plus becomes a minus when it is moved across the equals sign. Alternatively, children may use mental visualisations to represent these equations pictorially. This equation could be solved by visualising 4 blocks in one group and 9 blocks in another, and counting the difference between the groups (Lourenco et al., 2018). In this study, no significant difference in the efficacy of explicit and implicit mental rotation instruction was found. This contrasts the findings of Hawes *et al.* (2015) who did not report gains on Missing Term Problems following implicit mental rotation training. This highlights other factors, such as participant engagement during training, as possible explanations for the results reported by Hawes *et al.* (2015).

For the Number Line Estimation Task, a significant reduction in error was reported for children who completed spatial scaling training (d = 0.222). This far transfer of gains from spatial scaling to number line estimation may be explained by the fact that both tasks require proportional reasoning. If a child was asked to place the number 27 on a number line ranging from 0 to 100, they might reason that 27 is close to 25, which is one quarter of 100. By accurately dividing the number line into quarters, a child could place the number 27 on a number line, with relatively high accuracy (Newcombe et al., 2015; 2018; Rouder & Geary, 2014). Proportional reasoning is also required when comparing two spaces of different sizes (Newcombe et al., 2018). Alternatively, the Mental Number Line may be responsible for associations between spatial scaling and number line estimation. As outlined in section 1.3.2, this is the concept that numbers are represented spatially in the brain with smaller numbers on the left and larger numbers on the right (Barsalou, 2008; Lakoff & Núñez, 2000). Children may scale between a mental number line and the physical number line presented in number line estimation tasks (Dehaene, 1997; Fischer, 2003). Whilst spatial scaling has been associated with number line estimation in a number of studies (e.g., Mix et al., 2016), this is the first to show that spatial scaling training leads to improvements in number line estimation. An unexpected increase in error was reported following control training (d = 0.360). This could not be explained by motivational factors. Further investigation is needed to understand this effect.

Performance on the Geometry Task differed across item types. Gains on Geometry Symmetry Items were reported across time, but no effects of training mode or training type were found. Thus, effects in the experimental training conditions did not differ from those in the control conditions. This suggests significant practice effects for this task. In contrast, there was far transfer of training gains from both mental rotation (d = 0.308) and spatial scaling (d = 0.314), to Geometry Shape Items. From a theoretical perspective, children might use mental visualisation (also used in mental rotation tasks and possibly used in spatial scaling tasks) to picture and rotate the shapes presented to count the number of sides (faces) on the shape. Improved spatial scaling skills may have enabled participants to better use proportional reasoning to answer shape items. Instead of counting each individual side (face), participants may have first, segmented the shapes presented (all of which were symmetrical) into halves or thirds, then counted the sides (faces) in a single segment, and finally multiplied this to account for all segments.

4.4.2 Explicit vs. implicit instruction

For Geometry Shape Items there was a main effect of training mode. Gains were reported following implicit (d = 0.351) but not explicit (d = 0.069) instruction. Furthermore, there were no interactions with training type, i.e., those in experimental training conditions did not differ from those in the control condition. Therefore, gains following implicit instruction may be explained by the mode of feedback used in the delivery of implicit compared to explicit training. For Geometry Shape Items, participants were asked to count the number of sides (faces) on a shape. Errors can easily be made on this task if participants mistakenly count the same side (face) twice or if participants forget where on the shape they started counting. The use of a checking strategy may improve performance on this task, i.e., checking answers and repeating trials to confirm answers before submitting a response. The implicit instruction delivered in this study included feedback. Participants were required to carefully select responses and revise incorrect responses, thus modelling an effective self-monitoring (checking) strategy. This implicit instruction may have subsequently increased the likelihood of participants revising and rechecking their answers on the Geometry Task prior to submitting a response, which may in turn have increased task accuracy.

For all other measures, there were no main or interaction effects reported for training mode (explicit vs. implicit instruction). This suggests that explicit and implicit spatial instruction are largely similar in eliciting near, intermediate and far transfer of gains. These findings have importance given the practical considerations of delivering explicit and implicit instruction in the classroom. The delivery of instructional videos in a group context offers an easily implementable method of improving spatial thinking that does not require one-to-one student interaction or advanced IT facilities (such as a laptop for every student). This mode of instruction offers a feasible, cost-effective way of *spatialising* the primary school classroom.

4.4.3 Motivational Factors

There were no significant differences in participants' expectations of training across different training modes (η_p^2 = .013) or training types (η_p^2 = .002). There were no significant interactions between expectations and performance gains following training for any tasks. The similarities in expectations reported across groups and lack of significant interactions between expectations and performance gains, increase the reliability of the causal inferences made in this study (Boot et al., 2013).

For engagement, there were no differences reported for explicit training between training types. For implicit training, there was significantly higher engagement for participants in the control group compared to those who completed spatial scaling training. Participants who received implicit spatial scaling instruction completed additional trials of the Spatial Scaling Task, a task that they had previously completed as a pre-test at Time 1. For the implicit control group, the task completed was new, i.e., not completed at pre-test. As such, participants who completed spatial scaling training may have found their training less engaging as it was not novel. Although a significant difference in engagement was found for implicit instruction, the direction of the difference shows that the performance gains reported for spatial and mathematics skills were not attributable to engagement with training alone. As control training did not lead to training effects on any of the outcome measures, levels of engagement did not superficially align with training effects.

4.4.4 Implications, future directions and limitations

This study provides some of the first evidence that the association between spatial and mathematical performance reflects a causal influence of spatial ability on mathematics performance. The findings determine that the observed correlations between spatial and mathematical thinking cannot be solely explained by a common cause acting on both variables. The causal inferences drawn are further strengthened by the fact that this study controlled for motivational factors. Thus, it was possible to determine the extent to which cognitive training gains are attributable to training, over and above differences in participant engagement and expectation. Although a priori power analysis were completed, the results should be interpreted in light of the low power achieved for the ANOVA completed using the Number Line Estimation Task. Due to the relatively small effect size of this result, the power of this analysis was 0.68, below the recommended level of 0.8 (Cohen, 1988). Future research should replicate these findings using a larger sample.

A second limitation of this study was the short interval (2 minutes) between training and post-testing. The training completed in this study may have led to priming of certain strategies for task completion, and not conceptual change. However, even if this is the case, this is useful knowledge for teachers, given that priming led to enhanced performance on some mathematics tasks. From the results reported, it is not known whether training gains persisted. Further research is needed to investigate the durability of these gains. Importantly, the findings of other studies suggest that there is durability of spatial training gains. Uttal *et al.* (2013) compared spatial training studies with post-testing immediately following training, to studies that wait days, weeks or even months until post-testing. Uttal *et al.* (2013) found that spatial training gains were durable and that the timing of post-testing did not significantly influence the size of training gains reported following spatial training. Although priming cannot be ruled out, similarly to Cheng and Mix (2014), this study demonstrates shared cognitive processing in the completion of spatial and mathematics tasks, that is subject to modification through training. By extension, it is hypothesised that the gains reported following training in this study are not merely attributable to priming.

Due to the short interval between training and post-testing, findings that spatial training improves mathematics outcomes, suggest that spatial skills play a role in the execution of mathematics tasks. That is not to say that spatial skills are not important in the acquisition of novel mathematical skills (Mix et al., 2016). However, in this study the time difference between training and post-testing was too short for new mathematics skills to be learnt. Thus, transfer of gains in this study suggests that spatial skills are useful in the completion of mathematics tasks. To investigate this research question further, future studies should investigate the long term effects of spatial training on the acquisition of new mathematics skills.

Third, this study did not investigate dosage effects, i.e., whether differences in the duration of training or the number of training sessions influenced training gains. In this study, the dosage of training for both explicit instruction (3 to 4 minutes) and implicit instruction (3-6 minutes) was relatively short. This demonstrated that even short bouts of spatial training can lead to large transfer of gains to untrained domains. Future research is needed to explore whether the size of training gains is influenced by longer training sessions, or by repeated training over a series of training sessions.

4.4.5 Conclusion

This study demonstrated near, intermediate and far transfer of gains to both spatial and mathematical domains, following training in mental rotation and spatial scaling training at 8 years. Not only do these findings highlight the malleability of spatial skills, they also call attention to spatial ability as one domain in which cognitive training can lead to transfer effects. Explicit and implicit instruction led to similar gains in spatial and mathematical domains (except for geometry items). This emphasises the potential of explicit instruction as a practical means of eliciting far transfer of spatial training gains in the primary school classroom. The findings also highlight the importance of ensuring that the choice of cognitive training be determined by an understanding of the underlying cognitive mechanisms of training targets. In this study, mental visualisation was proposed as an underlying cognitive mechanism for mental rotation training, and proportional reasoning was proposed as an underlying cognitive mechanism for spatial scaling training. The gains reported highlight the importance of choosing task and age sensitive targets for spatial training. In turn, evidence from this training study clearly demonstrates the causal contribution of cognitive processes to mathematical cognition that was previously only inferred based on correlational evidence.

Chapter 5 General Discussion

5.1 Thesis Overview

Studies in adult and pre-school populations have reported a significant role for spatial thinking in mathematics outcomes (Verdine et al., 2014; Wai et al., 2009). However, few studies have attempted to assess transfer of spatial training gains to mathematics in children, and those that have report mixed success (Cheng & Mix, 2014; Hawes et al., 2015; Lowrie et al., 2017). Most studies do not address the fact that spatial and mathematical thinking are multi-dimensional constructs (Uttal et al., 2013; von Aster & Shalev, 2007). Therefore, variations in the efficacy of training studies, and differences in spatial-mathematical associations reported across studies, may be attributable to differences in the measures used across studies. It is unlikely that all spatial and mathematical sub-domains are related to the same degree and fine-scaled evaluation of spatial skills and their relations to particular mathematical sub-domains is an essential precursor to identifying effective training approaches. Findings to date are also limited by the fact that few studies explore the relationship between spatial and mathematical thinking in primary school children, even though spatial-mathematical relations may vary through development. Exploring the developmental aspects of the relationship would facilitate a better understanding of not just *if*, but *why* significant correlations are often reported between spatial and mathematics constructs. This understanding would increase the likelihood of developing successful training interventions and would enable determination of the causal relationship between different spatial and mathematical sub-domains.

The experimental studies presented in this thesis aimed to elucidate the relations between spatial and mathematical skills across development from 6 to 10 years. As previous literature in this domain does not suggest a linear coupling of all spatial and mathematical skills, the aims of this thesis were to provide detailed developmental profiles of spatial-mathematical associations, across a range of different spatial and mathematical sub-domains, accounting for other known predictors of mathematics. Throughout this thesis, the role of different spatial skills as predictors of mathematics was compared by classifying spatial skills using the Uttal *et al.* (2013) typology of spatial thinking. To enhance the practical applications of the findings, the role of spatial skills for mathematics was explored in the context of each of von Aster and Shalev's (2007) mathematical sub-domains (Chapter 3), and in the context of classroom-based mathematics performance (Chapter 2).

To explore differences in spatial-mathematical relations across developmental age, the role of spatial skills for mathematics outcomes was investigated from both longitudinal (Chapter 2) and cross-sectional (Chapter 3) perspectives. The importance of spatial thinking at the age at which children first begin primary school (5 years) was investigated (Chapter 2), as well as the role of spatial thinking throughout primary school, from 6 to 10 years (Chapter 3). Differences in the associations between spatial and mathematical skills for children of different genders and those in different SES groups were also assessed (Chapter 2). The findings of Chapter 2 and 3 provide detailed information on the nature of spatial-mathematical associations in primary school children. These outcomes formed the basis on which a spatial training intervention was developed (Chapter 4). The study presented in Chapter 4 compared the use of explicit and implicit instruction for training spatial skills at 8 years, and investigated the transfer of spatial training gains to other spatial and mathematical domains. This served to not only determine the malleability of spatial thinking in primary school aged children, but also to shed light on the causal relationship between specific spatial and mathematical outcomes.

This discussion chapter provides an overview of the main results reported in the experimental chapters of this thesis and outlines the theoretical conclusions drawn from these findings. The discussion considers the profiles of spatial task performance presented across Chapters 2 to 4 from a developmental perspective, in the context of the Uttal *et al.* (2013) typology of spatial thinking. The findings on spatial-mathematical relations reported in Chapters 2 to 4 are discussed. In particular, the discussion outlines the selectivity of the reported relations to specific spatial and mathematical sub-domains, and the sensitivity of the reported relations to developmental age. Arguments are put forward to support proportional reasoning, active processing and form perception, as three underlying mechanisms that may explain the observed spatial-mathematical relations. Expanding on the associational findings reported in Chapter 2 and Chapter 3, inferences are made on the causal relationship between spatial and mathematics skills, based on the results of the spatial training study presented in Chapter 4.

In this discussion the emerging conclusions are framed in the context of their theoretical, educational and economic implications. Finally, the limitations of this research are considered, and future research directions and emerging questions are explored.

5.2 Overview of findings

The findings of this thesis provide clarification of the associations between spatial and mathematical skills in childhood. Each chapter provides complementary insights into different aspects of the spatial-mathematical relationship.

Chapter 2 fine-tunes the current understanding of the complex relationship between spatial, mathematical and vocabulary skills. Without controlling for IQ, previous studies have been unable to elucidate whether there is a direct relationship between spatial and mathematical skills or whether these associations are attributable to the overlapping language demands of the tasks used, or to IQ (Alloway & Alloway, 2010; Mayes et al., 2009). Although language skills are a blunt measure of IQ, by comparing models that include and exclude language skills as predictors of mathematics, it was possible to estimate the true proportion of variation in mathematics explained by spatial skills in childhood. Spatial skills at 5 and 7 years explained between 8.8% (conservative result) and 22.6% (more liberal result) of the variation in mathematics achievement, based on models that included or excluded shared variance with language skills respectively. The models exploring spatial-mathematical relations in Chapter 2 also controlled for demographic factors including gender, ethnicity and SES. For the first time, it was determined that spatial thinking remains a significant predictor of mathematics, even after controlling for these demographic factors. Taken together, given the large-scale, generalisable nature of the study population in Chapter 2, it can be concluded with some confidence that observed spatial-mathematical associations in childhood do not merely reflect the underlying IQ demands of the tasks used, or differences in the demographic profiles of the participants tested.

The longitudinal design of the study in Chapter 2 allowed investigation of the role of spatial thinking for mathematics over an important developmental age range, the age at which children first enter formal schooling. This study provides the first evidence that intrinsic-dynamic spatial skills are significant concurrent and longitudinal predictors of general mathematics achievement at 5 and 7 years respectively. When compared to language ability, spatial skills were a weaker concurrent predictor of mathematics at 7 years. In contrast, spatial skills at 5 years were a stronger longitudinal predictor of mathematics than language skills. This suggests a particular longitudinal connection between spatial skills and mathematics performance, which may reflect the fact that intrinsic-dynamic spatial skills have a greater role for mathematical outcomes at earlier stages in development. This finding is interesting as it highlights a specific, positive role for early spatial skill training for later mathematics achievement. It supports previous findings that spatial thinking plays a greater role for the acquisition of new mathematics skills, compared to practiced ones (Ackerman, 1988; Uttal & Cohen, 2012). At 5 years, children in the UK begin formal schooling and thus are faced with large amounts of new mathematics material. The findings of this study support the concept that children with strong spatial skills at 5 years are better able to learn new mathematical concepts, which in turn influences their later mathematical performance.

These findings were extended in Chapter 3 to investigate the continuing role of spatial thinking for mathematics throughout the later primary school years. In this chapter, developmental profiles of spatial thinking across each of Uttal *et al.'s* (2013) spatial sub-domains were provided for children in consecutive age groups from 6 to 10 years. Performance accuracy increased across all spatial sub-domains with increasing age, with some subtle differences between intrinsic skills (disembedding and mental rotation) and extrinsic skills (spatial scaling and perspective taking). Intrinsic spatial skills showed rapid early development that slowed after age 8, and for some tasks (mental rotation) started to plateau. Extrinsic skills showed more gradual development that was reflected by a steady increase in performance from 6 to 10 years with no significant differences in performance between consecutive age groups. These profiles for children across the primary school years are the first charting spatial development across each of Uttal *et al.'s* (2013) sub-domains in children of this age. Detailing the developmental trajectories of each spatial sub-domain is informative to establishing an understanding of the relational structure of these skills. It also provides a set of benchmarks against which spatial development in atypical populations, and the development of other tasks (including both spatial and other cognitive tasks), can be compared. These developmental findings should also be interpreted in the context of the individual differences that were reported in spatial task performance at all ages. Environmental, biological and cultural factors may explain the differences in spatial task performance reported between children of the same age.

Having identified the developmental trajectories of these spatial skills, the role of each of Uttal *et al.'s* (2013) spatial sub-domains as predictors of mathematics were compared. Given the role of language skills in explaining mathematics outcomes that was outlined in Chapter 1, the models presented in this chapter also controlled for language ability. Overall, spatial skills explained 5 to 14% of the variation across three mathematics performance measures (standardised mathematics skills, ANS skills and number line estimation skills). Spatial scaling (extrinsic-static sub-domain) was identified as a particularly important predictor of all mathematics outcomes while disembedding (intrinsic-static sub-domain) was also a predictor of standardised mathematics performance. It is worth noting that there are few spatial scaling tasks suitable for administration to a wide age range of children in middle childhood. The Spatial Scaling Task designed for use in Chapter 3 of this thesis offers the first age-appropriate measure of spatial scaling for children aged 6 to 10 years. The study presented in Chapter 3 also found that some spatial-mathematical relations were developmentally sensitive and showed variation across age groups from 6 to 10 years. Mental rotation (intrinsic-dynamic spatial skill) was a predictor of standardised mathematics performance and 0-10 number line estimation at 6 and 7 years only. For the 0-1000 Number Line Estimation task, spatial scaling and disembedding were significant predictors at 7, 8 and 9 years, but not at 10 years. However, this was possibly due to ceiling effects in performance at 10 years and should be interpreted cautiously. Taken together, some spatial skills were significant predictors of mathematics across all age groups while others predicted mathematics outcomes in a task and age specific manner.

The findings of Chapters 2 and 3 were used to inform the design of the spatial training study presented in Chapter 4. Given the role of spatial scaling for mathematics across different tasks and age groups, this skill was chosen as a training target. The reported associations between mental rotation and mathematics in younger children, its use in previous studies, and the proposed theoretical explanations of associations between mental rotation and mathematics, made it suitable to be included as a training target. In this study, spatial training in both mental rotation and spatial scaling was administered and was effective in eliciting near transfer of gains. Spatial scaling training led to gains in spatial scaling performance, while mental rotation training led to gains in mental rotation performance. Although previous findings suggest that spatial thinking (intrinsic-dynamic sub-domain) is malleable in childhood (Uttal et al., 2013), these findings demonstrate, for the first time, the malleability of spatial scaling in children of this age. Intermediate transfer was reported from spatial scaling training to mental rotation performance; however, no similar transfer was reported between mental rotation training and spatial scaling performance. Far transfer of gains from spatial training to mathematics was task dependent. Mental rotation training led to gains in Missing Term Problems, spatial scaling training led to gains in accuracy on the Number Line Estimation Task, and both types of spatial training (mental rotation and spatial scaling) led to accuracy gains on Geometry Shape Items.

In Chapter 4, the effectiveness of implicit and explicit instruction as methods of training spatial thinking were compared. For most outcomes (except for the Geometry Task), there was no difference in the effectiveness of implicit (practice with feedback) compared to explicit instruction (instructional videos). For Geometry Shape Items, only implicit instruction rendered significant gains. No difference between implicit and explicit instruction was found for the spatial measures, Missing Term Problems or the Number Line Estimation Task. The study outlined in Chapter 4 provides some of the first evidence that explicit instruction using instructional videos can lead to improvements in spatial thinking with some transfer to mathematical domains. From a practical perspective, the instructional videos designed for use in this thesis, offer an easily implementable way of introducing spatial thinking into the classroom.

This training study is the first to explore the transfer of spatial training gains to mathematics while also controlling for motivational factors. There were no between training group differences in participants' expectations of training (measured pre-training). For implicit training, engagement was higher for participants in the control group compared to those who completed spatial scaling training (measured post-training). However, as outlined in Chapter 4, control training did not lead to training effects. Therefore, levels of engagement did not align with training effects and cannot explain the gains reported. In summary, spatial training led to near (to the specific spatial skills trained), intermediate (to untrained spatial skills) and far (to mathematics domains) transfer of gains. These gains were broadly similar for explicit and implicit instruction, except for Geometry Shape Items. Furthermore, the gains reported could not be attributed to motivational factors.

5.3 Theoretical conclusions

Overall, this thesis offers convincing evidence that spatial and mathematical thinking are associated. The complementary perspectives presented in each chapter provide refinement of the current understanding of spatial-mathematical relations leading to three main theoretical conclusions. First, relations between spatial and mathematical skills are sub-domain specific. Second, associations between spatial and mathematics skills are sensitive to developmental age. Third, spatial skills have a causal role in mathematical performance. Several other secondary conclusions can also be drawn from the findings presented in this thesis. These conclusions relate to the role of demographic and gender differences in spatial thinking, and the degree to which spatial skills uniquely explain mathematics performance, when considered in the context of language ability. Each of these conclusions are discussed in turn.

5.3.1 Specificity of spatial-mathematical relations

It has previously been argued that a single underlying cognitive mechanism, such as the Mental Number Line, explains all spatial-mathematical relations (Barsalou, 2008; Lakoff & Núñez, 2000). If this were the case, one would expect that different spatial skills would be similarly predictive of mathematics outcomes, as each association would be underpinned by the same process. This pattern of performance is not supported in this thesis, as not all spatial sub-domains were associated with mathematics to the same degree (Chapter 3). Furthermore, spatial training did not lead to uniform transfer of gains to all mathematics domains (Chapter 4) as would be expected if a single process underpinned all spatial-mathematical relations. For example, mental rotation training did not confer benefits for the Number Line Estimation Task and spatial scaling training was not beneficial for Missing Term Problems. Due to the reported selectivity of spatial-mathematical relations and the task specific transfer of spatial training gains to mathematics, no single known explanation is sufficient to explain the relationship between spatial and mathematical thinking. Instead, it is proposed that the underlying cognitive processes governing observed spatial-mathematical associations differ across mathematical and spatial sub-domains. Proportional reasoning, active processing, and form perception, are proposed as candidate mechanisms that may explain the spatial-mathematical relations reported in this thesis. For other spatial and mathematics skills, differing underlying mechanisms may be responsible. These conclusions are not intended to dispute the existence of a mental number line, but imply that no single known explanation, including the Mental Number Line, can explain the sub-domain specific results reported here and in other similar studies (Mix et al., 2016).

The thesis findings emphasise that spatial scaling (extrinsic-static sub-domain) is a particularly strong predictor of mathematics skills, accentuating the need to understand the mechanisms of this specific spatial-mathematical association. The findings support the theoretical prediction that proportional reasoning is the shared mechanism underpinning relations between spatial scaling and mathematics. As previously outlined, proportional reasoning is the ability to encode intensive quantities such as proportions or ratios. In spatial scaling tasks, proportional reasoning is required to assess the relative distances between two differently sized spaces (Newcombe et al., 2018). In this thesis, spatial scaling was only associated with mathematics skills that are also proposed to have proportional reasoning requirements. For the ANS Task, proportional reasoning facilitates ordinal comparisons of quantities. Mapping numbers onto a number line, which is a non-discrete (intensive) way to represent numbers, also requires proportional reasoning. For general mathematics achievement, proportional reasoning may be required for a range of tasks including reasoning about fractions and reading graphs. However, proportional reasoning is not theoretically useful for mathematics tasks that use extensive (exact) quantities such as the Missing Term Problems outlined in Chapter 4. In support of this, spatial scaling training did not lead to gains in Missing Term Problems. The patterns of performance reported in this thesis reflect and support these theoretical predictions, lending support to proportional reasoning as the candidate mechanism underpinning the role of spatial scaling for mathematics in childhood populations.

Active processing using mental visualisations was theoretically proposed as the underlying cognitive mechanism explaining relations between spatial tasks that require transformations, and mathematics outcomes. Intrinsic-dynamic spatial tasks including the Pattern Construction Task (Chapter 2) and the Mental Rotation Task (Chapters 3 and 4), and extrinsic-dynamic spatial tasks including the Perspective Taking Task (Chapter 3), each rely on mental transformations. Significant associations were found between the Pattern Construction Task (Chapter 2) and the Mental Rotation Task (Chapter 3), and standardised mathematics performance. It is proposed that within standardised mathematics tests, several items may be answered by using mental visualisation strategies. Questions presented as word problems may be solved using mental visualisation to imagine and cognitively manipulate the problem in pictorial format. Similarly, in multi-step mathematics problems, mental visualisations may be used to plan the steps needed to solve a given question. Mental rotation skills were also associated with performance on Missing Term Problems (Chapter 4). As outlined in Chapter 4, for Missing Term Problems, children may use mental visualisations to represent the equations presented pictorially.

Mental rotation performance was not a significant predictor of other mathematics outcomes including number line estimation and ANS skills, which are not theoretically predicted to recruit mental visualisation strategies (Chapter 3). The findings in this thesis provide evidence for the theoretical prediction that mental visualisations underpin associations between some spatial and mathematics tasks. However, it is noteworthy that not all tasks that require mental visualisation were significant predictors of mathematics. There were no significant associations between the Perspective Taking Task and mathematics skills. Thus, it is concluded that there are distinctions between the roles of different types of mental visualisation for mathematics. The implications of these findings are that object-based transformations such as those required for mental rotation are important for mathematics, while mental visualisations requiring imagined self-movement (e.g., perspective taking) do not appear to underpin spatial-mathematical relations. This distinction should be considered in the design of training studies targeting mental visualisation skills.

Finally, form perception, the ability to distinguish shapes from a more complex background or to break more complex pictures into parts, was also predicted to explain associations between certain spatial and mathematics tasks. This prediction was supported in this thesis by the finding that disembedding skills (intrinsic-static sub-domain) predicted both number line estimation and standardised mathematics performance (Chapter 3). Form perception is theoretically useful for identifying shapes and symbols, which is required for disembedding tasks like the CEFT. In the Number Line Estimation Task, form perception is also required for identifying numbers, in addition to interpreting demarcations on the number line itself (e.g., start and end points). For standardised mathematics, it is proposed that form perception skills allow individuals to distinguish symbols, such as + and × symbols, interpret charts and graphs, and understand the spatial relations between symbols and numbers in multi-digit numbers. The significant associations described between disembedding and certain mathematics tasks implies that form perception is an underlying mechanism that explains spatial-mathematical relations.

5.3.2 Developmental sensitivity of spatial-mathematical associations

This thesis has demonstrated that spatial thinking plays an important role for mathematics on entry into formal education and continues to be a significant predictor of mathematics across consecutive age groups in the primary school years. There is also reason to believe that the role of some spatial sub-domains as a predictor of mathematics varies through development from 6 to 10 years. As demonstrated in Chapter 2, intrinsic-dynamic spatial skills are a stronger longitudinal predictor (at 5 years) of later mathematics outcomes (at 7 years) when compared to language skills. However, this is not the case at age 7. Although it cannot be determined definitively, this suggests that the relative role of intrinsic-dynamic skills, compared to language skills, may be greater in younger childhood. Furthermore, as outlined in Chapter 3, mental rotation (intrinsic-dynamic sub-domain) is a predictor of mathematics for younger children (6 and 7 years) but not older children (8, 9, 10 years). Again, this suggests that the relative role for intrinsic-dynamic spatial skills in explaining mathematics outcomes decreases with age. These findings can be interpreted in light of other studies in this domain. Mix *et al*. (2016; 2017) also reported differential associations between spatial and mathematics skills at different ages and highlighted an age-specific role for mental rotation at 6, but not at 9 or 12 years. Taken together, it can be concluded that there is a developmental transition period during which there is a change in the impact of spatial skills on mathematics ability. Although the exact timing of this period is likely to vary between children, it is proposed to occur between the ages of 7 and 8.5 years, and is defined by a reduction in the role of intrinsic-dynamic spatial skills for mathematics performance.

These findings may be explained in several ways. First, certain spatial skills may play a greater role for the completion of novel, compared to practiced mathematics skills (Ackerman, 1988; Uttal & Cohen, 2012). As described above, intrinsic-dynamic spatial skills are proposed to require mental visualisation processes. These processes may be particularly useful in providing scaffolding during the learning of novel mathematical concepts. In the early primary school years, children are presented with substantial amounts of new mathematical material, compared to later years. Children with strong mental visualisation skills (reflected by high performance on intrinsic-dynamic spatial tasks) may apply these skills when learning novel mathematical concepts, which in turn may improve their mathematics performance. Second, through development, children may acquire new, more efficient strategies for learning and completing mathematics tasks. For example, children may rely more on memory techniques, or WM strategies, rendering the use of mental visualisation strategies redundant. Finally, the use of mental visualisation strategies may be constrained to certain sub-components of mathematics. As children get older, the types of mathematical concepts and tasks that they are required to learn and complete, may not be amenable to the use of mental visualisation. Having strong mental visualisation capabilities would not be expected to enhance an individual's ability to rote learn mathematical times tables for example, something that may be required in the later primary school years. To conclude, there is evidence that some spatial skills, particularly intrinsic-dynamic skills, have a particularly important role for mathematics in the early primary school years. However, there is evidence that this role decreases with developmental age.

5.3.3 Causal role of spatial skills on mathematics

The positive impact of spatial training on both spatial tasks and mathematical sub-domains provides evidence for a causal influence of spatial thinking (mental rotation and spatial scaling) on mathematics performance in children (see Figure 5.1). This conclusion is further strengthened by the fact that the training study outlined in this thesis controlled for motivational factors including expectations of, and engagement with training. Therefore, the training gains reported are not attributable to motivational factors alone.

This causal relationship between spatial skills and mathematics can be inferred because a manipulation in one variable (spatial skill) led to changes in the other variable (mathematics skill) (Pearl, 2000). It may be argued that a common cause such as a genetic influence, IQ, language skills or other cognitive skills such as WM may be influential for both spatial and mathematics outcomes. However, the findings reported in Chapter 4 indicate that the observed correlations between spatial and mathematical thinking cannot solely be explained by a common cause acting on both variables. As shown in Figure 5.1, without a direct cause between spatial and mathematical thinking, intervening on spatial skills would not lead to changes in mathematical outcomes. Thus, while a common cause such as a general cognitive factor or neural features may also exist between spatial and mathematical thinking (Oberauer, 2016), this study identified a specific, direct causal effect of spatial skills on mathematics performance.

Furthermore, these findings do not preclude a causal role of mathematical thinking on spatial skills, i.e., a bidirectional relationship (feedback loop) may exist between spatial and mathematical thinking. From a practical perspective, finding novel methods of improving mathematical thinking in children is an educational priority (National Audit Office UK, 2018) and this study aimed to determine the causal effect of spatial skills on mathematics. However, to establish whether a bidirectional relationship exists between spatial and mathematics skills, future research is needed investigating the effects of training mathematics skills on spatial performance.

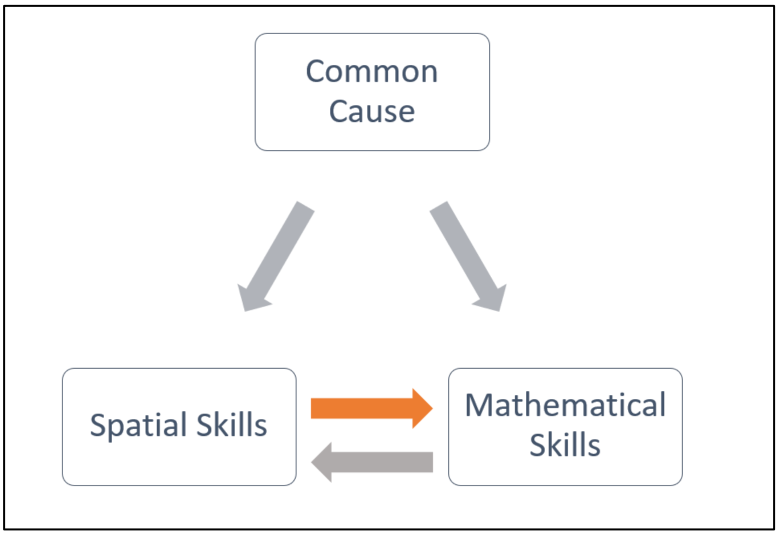

Figure 5.1. The causal relationship between spatial and mathematical thinking.
Note. Established and speculative causal relations are shown in orange and grey respectively.

5.3.4 Other theoretical conclusions

Although not a primary aim of this thesis, the results reported further extend the current understanding of whether there is variation in spatial skill across different demographic groups and between children of different genders. In Chapter 2, significant differences in spatial task performance at both 5 and 7 years were reported for children in different SES groups (income-based quintiles). Those from higher SES backgrounds outperformed their lower SES counterparts. Small to medium sized effects were reported. One factor that may contribute to these differences is that children from lower SES households may engage less regularly in spatial-related play, as they may have fewer resources and opportunities to do so. This may reduce their spatial experiences and subsequently lower their spatial task performance. This proposal is supported by work completed during the course of this PhD, in parallel with this thesis, that shows that children from lower SES groups have less regular engagement in early numeracy activities (including spatial activities such as block building) (Clerkin & Gilligan, 2018). Taken together, children from lower SES backgrounds have reduced spatial performance that may be attributable to early childhood experiences of spatial and numerical play. Further research is required to explore this hypothesis. It is also noteworthy that there were no significant interactions between SES and spatial skills in predicting mathematics outcomes. This suggests that although there are differences in spatial abilities across SES groups, the relationship between spatial and mathematical skills does not differ across groups. This shows that spatial skills are similarly important predictors of mathematics for children from high and low SES groups, a fact, as outlined later, that has important implications for the design of training studies.

It has remained an issue of debate whether there are gender differences in spatial ability in childhood. Many studies have outlined a male advantage in spatial task performance in the pre-school and primary school years (e.g., Carr et al., 2008; Casey et al., 2008). However, these differences may be attributable to gender biases in the types of tasks traditionally used to measure spatial thinking. More recently, and in contrast with previous findings, other studies have reported no significant male advantage on spatial measures (e.g., LeFevre et al., 2010; Manger & Eikeland, 1998;

Neuburger et al., 2011). All three experimental chapters in this thesis investigated the role of gender differences in spatial thinking. For the majority of tasks administered as part of this thesis, there were no significant gender differences in spatial performance. The only significant gender differences reported for spatial task performance favoured females (Chapter 2), and the effect sizes for these differences were small. Given the diversity of spatial tasks assessed, and the age range of children tested, the findings reported in this thesis add considerable weight to the argument against a male advantage in spatial thinking in childhood. Moving beyond a main effect of gender on spatial thinking, Chapter 2 also explored differences in the predictive role of spatial thinking for mathematics for boys and girls. Based on a nationally representative sample of children in the UK, the findings showed no interaction between spatial thinking and gender in predicting mathematics skills. This shows that spatial skills are similarly important predictors of mathematics for boys and girls, which as outlined for different SES groups above, has important implications for designing training studies.

Taken together, findings in this thesis provide evidence that individual variation in spatial thinking is greater than the variation explained by gender differences. This emphasises a need to move beyond gender-focused approaches to understanding differences in spatial thinking. Instead, possible targets for future intervention studies include the role of home-based experiences in the development of spatial skills. For example, the role of early play with spatial toys (Jirout & Newcombe, 2015; Ramani, Zippert, Schweitzer, & Pan, 2014) and the influence of exposure to spatial language (Pruden, Levine, & Huttenlocher, 2011). There is a need to elucidate what other factors, both genetic and environmental, shape the development of early spatial abilities.

This thesis also disentangles the relative roles of spatial and language skills in explaining mathematics outcomes. Although no measures directly measuring IQ were administered in this thesis, IQ is bluntly estimated by exploring shared variation between spatial, mathematics and language skills. It has been suggested previously that shared variance between these skills may be underpinned by their shared IQ demands (Alloway & Alloway, 2010; Mayes et al., 2009). In Chapter 2, it is apparent that some variation in mathematics is attributable to shared variance between spatial and language skills. This shared variation may be due to the overlapping IQ demands of the tasks used. It is also evident that spatial thinking explains unique variation in mathematics ability beyond language skills. This variation is unlikely to be attributable to IQ demands, as if it were, it would also be anticipated to overlap with language skills. These findings are interpreted to conclude that observed spatial-mathematical associations in childhood do not simply reflect underlying IQ demands of the skills measured. This is further supported by the study in Chapter 3, which showed that spatial skills continue to explain a significant proportion of the variation in mathematics outcomes even after controlling for language skills. The findings identify the distinct and overlapping contributions that language and spatial skills play in explaining mathematics achievement, highlighting an important role for both skills.

5.4 Implications

5.4.1 Educational implications

The evidence presented in this thesis strongly advocates for *spatialisation* of primary school mathematics curricula. Unfortunately, as outlined by Davis *et al.* (2015), mathematics curricula do not typically focus on spatial thinking. The current UK mathematics curriculum at Key Stage 2 explicitly refers to spatial thinking only once, in reference to the representations of large numbers (UK Department for Education, 2013). Hence, the findings reported here suggest that there is a need for *spatialisation* of the primary school classroom such that children are taught how to read diagrams and graphs, encouraged to sketch and draw, exposed to spatial language, and given hands-on opportunities to manipulate and explore with 3D materials (Newcombe, 2013). Enhancing spatial thinking in children may have both direct and indirect benefits for attainment. Given the ease with which they can be delivered, the findings from this thesis highlight the potential of instructional videos (explicit instruction) as a practical tool for *spatialising* the classroom.

In the Programme for International Student Assessment (PISA) study, an international assessment of mathematics, reading and science, at 15 years students in England (where the studies presented in Chapters 2, 3 and 4 were completed) perform at the international average (Jerrim & Shure, 2016). The mathematics scores of children in England have not increased significantly since 2006, which is worrying given that significant gains have been reported for other European countries over the same period of time, e.g., Portugal (Jerrim & Shure, 2016). In the most recent wave of the Trends in International Mathematics and Science Study (TIMSS), at 10 years students in England performed above the international average in mathematics. However, they had lower performance on the Geometry, Shapes and Measures sub-domain compared to other mathematics content domains including Number and Data Display (Greany et al., 2016). At 14 years, performance on the Geometry sub-domain was also significantly lower than overall mathematics performance (Greany et al., 2016). Comparable findings were reported for other countries in the UK. These findings are noteworthy as they highlight geometry and shape sub-domains of mathematics as mathematical content areas that children in the UK find challenging. Superficially, there are clear links between geometry sub-domains, where there is an emphasis on space and shape, and spatial skills. This supports the argument for targeting spatial thinking as a means of increasing mathematics attainment in the UK.

The introduction of spatial training and the use of spatial tools in mathematics classrooms may be a novel way of improving mathematics performance on international assessments, particularly in space and shape related domains. Other collaborative work completed during this PhD also highlighted a role of spatial thinking for science outcomes. More specifically, mental folding (an intrinsic-dynamic spatial task suitable for older children) predicted physics and chemistry outcomes, spatial scaling predicted chemistry and biology outcomes, and disembedding predicted chemistry outcomes, for children at 8 to 11 years (Hodgkiss et al., 2018). Although further research is required to elucidate the causal relations between spatial thinking and science, given the associations reported in Hodgkiss *et al.* (2018), it is possible that training spatial skills may also have benefits for science outcomes. Therefore, integrating spatial thinking into STEM classrooms may offer a novel way of improving students' academic outcomes in mathematics and science. Beyond direct benefits to individual student outcomes, raising mathematics and science attainment using spatial thinking may in turn improve the national standard of mathematics and science performance on an international stage.

As outlined in Chapter 2, no significant interactions were reported between gender and spatial thinking, nor SES and spatial thinking, in predicting mathematics outcomes. Given the size of the sample tested in Chapter 2, these findings offer convincing evidence that there are not differential relations between spatial and mathematics skills for children of different genders or for those in different SES groups. This suggests that targeting future training studies to distinct SES groups, or to males or females specifically, is not necessary. Thus, while there is evidence that the relations between spatial and mathematical skills are developmentally sensitive (Chapter 3), spatial-mathematical relations appear to be consistent across demographic groups. As such, it is proposed that the benefits of spatial training are not limited to specific sub-groups of children. From an educational perspective, this greatly enhances the ease with which spatial training tools can be designed and tested. It also increases the practicality of introducing spatial thinking into the classroom.

Taken together, the findings presented in this thesis have implications for educational outcomes at both an individual level and a national level. This thesis proposes that spatial training be introduced into the primary school classroom for all children as a means of improving spatial and mathematical thinking. The extended benefits of this may be an increase in student attainment on international mathematics assessments. There is some cross-sectional evidence that spatial training may also confer benefits to science domains.

5.4.2 Economic and societal implications

Outside the classroom, improving mathematics (and other STEM outcomes) may have a wider economic impact. As outlined in Chapter 1, many employers report difficulties recruiting suitably qualified STEM graduates (CBI, 2013) and improving STEM skills is a pressing economic priority (CEBR, 2015). Figures from the National Audit Office UK (2018) estimated 1.5 million STEM recruitment shortages in 2018, where employers were unable to hire suitable STEM employees. This shortfall emphasises a) a need to increase the number of individuals pursuing STEM careers, and b) a need to increase the quality of the mathematics and STEM skills of students entering and graduating STEM courses. One initiative that has been taken to combat the growing demand for STEM graduates has been to fund mathematics-focused interventions in primary schools. The UK Department for Education has invested £55 million into Maths Hubs in the UK (National Audit Office UK, 2018). Much of this funding has gone towards teaching "Maths Mastery" in schools. However, spatial thinking is not a key element of "Maths Mastery" and, as outlined in this thesis, engagement with and improvement of spatial thinking may lead to significant gains in mathematics. The transfer of spatial training gains highlighted in this thesis supports the use of spatial training in mathematics instruction. Although there is a need to replicate and extend these findings with larger samples, spatial training, such as that proposed in Chapter 4, may improve the quality of STEM graduates with consequent improvements for the STEM industry.

Beyond its role in education, spatial thinking is a valuable skill in everyday life. As outlined in Chapter 1, spatial thinking is required for a range of everyday activities such as stacking shelves, navigating around a shopping centre, parking a car and assembling furniture. However, for the vast majority of everyday skills that require spatial thinking, no formal training is provided (National Research Council of the National Academies, 2006). Spatial thinking plays a significant role in everyday activities that have health and safety implications, such as driving a car and operating machinery. Spatial training using simple instructional videos, as demonstrated in this thesis, may be an effective way to improve the accuracy of everyday spatial skills, which may have significant economic and societal impact, beyond its role in education.

5.5 Limitations and future directions

The results of this thesis should be interpreted in the context of their limitations, and the scope for future research. The study reported in Chapter 2 highlighted the value of using secondary data sets to explore cognitive development longitudinally. This study addressed recent appeals for wider utilisation of data from large-scale studies, particularly with regard to “articles exploring important aspects of the teaching and learning environment” (Lenkeit, Chan, Hopfenbeck, & Baird, 2015). This study draws attention to the wealth of educational and cognitive information that can be found in the MCS study. For international comparisons, future studies should use other large-scale studies such as the TIMSS and PISA studies. While these studies predominantly focus on educational achievement (including mathematics achievement), they also include cross-national psychological and sociological data suitable for investigating other influences on mathematical performance. As previously outlined, one example is the study by Clerkin and Gilligan (2018) which explored the role of pre-school numeracy play (including spatial play) on mathematics achievement and attitudes towards mathematics using the TIMSS data set.

The findings reported in Chapter 2 were limited in that they were isolated to a single age range of children. Using the Pattern Construction scores (intrinsic-dynamic sub-domain) investigated in Chapter 2, and the Spatial Working Memory task, the only spatial measure included in Wave 5 (age 11) of the MCS, future studies might link spatial skills in the primary school years with mathematics achievement at secondary school and beyond. Furthermore, while the MCS dataset enabled examination of the relationship between intrinsic-dynamic spatial skills and mathematics achievement at 7 years, the findings were limited to a single spatial sub-domain. This reflects one of the major limitations of using secondary data to answer novel research questions; the availability of suitable test measures. To expand these findings beyond the intrinsic-dynamic sub-domain there is a need for a) smaller cross-sectional studies such as that described in Chapter 3 and, b) the inclusion of a wider range of spatial measures in a large-scale, longitudinal project.

The study presented in Chapter 3 was the first to explicitly compare the role of Uttal *et al.'s* (2013) four sub-domains of spatial thinking in explaining mathematics outcomes. However, despite including all of Uttal *et al.'s* (2013) sub-domains, the study is limited in that it focuses on small-scale spatial thinking only. Small-scale spatial thinking involves table-top tasks, where there is no need for whole-body movement or for changing location (Broadbent, 2014). Future work might extend these findings to include large-scale spatial processes which require movement and observations from a number of vantage points, e.g., using real world or virtual navigation tasks (Kuipers, 1978; 1982). Furthermore, the study presented in Chapter 3 was the first to explore associations between spatial and mathematics skills in children aged 6 to 10 years, across each of Uttal *et al.'s* (2013) sub-domains, using a cross-sectional approach. However, the findings could be strengthened by longitudinal research following a single cohort of participants through development from 6 to 10 years.

In Chapter 4, it was shown that training spatial skills leads to near, intermediate and far transfer to mathematics. The duration of the spatial training delivered in this study was relatively short, demonstrating that even short bouts of spatial training lead to transfer of training gains to mathematics. However, this study did not investigate dosage effects, and future research is needed to investigate whether the amount of training delivered influences the size and durability of training gains. This study was also limited by the short interval between training and post-testing. On one hand, regardless of the durability of gains, the results demonstrated shared cognitive processing between spatial and mathematics skills, which was modified through intervention. Thus, this study led to the identification of a causal effect of spatial thinking on mathematics. On the other hand, it is possible that the gains reported in this study were due to priming of certain strategies for task completion, and not conceptual change. Hence, prior to national rollout of spatial training in the classroom, more research is needed to investigate the durability of spatial training gains in children. Training studies including neuroimaging could also be used to investigate whether neurophysiological changes occur during training. However, based on other studies investigating spatial training in adults, durability of spatial training gains in children are anticipated (Uttal et al., 2013). In short, even if the findings reported for spatial training reflect a priming effect, the results of this study have significant practical applications for teachers, given that priming enhanced performance on mathematics performance. Transfer of gains from spatial training to mathematical skills may reflect both priming and conceptual change. These two processes are necessarily inter-linked, as it is not possible to prime a process that you have not yet developed.

While most previous spatial training studies are based on mental rotation (or similar spatial tasks) (Uttal et al., 2013), the study presented in Chapter 4 demonstrated an important role for other spatial sub-domains, particularly spatial scaling. Mental rotation and spatial scaling were selected as training targets in this study, as these tasks specifically relate to mathematics outcomes at 8 years (Mix et al., 2016; 2017). However, future studies should explore whether spatial training using age appropriate targets might confer benefits to spatial and mathematics performance in older children, for example by training visuo-spatial thinking which has been associated with mathematics outcomes at 11 years (Mix et al., 2016;2017) respectively. Furthermore, given cross-sectional evidence that the role of spatial thinking extends beyond mathematics to other STEM domains (Hodgkiss et al., 2018; Wai et al., 2009), future studies could explore transfer of spatial training gains to performance in other STEM subjects.

Finally, although Chapter 2 and Chapter 3 control for the role of language skills (IQ), the results of this thesis did not control for other cognitive demands including working memory and executive functions, which may also contribute to associations between spatial and mathematical tasks (e.g., Gilmore et al., 2013; Hawes et al., 2017). However, previous studies suggest that spatial skills show specificity in predicting STEM outcomes (Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017; Wai et al., 2009). Factor-analysis studies also suggest that spatial-mathematical associations cannot be explained by other cognitive factors alone (Bailey, 2017). This evidence suggests that observed spatial-mathematical relations are not merely a reflection of other cognitive factors although, to confirm this, the study completed in Chapter 3 could be repeated controlling for other cognitive systems including WM, attention and executive functions.

5.6 Conclusion

The findings of this thesis elucidated and refined the relations between spatial and mathematical skills across development from 5 to 10 years. Spatial thinking was identified as a unique predictor of mathematics, distinct from language ability. Spatial scaling (extrinsic-static spatial skill) was highlighted in this thesis as a particularly strong predictor of mathematics, which is amenable to change through training, and leads to far transfer of training gains to mathematics. Disembedding was also a significant predictor of mathematics for some outcomes. That associations between spatial and mathematics skills are sensitive to developmental age also clearly emerged from the thesis. Mental rotation and pattern construction skills (intrinsic-dynamic sub-domain) were found to be stronger predictors of mathematics in younger children.

Including the first known study to control for motivational factors, the thesis allowed determination of the causal effect of spatial skills on mathematical performance. Training spatial skills led to near, intermediate and far transfer of gains and explicit instruction using specifically designed instructional videos was identified as a way of delivering spatial training in a classroom-based setting. The theoretical, educational and economic implications of the findings of the thesis are significant and emphasise the importance and value of developing novel ways to enhance the spatial skills of primary school aged children. Focusing on this area will serve to improve the mathematics skills of these children and will better equip them for a changing society in which mathematics and other STEM skills are becoming more important. This is a goal worth pursuing.

Appendices

Appendix A

The table below shows the mapping of spatial categories from previous models onto the Uttal *et al.* (2013) model of spatial skills (adapted from Uttal *et al.* 2013). A description of each task is included below the table.

Uttal et al. sub-domain (2013)	Description	Examples of measures	Linn & Petersen (1985)	Carroll (1993)
Intrinsic and static	Perceiving objects, paths, or spatial configurations amid distracting background information	Embedded Figures tasks[A], flexibility of closure[B]	Spatial visualization	Visuospatial perceptual speed
Intrinsic and dynamic	Piecing together objects into more complex configurations, visualizing and mentally transforming objects, often from 2-D to 3-D, or vice versa. Rotating 2-D or 3-D objects	Form Board[C], Block Design[D], Paper Folding[E], Mental Cutting[F], Mental Rotations Test[G], Cube Comparison[H], Perdue Spatial Visualization Test[I], Card Rotation Test[J]	Spatial visualization, mental rotation	Spatial visualization, spatial relations/speeded rotation
Extrinsic and static	Understanding abstract spatial principles, such as horizontal invariance or verticality	Water-Level[K], Water Clock[L], Plumb-Line[M], Cross-Bar[N], Rod and Frame Test[O]	Spatial perception	Not included
Extrinsic and dynamic	Visualizing an environment in its entirety from a different position	Piaget's Three Mountains Task[P], Guildford-Zimmerman spatial orientation[Q]	Not included	Not included

[A] Embedded Figures Tasks: Tasks of this type require identification of the spatial configuration of one object against a distracting background (Ekstrom, French, Harman, & Dermen, 1976; Okamoto et al., 2015; Witkin & Goodenough, 1981; Witkin, Otman, Raskin, & Karp, 1971).

[B] Flexibility of Closure Tasks: These tasks require participants to identify whether a series of complex drawings contain a more simplistic figure (Thurstone and Jeffrey, 1984).

[C] Form Board: In tasks of this type participants are shown an image in several disarranged parts and must determine which of several other pictures shows the pieces together, e.g., Minnesota Paper Form Board (Likert, 1970)

[D] Block Design: In block design tasks participants are shown a pattern and are asked to recreate the pattern by rearranging individual blocks

[E] Paper Folding: In paper folding tasks participants are shown an image of a piece of paper that has been folded. They are asked to determine what the paper would look like unfolded (e.g., Ekstrom, French, & Harman, 1976).

[F] Mental Cutting: In tasks of this type participants are shown a figure and a cut to the figure along a given plane. They must choose the resulting cross-section from a series of choices.

[G] Mental Rotations Test: Participants are asked to determine which images are rotated versions of a target image (e.g., Vandenberg & Kuse, 1978).

[H] Cube Comparison: In this type of task participants are shown images of two cubes with different letters and numbers on each face. The participant must determine whether the images could be of the same cube (Ekstrom, French, & Harman, 1976).

[I] Purdue spatial Visualisation Test: Participants are shown a reference object (unrotated) and the object again after undergoing a rotation. They are then shown a target object and are asked to determine what the target object would like if it underwent the same rotation (Guay, 1977).

[J] Card Rotation Test: In this task participants are presented with a target 2-D image and must determine which of five other images are a rotated version of the target image (Ekstrom, French, & Harman, 1976).

[K] Water-Level Test: Participants must draw a horizontal line in a tilted bottle or select which image, from a selection of images, shows a horizontal line in a titled bottle (e.g., Harris, Hanley, & Best, 1978).

[L] Water Clock Test: Participants are shown a water clock tilted at different angles. They must determine how water will move from one compartment of the clock to the other. They are also asked to identify the water level in the bottom compartment when it is filled to one third of its capacity (Roberts & Chaperon, 1989).

[M] Plumb-line Test: Participants must determine how a vertical line (i.e., a hanging light bulb) would look when hanging from the roof of a van, on a hill (Liben, 1978).

[N] Cross-bar Test: This task is similar to the water-level task. Participants are asked to identify which image shows a cross bar that is parallel to a horizontal plane, when the crossbar is attached to a movable rod (McGillicuddy-De Lisi, De Lisi & Youniss, 1978).

[O] Rod and Frame Test: In this test, participants must position a rod vertically in a frame that is oriented (Witkin, Dyk & Faterson, 1962).

[P] Piaget's Three Mountains Task: Participants are seated in front of a model of three distinct mountains (of different sizes and with distinguishable features). A doll is positioned at different positions around the model and participants are asked to determine which photograph shows what the doll can see (Piaget & Inhelder, 1976).

[Q] Guilford-Zimmerman Spatial Orientation Task: This task measures spatial orientation. Participants are required to identify the position of a boat that would give rise to a particular view of a landscape (Guilford & Zimmerman 1948).

Appendix B

Number line estimation analysis using PAE scores (Chapter 3)

Only children at 6 and 7 years completed the 0-10 block of the Number Line Estimation Task. No significant difference in performance was reported between these age groups, t (46) = .57, p =.570, d = 0.160. For the 0-100 block of the task, a significant effect of age on PAE scores was reported, F (4,131) = 17.86, p < .001, η_p^2 = .293. This could be explained by a linear contrast (p < .001). Games-Howell post-hoc tests indicated that error scores at 6 years were significantly higher than at 9 or 10 years (p < .001 for both). At 7 years, error scores were also significantly higher than at 9 (p < .005) and 10 years (p < .001). This pattern as also seen at 8 years, with significantly higher error compared to the two older age groups (p < .001 for both). There was no difference in error scores at 9 and 10 years (p = .411). For the 0-1000 block of the Number Line Estimation Task only participants aged 7 and older were included. There was a significant effect of age, F (3,104) = 9.48, p < .001, η_p^2 = .215, best explained by a linear contrast (p < .001). Games-Howell post-hoc tests indicated lower PAE scores at 10 years compared to all other groups including children at 7 years (p = .008), 8 years (p = .020) and 9 years (p = .041). No other significant group differences were found (p > .065).

Table S1.

Factors predicting* PAE scores *on the 0-10 Number Line Estimation Task (n = 48)

	b	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	R^2
Step 1									
Disembedding	-0.03	0.02	-0.20	-1.27	.212	1.95	43	.119	.075
Mental Rotation	0.03	0.02	0.21	1.29	.204				
Spatial Scaling	-0.04	0.02	-0.38	-2.06	.045				
Perspective Taking	0.02	0.02	0.12	0.78	.438				

Table S2:

Factors predicting* PAE scores *on the 0-100 Number Line Estimation Task (n = 136)

	b	***SE***	***ß***	***t***	**p**	***F***	***df***	***p***	R^2	ΔR^2
Step 1										
Age (months)	-0.03	0.01	-0.28	-2.87	.005	24.33	132	< .001	.341	
BPVS	-0.02	0.01	-0.16	-1.73	.087					
Gender	0.02	0.01	0.09	1.33	.186					
Step 2										
Disembedding	-0.01	0.01	-0.11	-1.26	.209	12.24	128	< .001	.368	.017
Mental Rotation	0.01	0.01	0.06	0.72	.471					
Spatial Scaling	-0.02	0.01	-0.17	-1.89	.062					
Perspective Taking	-0.01	0.01	-0.09	-1.04	.300					

Table S3.

Factors predicting* PAE scores *on the 0-1000 Number Line Estimation Task (n =108)

	b	*SE*	*ß*	*t*	p	*F*	*df*	*p*	R^2	Δ R^2
Step 1										
Age (months)	-0.07	0.02	-0.33	-3.05	.003	11.03	104	< .001	.220	
BPVS	-0.02	0.02	-0.10	-0.97	.334					
Gender	0.05	0.03	0.15	1.85	.067					
Step 2										
Disembedding	-0.01	0.02	-0.03	-0.32	.748	6.94	100	< .001	.280	.060
Mental Rotation	0.02	0.02	0.12	1.35	.181					
Spatial Scaling	-0.06	0.02	-0.33	-3.32	< .001					
Perspective Taking	-0.03	0.01	-0.16	-1.68	.096					
Step 3										
Scaling*Age	0.06	0.02	0.30	3.07	.003	7.77	99	< .001	.336	.136

Follow Up: Younger Group	b	*SE*	*ß*	*t*	p	*F*	*df*	*p*	*Adj. R^2*	*Δ Adj.R^2*
Step 1										
Age (months)	-0.06	0.03	-0.20	-1.72	.090	7.09	79	< .001	.182	
BPVS	-0.01	0.02	-0.06	-0.51	.614					
Gender	0.06	0.03	0.19	1.84	.070					
Step 2										
Disembedding	-0.02	0.02	-0.09	-0.79	.430	4.96	75	< .001	.253	.071
Mental Rotation	0.02	0.02	0.09	0.84	.402					
Spatial Scaling	-0.05	0.02	-0.27	-2.24	.028					
Perspective Taking	-0.02	0.02	-0.12	-1.11	.269					

Note. For all regression models, b = unstandardized coefficient; SE = Standard Error; ß = standardised coefficient; ANS = Approximate Number Sense; NFER PiM = National Foundation for Educational Research Progress in Mathematics; BPVS = British Picture Vocabulary Scale

Appendix C

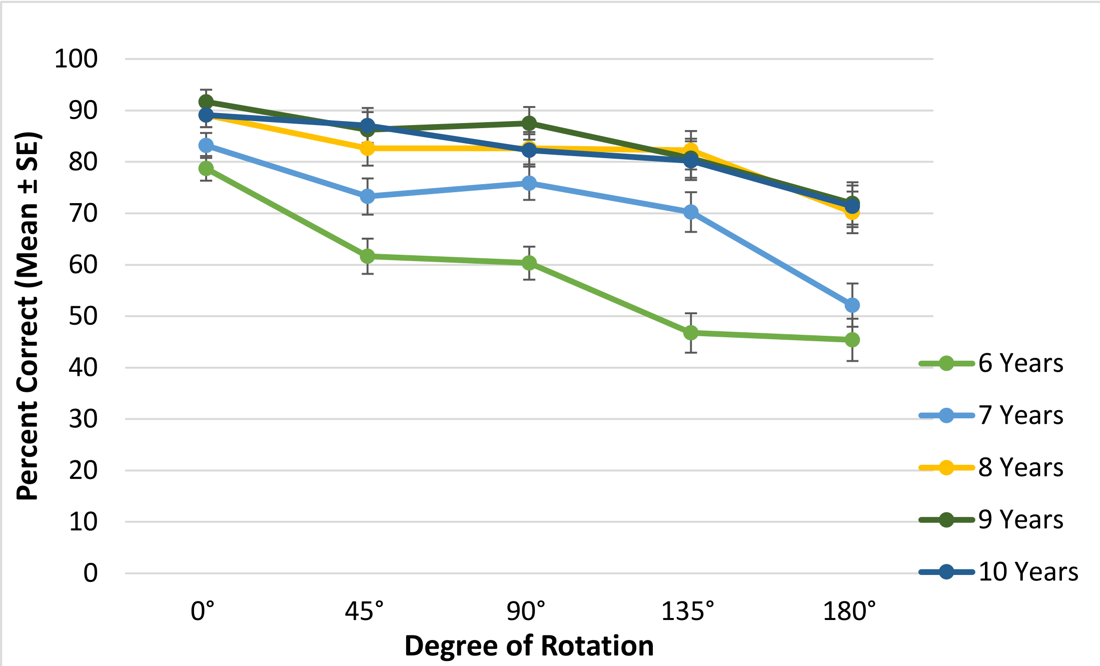

Figure S1. Percentage accuracy on the Mental Rotation Task at different degrees of rotation across age groups (Chapter 3)

Appendix D

Table S4.

Spatial task performance (percentage accuracy) across development (Chapter 3)

Task	Metric	6 Years	7 Years	8 Years	9 Years	10 Years
Disembedding	Mean ± SE	30.62 ± 2.28	35.87 ± 2.40	50.88 ± 2.91	50.71 ± 2.86	56.52 ± 3.22
	Max	56.00	64.00	88.00	92.00	84.00
	Min	4.00	16.00	20.00	20.00	20.00
Mental Rotation	Mean ± SE	52.40 ± 2.89	66.43 ± 3.31	78.52 ± 2.86	80.85 ± 2.39	77.42 ± 3.73
	Max	87.50	100.00	100.00	100.00	100.00
	Min	12.50	31.25	46.88	50.00	6.25
Spatial Scaling	Mean ± SE	37.78 ± 2.55	46.24 ± 3.51	56.77 ± 3.48	64.34 ± 2.85	68.46 ± 2.66
	Max	83.33	94.44	94.44	88.89	94.44
	Min	11.11	16.67	27.48	38.89	38.89
Perspective Taking	Mean ± SE	43.68 ± 2.52	48.75 ± 2.93	57.99 ± 3.14	66.48 ± 3.76	71.15 ± 3.66
	Max	77.78	88.89	94.44	100.00	100.00
	Min	16.67	22.22	27.78	27.78	38.89

Appendix E

Number line estimation analysis using R^2_{LIN} scores (Chapter 4)

A MANOVA was completed with time as a within participant variable (2 levels: pre-training, post-training). Training mode (2 levels: explicit, implicit) and training type (3 levels: mental rotation, spatial scaling, literacy) were included as between participant variables. Training analysis for R^2_{LIN} scores on the Number Line Estimation Task found no significant effect of training mode, F (1,237) = 0.06, p = .815, η_p^2 = .001, or training type, F (2,237) = 2.83, p = .061, η_p^2 = .023. However, the main effect of training type did approach significance. Viewing scores across Time 1 and Time 2, there was a reduction in the function of fit of R^2_{LIN} scores for the control group, and a slight increase in the function of fit of R^2_{LIN} scores for those completing spatial scaling training. No significant interaction between training type and training mode was reported, F (1,237) = 0.14, p = .869, η_p^2 = .001.

Appendix F

Assessing training effects using ANCOVA analysis with baseline performance as a covariate (Chapter 4)

As outlined in 4.3.3, one-way ANCOVAs with baseline performance as a covariate, can be used to explore training effects in studies with pre-post training designs. To investigate the effect of training on task performance, ANCOVAs were completed for each task in the test battery. Training mode (2 levels: explicit, implicit) and training type (3 levels: mental rotation, spatial scaling, literacy) were included as between participant variables. Post-training scores (Time 2) were included as the dependent variable and pre-training scores (Time 1) were included as a covariate.

Consistent with the MANOVA results in section 4.3.2, there was a main effect of training type for: the Mental Rotation Task, F (2,236) = 4.96, p = .008, η_p^2 = .040; the Spatial Scaling task, F (2,231) = 12.09, p < .001, η_p^2 = .094; Missing Term Problems, F (2, 208) = 3.85, p = .023, η_p^2 = .036; PAE scores on the Number Line Estimation Task, F (2,236) = 7.29, p = .001, η_p^2 = .058, and; Geometry Shape Items, F (1,218) = 4.91, p = .008, η_p^2 = .043. All significant differences between groups mirrored those reported in section 4.3.2. Consistent with the results reported in section 4.3.2, there was no main effect of training type for Geometry Symmetry Items, F (2,212) = 0.55, p = .877, η_p^2 = .005. As seen in Appendix B, there was also no main effect of training type on R^2_{LIN} scores on the Number Line Estimation Task, F (2,237) = 2.14, p = .121, η_p^2 = .018.

For training mode, there was a significant main effect for Geometry Shape Items, F (2,212) = 0.55, p = .877, η_p^2 = .005. This favoured implicit instruction. No other main effects of training mode were found for: the Mental Rotation Task, F (1,236) = 0.01, p = .969, η_p^2 = .001; the Spatial Scaling task, F (1,231) = 2.28, p = .133, η_p^2 = .010; Missing Term Problems, F (1, 208) = 2.43, p = .120, η_p^2 = .012, Geometry Symmetry Items, F (2,212) = 0.15, p = .701, η_p^2 = .001, and; PAE scores on the Number Line Estimation Task, F (1,236) = 2.99, p = .085, η_p^2 = .013. There were no significant interactions between training type and training mode for any task (p's > .391; η_p^2's < .008).

References

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General*, *117*(3), 288–318. https://doi.org/10.1037/0096-3445.117.3.288

Alloway, T. P., & Alloway, R. G. (2010). Investigating the predictive roles of working memory and IQ in academic attainment. *Journal of Experimental Child Psychology*, *106*(1), 20–29. https://doi.org/10.1016/j.jecp.2009.11.003

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. https://doi.org/10.1037/h0078874

Amador-Campos, J.A., & Kirchner-Nebot, T. (1997). Relations of scores on Children's Embedded Figures Test with age, item difficulty and internal consistency. *Perceptual and Motor Skills*, *85*, 675–682. https://doi.org/10.2466/pms.1997.85.2.675

Ansari, D. (2008). Effects of development and enculturation on number representation in the brain. *Nature Reviews. Neuroscience*, *9*(4), 278–291. https://doi.org/10.1038/nrn2334

Ansari, D., Garcia, N., Lucas, E., Hamon, K., & Dhital, B. (2005). Neural correlates of symbolic number processing in children and adults. *Neuroreport*, *16*(16), 1769–1773. https://doi.org/0.1097/01.wnr.0000183905.23396.f1

Antell, S. E., & Keating, D. P. (1983). Perception of numerical invariance in neonates. *Child Development*, *54*(3), 695–701. https://doi.org/10.2307/1130057

Ashcraft, M. H., & Moore, A. M. (2012). Cognitive processes of numerical estimation in children. *Journal of Experimental Child Psychology*, *111*(2), 246–267. https://doi.org/10.1016/j.jecp.2011.08.005

Baddeley, A.D., & Hitch, G. (1974). Working memory. In *G.H. Bower (Ed.), The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York: Academic Press.

Baenninger, M., & Newcombe, N. S. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, *20*(5–6), 327–344. https://doi.org/10.1007/BF00287729

Bailey, D. (2017). Causal Inference and the Spatial-Math Link in Early Childhood. *Monographs of the Society for Research in Child Development*, *82*(1), 7–30. https://doi.org/10.1111/mono.12280

Baillargeon, R., Spelke, E., & Wasserman, S. (1985). Object permanence in five-month old infants. *Cognition*, *20*, 191–218. https://doi.org/10.1016/0010-0277(85)90008-3

Banerjee, P. A. (2016). A systematic review of factors linked to poor academic performance of disadvantaged students in science and maths in schools. *Cogent Education*, *3*(1), 1178441. https://doi.org/10.1080/2331186X.2016.1178441

Barsalou, L. W. (2008). Grounded Cognition. *Annual Review of Psychology*, *59*(1), 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Barth, H.C., Kanwisher, N., & Spelke, E. S. (2003). The construction of large number representations in adults. *Cognition*, *86*(3), 201–221. https://doi.org/10.1016/S0010-0277(02)00178-6

Barth, H.C., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences*, *102*(39), 14116–14121. https://doi.org/10.1073/pnas.0505512102

Barth, H.C., & Paladino, A.M. (2011). The development of numerical estimation: evidence against a representational shift: Development of numerical estimation. *Developmental Science*, *14*(1), 125–135. https://doi.org/10.1111/j.1467-7687.2010.00962.x

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences*, *107*(5), 1860–1863. https://doi.org/10.1073/pnas.0910967107

Berliner, D. C. (1979). Tempus educare. In *P. L. Peterson & H. L. Walberg (Eds.), Conceptions of teaching.* Berkeley, CA: McCutchan.

Berressem, J. (2011). STEM - Getting to the root of the STEM problem. *Times Education Supplement*.

Blüchel, M., Lehmann, J., Kellner, J., & Jansen, P. (2013). The improvement in mental rotation performance in primary school-aged children after a two-week motor-training. *Educational Psychology*, *33*(1), 75–86. https://doi.org/10.1080/01443410.2012.707612

Boot, W. R., Simons, D. J., Stothart, C., & Stutts, C. (2013). The Pervasive Problem With Placebos in Psychology: Why Active Control Groups Are Not Sufficient to Rule Out Placebo Effects. *Perspectives on Psychological Science*, *8*(4), 445–454. https://doi.org/10.1177/1745691613491271

Broadbent, H. J. (2014). *Large-scale visuospatial cognition in Williams syndrome and typically developing children (unpublished doctoral dissertation).* London: Institute of Education, University of London.

Broadbent, H. J., Farran, E. K., & Tolmie, A. (2014). Object-based mental rotation and visual perspective-taking in typical development and Williams syndrome. *Developmental Neuropsychology*, *39*(3), 205–225. https://doi.org/10.1080/87565641.2013.876027

Brooks-Gunn, J., & Markman, L. B. (2005). The contribution of parenting to ethnic and racial gaps in school readiness. *The Future of Children*, *15*, 139–168.

Bruce, C. D., & Hawes, Z. (2015a). The role of 2D and 3D mental rotation in mathematics for young children: what is it? Why does it matter? And what can we do about it? *ZDM*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Bruce, C. D., & Hawes, Z. (2015b). The role of 2D and 3D mental rotation in mathematics for young children: what is it? Why does it matter? And what can we do about it? *ZDM*, *47*(3), 331–343. https://doi.org/10.1007/s11858-014-0637-4

Buckley, J., Seery, N., & Canty, D. (2018). A Heuristic Framework of Spatial Ability: a Review and Synthesis of Spatial Factor Literature to Support its Translation into STEM Education. *Educational Psychology Review*, *30*(3), 947–972. https://doi.org/10.1007/s10648-018-9432-z

Buckley, P. B., & Gillman, C. B. (1974). Comparisons of digits and dot patterns. *Journal of Experimental Psychology*, *103*(6), 1131–1136. https://doi.org/10.1037/h0037361

Burgess, N. (2006). Spatial memory: how egocentric and allocentric combine. *Trends in Cognitive Sciences*, *10*(12), 551–557. https://doi.org/10.1016/j.tics.2006.10.005

Busch, J. C., Watson, J. A., Brinkley, V., Howard, J., & Nelson, C. (1993). Preschool Embedded Figures Test Performance of Young Children: Age and Gender Differences. *Perceptual and Motor Skills*, *77*(2), 491–496. https://doi.org/10.2466/pms.1993.77.2.491

Bushnell, E. W., McKenzie, B. E., Lawrence, D. A., & Connell, S. (1995). The Spatial Coding Strategies of One-Year-Old Infants in a Locomotor Search Task. *Child Development*, *66*(4), 937–958. https://doi.org/10.1111/j.1467-8624.1995.tb00914.x

Butterworth, B. (1999). *The Mathematical Brain*. London: Macmillan.

Butterworth, B. (2010). Foundational numerical capacities and the origins of dyscalculia. *Trends in Cognitive Sciences*, *14*(12), 534–541. https://doi.org/10.1016/j.tics.2010.09.007

Butterworth, B, & Varma, S. (2013). Mathematical Development. In *Educational Neuroscience* (pp. 201–237). London: Wiley.

Butterworth, Brian. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, *46*(1), 3–18. https://doi.org/10.1111/j.1469-7610.2004.00374.x

Byrnes, J. P., & Wasik, B. A. (2009). Factors predictive of mathematics achievement in kindergarten, first and third grades: An opportunity–propensity analysis. *Contemporary Educational Psychology*, *34*(2), 167–183. https://doi.org/10.1016/j.cedpsych.2009.01.002

Cankaya, O., & LeFevre, J.-A. (2016). The Home Numeracy Environment: What Do Cross-Cultural Comparisons Tell Us About How to Scaffold Young Children's Mathematical Skills? In B. Blevins-Knabe & A. M. B. Austin (Eds.), *Early Childhood Mathematics Skill Development in the Home Environment* (pp. 87–

104). Springer International Publishing. https://doi.org/10.1007/978-3-319-43974-7_6

Carey, S. (1991). Knowledge acqusition: enrichment or conceptual change? In *S.Carey and R. Gelman, eds., The epigenesis of Mind: Essays on Biology and Cognition* (pp. 257–291). Hillsdale NJ: Lawrence Erlbaum.

Carey, S. (2004). Bootstrapping & the origin of concepts. *Daedalus*, *133*(1), 59–68. https://doi.org/10.1162/001152604772746701

Carey, S. (2009). Where Our Number Concepts Come From. *The Journal of Philosophy*, *106*(4), 220–254. https://doi: 10.5840/jphil2009106418

Carr, M., Alexeev, N., Wang, L., Barned, N., Horan, E., & Reed, A. (2017). The Development of Spatial Skills in Elementary School Students. *Child Development*. https://doi.org/10.1111/cdev.12753

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences*, *18*(1), 61–75. https://doi.org/10.1016/j.lindif.2007.04.005

Carroll, J. B. (1993). Human Cognitive Abilities: A Survey of Factor-Analytic Studies. New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511571312

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The Development of Spatial Skills Through Interventions Involving Block Building Activities. *Cognition and Instruction*, *26*(3), 269–309. https://doi.org/10.1080/07370000802177177

Casey, B. M., Pezaris, E., Fineman, B., Pollock, A., Demers, L., & Dearing, E. (2015). A longitudinal analysis of early spatial skills compared to arithmetic and verbal skills as predictors of fifth-grade girls' math reasoning. *Learning and Individual Differences*, *40*, 90–100. https://doi.org/10.1016/j.lindif.2015.03.028

Casey, B., Pezaris, E., & Nuttall, R. (1992). Spatial ability as a predictor of math achievement: The importance of sex and handedness patterns. *Neuropsychologia*, *30*(1), 35–45. https://doi.org/10.1016/0028-3932(92)90012-B

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction*, *34*, 11–21. https://doi.org/10.1016/j.learninstruc.2014.07.005

Centre for Economics and Business Research (CEBR). (2015). *The contribution of engineering to the UK economy- the multiplier impacts*. London.

Chatterjee, A. (2008). The Neural Organization of Spatial Thought and Language. *Seminars in Speech and Language*, *29*(03), 226–238. https://doi.org/10.1055/s-0028-1082886

Chen, Q., & Li, J. (2014). Association between individual differences in non-symbolic number acuity and math performance: a meta-analysis. *Acta Psychologica*, *148*, 163–172. https://doi.org/10.1016/j.actpsy.2014.01.016

Cheng, Y.-L., & Mix, K. S. (2014). Spatial Training Improves Children's Mathematics Ability. *Journal of Cognition and Development*, *15*(1), 2–11. https://doi.org/10.1080/15248372.2012.725186

Christoforou, C., Hatzipanayioti, A., & Avraamides, M. (2018). Perspective Taking vs Mental Rotation: CSP-Based Single-Trial Analysis for Cognitive Process Disambiguation. In Wang, S., Yamamoto, V., Jianzhong S., Yang Y., Jones, E., Iasemidis, L., Mitchell, T., (Eds.) *Proceedings of International Conference, Brain Informatics* (pp. 109-199). Arlington, TX, USA.

Clearfield, M., & Mix, K. S. (2001). Amount Versus Number: Infants' Use of Area and Contour Length to Discriminate Small Sets. *Journal of Cognition and Development*, *2*(3), 243–260. https://doi.org/10.1207/S15327647JCD0203_1

Clements, D., & Sarama, J. (2004). Building Blocks for early childhood mathematics. *Early Childhood Research Quarterly*, *19*, 181–89. https://doi.org/10.2307/30034954

Clerkin, A., & Gilligan, K. A. (2018). Pre-school numeracy play as a predictor of children's attitudes towards mathematics at age 10. *Journal of Early Childhood Research*. https://doi.org/10.1177/1476718X18762238

Cockcroft, W. (1982). *Mathematics Counts*. London: HMSO.

Cohen, C. A., & Hegarty, M. (2014). Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and*

Individual Differences, *33*, 63–71. https://doi.org/10.1016/j.lindif.2014.04.002

Cohen, D. (1988). *Statistical power analysis for the behavioural sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Earlbaum Associates.

Confederation of British Industry (CBI). (2013). *Changing the pace: CBI/Pearson education and skills survey 2013*. London: CBI.

Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*(4), 698–707. https://doi.org/10.3758/BF03196206

Cowan, R. (2015). *Number skills in primary school: Development and individual differences*. London: Institute of Education Press.

Cowan, R., & Donlan, C. (2010). *The Development and Importance of Proficiency in Basic Calculation. Full Research Report ESRC End of award report, RES-062-23-0667*. Swindon: ESRC.

Cowan, R., & Saxton, M. (2010). Understanding diversity in number development. *BJEP Monograph Series II, Number 7 - Understanding Number Development and Difficulties*, *1*, 1–14. https://doi.org/10.1348/97818543370009X12583699332339

Cragg, L., Keeble, S., Richardson, S., Roome, H. E., & Gilmore, C. (2017). Direct and indirect influences of executive functions on mathematics achievement. *Cognition*, *162*, 12–26. https://doi.org/10.1016/j.cognition.2017.01.014

Cutini, S., Scarpa, F., Scatturin, P., Dell’Acqua, R., & Zorzi, M. (2014). Number-Space Interactions in the Human Parietal Cortex: Enlightening the SNARC Effect with Functional Near-Infrared Spectroscopy. *Cerebral Cortex*, *24*(2), 444–451. https://doi.org/10.1093/cercor/bhs321

Davis, B., Drefs, M., & Francis, K. (2015). A history and analysis of current curriculum. In *Spatial Reasoning in the Early Years; Principles, Assertions and Speculations* (pp. 47–63). New York: Routledge.

de Hevia, M. D., Vallar, G., & Girelli, L. (2006). Visuo-spatial components of numerical representation. In *Imagery and Spatial Cognition: Methods, models and cognitive assessment*. John Benjamins Publishing.

De Smedt, B., Noël, M.P., Gilmore, C., & Ansari, D. (2013). How do symbolic and non-symbolic numerical magnitude processing skills relate to individual differences in children's mathematical skills? A review of evidence from brain and behavior. *Trends in Neuroscience and Education*, *2*(2), 48–55. https://doi.org/10.1016/j.tine.2013.06.001

Dean, A.L., & Harvey, W.O. (1979). An information-processing analysis of a Piagetian imagery task. *Developmental Psychology*, *15*, 474–475. https://doi.org/10.1037/0012-1649.15.4.474

Dean, A.L., & Sherzer, E. (1982). A comparison of reaction time and drawing measures of mental rotation. *Journal of Experimental Child Psychology*, *34*, 20–37. https://doi.org/10.1016/0022-0965(82)90028-5

Dehaene, S. (1997). *The Number Sense: How the mind Creates Mathematics*. New York: Oxford University Press.

Dehaene, S. (2011). *The Number Sense: How the Mind Creates Mathematics, Revised and Updated Edition* (Revised, Updated edition). New York: Oxford University Press.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371–396. https://doi.org/10.1037/0096-3445.122.3.371

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two - digit number comparison. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 626–641.https:// doi.org/10.1037/0096-1523.16.3.626

Dehaene, S., Piazza, M., Pinel, P., & Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, *20*(3–6), 487–506. https://doi.org/10.1080/02643290244000239

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25–32. https://doi.org/10.1016/S0160-2896(03)00061-8

DeLoache, J. S. (1987). Rapid change in the symbolic functioning of very young children. *Science*, *238*(4833), 1556–1557. https://doi.org/10.1126/science.2446392

DeLoache, J. S. (1989). Young children's understanding of the correspondence between a scale model and a larger space. *Cognitive Development*, *4*(2), 121–139. https://doi.org/10.1016/0885-2014(89)90012-9

Denham, C., & Lieberman, A. (1980). Time to Learn. A Review of the Beginning Teacher Evaluation Study. Retrieved from https://eric.ed.gov/?id=ED192454

Desforges, C., & Cockburn, A. (1987). *Understanding the Mathematics Teacher: A study of practice in first schools*. London: Falmer Press.

DeWind, N. K., and 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

Diamond, A. (1991). Neuropsychological insights into the meaning of object concept development. In *S. Carey and R. Gelman, eds., The Jean Piaget Symposium Series: The Epigenesis of Mind: Essays on Biology and Cognition* (pp. 67–110). Hillsdale NJ: Lawrence Erlbaum.

Dunn, L. M., Dunn, D. M., Styles, B., & Sewell, J. (2009). *The British Picture Vocabulary Scale - 3rd Edition.* London: GL Assessment.

Dweck, C. S. (2000). *Self-theories: Their Role in Motivation, Personality, and Development*. Psychology Press.

Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Elliott, C., Smith, P., & Mc Cullock, K. (1996). *British Ability Scales Second Edition (BAS II). Administration and Scoring Manual*. London: Nelson.

Elliott, C., Smith, P., & McCulloch, K. (1997). *British Ability Scales Second Edition (BAS II). Technical Manual.* London: Nelson.

Engineering UK. (2018). *Engineering UK 2018: The State of Engineering Synopsis.* London. UK.

Eysenck, H. J. (1939). Primary Mental Abilities. *British Journal of Educational Psychology*, *9*(3), 270–275. https://doi.org/10.1111/j.2044-8279.1939.tb03214.x

Farmer, G., Verdine, B. N., Lucca, K., Davies, T., Dempsey, R., Hirsh-Pasek, K., & Golinkoff, R. M. (2013). Putting the pieces together: Spatial skills at age 3 predict to spatial and math performance at age 5. Presented at the Society for Research in Child Development. , Seattle, WA.

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. *Journal of Experimental Child Psychology*, *123*, 53–72. https://doi.org/10.1016/j.jecp.2014.01.013

Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*(7), 307–314. https://doi.org/10.1016/j.tics.2004.05.002

Feigenson, L., Libertus, M. E., & Halberda, J. (2013). Links Between the Intuitive Sense of Number and Formal Mathematics Ability. *Child Development Perspectives*, *7*(2), 74–79. https://doi.org/10.1111/cdep.12019

Fias, W., & Bonato, M. (2018). Chapter 12 - Which Space for Numbers? In A. Henik & W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (pp. 233–242). Academic Press. https://doi.org/10.1016/B978-0-12-811529-9.00012-1

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE.

Finniss, D. G., Kaptchuk, T. J., Miller, F., & Benedetti, F. (2010). Placebo Effects: Biological, Clinical and Ethical Advances. *Lancet*, *375*(9715), 686–695. https://doi.org/10.1016/S0140-6736(09)61706-2

Fischer, M. (2003). Spatial representations in number processing--evidence from a pointing task. *Visual Cognition*, *10*(4), 493–508. https://doi.org/10.1080/13506280244000186

Fischer, M. H., & Shaki, S. (2014). Spatial associations in numerical cognition--from single digits to arithmetic. *Quarterly Journal of Experimental Psychology (2006)*, *67*(8), 1461–1483. https://doi.org/10.1080/17470218.2014.927515

Fisher, K. J., Borchert, K., & Bassok, M. (2011). Following the standard form: effects of equation format on algebraic modeling. *Memory & Cognition*, *39*(3), 502–515. https://doi.org/10.3758/s13421-010-0031-6

Fisher, S.E. (2006). Tangled webs: tracing the connections between genes and cognition. *Cognition*, *101*(2), 270–297. https://doi.org/10.1016/j.cognition.2006.04.004

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children’s knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99–103. https://doi.org/10.1037/0012-1649.17.1.99

Flombaum, J. I., Junge, J. A., & Hauser, M. D. (2005). Rhesus monkeys (Macaca mulatta) spontaneously compute addition operations over large numbers. *Cognition*, *97*(3), 315–325. https://doi.org/10.1016/j.cognition.2004.09.004

Foroughi, C. K., Monfort, S. S., Paczynski, M., McKnight, P. E., & Greenwood, P. M. (2016). Placebo effects in cognitive training. *Proceedings of the National Academy of Sciences*, *113*(27), 7470–7474. https://doi.org/10.1073/pnas.1601243113

Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age. *Cognitive Processing*, *14*(2), 117–127. https://doi.org/10.1007/s10339-012-0534-0

Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development*, *28*(4), 386–399. https://doi.org/10.1016/j.cogdev.2013.06.002

Frick, A., Möhring, W., & Newcombe, N. S. (2014a). Development of mental transformation abilities. *Trends in Cognitive Sciences*, *18*(10), 536–542. https://doi.org/10.1016/j.tics.2014.05.011

Frick, A., Möhring, W., & Newcombe, N. S. (2014b). Picturing perspectives: development of perspective-taking abilities in 4- to 8-year-olds. *Frontiers in Psychology*, *5*. https://doi.org/10.3389/fpsyg.2014.00386

Frick, A., & Newcombe, N. S. (2012). Getting the big picture: Development of spatial scaling abilities. *Cognitive Development*, *27*(3), 270–282. https://doi.org/10.1016/j.cogdev.2012.05.004

Frick, A., & Wang, S. (2014). Mental Spatial Transformations in 14- and 16-Month-Old Infants: Effects of Action and Observational Experience. *Child Development*, *85*(1), 278–293. https://doi.org/10.1111/cdev.12116

Friso-van den Bos, I., Kroesbergen, E. H., Van Luit, J. E. H., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. D. M. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of Experimental Child Psychology*, *134*, 12–29. https://doi.org/10.1016/j.jecp.2015.02.002

Friso-van den Bos, I., van der Ven, S. H. G., Kroesbergen, E. H., & van Luit, J. E. H. (2013). Working memory and mathematics in primary school children: A meta-analysis. *Educational Research Review*, *10*, 29–44. https://doi.org/10.1016/j.edurev.2013.05.003

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., … Schatschneider, C. (2010). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, *46*(6), 1731–1746. https://doi.org/10.1037/a0020662

Gallagher-Mitchell, T., Romero-Rivas, C., Rodriguez-Cuadrado, S., & Dackermann, T. (2017). *Age, cross-linguistic, and modality effects on children's number line estimation*. Presented at the European Society for cognitive Psychology Conference, Potsdam, Germany.

Gallagher-Mitchell, T., Simms, V., & Litchfield, D. (2018). Learning from where 'eye' remotely look or point: Impact on number line estimation error in adults. *Quarterly Journal of Experimental Psychology*, *71*(7), 1526–1534. https://doi.org/10.1080/17470218.2017.1335335

Gauvain, M. (1993). The development of spatial thinking in everyday activity. *Developmental Review*, *13*, 92–121. http://dx.doi.org/10.1006/drev.1993.1004

Gauvain, M. (1995). Thinking in niches: Sociocultural influences on cognitive development. *Human Development*, *38*, 25–45. http://dx.doi.org/10.1159/000278297

Geary, D. C., Hoard, M. K., Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, *78*(4), 1343–1359. https://doi.org/10.1111/j.1467-8624.2007.01069.x

Gelman, R., & Gaillistel, C.R. (1978). *The Child's Understanding of Number*. Cambridge: Harvard University Press.

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2018). The use of discrimination scaling tasks: A novel perspective on the development of spatial scaling in children. *Cognitive Development*, *47*, 133–145. https://doi.org/10.1016/j.cogdev.2018.04.001

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), e67374. https://doi.org/10.1371/journal.pone.0067374

Gilmore, C., Attridge, N., De Smedt, B., & Inglis, M. (2014). Measuring the approximate number system in children: Exploring the relationships among different tasks. *Learning and Individual Differences*, *29*, 50–58. https://doi.org/10.1016/j.lindif.2013.10.004

Gilmore, C. K., Attridge, N. & Inglis, M. (2011) Measuring the approximate number system. *Quarterly Journal of Experimental Psychology*, 64, pp. 2099-2109. https://doi.org/ 10.1080/17470218.2011.574710

Gilmore, C. K., McCarthy, S. E., & Spelke, E. S. (2007). Symbolic arithmetic knowledge without instruction. *Nature*, *447*(7144), 589–591. https://doi.org/10.1038/nature05850

Girelli, L., Lucangeli, D., & Butterworth, B. (2000). The Development of Automaticity in Accessing Number Magnitude. *Journal of Experimental Child Psychology*, *76*(2), 104–122. https://doi.org/10.1006/jecp.2000.2564

Goldhaber, D., Liddle, S., Theobald, R., & Walch, J. (2010). *Teacher Effectiveness and the Achievement of Washington's Students in Mathematics.* CEDR Working Paper 2010-6. Seattle, WA: University of Washington.

Greany, T., Barnes, I., Mostafa, T., Pensiero, N., & Swensson, C. (2016). *Trends in Maths and Science Study (TIMSS): National Report for England*. London: UCL Institute of Education.

Green, C.S., Bavelier, D., Kramer, A.F., Vinogradov, S., Ansorge, U., Ball, K.K., … Witt, C.M. (in press). Improving Methodological Standards in Behavioral Interventions for Cognitive Enhancement.

Guay, R. B. (1977). *Purdue spatial visualization test: Rotations*. Purdue Research Foundation, West Lafayette, IN.

Guilford, J.P., & Lacey, J.I. (1947). Printed Classification Tests, A.A.F. Aviation Psychological Progress Research Report, No.5, Washington, DC: U.S. Government Printing Office.

Guilford, J. P., & Zimmerman, W. S. (1948). *The Guilford–Zimmerman Aptitude Survey*. *Journal of Applied Psychology*, 32, 24-34.

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

Guisande, M., Fernanda Páramo, M., Tinajero Vacas, C., & Almeida, L. (2007). Field dependence-independence (FDI) cognitive style: An analysis of attentional functioning. *Psicothema*, *19*, 572–577.

Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology*, *48*(5), 1229–1241. https://doi.org/10.1037/a0027433

Gutman, L., & Feinstein, L. (2008). *Children's Well-Being in Primary School: Pupil and school effects*. London: Centre for Research on the Wider Benefits of Learning.

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*, *109*(28), 11116–11120. https://doi.org/10.1073/pnas.1200196109

Halberda, J., Mazzocco, M. 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

Hall, C. W., Davis, N. B., Bolen, L. M., & Chia, R. (1999) Gender and Racial Differences in Mathematical Performance. *The Journal of Social Psychology*, 139:6, 677-689, https://doi.org/ 10.1080/00224549909598248

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51. https://doi.org/10.1111/j.1529-1006.2007.00032.x

Hansen, K., Jones, E., & Budge, D. (2010). *Millennium Cohort Study Fourth Survey: A user's guide to Initial Findings (2nd ed.)*. Centre for Longitudinal Studies, Institute of Education, University of London, London.

Hansen, K., & Joshi, H. (2008). *Millennium Cohort Study Third Survey: A user's guide to Initial Findings.* Centre for Longitudinal Studies, Institute of Education, London.

Harris, L. J., Hanley, C., & Best, C. T. (1978). Conservation of horizontality: Sex differences in sixth graders and college students. In M. S. Smart & R. C. Smart (Eds.), *Adolescence: Development and relationships* (pp. 95-107). New York: Macmillan.

Harris, J., Newcombe, N. S., & Hirsh-Pasek, K. (2013). A New Twist on Studying the Development of Dynamic Spatial Transformations: Mental Paper Folding in Young Children. *Mind, Brain, and Education*, *7*(1), 49–55.

Hart, R. A., & Moore, G. T. (1973). The development of spatial cognition: a review. In *Image and Environment: Cognitive Mapping and Spatial Behavior*. Chicago: Aldine Publishing Company.

Hauser, M., & Carey, S. (1998). Building a cognitive creature from a set of primitives: Evolutionary and developmental insights. In *The Evolution of Mind* (pp. 51–106). Oxford: Oxford University Press.

Hawes, Z., Moss, J., Caswell, B., Naqvi, S., & MacKinnon, S. (2017). Enhancing Children's Spatial and Numerical Skills through a Dynamic Spatial Approach to Early Geometry Instruction: Effects of a 32-Week Intervention. *Cognition and Instruction*, *35*(3), 236–264. https://doi.org/10.1080/07370008.2017.1323902

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education*, *4*(3), 60–68. https://doi.org/10.1016/j.tine.2015.05.001

Hawes, Z., Tepylo, D., & Moss, J. (2015). Developing spatial thinking. In *Spatial Reasoning in the Early Years: Principles, Assertions and Speculations* (pp. 29–45). New York and London: Routledge.

Haywood, K., Teeple, J., Givens, M., & Patterson, J. (1977). Young children's Rod-and-Frame Test performance. *Perceptual and Motor Skills*, *45*(1), 163–169. https://doi.org/10.2466/pms.1977.45.1.163

Heckman, J. J., & Masterov, D. V. (2007). *The Productivity Argument for Investing in Young Children* (Working Paper No. 13016). National Bureau of Economic Research. https://doi.org/10.3386/w13016

Hegarty, M. (2014). Spatial Thinking in Undergraduate Science Education. *Spatial Cognition & Computation*, *14*(2), 142–167. https://doi.org/10.1080/13875868.2014.889696

Hegarty, M., Keehner, M., Cohen, C., Montello, D. R., & Lippa, Y. (2007). The Role of Spatial Cognition in Medicine: Applications for Selecting and Training Professionals. In *Applied spatial cognition: From research to cognitive technology* (pp. 285–315). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and*

Individual Differences, *19*(1), 61–70. https://doi.org/10.1016/j.lindif.2008.04.006

Hegarty, M., & Kozhevnikov, M. (1999). Types of Visual-Spatial Representations and Mathematical Problem Solving. *Journal of Educational Psychology*, *91*, 684–689. https://doi.org/10.1037//0022-0663.91.4.684

Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., & Lovelace, K. (2006). Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, *34*(2), 151–176. https://doi.org/10.1016/j.intell.2005.09.005

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Hill, V. (2005). Through the past darkly: A review of the British Ability Scales Second Edition. *Child and Adolescent Mental Health*, *10*, 87–98. https://doi.org/10.1111/j.1475-3588.2004.00123.x

Hodgkiss, A., Gilligan, K. A., Tolmie, A., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology*. https://doi.org/10.1111/bjep.12211

Höffler, T. (2010). Spatial Ability: Its Influence on Learning with Visualizations—a Meta-Analytic Review. *Educational Psychology Review*, *22*, 245–269. https://doi.org/10.1007/s10648-010-9126-7

Holloway, I. D., & Ansari, D. (2008). Domain-specific and domain-general changes in children’s development of number comparison. *Developmental Science*, *11*(5), 644–649. https://doi.org/10.1111/j.1467-7687.2008.00712.x

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4), F9–F15. https://doi.org/10.1111/j.1467-7687.2009.00848.x

Honoré, N., & Noël, M.-P. (2016). Improving Preschoolers’ Arithmetic through Number Magnitude Training: The Impact of Non-Symbolic and Symbolic Training. *PLOS ONE*, *11*(11), e0166685. https://doi.org/10.1371/journal.pone.0166685

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435–448. https://doi.org/10.1038/nrn1684

Hutchins, E. (1995). *Cognition in the Wild*. Cambridge: MIT Press.

Huttenlocher, J., Newcombe, N.S., & Vasilyeva, M. (1999). Spatial Scaling in Young Children. *Psychological Science*, *10*(5), 393–398. https://doi.org/10.1111/1467-9280.00175

Huttenlocher, J., Jordan, N. C., & Levine, S. C. (1994). A mental model for early arithmetic. *Journal of Experimental Psychology: General*, *123*(3), 284–296. https://doi.org/10.1037/0096-3445.123.3.284

Inglis, M., and Gilmore, C. K. (2014). Indexing the approximate number system. *Acta Psychologica.* 145, 147–155. doi: 10.1016/j.actpsy.2013.11.009

Ireson, J., Hallam, S., & Hurley, C. (2005). What are the effects of ability grouping on GCSE attainment? *British Educational Research Journal*, *31*(4), 443–458. https://doi.org/10.1080/01411920500148663

Izard, V., Sann, C., Spelke, E. S., & Streri, A. (2009). Newborn infants perceive abstract numbers. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(25), 10382–10385. https://doi.org/10.1073/pnas.0812142106

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*. https://doi.org/10.1073/pnas.0801268105

Jaeggi, S. M., Buschkuehl, M., Shah, P., & Jonides, J. (2014). The role of individual differences in cognitive training and transfer. *Memory & Cognition*, *42*(3), 464–480. https://doi.org/10.3758/s13421-013-0364-z

Jerrim, J., & Shure, N. (2016). *Achievement of 15-YearOlds in England: PISA 2015 National Report*. London: UCL Institute of Education.

Jia, S., Zhang, Q., & Li, S. (2014). Field dependence–independence modulates the efficiency of filtering out irrelevant information in a visual working memory task. *Neuroscience*, *278*, 136–143. https://doi.org/10.1016/j.neuroscience.2014.07.075

Jirout, J. J., & Newcombe, N. S. (2015). Building Blocks for Developing Spatial Skills Evidence From a Large, Representative U.S. Sample. *Psychological Science*, 0956797614563338. https://doi.org/10.1177/0956797614563338

Johnson, E. S., & Meade, A. C. (1987). Developmental Patterns of Spatial Ability: An Early Sex Difference. *Child Development*, *58*(3), 725. https://doi.org/10.2307/1130210

Johnson, S. (2009a). Developmental origins of object perception. In *A. Woodward & A. Needham (Eds.), Learning and the infant mind* (pp. 47–65). New York: Oxford University Press.

Katz, B., Jaeggi, S., Buschkuehl, M., Stegman, A., & Shah, P. (2014). Differential effect of motivational features on training improvements in school-based cognitive training. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00242

Kellman, P., & Spelke, E. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology*, *15*, 483–524. http://dx.doi.org/10.1016/0010-0285(83)90017-8

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75–86. https://doi.org/10.1207/s15326985ep4102_1

Kohs, S. C. (1919). *Block Designs Test*. Wood Dale, IL: Stoelting.

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age Differences in Imagery Abilities. *Child Development*, *61*(4), 995–1010. https://doi.org/10.2307/1130871

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the Visualizer-Verbalizer Dimension: Evidence for Two Types of Visualizers. *Cognition and Instruction*, *20*(1), 47–77. https://doi.org/10.1207/S1532690XCI2001_3

Kozhevnikov, M., Kosslyn, S., & Shephard, J. (2005). Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, *33*(4), 710–726. https://doi.org/10.3758/BF03195337

Kozhevnikov, M., & Thornton, R. (2006). Real-Time Data Display, Spatial Visualization Ability, and Learning Force and Motion Concepts. *Journal of Science Education and Technology*, *15*, 111–132. https://doi.org/10.1007/s10956-006-0361-0

Kuipers, B. (1978). Modeling Spatial Knowledge. *Cognitive Science*, *2*(2), 129–153. https://doi.org/10.1207/s15516709cog0202_3

Kuipers, B. (1982). The "Map in the Head" Metaphor. *Environment and Behavior*, *14*(2), 202–220. https://doi.org/10.1177/0013916584142005

Lachance, J. A., & Mazzocco, M. M. 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. https://doi.org/10.1016/j.lindif.2005.12.001

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodies brings mathematics into being.* New York: NY: Basic Books.

Landy, D., & Goldstone, R. L. (2007). How abstract is symbolic thought? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *33*(4), 720–733. https://doi.org/10.1037/0278-7393.33.4.720

Landy, D., & Goldstone, R. L. (2010). Proximity and precedence in arithmetic. *The Quarterly Journal of Experimental Psychology*, *63*(10), 1953–1968. https://doi.org/10.1080/17470211003787619

Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences*, *23*, 123–130. https://doi.org/10.1016/j.lindif.2012.08.001

Laski, E. V., & Siegler, R. S. (2007). Is 27 a Big Number? Correlational and Causal Connections Among Numerical Categorization, Number Line Estimation, and Numerical Magnitude Comparison. *Child Development*, *78*(6), 1723–1743. https://doi.org/10.1111/j.1467-8624.2007.01087.x

Le Corre, M., & Carey, S. (2007). One, two, three, four, nothing more: an investigation of the conceptual sources of the verbal counting principles. *Cognition*, *105*(2), 395–438. https://doi.org/10.1016/j.cognition.2006.10.005

LeFevre, J.-A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: longitudinal predictors of performance. *Child Development*, *81*(6), 1753–1767. https://doi.org/10.1111/j.1467-8624.2010.01508.x

Lenkeit, J., Chan, J., Hopfenbeck, T. N., & Baird, J.-A. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, *16*, 102–115. https://doi.org/10.1016/j.edurev.2015.10.002

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*(4), 940–949.

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology*, *53*(1), 72–103. https://doi.org/10.1016/S0022-0965(05)80005-0

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic Status Modifies the Sex Difference in Spatial Skill. *Psychological Science*, *16*(11), 841–845. https://doi.org/10.1111/j.1467-9280.2005.01623.x

Liben (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill-Palmer Quarterly,* 24, 97-100.

Liben, L. S., & Downs, R. M. (1993). Understanding person space map relations: Cartographic and developmental perspectives. *Developmental Psychology*, *29*(4), 739–752. https://doi.org/10.1037/0012-1649.29.4.739

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, *13*(6), 900–906. https://doi.org/10.1111/j.1467-7687.2009.00948.x

Likert, R. (1970). *The Revised Minnesota Paper Form Board Test Manual*. The Psychological Corporation, New York.

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New Trends in Gender and Mathematics Performance: A Meta-Analysis. *Psychological Bulletin*, *136*(6), 1123–1135. https://doi.org/10.1037/a0021276

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child Development*, *56*(6), 1479–1498. https://doi.org/10.2307/1130467

Lipton, J. S., & Spelke, E. S. (2003). Origins of Number Sense: Large-Number Discrimination in Human Infants. *Psychological Science*, *14*(5), 396–401. https://doi.org/10.1111/1467-9280.01453

Lizarraga, M. L., & Ganuza, J. M. G. (2003). Improvement of Mental Rotation in Girls and Boys. *Sex Roles*, *49*(5), 277–286. https://doi.org/10.1023/A:1024656408294

Lohman, D.F. (1988). Spatial abilities as traits, processes, and knowledge. In *R. J. Sternberg (Ed.), Advances in the psychology of human intelligence* (pp. 181–248). Hillsdale, NJ: Erlbaum.

Lourenco, S. F., Cheung, C.-N., & Aulet, L. S. (2018). Chapter 10 - Is Visuospatial Reasoning Related to Early Mathematical Development? A Critical Review. In A. Henik & W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (pp. 177–210). Academic Press. https://doi.org/10.1016/B978-0-12-811529-9.00010-8

Lowrie, T., Logan, T., & Ramful, A. (2017). Visuospatial training improves elementary students' mathematics performance. *British Journal of Educational Psychology*, *87*(2), 170–186. https://doi.org/10.1111/bjep.12142

Lyons, I. M., Ansari, D., & Beilock, S. L. (2012). Symbolic Estrangement: Evidence against a Strong Association between Numerical Symbols and the Quantities They Represent. *Journal of Experimental Psychology: General*, *141*(4), 635–641. https://doi.org/10.1037/a0027248

Lyons, I. M., Price, G. R., Vaessen, A., Blomert, L., & Ansari, D. (2014). Numerical predictors of arithmetic success in grades 1-6. *Developmental Science*, *17*(5), 714–726. https://doi.org/10.1111/desc.12152

Manger, T., & Eikeland, O.-J. (1998). The effects of spatial visualization and students' sex on mathematical achievement. *British Journal of Psychology*, *89*(1), 17–25. https://doi.org/10.1111/j.2044-8295.1998.tb02670.x

Markey, S.M. (2010). *The relationship between visual-spatial reasoning ability and math and geometry problem solving.* Retrieved from Dissertation Abstracts International: Section B. The Sciences and Engineering.

Marmor, G. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Psychology*, *7*(4), 548–559. https://doi.org/10.1016/0010-0285(75)90022-5

Marmor, G. (1977). Mental rotation and number conservation: are they related? *Developmental Psychology*, *13*, 320–325.

Masangkay, Z. S., McCluskey, K. A., McIntyre, C. W., Sims-Knight, J., Vaughn, B. E., & Flavell, J. H. (1974). The Early Development of Inferences about the Visual Percepts of Others. *Child Development*, *45*(2), 357–366. https://doi.org/10.2307/1127956

Matejko, A. A., & Ansari, D. (2016). Trajectories of Symbolic and Nonsymbolic Magnitude Processing in the First Year of Formal Schooling. *PLOS ONE*, *11*(3), e0149863. https://doi.org/10.1371/journal.pone.0149863

Mayes, S. D., Calhoun, S. L., Bixler, E. O., & Zimmerman, D. N. (2009). IQ and neuropsychological predictors of academic achievement. *Learning and Individual Differences*, *19*(2), 238–241. https://doi.org/10.1016/j.lindif.2008.09.001

Mazzocco, M. M. M., Feigenson, L., & Halberda, J. (2011). Preschoolers' Precision of the Approximate Number System Predicts Later School Mathematics Performance. *PLoS ONE*, *6*(9), e23749. https://doi.org/10.1371/journal.pone.0023749

Mazzocco, M. M. M., & Myers, G. F. (2003). Complexities in identifying and defining mathematics learning disability in the primary school-age years. *Annals of Dyslexia*, *53*(1), 218–253. https://doi.org/10.1007/s11881-003-0011-7

McGillicuddy-De Lisi, A. V., De Lisi, R., & Youniss, J. (1978) Representation of the horizontal coordinate with or without the liquid. *Merrill-Palmer Quarterly*, 24, 199-208.

McConney, A., & Perry, L. B. (2010). Socioeconomic status, self-efficacy, and mathematics achievement in Australia: a secondary analysis. *Educational*

Research for Policy and Practice, *9*(2), 77–91. https://doi.org/10.1007/s10671-010-9083-4

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychological Bulletin, 86(5), 889-918.http://dx.doi.org/10.1037/0033-2909.86.5.889

McNeil, N. M., & Alibali, M. W. (2004). You'll see what you mean: Students encode equations based on their knowledge of arithmetic. *Cognitive Science*, *28*(3), 451–466. https://doi.org/10.1016/j.cogsci.2003.11.002

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, *49*(2), 270–291. https://doi.org/10.1037/a0028228

Melby-Lervåg, M., Redick, T. S., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer": Evidence From a Meta-Analytic Review. *Perspectives on Psychological Science*, *11*(4), 512–534. https://doi.org/10.1177/1745691616635612

Menard, S. W. (1995). *Applied logistic regression analysis*. Sage Publications.

Merrell, C., & Tymms, P. B. (2001). Inattention, hyperactivity and impulsiveness: their impact on academic achievement and progress. *The British Journal of Educational Psychology*, *71*(Pt 1), 43–56. https://doi.org/10.1348/000709901158389

Michael, W. B., Guilford, J. P., Fruchter, B., & Zimmerman, W. S. (1957). The description of spatial-visualization abilities. Education and Psychological Measurement, 17, 185–199. https://doi.org/10.1177/001316445701700202

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences*, *26*, 141–152. https://doi.org/10.1016/j.lindif.2012.03.012

Miller, K. F., & Stigler, J. W. (1987). Counting in Chinese: Cultural variation in a basic cognitive skill. *Cognitive Development*, *2*(3), 279–305. https://doi.org/10.1016/S0885-2014(87)90091-8

Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: developmental and educational implications. *Advances in Child Development and Behavior*, *42*, 197–243.https://doi.org/10.1016/B978-0-12-394388-0.00006-X

Mix, K. S., Hambrick, D. Z., Satyam, V. R., Burgoyne, A. P., & Levine, S. C. (2018). The latent structure of spatial skill: A test of the 2 × 2 typology. *Cognition*, *180*, 268–278. https://doi.org/10.1016/j.cognition.2018.07.012

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C. J., Hambrick, D. Z., & Konstantopoulos, S. (2017). The Latent Structure of Spatial Skills and Mathematics: A Replication of the Two-Factor Model. *Journal of Cognition and Development*, *18*(4), 465–492. https://doi.org/10.1080/15248372.2017.1346658

Mix, K. S., Levine, S. C., Cheng, Y.-L., Young, C. J., Hambrick, D. Z., Ping, R., & Konstantopoulos, S. (2016). Separate but correlated: The latent structure of space and mathematics across development. *Journal of Experimental Psychology. General*, *145*(9), 1206–1227. https://doi.org/10.1037/xge0000182

Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: another look. *Developmental Psychology*, *33*(3), 423–428.

Mix, K. S., Prather, R. W., Smith, L. B., & Stockton, J. D. (2014). Young Children’s Interpretation of Multidigit Number Names: From Emerging Competence to Mastery. *Child Development*, *85*(3), 1306–1319. https://doi.org/10.1111/cdev.12197

Möhring, W., Newcombe, N. S., & Frick, A. (2014). Zooming in on spatial scaling: Preschool children and adults use mental transformations to scale spaces. *Developmental Psychology*, *50*(5), 1614–1619. https://doi.org/10.1037/a0035905

Möhring, W., Newcombe, N. S., & Frick, A. (2015). The relation between spatial thinking and proportional reasoning in preschoolers. *Journal of Experimental Child Psychology*, *132*, 213–220. https://doi.org/10.1016/j.jecp.2015.01.005

Möhring, W., Newcombe, N. S., & Frick, A. (2016). Using Mental Transformation Strategies for Spatial Scaling: Evidence From a Discrimination Task. *Journal* *of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.1037/xlm0000240

Möhring, W., Newcombe, N. S., Levine, S. C., & Frick, A. (2016). Spatial Proportional Reasoning Is Associated With Formal Knowledge About Fractions. *Journal of Cognition and Development*, *17*(1), 67–84. https://doi.org/10.1080/15248372.2014.996289

Moll, H., & Tomasello, M. (2007). How 14- and 18-month-olds know what others have experienced. *Developmental Psychology*, *43*(2), 309–317. https://doi.org/10.1037/0012-1649.43.2.309

Moll, K., Snowling, M. J., Göbel, S. M., & Hulme, C. (2015). Early language and executive skills predict variations in number and arithmetic skills in children at family-risk of dyslexia and typically developing controls. *Learning and Instruction*, *38*, 53–62. https://doi.org/10.1016/j.learninstruc.2015.03.004

Moore, A. M., & Ashcraft, M. H. (2015). Children's mathematical performance: Five cognitive tasks across five grades. *Journal of Experimental Child Psychology*, *135*, 1–24. https://doi.org/10.1016/j.jecp.2015.02.003

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*(5109), 1519–1520.

Muldoon, K., Towse, J., Simms, V., Perra, O., & Menzies, V. (2013). A longitudinal analysis of estimation, counting skills, and mathematical ability across the first school year. *Developmental Psychology*, *49*(2), 250–257. https://doi.org/10.1037/a0028240

Mundy, E., & Gilmore, C. K. (2009). Children's mapping between symbolic and nonsymbolic representations of number. *Journal of Experimental Child Psychology*, *103*(4), 490–502. https://doi.org/10.1016/j.jecp.2009.02.003

Mussolin, C., Nys, J., Content, A., & Leybaert, J. (2014). Symbolic Number Abilities Predict Later Approximate Number System Acuity in Preschool Children. *PLOS ONE*, *9*(3), e91839. https://doi.org/10.1371/journal.pone.0091839

Myers, R. H. (1990). *Classical and Modern Regression with Applications*. Duxbury/Thompson Learning.

Nardini, M., Jones, P., Bedford, R., & Braddick, O. (2008). Development of cue integration in human navigation. *Current Biology: CB*, *18*(9), 689–693. https://doi.org/10.1016/j.cub.2008.04.021

National Audit Office UK. (2018). *Delivering STEM (science, technology, engineering and mathematics) skills for the economy*. London: Department for Business, Energy & Industrial Strategy Department for Education.

National Foundation for Educational Research (NFER). (2004). *NFER Progress in Maths 4-14 Test Series*. London: NFER-Nelson.

National Research Council of the National Academies. (2006). *Learning to Think Spatially*. Washington DC: The National Academies Press. https://doi.org/10.17226/11019

Needham, A. (2009). Learning in Infants' Object Perception, Object-Directed Action, and Tool Use. In *Learning and the Infant Mind* (pp. 208–226). New York: Oxford University Press.

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. https://doi.org/10.1016/j.paid.2011.02.017

Newcombe, N. S., & Huttenlocher, J. (1992). Children's early ability to solve perspective-taking problems. *Developmental Psychology*, *28*(4), 635–643. https://doi.org/10.1037/0012-1649.28.4.635

Newcombe, N. S. (2013). Seeing Relationships: Using Spatial Thinking to Teach Science, Mathematics, and Social Studies. *American Educator*, *37*(1), 26.

Newcombe, N. S. (2017). Building spatial skills in preschool. *The Psychologist*, 48–51.

Newcombe, N. S. (2018). Three kinds of spatial cognition. In *J. Wixted (Ed.), Stevens' Handbook of Experimental Psychology and Cognitive Neuroscience* (4th ed.).

Newcombe, N. S., & Huttenlocher, J. (2000). *Making Space: The Development of Spatial Representation and Reasoning*. London: MIT Press.

Newcombe, N. S., & Huttenlocher, J. (2006). Development of spatial cognition. In *D. Kuhn & R. S. Siegler (Eds.), Handbook of child psychology* (6th edition, pp. 734–776). John Wiley.

Newcombe, N. S., Huttenlocher, J., & Learmonth, A. (1999). Infants' coding of location in continuous space. *Infant Behavior and Development*, *22*(4), 483–510. https://doi.org/10.1016/S0163-6383(00)00011-4

Newcombe, N. S., Möhring, W., & Frick, A. (2018). Chapter 9 - How Big Is Many? Development of Spatial and Numerical Magnitude Understanding. In A. Henik & W. Fias (Eds.), *Heterogeneity of Function in Numerical Cognition* (pp. 157–176). Academic Press. https://doi.org/10.1016/B978-0-12-811529-9.00009-1

Newcombe, N. S., & Shipley, T. (2015). Thinking About Spatial Thinking: New Typology, New Assessments. In *Studying Visual and Spatial Reasoning for Design Creativity* (pp. 179–192). Springer, Dordrecht. https://doi.org/10.1007/978-94-017-9297-4_10

Newcombe, N. S., Sluzenski, J., & Huttenlocher, J. (2005). Preexisting Knowledge Versus On-Line Learning. *Psychological Science*, *16*(3), 222–227. https://doi.org/10.1111/j.0956-7976.2005.00807.x

Newcombe, N. S., Uttal, D. H., & Sauter, M. (2013). Spatial Development. In P. D. Zelazo (Ed.), Oxford Handbook of Developmental Psychology. *New York: Oxford University Press.*, 564–590.

Newcombe, N.S, Levine, S. C., & Mix, K. S. (2015). Thinking about quantity: the intertwined development of spatial and numerical cognition. *Wiley Interdisciplinary Reviews. Cognitive Science*, *6*(6), 491–505. https://doi.org/10.1002/wcs.1369

Nunes, T., Bryant, P., Evans, D., Bell, D., Gardner, S., Gardner, A., & Carraher, J. (2007). The contribution of logical reasoning to the learning of mathematics in primary school. *British Journal of Developmental Psychology*, *25*(1), 147–166. https://doi.org/10.1348/026151006X153127

Oakes, J. (2005). *Keeping Track: How Schools Structure Inequality, Second Edition* (2nd Revised ed. edition). New Haven, Conn.; London: Yale University Press.

Oberauer, K. (2016). Parameters, Not Processes, Explain General Intelligence. *Psychological Inquiry*, *27*(3), 231–235. https://doi.org/10.1080/1047840X.2016.1181999

Obersteiner, A., Reiss, K., & Ufer, S. (2013). How training on exact or approximate mental representations of number can enhance first-grade students' basic number processing and arithmetic skills. *Learning and Instruction*, *23*, 125–135. https://doi.org/10.1016/j.learninstruc.2012.08.004

Office of Communication (OFcom). (2015). *Children and Parents: Media Use and Attitudes Report*. London: OFcom.

Okamoto, Y., Kotsopoulos, D., McGarvey, L., & Hallowell, D. (2015). The development of spatial reasoning in young children. In *Spatial Reasoning in the Early Years: Principles, Assertions, and Speculations* (pp. 15–29). New York: Routledge.

Okamoto-Barth, S., & Call, J. (2008). Tracking and inferring spatial rotation by children and great apes. *Developmental Psychology*, *44*(5), 1396–1408. https://doi.org/10.1037/a0012594

Örnkloo, H., & von Hofsten, C. (2007). Fitting objects into holes: On the development of spatial cognition skills. *Developmental Psychology*, *43*(2), 404–416. https://doi.org/10.1037/0012-1649.43.2.404

Paas, F., & Sweller, J. (2012). An Evolutionary Upgrade of Cognitive Load Theory: Using the Human Motor System and Collaboration to Support the Learning of Complex Cognitive Tasks. *Educational Psychology Review*, *24*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2

Palmer, S. (1978). Fundamental aspects of cognitive representation. In *E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization* (pp. 259–303). Hillsdale, NJ: Erlbaum.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge University Press.

Penner, A. M., & Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, *37*(1), 239–253. https://doi.org/10.1016/j.ssresearch.2007.06.012

Phelps, B. I. (2011). Dishabituation. In S. Goldstein & J. A. Naglieri (Eds.), *Encyclopedia of Child Behavior and Development* (pp. 509–512). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-79061-9_867

Piaget, J. (1941). *The child's Conception of Number*. New York: Norton.

Piaget, J. (1951). *Play, dreams, and imitation in childhood*. New York: Norton.

Piaget, J. (1952). *The Origins of Intelligence in Children*. New York: Norton.

Piaget, J. (1954). *The Construction of Reality in the Child.* New York: Basic Books.

Piaget, J., & Inhelder, B. (1948). *The Childs Conception of Space*. New York: Norton.

Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*(12), 542–551. https://doi.org/10.1016/j.tics.2010.09.008

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D., … Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33–41. https://doi.org/10.1016/j.cognition.2010.03.012

Piazza, M., Pica, P., Izard, V., Spelke, E. S., & Dehaene, S. (2013). Education Enhances the Acuity of the Nonverbal Approximate Number System. *Psychological Science*, *24*(6), 1037–1043. https://doi.org/10.1177/0956797612464057

Pica, P., Lemer, C., Izard, V., & Dehaene, S. (2004). Exact and Approximate Arithmetic in an Amazonian Indigene Group. *Science*, *306*(5695), 499–503. https://doi.org/10.1126/science.1102085

Pillow, B. H., & Flavell, J. H. (1986). Young Children's Knowledge about Visual Perception: Projective Size and Shape. *Child Development*, *57*(1), 125–135. https://doi.org/10.2307/1130644

Pinel, P., Dehaene, S., Rivière, D., & LeBihan, D. (2001). Modulation of parietal activation by semantic distance in a number comparison task. *NeuroImage*, *14*(5), 1013–1026. https://doi.org/10.1006/nimg.2001.0913

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin*, *131*(4), 592–617. https://doi.org/10.1037/0033-2909.131.4.592

Praet, M., & Desoete, A. (2014). Number line estimation from kindergarten to grade 2: A longitudinal study. *Learning and Instruction*, *33*, 19–28. https://doi.org/10.1016/j.learninstruc.2014.02.003

Price, G. R., & Fuchs, L. S. (2016). The Mediating Relation between Symbolic and Nonsymbolic Foundations of Math Competence. *PLOS ONE*, *11*(2), e0148981. https://doi.org/10.1371/journal.pone.0148981

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: does talk about the spatial world matter?: Children's spatial thinking. *Developmental Science*, *14*(6), 1417–1430. https://doi.org/10.1111/j.1467-7687.2011.01088.x

Purpura, D. J., Schmitt, S. A., & Ganley, C. M. (2017). Foundations of mathematics and literacy: The role of executive functioning components. *Journal of Experimental Child Psychology*, *153*, 15–34. https://doi.org/10.1016/j.jecp.2016.08.010

Purpura, D. J., & Simms, V. (2018). Approximate number system development in preschool: What factors predict change? *Cognitive Development*, *45*, 31–39. https://doi.org/10.1016/j.cogdev.2017.11.001

Ramani, G. B., Zippert, E., Schweitzer, S., & Pan, S. (2014). Preschool children's joint block building during a guided play activity. *Journal of Applied Developmental Psychology*, *35*(4), 326–336. https://doi.org/10.1016/j.appdev.2014.05.005

Rattan, A., Good, C., & Dweck, C. S. (2012). "It's ok — Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of Experimental Social Psychology*, *48*(3), 731–737. https://doi.org/10.1016/j.jesp.2011.12.012

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's Working in Working Memory Training? An Educational Perspective. *Educational Psychology Review*, *27*(4), 617–633. https://doi.org/10.1007/s10648-015-9314-6

Reichenbach, H. (1956). *The Direction of Time*. University of California Press.

Reuhkala, M. (2001). Mathematical Skills in Ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, *21*(4), 387–399. https://doi.org/10.1080/01443410120090786

Rips, L. J., Bloomfield, A., & Asmuth, J. (2008). From numerical concepts to concepts of number. *Behavioral and Brain Sciences*, *31*, 623–642.

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: interpretations and misinterpretations. *Nature Reviews Neuroscience*, *11*(4), 264–274.

Roberts, M., & Chaperon, H. (1989). Cognitive and exemplary modelling on the Piagetian water-level task. *International Journal of Behavioural Development*, 12, 453-473.

Robinson, J. P., & Lubienski, S. T. (2011). The Development of Gender Achievement Gaps in Mathematics and Reading During Elementary and Middle School: Examining Direct Cognitive Assessments and Teacher Ratings. *American Educational Research Journal*, *48*(2), 268–302. https://doi.org/10.3102/0002831210372249

Rogoff, B. (1990). *Apprenticeship in thinking: Cognitive development in social context* (Vol. xiv). New York, NY, US: Oxford University Press.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science*, *17*(4), 525–536. https://doi.org/10.1111/desc.12166

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The Development of Internal Representations of Magnitude and Their Association with Arabic Numerals. *Journal of Experimental Child Psychology*, *81*(1), 74–92. https://doi.org/10.1006/jecp.2001.2645

Salatas, H., & Flavell, J. H. (1976). Perspective Taking: The Development of Two Components of Knowledge. *Child Development*, *47*(1), 103–109. https://doi.org/10.2307/1128288

Sasanguie, D., Göbel, S. M., Moll, K., Smets, K., & Reynvoet, B. (2013). Approximate number sense, symbolic number processing, or number–space mappings: What underlies mathematics achievement? *Journal of Experimental Child Psychology*, *114*(3), 418–431. https://doi.org/10.1016/j.jecp.2012.10.012

Sattler, J. M. (1988). *Assessment of children, 3rd ed*. San Diego, CA, England: Jerome M. Sattler.

Saxe, G. B., Guberman, S. R., Gearhart, M., Gelman, R., Massey, C. M., & Rogoff, B. (1987). Social Processes in Early Number Development. *Monographs of the Society for Research in Child Development*, *52*(2), i–162. https://doi.org/10.2307/1166071

Schneider, M., Beeres, K., Coban, L., Merz, S., Susan 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). https://doi.org/10.1111/desc.12372

Schneider, M., Merz, S., Stricker, J., Smedt, B. D., Torbeyns, J., Verschaffel, L., & Luwel, K. (2018). Associations of Number Line Estimation With Mathematical Competence: A Meta-analysis. *Child Development*, *89*(5), 1467–1484. https://doi.org/10.1111/cdev.13068

Schwaighofer, M., Fischer, F., & Bühner, M. (2015). Does Working Memory Training Transfer? A Meta-Analysis Including Training Conditions as Moderators. *Educational Psychologist*, *50*(2), 138–166. https://doi.org/10.1080/00461520.2015.1036274

Sekuler, R., & Mierkiewicz, D. (1977). Children's Judgments of Numerical Inequality. *Child Development*.

Shah, P., Buschkuehl, M., Jaeggi, S., & Jonides, J. (2012). Cognitive training for ADHD: The importance of individual differences. *Journal of Applied Research in Memory and Cognition*, *1*, 204–205. https://doi.org/10.1016/j.jarmac.2012.07.001

Shea, D. L., Lubinski, D., & Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614. https://doi.org/10.1037/0022-0663.93.3.604

Shusterman, A., Ah Lee, S., & Spelke, E. S. (2008). Young children's spontaneous use of geometry in maps. *Developmental Science*, *11*(2), F1–F7. https://doi.org/10.1111/j.1467-7687.2007.00670.x

Shutts, K., Ornkloo, H., von Hofsten, C., Keen, R., & Spelke, E. S. (2009). Young children's representations of spatial and functional relations between objects. *Child Development*, *80*(6), 1612–1627. https://doi.org/10.1111/j.1467-8624.2009.01357.x

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428–444. https://doi.org/10.1111/j.1467-8624.2004.00684.x

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237–243. https://doi.org/10.1111/1467-9280.02438

Siegler, R. S., Thompson, C. A., & Opfer, J. E. (2009). The Logarithmic-To-Linear Shift: One Learning Sequence, Many Tasks, Many Time Scales. *Mind, Brain, and Education*, *3*(3), 143–150. https://doi.org/10.1111/j.1751-228X.2009.01064.x

Simms, V., Clayton, S., Cragg, L., Gilmore, C., & Johnson, S. (2016). Explaining the relationship between number line estimation and mathematical achievement: The role of visuomotor integration and visuospatial skills. *Journal of Experimental Child Psychology*, *145*, 22–33. https://doi.org/10.1016/j.jecp.2015.12.004

Simon, T. J. (1997). Reconceptualizing the origins of number knowledge: A "non-numerical" account. *Cognitive Development*, *12*(3), 349–372. https://doi.org/10.1016/S0885-2014(97)90008-3

Skwarchuk, S., Vandermaas-Peeler, M., & LeFevre, J.-A. (2016). Optimizing the Home Numeracy Environments of 3- to 6-Year-Old Children in the USA and Canada. In B Blevins-Knabe & A. Austin (Eds.), *Early Childhood Mathematics Skill Development in the Home Environment* (pp. 127–146). Springer International Publishing. https://doi.org/10.1007/978-3-319-43974-7_8

Sodian, B., Thoermer, C., & Metz, U. (2007). Now I see it but you don't: 14-month-olds can represent another person's visual perspective. *Developmental Science*, *10*(2), 199–204. https://doi.org/10.1111/j.1467-7687.2007.00580.x

Sonnenschein, S., and C. Galindo. 2015. Race/Ethnicity and Early Mathematics Skills: Relations between Home, Classroom, and Mathematics Achievement. *Journal of Educational Research.* 108 (4): 261–277.

Sorby, S. A., & Baartmans, B. J. (2000). The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students. *Journal of Engineering Education*, *89*(3), 301–307. https://doi.org/10.1002/j.2168-9830.2000.tb00529.x

Spelke, E., & Newport, E. (1998). Nativism, empiricism, and the development of knowledge. In *Handbook of Child Psychology:Theoretcial Models of Human Development* (Vol. 1). New York: Wiley.

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89–96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology*, *14*(2), 92–104. https://doi.org/10.1037/a0019491

Starkey, P. (1992). The early development of numerical reasoning. *Cognition*, *43*(2), 93–126. https://doi.org/10.1016/0010-0277(92)90034-F

Starkey, P., & Cooper, R. G. (1980). Perception of numbers by human infants. *Science*, *210* (4473), 1033–1035.

Starkey, P., Spelke, E. S., & Gelman, R. (1990). Numerical abstraction by human infants. *Cognition*, *36*(2), 97–127. https://doi.org/10.1016/0010-0277(90)90001-Z

Strobach, T., & Karbach, J. (2016). Introduction In T. Strobach & J. Karbach (Eds.), *Cognitive Training: An Overview of Features and Applications* (pp. 1–5). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-42662-4_1

Stull, A. T., Hegarty, M., Dixon, B., & Stieff, M. (2012). Representational Translation With Concrete Models in Organic Chemistry. *Cognition and Instruction*, *30*(4), 404–434. https://doi.org/10.1080/07370008.2012.719956

Talmy, L. (2000). How Language Structures Space. In *Towards a Cognitive Semantics* (Vol. 1, pp. 177–254). MIT Press.

Taylor, H. A., & Hutton, A. (2013). Think3d!: Training Spatial Thinking Fundamental to STEM Education. *Cognition and Instruction*, *31*(4), 434–455. https://doi.org/10.1080/07370008.2013.828727

Tettamanti, M., Buccino, G., Saccuman, M. C., Gallese, V., Danna, M., Scifo, P., … Perani, D. (2005). Listening to Action-related Sentences Activates Fronto-parietal Motor Circuits. *Journal of Cognitive Neuroscience*, *17*(2), 273–281. https://doi.org/10.1162/0898929053124965

Thompson, J. M., Nuerk, H.-C., Moeller, K., & Cohen Kadosh, R. (2013). The link between mental rotation ability and basic numerical representations. *Acta*

Psychologica, *144*(2), 324–331. https://doi.org/10.1016/j.actpsy.2013.05.009

Thurstone, L. L. (1938). *Primary mental abilities*. University of Chicago Press: Chicago.

Thurstone, L. L., Jeffrey, T.E. (1984). *Closure Flexibility (Concealed Figures): Interpretation and Research Manual*. Reid London House: London.

Tosto, M. G., Hanscombe, K. B., Haworth, C. M. A., Davis, O. S. P., Petrill, S. A., Dale, P. S., … Kovas, Y. (2014). Why do spatial abilities predict mathematical performance? *Developmental Science*, *17*(3), 462–470. https://doi.org/10.1111/desc.12138

Träff, U. (2013). The contribution of general cognitive abilities and number abilities to different aspects of mathematics in children. *Journal of Experimental Child Psychology*, *116*(2), 139–156. https://doi.org/10.1016/j.jecp.2013.04.007

Tversky, B., Morrison, J. B., Franklin, N., & Bryant, D. J. (1999). Three Spaces of Spatial Cognition. *The Professional Geographer*, *51*(4), 516–524. https://doi.org/10.1111/0033-0124.00189

Twyman, A., & Newcombe, N. S. (2010). Five reasons to doubt the existence of a geometric module. *Cognitive Science*, *34*, 1315–1356. https://doi: 10.1111/j.1551-6709.2009.01081

UK Department for Education. (2013). Mathematics programmes of study: key stages 1 and 2 National curriculum in England. Retrieved from https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study

UK Department for Education (2017). National curriculum assessments: key stage 2, 2017 (revised). Retrieved from https://www.ethnicity-facts-figures.service.gov.uk/education-skills-and-training/7-to-11-years-old/maths-attainments-for-children-aged-7-to-11-key-stage-2/latest

UK Department for Education. (2017b). Schools, pupils and their characteristics: January 2017. Retrieved from https://www.gov.uk/government/statistics/schools-pupils-and-their-characteristics.

Uttal, D. H. (1996). Angles and distances: children's and adults' reconstruction and scaling of spatial configurations. *Child Development*, *67*(6), 2763–2779. http://dx.doi.org/10.2307/1131751

Uttal, D. H., & Cohen, C. (2012). Spatial Thinking and STEM Education. In *Psychology of Learning and Motivation* (Vol. 57, pp. 147–181). Elsevier. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, *139*(2), 352–402. https://doi.org/10.1037/a0028446

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and Enhancing Spatial Thinking: Links to Achievement in Science, Technology, Engineering, and Mathematics? *Current Directions in Psychological Science*, *22*(5), 367–373. https://doi.org/10.1177/0963721413484756

Van Breukelen, G. J. (2006). ANCOVA versus change from baseline had more power in randomized studies and more bias in nonrandomized studies. Journal of Clinical Epidemiology, 59(9), 920-925. https://doi.org/10.1016/j.jclinepi.2006.02.007

Van Herwegen, J., Costa, H. M., Nicholson, B., & Donlan, C. (2018). Improving number abilities in low achieving preschoolers: Symbolic versus non-symbolic training programs. *Research in Developmental Disabilities*, *77*, 1–11. https://doi.org/10.1016/j.ridd.2018.03.011

Vanbinst, K., Ceulemans, E., Peters, L., Ghesquière, P., & De Smedt, B. (2018). Developmental trajectories of children's symbolic numerical magnitude processing skills and associated cognitive competencies. *Journal of Experimental Child Psychology*, *166*, 232–250. https://doi.org/10.1016/j.jecp.2017.08.008

Vandenberg, S. G. & Kuse, A.R. (1978). Mental rotations: a group test of three dimensional spatial visualization, *Perceptual and Motor Skills*, Vol.47, pp. 599-604.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and mathematical skills across the pre-school years.

Monographs of the Society for Research in Child Development, *82*. https:// doi: 10.1111/mono.12280.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing Building Blocks: Preschoolers' Spatial Assembly Performance Relates to Early Mathematical Skills. *Child Development*, *85*(3), 1062–1076. https://doi.org/10.1111/cdev.12165

Verguts, T., & Fias, W. (2008). Symbolic and nonsymbolic pathways of number processing. *Philosophical Psychology*, *21*, 539–554. http://dx.doi.org/10.1080/09515080802285545

von Aster, M. G., & Shalev, R. S. (2007). Number development and developmental dyscalculia. *Developmental Medicine & Child Neurology*, *49*(11), 868–873. https://doi.org/10.1111/j.1469-8749.2007.00868.x

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. http://dx.doi.org/10.1037/0033-2909.117.2.250

Wai, J., Lubinski, D., & Benbow, C. P. (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. https://doi.org/10.1037/a0016127

Wang, Z., Zhou, R., & Shah, P. (2014). Spaced cognitive training promotes training transfer. *Frontiers in Human Neuroscience*, *8*. https://doi.org/10.3389/fnhum.2014.00217

Werner, K., & Raab, M. (2014). Moving your eyes to solution: Effects of movements on the perception of a problem-solving task. *The Quarterly Journal of Experimental Psychology*, *67*(8), 1571–1578. https://doi.org/10.1080/17470218.2014.889723

Wiese, H. (2007). The co-evolution of number concepts and counting words. *Lingua*, *117*(5), 758–772. https://doi.org/10.1016/j.lingua.2006.03.001

Winter, B., Matlock, T., Shaki, S., & Fischer, M. H. (2015). Mental number space in three dimensions. *Neuroscience & Biobehavioral Reviews*, *57*, 209–219. https://doi.org/10.1016/j.neubiorev.2015.09.005

Witkin, H. A., Dyk, R. B., & Faterson, H. F. (1962). *Psychological differentiation.* New York: Wiley.

Witkin, H. A., & Goodenough, D. R. (1981). Cognitive styles: essence and origins. Field dependence and field independence. *Psychological Issues*, (51), 1–141.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology*, *7*(3), 291–300. http://dx.doi.org/10.1037/h0025070

Witkin, H. A., Otman, P. K., Raskin, E., & Karp, S. (1971). *A manual for the embedded figures test*. California: Consulting Psychologists Press.

Wolfgang, C., Stannard, L., & Jones, I. (2001). Block play performance among preschoolers as a predictor of later school achievement in mathematics. *Journal of Research in Childhood Education*, *15*(2), 173-181. https://doi.org/10.1080/02568540109594958

Wynn, K. (1992). Addition and subtraction by human infants. *Nature*, *358*(6389), 749–750. https://doi.org/10.1038/358749a0

Xenidou-Dervou, I., Molenaar, D., Ansari, D., van der Schoot, M., & van Lieshout, E. C. D. M. (2017). Nonsymbolic and symbolic magnitude comparison skills as longitudinal predictors of mathematical achievement. *Learning and Instruction*, *50*, 1–13. https://doi.org/10.1016/j.learninstruc.2016.11.001

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1–B11. http://dx.doi.org/10.1016/S0010-0277(99)000669

Young, C. J., Levine, S. C., & Mix, K. S. (2018). The Connection Between Spatial and Mathematical Ability Across Development. *Frontiers in Psychology*, *9*. https://doi.org/10.3389/fpsyg.2018.00755

Zacks, J. M., Mires, J., Tversky, B., & Hazeltine, E. (2000). Mental spatial transformations of objects and perspective. *Spatial Cognition and Computation*, *2*(4), 315–332. https://doi.org/10.1023/A:1015584100204

Zacks, J. M., Vettel, J. M., & Michelon, P. (2003). Imagined viewer and object rotations dissociated with event-related FMRI. *Journal of Cognitive Neuroscience*, *15*(7), 1002–1018. https://doi.org/10.1162/089892903770007399

Zhang, X., Koponen, T., Räsänen, P., Aunola, K., Lerkkanen, M.-K., & Nurmi, J.-E. (2014). Linguistic and Spatial Skills Predict Early Arithmetic Development via Counting Sequence Knowledge. *Child Development*, *85*(3), 1091–1107. https://doi.org/10.1111/cdev.12173

Zimmerman, W. S. (1954). Hypotheses Concerning the Nature of the Spatial Factors. *Educational and Psychological Measurement*, *14*(2), 396–400. https://doi.org/10.1177/001316445401400220