University of Southampton

FACULTY OF SOCIAL, HUMAN AND MATHEMATICAL SCIENCES

School of Psychology

Children's Mental Representation of Number, Their Number Line Estimations and Maths Achievement: Exploring the Role of 3D Mental Rotation Skills.

by

Lesley Anne Honour

Thesis for the degree of Doctor of Educational Psychology

July 2019

Word count: 19168

University of Southampton

Abstract

FACULTY OF SOCIAL, HUMAN AND MATHEMATIC SCIENCES

School of Psychology

Thesis for the degree of Doctor of Educational Psychology

Children's Mental Representation of Number, Their Number Line Estimations and Maths Achievement: Exploring the Role of 3D Mental Rotation Skills.

Lesley Anne Honour

As children's early grasp of number is a reliable indicator of their future mathematical competence, any internal representations they develop to encode conceptual understanding, or as a framework for problem solving, are of particular interest to researchers. Although the exact nature and form of these mental models are difficult to establish, prior research has identified a directional left to right number line as the most probable internal representational schema for number magnitude. However, the extent to which these types of mental representations influence maths achievement is not clear, as there has yet to be a systematic review of the literature evidencing such links. This review set out to systematically gather, and critically examine, findings in one small area of a potentially large field of research; specifically the relationship between children's mental representation of number magnitude and early maths achievement. Ten studies met the criterion for inclusion, each exploring the influence of mental representations through correlational data, or through training designed to enhance any such internal symbolic framework. Results indicate that internal representations of number are important for mathematical competence particularly in the early years, as notation and calculation with integers becomes increasingly symbolic. Implications for EPs were discussed including best approaches that model and encourage precision of the mental number line, such as linear board games, and awareness of specific groups that are most likely to benefit from any such intervention.

Number magnitude knowledge is a foundational concept within mathematics, observable early in life through behavioural phenomena and linked with spatial-numerical associations. The internal representation of symbolic number magnitude is thought to be a directional left to right mental number line, acting as a framework to encode conceptual understanding and to support problem solving. Accurate performance on the associated metric, the number line estimation task (NLE), has been interpreted as improved understanding of number magnitude and is reliably related to maths achievement. However, the exact role spatial skills play in the relationship between number magnitude understanding and NLE tasks is still unclear, particularly as these skills are themselves independently related to maths achievement, whilst also being influential in the proportional judgement strategies used to complete the NLE task. To investigate these relationships, 98 primary children were recruited for a RCT training 3D mental rotation skills using a computer-based tool, over ten sessions. Spatial skill and NLE performance strongly correlated with maths achievement, and although spatial training improved spatial ability, this was not significant. Unexpectedly, spatial training did not influence NLE performance. Implications for EPs include effective use of number line tasks to target weak number magnitude understanding, and benefits of providing spatial tasks to support positive maths outcomes.

Table of Contents

List of Tables

List of Figures

Research Thesis: Declaration of Authorship

Print name:	Lesley Anne Honour

Title of thesis:	Children's Mental Representation of Number, Their Number Line Estimations and Maths Achievement: Exploring the Role of 3D Mental Rotation Skills.

I declare that this thesis and the work presented in it are my own and has been generated by me as the result of my own original research.

I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at this University;
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
3. Where I have consulted the published work of others, this is always clearly attributed;
4. Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
5. I have acknowledged all main sources of help;
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
7. None of this work has been published before submission

Signature:		Date:	

Acknowledgements

With grateful thanks to Malcolm, Katy, Tom, Rachel, Aaron, Dency, Joe, Kate, Sophie, Kim and Pippa. You have each offered tremendous support throughout, and kept my motivation levels high.

Also to Sarah, Ed and Christian, for the knowledge and expertise which you have been freely willing to share. The listening ear, advice and encouragement have been invaluable.

Definitions and Abbreviations

ANS	Approximate Number System
DfE	Department for Education
EEG	Electroencephalogram
EP	Educational Psychologist
NLE	Number Line Estimation task
PAE	Percentage Absolute Error
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
R^2_{LIN}	The extent of the linear relationship between estimated position and actual position of numbers along a line.
RCT	Randomised Controlled Trial
SATs	Standardised End of Key Stage Tests and Assessments
SEN	Special Educational Needs
SENCo –	Special Educational Needs Co-ordinator
SD	Standard Deviation
SNARC	Spatial-Numerical Association of Response Codes
SPSS	Statistical Package for the Social Sciences

Chapter 1 The Relationship Between Children's Mental Representation of Number and Maths Achievement: A Systematic Review of the Literature.

1.1 Introduction

The idea that children's early grasp of number is a reliable indicator of their future competence within mathematics has been explored by a variety of researchers (Booth & Siegler, 2008; Geary, Hoard, Nugent, & Bailey, 2013). Within this research particular interest has been given to how children assimilate concepts of number magnitude, and the nature and influence of any internal models they develop (Schneider, Grabner, & Paetsch, 2009; Thompson & Siegler, 2010). These mental representations are important as magnitude is an influential feature of number affecting estimates and calculation. They are posited to have a dual role, as a mechanism for encoding conceptual understanding but also acting as a framework for manipulating number during problem solving (Park, Bermudez, Roberts, & Brannon, 2016; Skemp, 1978). Consolidating the internal representation of number magnitude, through training skills in the associated external representation, is reported to support children's mental calculation abilities, increase their accuracy during arithmetical problem solving and improve their magnitude comparisons of fractions (Hamdan & Gunderson, 2017; Kucian et al, 2011; Sella, Tressoldi, Lucangeli, & Zorzi, 2016). Mental representations of number have also been investigated as the basic building blocks necessary for more advanced abstract mathematics (Case & Okamoto, 1996). These suggested benefits indicate a need for a systematic review of the literature to explore evidence of any connection between children's mental representation of number and early maths achievement. Through evaluating any links, this review may also offer insight into the strength of any association, notice the form representations may take, how they are acquired, and discuss implications for how improved maths outcomes are best supported.

1.1.1 Representations

Representations in the broad sense can be thought of as symbolic; the way that one thing stands for another. Representations can be external, such as gestures, written text, graphs, paintings or maps. They can also be internal, when they might take the form of a visual image that can replicate or give an impression, propositional representation (where language cues are used), or mental models which are structural analogues of ideas (Johnson-Laird, 1983). However, mental representations should not be considered solely as discrete pictures or words in the mind. Kosslyn and Pomerantz (1977) discuss how other senses can be involved in the encoding process suggesting representations do not occur in isolation but rather belong to highly structured systems similar to a computer based structure, where hierarchical links join different concepts, reflecting reality in a personalized web of representations.

Goldin and Kaput (1996) considered the difficulties in trying to describe how representations ‘look’ inside the mind and suggested that observable behaviours are the chief clues that allow us to infer what might be encoded in the brain and nervous system. With this in mind, researchers have been interested in how external models can influence the acquisition and structure of internal representations, and how the process works in reverse as internal representations are retrieved and reproduced in some external form (e.g. Rapp & Uttal, 2006). Although there may not be an exact reflection between mind and action, knowing more about any ‘match’ between the internal and external (whether intentionally created or more automatically produced) would help teachers provide better models to aid learning, and would also support the development of tasks that more accurately measure internal understanding.

1.1.2 Ontology and Epistemology

In trying to describe how information is represented internally, both ontological and epistemological questions begin to surface, as the enquiry centres on the fundamental question of memory, both form and function. What is the nature of representations? Are they related to thoughts and ideas? To what extent can we accurately represent a shared conceptual understanding of an idea, even if there is such a thing? These questions are beyond the scope of what can be covered here but they provide a backdrop within which the review question has been considered. In terms of epistemology, the approach taken is to focus on the way knowledge of internal representations has been gained, and how that method of acquiring knowledge may support conclusions about the validity of those representations. The genetic epistemology described by Piaget fits well here as it considers developmental change and suggests knowledge consists of structures which come into being through adaptation of these structures with the environment.

1.1.3 Representations within Mathematics

Symbolic representations are a key part of communicating ideas and concepts within mathematics. Think for a moment of the operations denoted by four simple marks that refer to different ways quantities can be grouped and manipulated, or the relative position of numerals that indicates value above that which the numeral carries (place value). The many and varied external representations within mathematics form a language that aims to convey a shared understanding of concepts and procedures that a child can assimilate into an internal schema. Bruner's (1966) framework of learning suggested children create representational understanding as they move through a stepped sequence of three stages: enactive, iconic and symbolic. First, an idea is defined through participation and doing, where actions are the encoding of an experience (e.g. a baby shakes a rattle, hears a noise and subsequently shakes her hand even without the rattle as the representation of the noise is the action, not the object). In the iconic stage repeated manipulation of concrete materials generates a pattern of understanding that is represented as an image (e.g. the way a pizza is divided to represent fractions). At around age 6 or 7, symbolic notation can be used to represent ideas, allowing a level of abstraction to develop.

Bruner's (1966) model has been built on by other researchers as they have explored the importance of representations in developing mathematical competence. Skemp (1978) identified two types of mathematical understanding: instrumental and relational. Instrumental was described as procedural memories - the 'how' or rules of completing a task. In contrast, relational understanding involved the building up of a cognitive structure of ideas (including iconic and symbolic representations) to form a framework schema which can be adapted as new understanding is assimilated. Similarly, Hiebert and Carpenter (1992) explored the purpose of external and internal representations, suggesting communicating ideas in mathematics relies on external representations which are then transformed into an internal representation which can be drawn upon for subsequent reference and problem solving. They suggested that any errors in this transference can lead to misconceptions which may persist, as they are part of an internal model of understanding upon which new ideas are built. The idea that notation and symbols can constrain or support our understanding in this way was also put forward by Kaput (1991). He maintained that making connections between external and internal representations is vital for maths competence, as it helps extend understanding from concrete to more abstract systems.

1.1.4 Representations of Number

As this review is focused specifically on how children internally represent number magnitude within mathematics, it will be helpful to outline theories of how this concept develops. Following on from early research into subitizing (perceiving the number of objects without counting each item; Kaufmann, Lord, Reese, and Volkmann, 1949), numerous infant studies suggest an early visual appreciation of differing quantities (Strauss & Curtis, 1981; van Loosbroek & Smitsman, 1990; Xu & Spelke, 2000). This appears to be underpinned by two innate cognitive systems, the Parallel Individuation System which is triggered to track individual items when there are fewer than 4, and the Approximate Number System (ANS) that perceptually estimates difference, for example in a dot array (Piazza, 2011; Gallistel & Gelman, 2000). There is general agreement that the ANS improves in precision throughout childhood and adolescence (Halberda & Feigenson, 2008; Libertus & Brannon, 2010; Xu & Spelke, 2000) although Webber's law appears to hold across all stages, (it is easier to detect differences of equal absolute value in lower numbers than higher numbers). Both systems indicate an initial, innate, non-symbolic awareness (or representation) of relative magnitude.

As children experience the symbolic language of the Arabic numeral system through practising the number sequence with words and numerals, such as counting 'how many' and using numerals to compare group quantities, researchers such as Dehaene (2004) and Li and Baroody (2014) suggest a mapping occurs between the non-symbolic and symbolic representations, with each enhancing the other. The ANS is thought to become more accurate as symbolic digits support discrimination of near quantity difference (Mussolin, Nys, Leybaert, & Content, 2015), while numerical knowledge (including multi digit base 10 understanding) improves through awareness of relational magnitude that the ANS offers (Feigenson, Dehaene, & Spelke, 2004). However this developmental trajectory has been challenged by Gunderson, Spaepen, and Levine (2015) who suggest that exact symbolic and approximate non-symbolic number knowledge may develop in parallel without such causal interaction. They base this claim on work exploring whether knowledge of the cardinal principle (the last number counted gives the quantity of the whole set) was a prerequisite for approximate number word knowledge, or vice versa. They found no such relationship (either way) and so concluded that the cognitive systems underlying these two aspects of numerical development were likely to be quite distinct.

1.1.5 Spatial-Numerical Associations (SNAs)

Woven into the literature exploring internal representations in any general domain there are references to the role of spatial awareness, including positional relations between linked ideas (Reed, 1974), the role of embodiment and movement in encoding a representation (Moeller, Martignon, Wessolowski, Engel, & Nuerk, 2011) and possible brain areas where domain specific representations may be stored (Piazza, 2011). This spatial component involved in creating and retrieving mental representations has additional relevance within mathematics, as links between the spatial and mathematical domains have long been suggested. Spatial connections seem naturally associated with geometry but are also pertinent to number magnitude, as demonstrated by the Cartesian coordinate system developed by René Descartes. Within this system, a uniform distance along the horizontal or vertical plane represents a unit increase in magnitude, from left to right and from bottom to top. A simplified model is used with children from an early age where the positive x-axis becomes a number line with integers placed in ascending order from left to right. These type of external spatial representations of number seem intuitively helpful, but is this because they align well with internal representations for which humans are already primed?

Research into spatial-numerical associations (SNAs) through observing behavioural phenomena, offers some evidence that particular types of SNAs are present in infants and children, and that some persist throughout development. This suggests a possible innate integration of number and space. One example is the Spatial-Numerical Association of Response Codes (SNARC; Dehaene, Bossini, & Giraux, 1993). This behavioural phenomenon demonstrates that given a parity judgment task, left to right readers associate small numbers more with the left hand side and larger numbers with the right. This has been observed in a variety of populations including Chinese children from age 4 (Dehaene, Bossini, & Giraux, 1993; Yang et al., 2014; van Galen & Reitsma, 2008). Other SNAs have also been evidenced from observations including the distance effect, where the greater the distance between the two numbers compared, the better the performance (Moyer & Landauer, 1967), and the ratio effect where the numerical ratio between two numbers to be compared predicts performance (e.g. in college students; Buckley & Gillman, 1974).

1.1.6 The Mental Number Line and the Number Line Estimation Task.

SNAs are used as supportive evidence for an innate number-space mapping system in the form of a mental number line where numerical magnitude is represented as increasing from left to right along a horizontal plane. Two key models have been put forward, the logarithmic model (where magnitudes are internally represented with a constant variability following a logarithmic pattern; Dehaene, 1997) and the accumulator model (where magnitudes are represented linearly but accuracy decreases for larger magnitudes and has scalar variability; Gibbon & Church, 1981).

As both posited models are internal representations, they are, by their nature, difficult to measure. However, a task called the number line estimation task (NLE) has been put forward as an external measure thought to reflect the internal representation of the mental number line (Booth & Siegler, 2006; Opfer & Siegler, 2007; Siegler & Booth, 2004; Siegler & Opfer, 2003; Siegler & Ramani, 2009). The NLE task involves placing numbers on a line where the only stimuli are the initial and final numbers within a set scale along a left to right plane (the bounded NLE task). Researchers have consistently found that younger children place numbers in a logarithmic arrangement (large numbers compressed) while older children (approximately over 8 years) tend to place numbers in an increasingly linear arrangement, with distances between numbers more equal along the line (Booth & Siegler, 2006; Opfer & DeVries, 2008). Siegler, Thompson, and Opfer (2009), suggested this 'log-linear shift' occurs more than once, with estimates becoming more linear as each scale becomes familiar (e.g. 1 to 100 will produce linear estimates before 1 to1000). This shift has been interpreted as a marker for improved conceptual understanding of number magnitude, especially as improved accuracy of estimates (percentage absolute error; PAE) and linear performance (relationship between estimated position and actual position; R^2_{LOG} /R^2_{LIN}) has been shown to positively correlate with maths achievement (Booth & Siegler, 2006; Fazio, Bailey, Thompson, & Siegler, 2014; Geary, 2011).

The premise of an internal representation of number in the form of a mental number line, and the accompanying bounded NLE task, is prevalent throughout number magnitude research. However, there is ongoing discussion about how each should be interpreted, with two areas of debate concerning maths achievement. Firstly, if the mental number line is not a matured culmination of innate number sense skills (i.e. not innate itself) but gleaned from repeated exposure to cultural and educational conventions, then acquiring skills necessary for maths achievement, such as arithmetical competency, might be built on domain general cognitive factors rather than domain specific number sense (Núñez, 2011). This debate is not suggesting that the mental number line (whether innate or acquired) is an unhelpful framework for conceptualising number or to support problem solving, but rather the mental number line (and its refinement towards linearity) may not be the only basis upon which number understanding is built. For example, Schneider, Grabner, and Paetsch (2009), identified domain general factors (such as memorized algorithms) had more influence on school maths achievement of 110 children aged 11, than domain specific skills (e.g. SNARC effect; Dehaene, Bossini, & Giraux, 1993) which are posited to underpin the mental number line. The second area of debate is whether the bounded NLE task is a valid measure of mental representation of number magnitude. Some researchers suggest gains in accurate placement of numbers along the line (log-linear shift), are not due to improved perception of difference between integers (fundamental to numerical cognition theories), but result from development of domain general cognitive skills, proportion-judgment strategies, or mastery of place value within the base 10 Arabic number system (Barth & Paladino, 2011; Slusser, Santiago, & Barth, 2013; LeFevre et al, 2013). This is of interest because if number magnitude representation and understanding is key for early maths achievement, any measurement tool needs to have validity, i.e. measure what is intended, not some other construct or skill. Cohen and Blanc-Goldhammer (2011) addressed this question, creating an unbounded version of the NLE task that prevents the 'whole' being seen and a proportion judgement being made. After extensive analysis, they suggested that this task provided a purer measure of integer estimation (e.g. 50 is 50 units to the right rather than 50 is halfway between 0 and 100). This position has been supported by research into eye tracking (e.g. Reinert, Huber, Nuerk, & Moeller, 2015).

It seems then that the use of the bounded NLE task as a measure of children's internal representation of number magnitude is controversial. However, the bounded NLE task persists within research into number magnitude and appears to be regarded as a useful metric to assess children's understanding of the symbolic number system. Even if it does not directly mirror a mental number line, it is a task that evaluates understanding of the relationship of integers with each other (especially in the unbounded form), and as such may still provide a window into the quality of a child's internal representation of number. It has been important to recognise and discuss any limitations of the bounded NLE task, as it features as a measure in many of the studies included within this review.

1.1.7 The Aim of This Review

If learning is the acquisition of knowledge and the subsequent use of that knowledge during problem solving, then the storage and retrieval of information from memory is an important area of research. While external expressions of this learning can be observed through verbal and written assessments, much less is known about how knowledge is encoded, what information and conceptual understanding looks like in the brain and how it is arranged. As more is found out about these representations, mental models, mappings, networks or schemas, we become more able to appreciate how they support retrieval and manipulation of information, how they accommodate new conceptual ideas within prior learning, the influence they may exert on educational achievement, and the types of teaching approaches that best support them.

Within this broad context, this review considers one specific area of learning, number magnitude. The author has a keen interest in how children become fluent within numeracy, and number magnitude has been identified as a basic building block of mathematics. Number magnitude refers to comparison of quantity, both in symbolic (numerals) and non-symbolic form (e.g. dot arrays), and is not synonymous with counting or performing calculations. In addition the choice of number magnitude is a helpful lens through which to consider mental representations as it lends itself to exploring a developmental mapping relationship; from an innate non-symbolic representation (such as ANS) to a more symbolic representation using Arabic numerals.

There has been much interest in the relationship between how children represent number magnitude and any connections with maths outcomes. Measures of these representations have most often been operationalised through dot array discrimination for non-symbolic representations (e.g. exploring the relationship between pre-schoolers' precision of the approximate number system and maths performance; Mazzocco, Feigenson and Halberda, 2011), and the bounded NLE task for symbolic representations (e.g. targeting numerical performance through training number magnitude skills; Sella, Tressoldi, Lucangeli, & Zorzi, 2016). These type of studies have produced disparate results. Although Fazio et al. (2014) attempted a to gauge the size of the relation between non-symbolic numerical magnitude understanding and general mathematics achievement with a meta-analysis of 19 studies, there has been less synthesis around findings describing symbolic representation of number and achievement in maths, especially in the early years. Additionally many studies rely on the premise of the bounded NLE task being a valid measure of internal representation of number magnitude despite recent findings which suggest otherwise. This premise will be taken into consideration as this review included studies that use both types of NLE task (bounded and unbounded), as well as other measures that appropriately operationalised internal representation of number magnitude.

1.2 Method

1.2.1 Search Strategy

Searches were conducted within three electronic databases: PsycINFO (via EBSCO: 1990-2019), Scorpus (1990-2019) and ERIC (via ProQuest: 1990-2019). The databases were searched within the domains of abstract, title and keyword. The earliest date chosen for the search was 1990, as the seminal paper from Dehaene, Bossini, and Giraux (1993) detailing the SNARC effect (Spatial-Numerical Association of Response Codes) initiated an increased level of research in spatial-numerical representations that is pertinent to the review question. The search was based on key terms generated by the review question, with each term exploded as guided by scoping searches prior to the review question being finalised. The search terms were: (child* OR pupil* OR "school student*" OR "grade student*") AND (num*) AND (mental OR internal*) AND (represent*) AND ("math* outcome" OR "math* achieve*" OR "math* skills" OR "math* competen*" OR "math* proficien*" OR "math* performance" OR "arithmetic* competen*" OR "arithmetic* proficien*" OR "arithmetic* outcome" OR "arithmetic* achieve*" OR "arithmetic* skill*" OR "arithmetic* performance*"). Papers obtained were then screened according to the inclusion criteria and relevant papers had their reference lists searched to identify any additional papers that also met the inclusion criteria. See Figure 1 for the PRISMA flow diagram.

The relationship between children's mental representation of number and maths achievement: A systematic review of the literature.

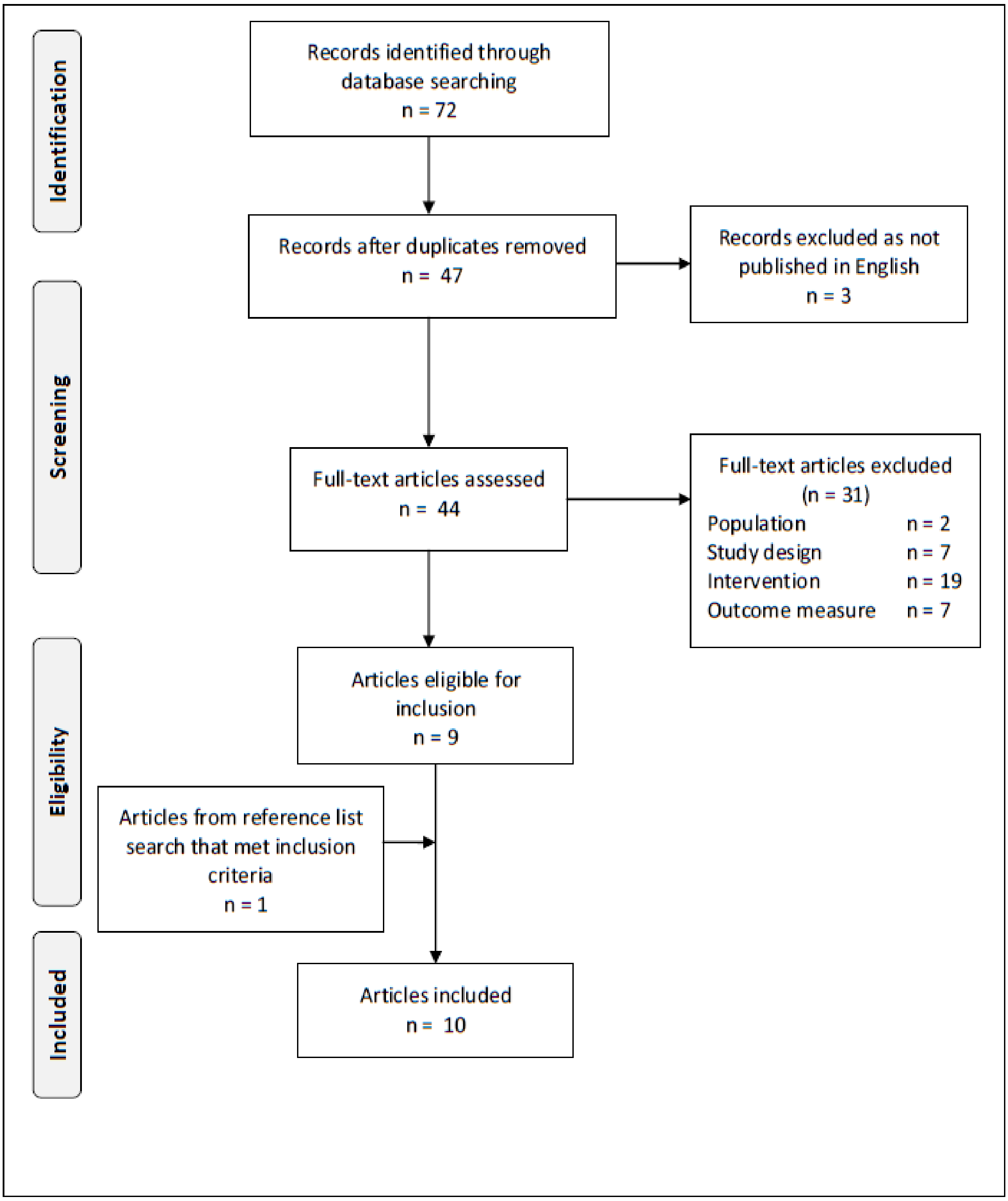


Figure 1. PRISMA Flow Diagram

1.2.2 Included Studies

Table 1 details the inclusion and exclusion criteria that were applied to ensure the final articles were relevant to the review question. Using these criteria the initial 72 studies yielded 10 to be included in the review.

Table 1
Inclusion and Exclusion Criteria for Studies

	Inclusion Criteria	Exclusion Criteria
Population	Children aged between birth and 12 years	Students over 12 years
Study Design	Quantitative designs including • observed correlations • comparison of groups	Qualitative studies Meta-analyses
Intervention	Any study which primarily explores the relationship between mental representation of number and maths achievement. Number magnitude should be the focus (in the form of integers).	Studies that explore links between • maths achievement and domain general cognitive factors, such as working memory • maths achievement and external representations only. • mental representation of number and language based skills. Studies that focus on common or decimal fractions.
Outcome Measures	Mental representation of number operationalised as • bounded or unbounded number line task • other spatial-numerical association tasks • approximate number discrimination tasks Maths progress / achievement data from any tests which include an arithmetic element.	Any other outcome measure (e.g. differences in brain activity as measured by EEG)
Language	English	Published in any language other than English

1.2.3 Data Extraction

The data extracted from eligible papers included: descriptive information about the participants (including age, study location and any other significant group identifiers); study design; information about the intervention (type, duration, frequency, delivery) and control condition; outcome measures used; and findings in the form of intervention effectiveness and / or correlations. See Appendix A for findings from data extraction.

1.2.4 Quality Assessment: Criteria and Ratings

The studies that met the inclusion criteria were screened for quality and although this did not allow for direct comparison across studies, it did indicate the relative methodological strengths of each. This screening considered three strands of evidence (A, B and C), with the weight of the strands combined in an overall judgement (D) reflecting the study's quality and contribution towards answering the review question (Gough, 2007). As all studies were quantitative, the first strand of evidence used a rating framework developed by Woods for his own research (Tyrell & Woods, 2018; Flitcroft & Woods, 2018) and based on quantitative research guidelines (Choi, 1998; Cohen, Manion, & Morrison, 2007; Genaidy et al., 2007; Wallace & Wray, 2011). This framework supports the scrutiny of the study design; specifically the gathering, analysis and interpretation of data. The framework was adapted in the current study as one criterion (appropriate measurement instrumentation) was explicitly addressed by an inclusion / exclusion statement. The original and adapted frameworks are included in Appendix B and C respectively.

Ratings of 0, ½ or 1 were given for each of 13 criteria, with the quality recorded as high (9.5 -13), medium (4.5 -9) or low (0-4) for each study. The other two strands of evidence considered how appropriate the study's methodology suited the review question (in its aims, selection of participants and method of data collection) and also how relevant the focus of evidence was in answering the review question (specifically with regard to how mental representation of number had been operationalised, and how findings map onto developmental trajectories). Both these strands were rated high, medium or low quality, and were given equal weight towards the overall judgement of the quality of each study. As only the author of this review rated the studies, the outcomes of the quality assessment may be subject to researcher bias. Overall ratings are shown in Table 2, with full quality assessment details for strand A included in Appendix C, and weight of evidence strands B and C found in Appendix D.

Table 2

Weight of Evidence of Included Studies

	Weight of Evidence A Gathering, analysis and interpretation of data. 9.5-13 = High 4.5-9 = Medium 0-4 = Low	Weight of Evidence B Review specific appropriateness of method High Medium Low	Weight of Evidence C Review specific focus of study High Medium Low	Weight of Evidence D Overall extent to which the study contributes evidence to answering review question. Combining strands A, B and C.
Siegler & Booth (2004)	10.5 High	High	High	High
Booth & Siegler (2008)	9 Medium	High	High	High
Siegler & Ramani (2009)	9.5 High	High	Medium	High
Kucian et al. (2011)	11.5 High	Medium	Medium	Medium
Mazzocco et al. (2011)	12 High	High	High	High
Gunderson et al. (2012)	10.5 High	Medium	Medium	Medium
Obersteiner et al. (2013)	10.5 High	High	High	High
Link, Nuerk & Moeller (2014)	9 Medium	High	High	High
Sella et al. (2016)	11 High	Medium	Medium	Medium
Aulet & Lourenco (2018)	12 High	High	High	High

1.3 Results

The 10 studies under review span a time period from 2004 until 2018. Four focused on correlations and predictive factors while six looked at aspects of group difference alongside correlations. Five delivered some kind of training to promote better internal representation of number, with each using a design that allowed for a control group. These representations were primarily measured by the bounded number line task, with only one additionally using the unbounded version. Magnitude comparison tasks using various SNAs were employed in three of the studies and all 10 used various mathematical measures that had an arithmetic element, some more standardised than others. Participants ranged from 4 ½ to 10 years old. Although these factors offer alternate ways to group and review the studies, it was felt a chronological approach was best suited, as the studies build upon each other's findings, with any questions raised being addressed by subsequent work. Overall the studies were of a good quality, with seven achieving a weighting of high and three of medium (WoE D). Nine were judged to have high methodological quality (WoE A), seven to have a method highly appropriate to the review question (WoE B), and six found to have high relevance to the focus of the review question (WoE C).Those that did not achieve an overall rating of 'high' generally explored the review question indirectly, but subsidiary findings were found to be helpful and methodically sound. As each study was evaluated the overall weight of evidence was kept in mind so that any WoE strand not achieving 'high' could be addressed and impact assessed.

Three of the studies included within this review were collaborations between Robert Siegler and his colleagues, whose interest in mental number representations has generated a large body of research on magnitude knowledge. His research suggests the mental number line should be viewed as a common core of numerical development, and this idea is formalized in Siegler's Integrated Theory of Numerical Development, which details developmental progress of the mental number line (Siegler, 2016). Developing from the innate 'kernel' of representations such as the ANS, Siegler suggests more formal symbolic coding of magnitude maps onto this discriminatory system through increasingly accurate left to right placement of small whole numbers, larger numbers, fractions, decimals and, from right to left, negative numbers. He suggests automated access to this structure forms the basis of competency in most aspects of mathematics, including arithmetic, and therefore has an important role in maths achievement. Although these ideas have been discussed during the introduction to this review, it is helpful to revisit them here, as they form the emerging strands of a theoretical framework within which his studies are positioned. This is particularly relevant in terms of Siegler's view of the bounded NLE task as a valid, robust measure of the mental number line construct, with the unbounded version not developed by Cohen and Blanc-Goldhammer until 2011.

The earliest study from Siegler and Booth (2004) combined two experiments, only the first of which is relevant here. The researchers looked at associations between estimations on the bounded number task (accuracy and linearity) and a standardised maths achievement test score, analysing the results of correlations for three different year groups of children (kindergarten, first and second grade). As hypothesized, year on year accuracy error (PAE) reduced whilst linearity (R^2_{LIN}) improved. Additionally within each year group, PAE predicted maths achievement levels and the higher the test scores, the more linear an individual child's estimates were found to be (refer to Table 3 for effect sizes of mental number line training on maths achievement across all studies). The researchers suggest these improvements indicate a specific developmental change in understanding and representing number magnitude, i.e. a log – linear shift. They do not conclude that this change occurs regardless of mathematical experience, rather they point to the types of mathematical models that support the change and begin to hint at a casual element for maths achievement in developing a secure linear representation of number. The researchers do not go on to consider the influence of any domain general skills may have contributed to performance on the bounded NLE task. This is understandable as questions around the bounded number line as a valid measure post-date this study.

Following up this hint at causality, Booth and Siegler (2008) structured a RCT where computer generated, supportive external models of the internal number line were available to children as they tried to 'learn' the answers to simple arithmetic problems (in the form of coloured bars and circled numbers that visually illustrated calculations along a left to right line). The support ranged from a complete visual model with verbal support, through a less supportive process, down to the control condition where children received least stimuli. The hypothesis focused on the use of visual external representations acting as a mirror of the mental number line, with explicit taught strategies rehearsing how to access the framework during problem solving expected to lead to greater calculation proficiency. Again, using the bounded number line estimation task, PAE and R^2_{LIN} were measures of accuracy and linearity, confirming that children's estimates were positively correlated with existing addition knowledge and predictive of how well they could tackle novel calculations. Looking at between group differences, the researchers found accurate, full visual representations enhanced children's ability to solve novel problems beyond the level of the group who did not receive supportive visual modelling. With the researchers relying on the bounded number line task as an accurate measure of an internal representation, they concluded that mental representation of number magnitude is casually related to arithmetic learning. This conclusion may be questionable, as the nature of a 'novel' arithmetic problem within this study described an addition calculation that was inaccurate at pre-test but correct after it had been specifically trained during the intervention.

Table 3

Correlations and Group Effect Sizes of Mental Number Line Skill on Maths Achievement.

	Correlation / regression between mental number line task performance (bounded)and maths achievement (where multiple achievement measures, the addition component has been reported)	Effect size of training mental number line performance on group maths achievement
Seigler and Booth (2004)	Kindergarten *pr*= -.45 *p*<.05 First grade *pr*= -.66 *p*<.01 Second grade *pr*= -.37 *p*<.05	
Mazzocco (2011)	r^2= .28 *p*=.030	
Gunderson et al. (2012)	*pr* = .66 *p*<.001	
Link, Nuerk & Moeller (2014)	*pr* = -.36 *p*<.05	
Aulet & Lourenco (2018)	*pr* = -.29 *p*= 0.041	
Booth and Siegler (2008)		d= 1.51 *p*<0.01
Seigler and Ramani (2009)		d= 0.75 *p*<0.01
Kucian et al. (2011)		d = 0.65 *p*<.01
Obersteiner et al. (2013)		d = 0.11 *p*=.533
Sella et al. (2016)		d = 1.62 *p*=0 .002

With Siegler's (2016) suggestion that the mental number line is a core operating system for managing number, it is unsurprising that he continued to explore how children acquire such a linear representations of number magnitude, and any accompanying influence on maths outcomes. As the mental number line is made up of symbolic numerals acquired from conventions of the mathematical environment (via mapping onto more innate competencies), Siegler and Ramani (2009) considered experiential opportunities that could benefit children's mathematical progress. The researchers looked specifically at board games, and planned an experiment with 88 pre-schoolers (M = 4y 8m) from disadvantaged backgrounds who were reported to have had less opportunity to play these types of games than some of their more affluent peers. They hypothesized that exposure and familiarization with linear board games (but not circular ones) would numerical knowledge and facilitate the acquisition of new numerical knowledge through referencing an improved internal number magnitude representation. This was similar to the ‘mirror’ idea from the previous study; that exposure to a replica template would encourage mathematical gains. Specifically in this study the ‘counting on’ strategy was modelled, with a linear format offering left to right space-number mapping practise that was not available in a circular format. The games involved counting on in a straight line after a throw of the die. So if a counter was at square 6 and the player rolled a 3, they would be encouraged to say “7, 8, 9” as the counter was moved forward. Random assignment to play board games on a linear or circular board was controlled with a third group of children who only counted number and objects, and identified numerals. Researchers found children who played the linear games showed significantly greater improvement than the children in the other groups, for linearity of estimates, for number magnitude comparisons, and also for children’s performance on the arithmetic tasks in the study. As well as looking at results for the main hypotheses the researchers also explored the errors made by each group, finding that more of the inaccurate answers from the linear group became ‘near misses’ after playing with a linear board. They interpreted these results as an indication that the retrieval structure for numerical information had been enhanced, with encoding of magnitude becoming more precise and retrieval more fluent. They conclude by reiterating their findings that the mental number line is a valuable framework that supports achievement in mathematics.

Following on from Siegler and Ramani (2009), Kucian et al. (2011) identified that children experiencing difficulties with calculation were a helpful group within which to explore how the mental number line can support mathematics achievement. Number line estimates of children with dyscalculia are less accurate and less linear (numerals not evenly spaced along a scale, but compacted at the higher end) than their more typical counterparts (Piazza et al., 2011). The researchers suggested this reflects poor mental representations of symbolic number which in turn exerts a negative effect on arithmetic competence. Although the study focused on measures of brain activity within specific regions, and possible plasticity in response to training, other measures more relevant to this review are also included such as mathematical performance and number line estimations. The 36 boys and girls (*M* = 9.6; 22 diagnosed with dyscalculia) all received the same training (15 mins a day for 25 days), at home via a computer game ‘Rescue Calcularis’ where the player needs to steer a rocket to the exact point in a number line that matches with a symbolic or non-symbolic stimulus. The game provided an external model of the internal left to right number line, with rehearsal of number points alongside corrective feedback, and was hypothesised to improve accuracy and linearity on number line tasks (reduce PAE, increase R^2_{LIN}), and promote mathematical gains. Some of these improvements were found, with a notable result showing children with dyscalculia ‘catching up’ with the control group in the linearity of their estimations. Although training significantly improved both groups percentage of correctly solved arithmetic problems, the researchers acknowledged that they could not assume that this was through the mechanism of an improved mental representation of number, as practise and rehearsal may also have been influential factors. Additionally, practise was carried out at home which raises parental involvement as a possible influential factor on any findings.

Mazzocco, Feigenson, and Halberda’s (2011) research focused on the approximate number system and considered whether pre-schoolers’ skills in approximating non-symbolic number could predict future maths performance. They carried out a 2 year longitudinal study beginning when the children were 3 to 4 years old which offers insight into the bridge between non-symbolic representations and any that develop later, with symbolic numerals and structures acquired from the environment overlaying the innate systems. The researchers found ANS precision at preschool age significantly accounted for 28% of the variance in maths achievement scores at age 6, which rose to 35% when four children, whose scores were not greater than chance on the ANS, were removed from the analysis. This seems to suggest that even very early non-symbolic representations of number magnitude play a role in later maths achievement.

Gunderson, Ramirez, Beilock and Levine’s (2012) were interested in identifying the proximal link between mental representation of number and maths achievement and focused their research around the role of spatial skill. Their research question focused on how differences in spatial ability influences how quickly children develop a linear mental representation of numerals, which in turn supports performance on symbolic numerical tasks that are not explicitly spatial. This route to maths achievement was tested longitudinally, with two specific hypotheses. Firstly, that children’s spatial skill would predict future performance on number line knowledge (in the form of the bounded task), and secondly that these same skills would also predict achievement within a maths task that was not obviously spatial (but mediated by spatial ability in the form of an improved mental number line). The researchers chose a mental rotation task as indicative of general spatial skill, as it had been shown to be related to a variety of spatial tasks. Analysis of the data from 42 children aged 7 showed those with better mental rotation skills displayed greater gains in number line estimation over time than children whose skills were poor. Also those same skills were a significant predictor of later number line performance after controlling for previous estimation scores and prior achievement within mathematics. During a second study Gunderson et al. (2012) followed children between the ages of 5 and 8 years, tracking their performance on visuo-spatial skills in the form of a task that required them to match a shape with its two components (essentially a task requiring mental rotation and transformation). Here the results showed accuracy in approximate symbolic calculations at age 8 could be predicted from spatial skills at age 5, and that this was mediated by number line knowledge at age 6. This indicates mental rotation might have a unique role in the development of the mental number line as a spatial-numerical mapping that supports calculation. In addition, Gunderson et al., found no comparable association when approximate non-symbolic task results were analysed. The researchers suggest this discrepancy confirmed that the mental number line is a representation of numerals rather than non-symbolic magnitudes, and as such is important to numerical development and achievement. They also reflected on the approximate nature of the calculation task (i.e. naming who had the most, rather than giving an exact total), suggesting exploration with exact calculations could be the next step for further study. It is important to note that this study used only 6 estimation items to secure the bounded NLE task data at age 8 years, which is the fewest amongst the six studies that used NLE task measures (most others using 20-25 items). This number seems particularly low to support the validity of any findings.

Obersteiner, Reiss and Ufer (2013) base their research on the two very early number magnitude systems (observed even in infants) which respectively allow for discrimination of very small exact quantities and larger approximate ones. The focus on these two systems is unique within this review, as the other studies discuss differences in terms of symbolic and non-symbolic factors. The researchers suggest both the exact and approximate systems may have internal representations that are separate, and thus should be trained at an individual level, rather than to improve a combined representation, such as the mental number line. The study used a random controlled design with combinations of exact and approximate training delivered via an adapted version of a computer game, ‘The Number Race’. The 147 children aged 6 and 7 (M = 6.91 years), were shown a screen where two quantities of diamonds were shown (verbally, pictorially or numerically) and then had to compare the represented magnitudes, either through an approximate comparison, or through an exact calculation. They used this number to track through a number line to the finish. The researchers were able to show that participant's exact or approximate numerical processing skills improved only in the area in which they had been trained. There was no crossover effect which the researchers interpreted as an indication that both types of instructional approach are necessary for optimum and flexible numerical understanding. They also found that the effect of number magnitude training on arithmetical performance was quite small ($\eta^2 = 0.003$).

The research from Link, Nuerk and Moeller (2014) explored whether an accurate linear mental number line is the basis for numerical development by studying the tools that measure the internal construct. In response to Cohen and Blanc-Goldhammer's (2011) suggestion that accurate bounded number line estimations rely on proportional judgement and place value skill rather than pure integer knowledge, they directly compared it with an alternative; the unbounded NLE task. With no initial or final anchor points, the 'whole' is not easily seen, which limits a visual chunking of the line, requiring understanding of magnitude to be used instead (e.g., 500 is 50 units to the right of zero rather than 50 is halfway between 0 and 100). Their findings, with regard to associations between maths competencies and the bounded and unbounded NLE tasks, confirmed previous findings suggesting a high correlation between bounded estimation accuracy and arithmetic skill (Booth & Siegler, 2006; Geary, 2011; Fazio, Bailey, Thompson, & Siegler, 2014). It surprised the researchers that this association was not evident for the unbounded task, as this was supposed to be a purer measure of integer magnitude understanding. Additionally the results confirmed that proportional strategies were used in the bounded, but not the unbounded task (i.e. error variability decreased around reference points such as start, midpoint and end for the bounded task). The researchers drew two important conclusions from these findings. First, it is unlikely that the unbounded and bounded number line estimation tasks assess the same underlying construct (representation of number magnitude) as the reliability of both tasks was good. Second, as the unbounded task was not confounded by proportional techniques, it is a more accurate measure of the internal spatial-numerical representation. Therefore, the researchers frame a final conclusion, that if arithmetic competency is driven by a superior mental number line, it should be the purer measure of the mental number line (unbounded task) that more positively correlates with maths achievement. As it does not, it cannot be inferred that the mental number line is requisite for positive maths outcomes.

Like the Obersteiner, Reiss and Ufer (2013) study, Sella, Tressold, Lacangeli and Zorzi's (2016) research utilizes 'The Number Race', a game which allows pre-schoolers to engage in interactive number comparisons in non-symbolic and symbolic form. Using a randomized controlled trial the researchers hypothesized that the basic numerical skills of children in the training group would be enhanced, specifically that improvements in number comparison and arithmetic would exceed any improvements in the control group. They suggested any gains would be as a result of the reinforcement to the mental number line that the rehearsal of number magnitude positioning within the game allowed, and used the bounded number line estimation task as an appropriate measure of this effect. The findings were in line with the hypotheses, as the training group demonstrated significant large reductions in the percentage absolute error made on the number line tasks when compared to the control group. There was also a significant large enhancement of basic mental calculation skills for the intervention group over the control. During the discussion, the researchers reflected on the positive effect of computer game play that models the mental number line as a problem solving framework and how this practise may have driven the gains. Unfortunately they did not address any implications of using the bounded NLE task in preference to the unbounded task, such as the possibility that video game play may have enhanced visuo-spatial skills that support proportional judgements, which in turn benefitted number line estimations.

Aulet and Lourenco's (2018) study conjectured that the directionality of the number line could underpin 'operational momentum' which is the spatial-directional biases associated with moving towards the right and left for addition and subtraction respectively. Using two SNA tasks to explore this idea (one symbolic using numerals, and one non-symbolic comparing dot-arrays); they hypothesised that if both tasks relied on a directional mental number line, then they should correlate with each other. Additionally, they explored any correlation with maths achievement, assessing the 5 to 7 year olds skills in addition and subtraction using three tasks; a symbolic exact task (numeric calculations), a symbolic approximate task (verbal presentation of addition / subtraction problems requiring a comparison to identify 'more') and a non-symbolic approximate task delivered across two modalities (sight and sound). The researchers noted a significant correlation between performances on the two SNA tasks which they took as evidence for a robust left to right mental number line upon which children drew during the tasks. The other main finding was a significant negative correlation between the magnitude comparison task (the non-symbolic SNA) and the approximate cross-modal arithmetic task (ACA task), which is a task using visual dots and audible tones to represent individual units within simple addition and subtraction problems. It also uses concealed images to encourage the children to reference their own representation of number before their answers are visually revealed. This task allowed the researchers to support calculation across modalities, increasing the children's access to each test item. The researchers concluded that in light of the body of evidence that positively links the mental number line to maths achievement, it was unlikely that their finding of a negative correlation (i.e. those with stronger magnitude comparison skills performed worse on the ACA maths task, after controlling for age and general cognitive factors) implied the mental number line impeded maths performance. Looking for another explanation, they surmised that the ACA task engaged the children to such an extent that they found it difficult to inhibit a response until the full item was revealed, thus leading to inaccurate responses and a negative correlation. Although this study added evidence supporting links between spatial-numerical associations (SNAs) and a left to right directional mental number line, it did not demonstrate any positive relationship between this same line and early maths proficiency.

1.4 Discussion

1.4.1 Summary of Findings and Conclusion

The ten studies included within this review explored the relationship between children's mental representation of number and maths achievement. At the heart of the research is a key question about how children learn to encode number in a symbolic framework within which they can successfully operate mathematically. The studies build on the premise that innate non-symbolic SNAs intuitively support the development of a symbolic mental representation of number that is spatial i.e. a schematic representation where position in space is an essential element of how number magnitude is conceptualized. Each study then explores how these developing symbolic mental frameworks may be acquired from the environment, either through interpretation of significant correlations between associated factors, or through specific training that may enhance the development of mental representations. All consider the relevance of their findings with respect to wider mathematical outcomes for the individual.

The quality of any internal frameworks that develop are difficult to evaluate and this review has highlighted the continuing reliance on, and preference for, the bounded NLE task as the most relevant measure, despite some early indications that the unbounded NLE task may be more precise. The unbounded task has only been referenced once amongst the six studies that post-date it (Link, Nuerk and Moeller, 2014) and this may be problematic for any conclusions this review draws, as any evidence base may be confounded by the use of a measure that has suspected reduced validity. Three of the studies avoid this difficulty by using measurement tasks that purport to assess the ANS and exact number systems (e.g. non-symbolic magnitude comparisons and subitizing). However, this too might be considered problematic in light of research that suggests these early innate non-symbolic systems are subsumed by later symbolic ones, with correlation coefficients between performance on these types of tasks and maths outcomes decreasing with age (presumably as they have limited application for calculations involving large numbers; Cipora & Nuerk, 2013). Despite these concerns, the majority of studies refer to the large body of evidence detailing the bounded number line as robust and therefore the reviewer remained mindful of this confidence when making quality assurance decisions.

Overall the findings of this review reveal that the relationship between a child's mental representation of number and their maths achievement is an important one, particularly in the early years as they begin to transfer their understanding of number magnitude from non-symbolic to symbolic representations. Specifically three key findings will be discussed; the mental number line as a linear framework for early understanding of number; approaches that successfully support the acquisition of a linear representation of number; and aspects of mathematics influenced by mental representation of number including the size of any effect.

The studies' participants were aged from 4.5 to 10 years old with children across this age range demonstrating improved performance on the mental number line (PAE, R^2_{LIN} or both) in all of the five studies that offered training to promote mathematical achievement via gains in mental representation of number. These improvements in accurate representation of integer magnitude (e.g. 10 is 10 units to the right), including the gradual developmental move from a logarithmic pattern to a linear one, were demonstrated across age and maths abilities. Collectively these findings suggest a left to right number line is a common internal framework within which children represent number magnitude. Results from Kucian et al. (2011) and Siegler and Ramani (2009) support this conclusion, recruiting groups that had less established mental number representations and deficits in number knowledge so they could directly explore effects of exposure to external models that were linear. Using trainings that rehearsed a left to right linear representation of number, the researchers were able to show how children's mental models improved alongside mathematical gains. The researchers also suggested that the rapidity with which improved representations were acquired suggested linear symbolic representations were a natural mapping from innate spatial-numerical left to right associations, and more helpful than, for example, circular representations. Gunderson et al. (2012) specifically states that as only symbolic maths task performance at age 8 could be predicted by number line estimates at age 6 (with no such association with non-symbolic task performance) the mental number line is likely to be a representation of numerals. This ideas is supported by Siegler and Booth (2004), Booth and Siegler (2008) and Siegler and Ramani (2009), who suggest placement of integers comes first, followed by fractions, decimals and then negative numbers when the child is developmentally ready to represent them along the same line.

All five studies that delivered training used approaches that were direct models of mental representation of number in the form of a left to right number line. Gains in performance on the number line estimation task were noted in every group that received an intervention, with four of the five producing small to medium significant effect sizes over the control groups. Four trainings were computer based, while the approach that generated the largest effect on estimation accuracy and / or linearity was the linear board games devised by Siegler and Ramani (2009). Training was particularly helpful for groups who may have experienced difficulties acquiring a mental representation of number, as demonstrated by the children with dyscalculia whose linearity performance on the bounded NLE task caught up with the control group without dyscalculia (Kucian et al., 2011). Increased rates of progress were not seen in their maths achievement tests but the researchers point to the short study time which if lengthened, may have had a larger effect.

The final key finding focuses on aspects of mathematics influenced by mental representation of number and the size of any effect. Although the inclusion criteria for this review did not limit mathematical achievement to only one measure, arithmetic was identified as a necessary element for mathematical competency, as calculation skills have been shown to significantly contribute to maths achievement outcomes (Cowan et al., 2011). This is not to suggest that representations of number magnitude cannot be supportive of other competencies and indeed some of the studies included in this review focused on skills such as counting and numeral identification, alongside arithmetic.

Looking first at associations, the majority of the studies found medium to strong correlations between arithmetic tasks and PAE or R^2_{LIN}. However not all studies found these associations, with Link et al. (2016) finding no correlation between the unbounded NLE task and arithmetical competencies and Gunderson et al. finding bounded NLE tasks only correlated with symbolic arithmetic tasks and not non-symbolic ones. Both of these results are interesting as they seem to counter the overall trend in the findings showing accurate number line estimates positively correlate with performance on arithmetic. The lack of a positive correlation between the unbounded NLE task estimates and arithmetic scores is interpreted by Link et al. (2014) as evidence that the mental number line is not necessary for arithmetic competency. This is based on the premise that the unbounded task is a purer measure and should therefore show a stronger correlation. However, this interpretation seems a little ‘all or nothing’ and is not followed up by any nuanced discussion around limitations (e.g. the sole use of PAE as a measure of accurate estimates, without the inclusion of R^2_{LIN} analysis). The discrepancy Gunderson found (symbolic arithmetic tasks positively correlating with estimation accuracy whilst non-symbolic did not) is easier to explain. The finding supports rather than challenges the notion of an association between the mental number line and maths achievement if it is accepted that the mental number line itself is a symbolic representation where numerals and space combine to reflect magnitude relationships. Two studies looked at the predictive power of mental representations of number, with Mazzocco et al. (2011) finding preschool precision on the approximate number system significantly predicted maths performance (r^2=.278) and Gunderson et al. (2012) finding children’s number line knowledge at 6 predicted their performance on an approximate symbolic calculation at age 8.

Looking at group differences in response to training (over the control group) only two studies out of the six found a significant effect on maths tasks with an arithmetic element. The effect sizes were large (Hedges g_c=1.58 and Cohens d= 1.51) and were accompanied by significant medium effects on number line estimation performance. Looking at these two effects in detail, the first was obtained from 45 pre-schoolers’ performance on six single digit addition and subtraction problems pre and post training via the intervention of the computer game ‘The Number Race’ (Sella et al., 2016). The second was from 105 seven year olds who received direct training on the four of 13 single and double digit addition problems that had been answered least well at pre-test. Their training was also computer based and used visual coloured bars to rehearse the use of the mental number line as a supportive tool for calculation (Booth & Siegler, 2008). Although these effect sizes are discussed in the context of being larger than produced by the other studies, it may be helpful here to keep in mind Simpson’s (2017) observation that any comparison of effect sizes should be made with due caution, as they are open to researcher manipulation at the time the study was designed, carried out or during interpretation of findings.

Even though there is evidence of positive correlations between mental representation of number and maths achievement across age, and support for training approaches that promote the acquisition of an accurate mental number line, the findings within this review do not add conclusive evidence to the idea that mental representation of number is a cornerstone of more advanced mathematical operations. However, taken together these findings do appear to show that the influence of mental representation of number on maths achievement is evidenced at an early stage of single digit integer operation, and mainly through addition, with a few researchers also including evidence for links with subtraction and multiplication (e.g. Kucian et al., 2011). The findings also show that groups who were less likely to develop an early secure linear representation of symbolic magnitude, and who appear to experience associated difficulties in making expected progress within arithmetic, can benefit from a rehearsal type training model that closely replicates a linear representation of number in the form of a mental number line.

1.4.2 Strengths and Limitations

It has been a strength of this systematic literature review that the search yielded studies of a reasonable quality. All were found to have a medium to high overall rating, with none being judged as low. This, in part, was due to the robust process of identifying and quality assuring the articles. This review was not limited to work published in journals, and a variety of dissertations and other grey literature were screened for suitability. In addition, the selected articles were international and representative of research taking place in a variety of cultural environments. Finally, the review brings together studies from different fields of research, including contributions from cognitive psychology, education, and researchers with an interest in neuroscience.

To operationalise the review question it was necessary to limit factors which may, on reflection, have yielded helpful information. The key limiting factor has been the focus on integers, with the exclusion of fractions and negative numbers possibly skewing the findings on the impact of mental representation of number on maths achievement. This also prevented a more longitudinal overview of the development of mental representation, where new numerical information is assimilated within a current framework over time. Additionally the choice of search terms may have created a bias towards the mental number line as the main type of representation, as it has been noticed during the course of the review that some types of internal representation of number were not identified in the initial search (e.g.in the form of an abacus; Frank & Barner, 2012).

Aside from the limitations of the review process itself, consideration must also be given to the limitations arising from an attempted synthesis of a group of studies that operationalise their research through diverse methodology. The question around the validity of the bounded NLE task has already been discussed, with the reviewer concluding that there is enough evidence to accept the view that the task does (to some extent) measure what it sets out to measure; the precision of the mental number line. However, this judgement may be open to question as further research explores the claim that the unbounded task might be a purer measure.

Similarly, operationalising the construct of mathematical achievement might be considered a limitation, as different researchers have used contrasting measures. Six of the studies used a standardised test, or sub-test, to determine maths competency, none of which were the same due to the different ages of participants and varying locations. These types of test were mostly used for correlation purposes and considered less sensitive for measuring any mathematical gains after a short training period. Instead, the training studies focused mostly on any benefits to calculation skills, with measures created by the researchers. For example Booth and Siegler’s (2008) study found a significant effect size over the control group for maths improvement after mental number line training, yet the post intervention maths assessment consisted of single digit addition items that had been used during training. This type of measure can be contrasted with Obersteiner, Reiss and Ufer (2013) who used 16 novel addition and subtraction items to establish any gain. These discrepancies may have limited the reviewer’s attempt to compare effects and also made it difficult to establish a consensus regarding the type of new learning that illustrates mathematical progress.

1.5 Recommendations

1.5.1 Implications for Educational Psychologists

This systematic review has identified helpful areas of focus for Educational Psychologists who are working to support school staff help children overcome barriers to acquiring skills for numeracy. Sharing understanding about early, more innate number magnitude skills might be a useful starting point for discussion, as a hypothesis woven through the findings from this review suggests difficulties arise when symbolic information around number is not carefully mapped onto this early framework. Discussions that identify a child’s skill with non-symbolic number may give insight into misconceptions that have developed around symbolic number, which, if left unresolved, may be problematic for more advanced arithmetic.

Additionally this review has confirmed that the left to right mental number line, as an accepted way of representing number magnitude in the mind, is a useful model that helps children encode, manipulate and recall information. Even if the mental number line is only borrowed from an external model acquired from the environment (i.e. not innate or mapped onto early intuitive spatial-numerical associations) the current findings have shown it to be a helpful tool within which to perform arithmetic, with children making some progress when they use it. It is also a model which provides a schema that can expand to encompass further mathematical number magnitude concepts, such as fractions and negative numbers. Educational Psychologists can promote this model as a useful tool, where rehearsal of placing numbers in a linear format aids understanding of the relationship between numbers (linearity) and place value.

The findings will also allow Educational Psychologists to draw on an evidence base to recommend the types of interventions that support children acquire an accurate mental representation of number that can contribute to achievement in mathematics. Computer games appeared to be effective in modelling the number line, but a stand out intervention is the use of a linear board game which promoted an improved mental number line and significant gains within addition. Recommending time playing with games such as Snakes and Ladders, and other homemade games in a linear format, is a cost effective intervention. As a multisensory experience it also links to Kosslyn and Pomerantz's (1977) emphasis on multiple senses facilitating an enriched encoding process as it involves kinaesthetic (moving the counter along the line from left to right), auditory (counting on) and visual elements.

In summary this systematic review points to six clear implications for Educational Psychologists.

- When exploring children's difficulties with maths it is important to backtrack to assess early non-symbolic number magnitude skills as these influence later symbolic performance.
- As misconceptions about number are likely to stem from a confused representation of number magnitude, it will be helpful to use concrete materials to allow pupils to demonstrate any internal frameworks they have

created. Having the opportunity to externalise internal understanding will give insight into aspects of number magnitude that may need to be revisited.

- It is important for Educational Psychologists to work from the premise that the left to right mental number line is a helpful way of representing number magnitude in the mind, and should be promoted as a useful model that helps children encode, manipulate and recall information.
- The left to right mental number line is a representation that can be acquired by direct modelling and is particularly helpful for children between 5 and 8 years who have not had experiences that would help develop such a model e.g. by playing linear board games.
- Interventions that are particularly effective in supporting the development of a mental number line that can contribute to achievement in mathematics include games that explicitly model and rehearse number magnitude placement. Both computer representations and board games are effective in achieving these gains, with linear board games such as Snakes and Ladders proving most cost effective. It is important to encourage children to count on when taking a turn, e.g. if a 2 is rolled and the counter is on 16, the child should verbalise "17, 18"
- Educational Psychologists should keep in mind that evidence in this review suggests the mental number line as a schema that may expand to encompass further mathematical number magnitude concepts, such as fractions and negative numbers. This may prove useful when supporting older pupils experiencing difficulty with more advanced number concepts.

Despite the positive findings within this review, Educational Psychologists should also be mindful of the limitations, which point to these interventions being more suitable for younger children, or children in primary school who are struggling with number concepts. Although this review has not included participants who are over ten years old, the information here can help inform understanding of the developmental basis of number magnitude, and this should prove useful to Educational Psychologists when involvement is requested to address difficulties in numeracy.

1.5.2 Future Research

It is surprising that although six of the studies post-date the Cohen and Blanc-Goldhammer (2011) research suggesting the unbounded NLE task as a possible purer assessment of number magnitude, only one of the studies uses it as a measure. It seems important then that claims about this task are explored further, with future research into mental representations of number at least including both types of task so comparisons can be made.

Additionally, at the root of the discussion around the bounded and unbounded NLE tasks as measures of number magnitude representation is the idea that the bounded task uses visual proportional skills to determine accurate placement of estimates. This is interesting as there is a large body of research that shows visuo-spatial skills themselves are positively correlated with maths achievement (Alloway & Passolunghi, 2011; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Delgado, & Prieto, 2004; Rasmussen & Bisanz, 2005; Carlson, Rowe, & Curby, 2013; Mix & Cheng, 2012). Included within this review is an article exploring whether these spatial skills may be an influential link between mental representation of number and maths competence. Gunderson et al. (2012) explored the hypothesis that children's spatial skill would predict future performance on number line knowledge which in turn would predict calculation competence, and found that spatial skill likely benefits numerical knowledge by supporting the acquisition of a linear spatial representation of numbers. Therefore, it may be helpful to further explore these links to establish the extent to which spatial skill influences performance on the number line task. For example, a design exploring group differences, where spatial training is an intervention used to establish any gains in precision of participant's estimates might add insight into exactly what the bounded NLE task measures. The type of spatial training would have to be carefully considered, particularly as it has not been described as a unitary construct (Uttal et al., 2013). This type of research may also be a stepping stone that informs the wider picture, for the accuracy of Gunderson et al. (2012) assertion that spatial skills act on maths achievement via enhancing the mental number line may be questionable, particularly as she only used the bounded NLE task, which itself is suspected of being confounded by spatial skill. A spatial training intervention employing both NLE tasks might generate findings that further unpick the role of spatial skills so that researchers can more fully understand the relationship between children's number line estimations, mental representation of number magnitude, and maths achievement. In chapter 2 this idea is explored through a randomised controlled design using just such an intervention.

Chapter 2 The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

2.1 Introduction

Research on how children develop competency within mathematics has focused on early number skills, with ongoing investigation into what these skills look like, the extent to which they build on innate or acquired models of understanding, and how best they can be fostered during teaching and learning (e.g. Shanley, Clarke, Doabler, Kurtz-Nelson, & Fien, 2017). Awareness that maths achievement contributes to life chances of the individual is a key driver for research (Williams, Clemens, Oleinikova & Tarvin, 2003). In addition early maths competence is also set within a wider context where young people's mathematical performance is nationally graded (Ofsted, 2015), internationally ranked (OECD, 2014), and where employers voice concerns over mathematical abilities of school leavers (Education and Training Foundation, 2016).

Spatial skills predict mental representation of number magnitude and math achievement, and mental representation of number magnitude predicts math achievement. Chapter 1 considered the association between mental representation of number magnitude and math achievement. Chapter 2 considered the association between spatial skills and mental representation of number magnitude. In addition Chapter 2 considered all three of the concepts in the conceptual model and the relationships between them.

2.1.1 Domain Specific Number Magnitude Skills

Research suggests the earliest number skills are innate (Strauss & Curtis, 1981; van Loosbroek & Smitsman, 1990; Xu & Spelke, 2000). Butterworth (2005) used the term 'number sense' to identify this intuitive capacity for processing number. Very young infants have shown recognition of non-symbolic number magnitude, and an ability to visually discern between groups of objects of varying quantities (Strauss & Curtis, 1981; van Loosbroek & Smitsman, 1990; Xu & Spelke, 2000). It is thought that two cognitive systems underpin these skills, the parallel individuation system which is triggered to track individual items when there are fewer than 4, and the Approximate Number System (ANS) that perceptually estimates difference between larger quantities (Piazza, 2011; Gallistel & Gelman, 2000). These abilities seem to be domain specific to mathematics. Number magnitude then, appears to be a construct for which humans are primed and has a central place in symbolic maths schema, supporting understanding of counting, number relations, number operations and basic arithmetic.

2.1.2 Maths Competence and Domain General Factors

An alternative explanation for number competence highlights the role of domain general abilities, including individual differences in working memory, long term memory, processing speed, metacognitive capacity, phonological skills, and visuospatial processing (Krajewski & Schneider, 2009). Amongst these skills, visuospatial expertise has received much attention, being consistently linked with high achievement in maths (Alloway & Passolunghi, 2011; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Delgado, & Prieto, 2004; Rasmussen & Bisanz, 2005; Carlson, Rowe, & Curby, 2013; Mix & Cheng, 2012). These types of research evidence that spatial skills support not only visual tasks such as geometry, measurement and symmetry, but also benefit other areas of mathematics, including number tasks.

2.1.3 Number and Space

Domain specific early number magnitude skills have been shown to link with domain general visuo-spatial abilities as they are both involved in spatial-numerical associations (SNAs; Siegler & Opfer, 2003). SNAs are evidenced through behavioural phenomena present in infancy, childhood, and also in adults. One example is the Spatial-Numerical Association of Response Codes, or SNARC effect, where small numbers are more associated with the left hand side and larger numbers with the right (Dehaene, Bossini, & Giraux, 1993). As with ANS abilities, the SNAs are thought to be intuitive, and their presence has given rise to a widely held view that number magnitude is symbolized internally as a left to right mental number line (Dehaene, et al., 2004; de Hevia, & Spelke, 2009). In Siegler's (2016) Integrated Theory of Numerical Development, he surmises that the development of the mental number line occurs as symbolic information is mapped onto the 'innate kernel' of early intuitive capacities such as the ANS. This internal representation becomes increasingly accurate, with the placement of small whole numbers (as symbols for number magnitude) followed by larger numbers, fractions, and then negative numbers as the child becomes developmentally ready for each scale. This theory is in line with Case and Okamoto's (1996) view, that the mental number line is referenced as a spatial framework for problem solving and that automated access to this structure forms the basis of proficiency in most aspects of mathematics. However, not all researchers conclude that maths competency relies so heavily on such a hardwired internal representation, particularly when calculation involves numbers other than small integers (Schneider, Grabner, & Paetsch, 2009; Núñez, 2011). Instead they consider the influence of external spatial models often used within the classroom, such as left to right number lines supporting counting and ordering, and the Cartesian plane to symbolise direction and magnitude. Núñez (2011) suggests these types of models impact maths achievement not by enhancing a matching innate model that is already present, but through facilitating understanding through domain general mechanisms, such as building up conceptual mappings of taught information, activating imagination and providing memory hooks that facilitate information retrieval.

2.1.4 Number Line Estimations

By their nature, the form and quality of internal representations are difficult to describe and assess. The accepted measure that reflects the precision of the mental number line is an external representation called the number line estimation task (NLE task; Siegler & Opfer, 2003). The NLE task involves placing numbers on a line where the only stimuli are the initial and final numbers within a set scale along a left to right plane (the bounded NLE task). Researchers have consistently found that younger children place numbers in a logarithmic arrangement (large numbers compressed) while older children tend to produce a more linear placement with distances between numbers more equal along the line (Booth & Siegler, 2006; Opfer & DeVries, 2008). Siegler, Thompson, and Opfer (2009) suggested this 'log-linear shift' occurs more than once, with estimates becoming more linear as each scale becomes familiar (e.g. 1 to 100 will produce linear estimates before 1to1000). This shift has been interpreted as a marker for improved conceptual understanding of number magnitude, especially as improved accuracy of estimates (percentage absolute error; PAE) and linear performance (relationship between estimated position and actual position; R^2_{LOG}/R^2_{LIN}) has been shown to positively correlate with maths achievement (Booth & Siegler, 2006; Fazio, Bailey, Thompson, & Siegler, 2014; Geary, 2011; Booth & Siegler, 2008; Muldoon et al., 2013; van den Bos et al, 2015).

2.1.5 The Influence of Spatial Skills

With early links between space and number evidenced by SNAs and a proposed mental number line that combines spatial direction with magnitude, it seems reasonable to suggest that spatial skills are likely to impact on maths achievement. Research has consistently found that visuo-spatial skills correlate with and predict general maths performance (Alloway & Passolunghi, 2011; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Delgado, & Prieto, 2004; Rasmussen & Bisanz, 2005; Carlson, Rowe, & Curby, 2013; Mix & Cheng, 2012). However, visuo spatial skills are not described in the literature as a unitary construct. Linn and Peterson (1985) identified spatial visualization, spatial perception and mental rotation as three distinct elements of spatial skills, while Carroll (1993) distinguished only two; visualization and orientation. Through a largescale meta-analysis Uttal et al. (2013) suggested a more nuanced third model, a 2 x 2 framework, differentiating between intrinsic and extrinsic skills employed in either a dynamic or static context. 3D mental rotation skills, which involve both intrinsic (within one object,) and dynamic (moving) factors, have been identified as highly influential to maths achievement. Examples of this finding include research from Verdine et al. (2014) who reported intrinsic-dynamic spatial ability in 3 year olds predicted performance on the problem solving subtest of the Wechsler Individual Achievement Test (WIAT) at age 4. Similarly, Gilligan, Flouri and Farran's (2017) findings, based on data from over 12000 participants, reported intrinsic-dynamic spatial skills accounted for a significant proportion of the variance in children's mathematics achievement.

In trying to find out more about these associations, an initial focus of research was to establish the malleability of spatial skill, to find if training a specific spatial skill would improve it, with any gains sustained over time. Uttal et al. (2013) conducted a comprehensive review of 217 studies that had used spatial interventions to train children, adolescents or adults; concluding that spatial skills do benefit from direct and indirect training (e.g. direct practice with tasks that are spatial, or playing video games). Those that received the most benefit were those with poorer skills at the start, with the effect being moderately large and relatively persistent over time. These effects were not limited by age, gender or type of training, as demonstrated by de Lisi and Wolford (2002) who studied effects of video game training on children under 13; and Edd (2001) who assessed the impact on adults spatial skills when they were given increased opportunities to rotate and handle 3D models. In addition, improvements in certain spatial skills transferred to other non-trained spatial skills (e.g. the use of Logo, a small programmable floor robot that can follow directional instructions, improved children's performance on a mental rotation test; Eikenberry, 1988). This raises the question whether training a spatial skill can benefit a maths competence that is not overtly spatial, e.g. a calculation task.

Research exploring the impact spatial training exerts on a non-spatial tasks is limited, with initial studies focusing on adult participants. Hsi, Linn, and Bell (1997) found improvements in undergraduates' performance on tasks in their engineering course after participating in spatial training. Attendance at the intervention was voluntary, and this suggests those students benefitting may already have had a greater motivation to achieve than the 'control' group. Other similar studies using undergraduate participants have found similar effects (e.g. Sorby, 2009; Sorby, Casey, Veurink, & Dulaney, 2013). The findings from research with children began with Cheng and Mix (2014) reporting a significant, though small (d = 0.20), effect of spatial training on the basic arithmetical ability of 58 children aged 6 to 8 years. The intervention involved a single 40 minute session of rehearing mental rotation and translation tasks, and was found to have most effect on 'missing term' calculations. Despite this positive start, subsequent studies have found no such effect with Hawes, Moss, Caswell, and Poliszczuk (2015) finding mental rotation training did not impact on arithmetical ability of a group of primary school children. Similarly, Xu and LeFevre (2016) compared the effects of numerical sequential training with spatial training on 84 children aged 3 to 5 years old. After a single session of training identifying constituent parts of different shapes, it was reported that spatial training did not enhance performance on number ordering or accuracy of estimates on a bounded number line task, even though spatial skills were correlated with performance on these tasks. It may be that these are examples demonstrating that very brief spatial training is not adequate to promote transfer to a more general maths competence, or that the type of training used did not employ the spatial element that would be most effective in improving mathematical performance.

With limited evidence that improving spatial competence can improve performance on calculation, but with robust evidence that the two are correlated (Alloway & Passolunghi, 2011; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Delgado, & Prieto, 2004; Rasmussen & Bisanz, 2005; Carlson, Rowe, & Curby, 2013; Mix & Cheng, 2012), it is interesting to reflect further on the links between number and space, particularly with reference to the mental number line that is reported to integrate domain general and domain specific factors. SNAs suggest visuo-spatial skills are integral to the development of a left right linear representation of number magnitude, a domain specific cognitive construct, yet spatial skill has been identified as a domain general cognitive skill. With these two factors coming together in the mental number line, it is surprising that there is such limited evidence of their impact on each other. This also raises a question about the tool that measures the mental number line. If accuracy on the bounded NLE task employs general cognitive skills, including spatial ones, then the measure may not be assessing pure number magnitude understanding, a domain specific skill. Researchers exploring strategies used to complete the bounded NLE task have identified possible confounding factors, specifically the use of proportional reasoning skills, where the estimate is made by visually dividing up the line using the start, middle and end of the line as anchor points (e.g. Barth & Paladino, 2011; Rouder & Geary, 2014). The use of a strategy based on visuo-spatial skills questions the validity of the bounded NLE task as a measure of pure integer estimation and this is why the introduction of the unbounded NLE task may be helpful, as it removes the final anchor point in an attempt to prevent the 'whole' being seen and visual proportional judgements being made. However, with or without the anchor points, the full extent to which visuo-spatial skills exert an influence on performance is not yet clear, firstly because spatial factors are closely linked with the formation of the mental linear framework, and secondly because spatial skill is not thought to be a unitary construct.

2.1.6 The Role of Mental Rotation Skills

It seems then that spatial skills (particularly 3D rotation skills) may support maths competency across domain specific and domain general processes, although exploration into exactly how spatial skill acts on maths outcomes is ongoing. Hubbard, Piazza, Pinel and Delaene (2005) identified that neural networks for spatial and numerical processing partly overlap and this lends support to the idea that spatial-numerical links are important for maths competence. Gunderson, Spaepen, and Levine (2015) explored this relationship finding that children's spatial skill (2D mental rotation and transformation ability) predicted future performance on number line knowledge which in turn predicted calculation competence, suggesting that spatial skills act on maths achievement via enhancing the mental number line. Similarly, Thompson, Nuerk, Moeller and Cohen Kadosh (2013) found the 3D mental rotation skills of adult participants significantly influenced performance on a bounded NLE task. LeFevre et al. (2013) undertook a longitudinal study with over 500 children from 5 to 9 years to explore how visuo-spatial skills, number line estimation and mathematical achievement interact. Moderate, significant correlations were found between spatial ability and number line task performance, arithmetic, and number system knowledge. In addition spatial ability predicted growth in number line knowledge. However, it would have been interesting to see if these correlations remained if the unbounded NLE task had been used, yet only the bounded task was employed to assess performance. In addition there was no evidence to support Gunderson, Spaepen, and Levine’s (2015) finding that performance on the NLE task predicted growth in arithmetic knowledge. As the spatial measure was composed of different elements including visuo-spatial working memory, mental rotation, and analogical reasoning, LeFevre et al. (2013) suggests further research would benefit from more focus on individual components, to help identify the relevant relationships.

2.1.7 Rationale for this Research

There is much evidence that both the bounded NLE task and spatial skills correlate with achievement in mathematics. Currently much of the research studying links between spatial skills, performance on the bounded NLE task and maths achievement has focused on predictive relationships, with less exploration into any casual factors. In addition, there is some speculation, although with limited evidence, that spatial skills influence maths achievement via enhancing precision of the mental number line, as measured by improved performance on the bounded NLE task. Any influence on the mental number line is important, as it is a framework purported to support understanding of number magnitude, a foundational concept within mathematics. However, the extent to which spatial skill influences the mental number line is not clear, particularly as spatial skill is not a unitary construct. Evidence points to 3D rotation skills being the component of spatial skill most associated with maths achievement and therefore this skill is of interest in relation to any effect it may have on the mental number line and its associated metric, the bounded NLE task.

Building on findings in Chapter 1, where the association between mental representation of number magnitude and math achievement was evaluated, this chapter aims to further consider how spatial skills predict mental representation of number magnitude and math achievement, and mental representation of number magnitude predicts math achievement. Therefore Chapter 2 will focus on the association between spatial skills and mental representation of number magnitude and the relationship between all three of the concepts in the conceptual model.

This research will offer an opportunity to replicate some of the correlations previously found in this field of study. In addition, using a randomised controlled trial, this study will follow recent research that has investigated how the mental number line responds to training of 3D spatial skills, helping to contribute to the body of research exploring links between spatial ability, representation of number magnitude, and achievement in mathematics (Cheng & Mix, 2014; Hawes, Moss, Caswell, & Poliszczuk, 2015; Xu & LeFevre, 2016).

As this research is focused on any effect spatial training might have on the mental number line, it is important to respond to concerns that the bounded NLE task may not be the best measure of pure integer estimation. If, as suspected by some researchers, it relies on proportional reasoning underpinned by visuo-spatial skills, there is a possibility that findings may be confounded. To address this issue the unbounded NLE task will be used as an additional metric, alongside the bounded NLE task. This task is similar to the bounded task but has no anchor numerals that support the 'whole' being seen and proportional judgements being made.

2.1.8 Research Questions

This study addresses the following research questions:

- Are 3D spatial skills and performance on bounded and unbounded NLE tasks reliably correlated with maths achievement?
- Does training 3D spatial skills improve 3D spatial ability and are changes sustained over time?
- Does training 3D spatial skills improve precision of the mental number line, as measured by bounded and unbounded NLE tasks?

2.2 Method

2.2.1 Design

Using a mixed design with random assignment, quantitative data from the intervention and control groups was gathered over three time intervals. The intervention was 3 weeks long, with 10 school days of intervention followed by a rest week to establish any persistent effect. This duration was chosen for practical reasons to fit with school timetables, but also to reflect the typical parameters noted for interventions in Chapter 1(e.g Obersteiner, Reiss and Ufer, 2013) where the average was approximately twelve 15-20 minute sessions, most often delivered daily. The between participants variable was a computer based spatial intervention delivered across training sessions, each of 15 minutes duration. The two dependent variables were 3D spatial scores, and precision of estimations across the bounded and unbounded NLE, with measures taken pre, post and one week after training had finished. Additionally, maths achievement scores were collected for correlation purposes, with the intention of replicating associations found by previous researchers.

2.2.2 Participants

To establish the optimum number of participants for this study a power analysis (where 1- Beta = 0.95, alpha level = 0.05 and $\eta^2 = 0 .14$) was conducted using G*Power (Faul et al. 2007). This analysis suggested a sample size of 82 participants. The effect size ($\eta^2 = 0 .14$, $p = 0.005$) for this calculation was taken from Cheng and Mix's (2014) research as it was assessed to be the most relevant to the proposed study. It used a mixed design with two randomly assigned groups of junior school aged children participating in spatial training or control activities, to establish any effect on maths performance. Cheng and Mix (2014) collected data across three time points using an analysis of variance (ANOVA) for analysis whilst controlling for prior maths achievement.

Children over 8 years have been shown to achieve reasonable accuracy and linearity when making estimations on a bounded 0 to100 number line (Siegler & Booth, 2004) but find the unbounded task more difficult (Kim & Opfer, 2017). As both were to be used in this study, Year 6 pupils were targeted as appropriate participants as they are over 8 years and because they also take externally moderated mathematics achievement tests which are standardised across the national cohort.

In line with the power analysis described, and after obtaining all necessary ethical approval via Southampton University, an initial 98 participants aged 10 and 11 years were recruited from 6 state primary schools within Hampshire. Recruitment was via an email approach to 18 schools, followed by personal phone calls to each head teacher. As one of the schools (21 children) were unable to schedule full training sessions they agreed to be a pilot, with initial data and procedures being evaluated before commencing the full study. The remaining 77 participants ($M = 10.5$ years) consisted of 34 boys and 43 girls of mixed ability, with 37 in the intervention group and 40 in the control condition. Data from all 77 participants gathered during the initial visit was used for correlation analysis. However, six participants were excluded from group analyses due to missing data (six or more test items on one of the NLE tasks not completed) or absence (three or more training sessions missed). In addition, the data from one participant repeatedly appeared as an extreme outlier during analyses. Further investigation found this participant to have severe special educational needs and so this data was only included for correlation purposes.

2.2.3 Materials

2.2.3.1 Measure of 3D Mental Rotation Skill

Skill in 3D mental rotation was measured using a computer based assessment tool adapted by Bokhove and Redhead (2017) which Ganis and Kievit's (2015) had redesigned from a mental rotation skill task originally created and standardised by Shepard and Metzler (1971). Two static 2D representations of 3D cube formations were viewed on screen which prompted mental rotation strategies to establish if the two were the 'same' or 'different' (see Figure 2). In total 24 pairs were offered sequentially, and in different orders for the pre, post and subsequent testing. Bokhove and Redhead's (2017) validation of this tool was found to be largely in line with Ganis and Kievit's (2015) results, with a similar linear relationship of response time and error rate. They also found that judgements on 'different' pairs were made slower (*M* =13.47) than 'same' pairs, and with a much higher error rate (21.2% versus 7.56%).

2.2.3.2 3D Mental Rotation Skill Training Tool

In addition to providing a metric for 3D mental rotation skills, the computer based tool was also used for the intervention training. The training mode required a cube formation to be built on an empty 5 x 5 base, by adding individual cubes with a single click in the chosen square (or on a side of a chosen cube). This formation had to match the target, shown only through a front, side and plan view, with checking allowed through full rotation of the base plane in any desired direction (see Figure 3). In total there were 20 practise items with each having two levels (matching, and matching with the least number of cubes required). The items could be revisited and completed in any order.

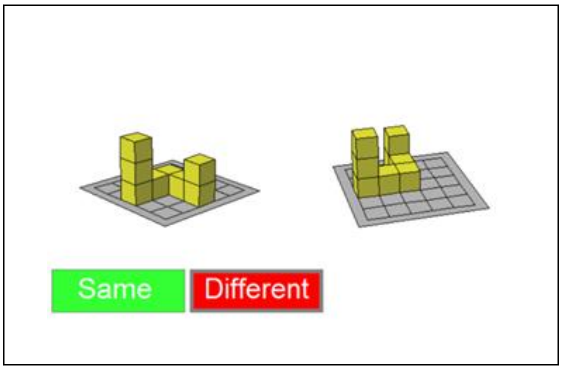


Figure 2. Example 3D mental rotation skill assessment item

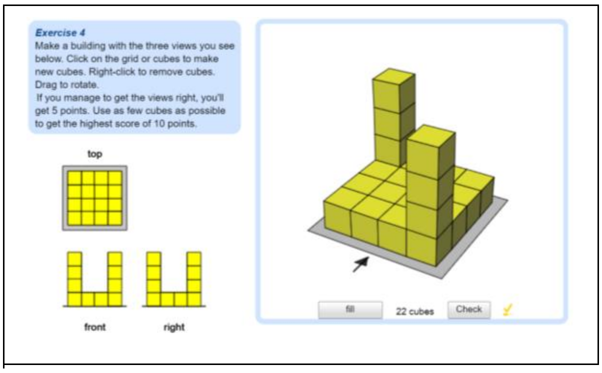


Figure 3. Example 3D mental rotation skill training item

2.2.3.3 Measure of Mental Number Line Precision

Precision of the mental number line was measured using the bounded and unbounded number line task, with both metrics providing data that was used to assess accuracy (PAE). It was important the unbounded NLE task was included in this research as a conclusion drawn from the review in chapter one identified that as a possible purer measure of integer representation, it had been somewhat neglected by researchers, limiting comparisons being drawn between the two NLE tasks. In addition the decision was made not to measure linearity (R^2_{LIN}) of the estimates, as it is PAE that appears to be less confounded with visuo-spatial skills and visuo-motor integration. R^2_{LIN} values have been shown to specifically correlated with global visuo-motor integration ability, an important skill when attempting to use proportional judgement to achieve an even spread of numbers across a number line (Wai, Lubinski, & Benbow, 2009).

The number line tasks were presented in an A4 booklet consisting of 20 bounded and 15 unbounded items (see Figure 4 for an example of both). This number of estimation trials was chosen to support the validity of any finding, as the reviewer had noted in chapter one that using only six items for this metric may be viewed as a limitation (e.g. Gunderson et al., 2012). In addition, the scale used in both NLE tasks needed to be matched to the age of the participants to allow adequate variation in performance. The choice of scale was considered in light of previous research suggesting some disparity in performance between the two NLE tasks. Children over 8 years have generally been shown to achieve reasonable accuracy and linearity when making estimations on a bounded 0 to100 number line (Siegler & Booth, 2004). However, using the unbounded task, good performance (in terms of both PAE and R^2_{LIN}) has been shown to be more difficult to achieve at the same age (Kim & Opfer, 2017). As the pilot study with 29 children (M = 10.8 years) yielded a good range of responses on both tasks using the 0-100 number line, the decision was made to proceed with this scale (see Appendix E for pilot study information).

Bounded NLE Task Item

Unbounded NLE Task Item

Figure 4. Example Bounded and Unbounded NLE Task Items

2.2.3.4 Control Group Activity

The control group were given a word based activity booklet that avoided any overtly spatial components. The answers were available at the end of each session (see Figure 5 for an example).

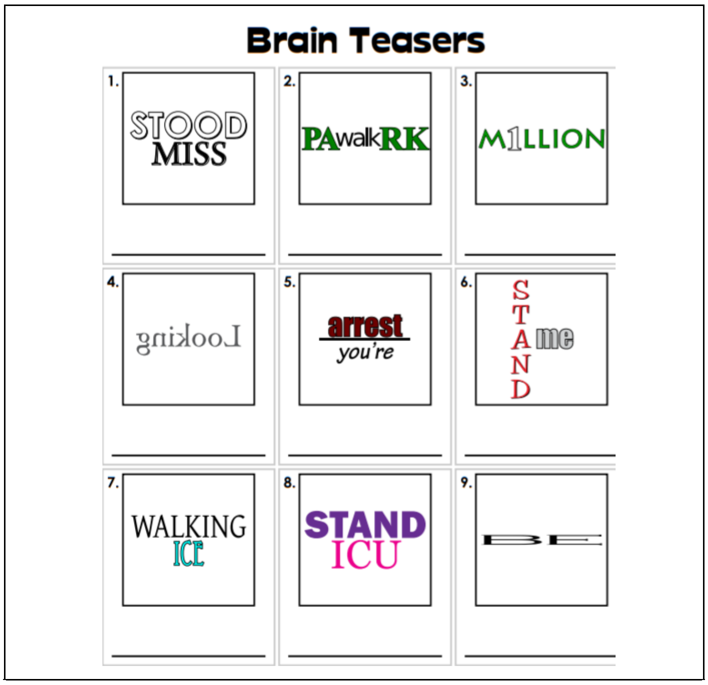


Figure 5. Example Filler Item for Control Group

2.2.3.5 Maths Achievement

Maths achievement scores were obtained for correlation purposes using Standardised End of Key Stage Tests and Assessments in mathematics (SATs). Scores for 47 participants had been externally moderated, while data for the remaining 30 participants came from teacher assessment scores, gathered through 'mock' tests taken up to 1 month prior to the actual test, and using previous SAT question books and mark schemes.

2.2.4 Procedure

After sharing study information with all parties and obtaining all permissions from the Psychology Ethics Committee, University Research Governance Office, and senior leaders at 6 schools within Hampshire, parent consent via an opt-in protocol secured 98 participates, 21 of whom only completed the first session data gathering as a pilot (receiving no intervention or follow up as requested by school staff). The remaining 77 participants received three visits from the researcher on days 1, 10 and 15 of the study. Prior to the first day, the researcher met with a member of school staff to check all consents were in order, allocate code names which would also act as the program log-ins, randomly allocate code names to intervention and control conditions, and demonstrate the computer program to ensure full familiarity and functionality within school systems.

On Day 1, the participants met with the researcher to review the information sheet and sign assent forms (see Appendix F for all child / parent written communications from the researcher). They then proceeded to complete the NLE task booklet (including bounded and unbounded items). This included 3 practice items which the research demonstrated to ensure all participants were clear how to complete the task, and followed the procedure outlined by Siegler and Opfer, 2003 for the bounded task and the protocol described by Link, Nuerk and Moeller (2014) for the unbounded task (see Appendix G). After a short break the children were then shown how to log on to the computer program via an on-line system capturing scores which could only be accessed by the researcher as administrator (https://app.dwo.nl/en/student/). Here they received instructions on how to access the task, watched a demonstration of one item, and then completed 24 trials before logging off. At this point the participants divided into their respective conditions for a 15 minute training session; the control group working on the literacy based filler, and the spatial training group practising with the spatial training program after receiving initial instructions.

Separate training for the control and intervention groups continued over days 2 to 10, overseen by member of staff trained by the researcher. After training on day 10, the researcher returned to be present during data collection for the NLE task and spatial skill, using the same protocol as day 1 but with different target numbers in the NLE task booklet, and different comparison items within the computer program. There was no training on days 11-15. However the researcher returned on day 15 for the final data point, again providing novel items for each measure. The participants were thanked, and received a debriefing sheet for themselves, and one to take to parents (see Appendix F for all child / parent written communications from the researcher).

2.3 Results

2.3.1 Preparing Data for Analysis

The NLE tasks needed to be assessed for accuracy, measured through percentage absolute error (PAE), calculated for each item. If a mark was placed at 83 on a 0 to100 number line, when the correct placement should have been at 56 then the absolute error would be (83-56) / 100 = 27 %. After PAE was calculated for all items, an average was taken as the participant's score.

A first look at the data from the 77 participants in the main study revealed one child's scores were markedly lower than any other participant (e.g. PAE 67.5%, when next nearest was16.6%). On closer scrutiny, and discussion with school staff, it was found this participant's additional needs were likely to have been unique amongst the group (his work had to be extensively differentiated). As his data appeared as an outlier in every analysis and skewed normality across groups, the decision was taken to remove his data from all analyses.

In addition, six participants had extensive missing data due to absence on data collection days 10 or 15. The data from these participants was included for correlation purposes but removed when group differences were being analysed.

2.3.2 Correlations

A visual inspection of box plots confirmed data was normally distributed with no outliers. In addition, the preliminary analysis also confirmed linear relationships between SAT scores, spatial scores, and PAE scores for the bounded and unbounded NLE task taken at time point 1 (see Appendix H for SPSS output).

Using Pearson's correlation coefficient analysis to explore the relationships between the variables revealed a statistically significant, strong positive correlation between SAT scores and spatial scores, r (76) = .544, p < .0001, 95% CI [0.363, 0.685]. This relationship is in line with previous research which has found better spatial skill amongst those whose mathematical achievement is high. Similarly the analysis revealed a statistically significant, large negative correlation between SAT scores and PAE for both the bounded task r (76) = -.562, p < .0001, 95% CI [-0.699, 0.385] and unbounded task r (76) = -.479, p < .0001, 95% CI [-0.636, 0.284]. This confirms previous findings that suggest the internal representation of number magnitude (via the mental number line) is more accurate amongst those who do well in mathematics. These results build on the findings in Chapter 1 as they suggest the positive relationship found between maths achievement and accurate mental representation of number is one that persists from the early years into upper juniors.

Turning to the association between spatial skill and internal representation of number magnitude, the analysis revealed a statistically significant, small negative correlation between spatial scores and PAE on the bounded NLE task, r (76) = -0.282, p < .016, 95% CI [-0.477, -0.060] while there was a statistically significant, medium negative correlation between spatial scores and PAE on the unbounded NLE task, r (76) = -0.428, p < .0001, 95% CI [-0.596, -0.224]. This shows that good spatial skills are linked with accurate mental representations of number magnitude, especially when using the unbounded NLE task as a measure of the mental number line.

The final correlational analysis revealed a statistically significant, small positive correlation between the two types of NLE task r (76) = .286, p < .014, 95% CI [0.065, 0.48]. This points to both metrics assessing a similar construct, but the lack of a strong correlation does raise questions about which is the more accurate measure.

2.3.3 Group Differences

Before analysing any differences resulting arising from group training, it was important to establish the extent to which prior maths attainment might influence the results. After a visual inspection of boxplots showed SAT scores in both groups had a normal distribution with no outliers, and with homogeneity of variances confirmed by Levene's test (p = .716), it was found that SAT scores were slightly higher in the control group (103.5 ± 5.43) than in the intervention group (101.7 ± 5.53). However, using an independent t test, there was no statistically significant difference between these means, t (68) = 1.368, p = .176 95% CI [-4.412, 0.824].

Next, preliminary analyses, as assessed by inspection of boxplots and histograms, showed some evidence of a negative skew within the spatial scores across groups. As this appeared to be due to a ceiling effect of the measurement tool, it was felt necessary to analyse studentized residuals and their Q-Q plots to establish acceptable normality of the data. All descriptive statistics can be viewed in Table 3.

Table 3. *Mean (SD) Spatial and PAE Scores By Training Condition*

Condition	Spatial Training			Control		
	Time 1	Time 2	Time 3	Time 1	Time 2	Time 3
Spatial Score	17.88	19.26	19.44	18.44	18.58	18.58
	(3.08)	(3.89)	(4.06)	(3.56)	(4.77)	(4.76)
PAE Bounded	8.27	10.21	12.67	8.5	8.98	10.52
	(3.32)	(5.48)	(4.42)	(3.2)	(4.39)	(4.91)
PAE Unbounded	14.17	14.01	14.34	13.58	12.71	14.06
	(10.26)	(8.97)	(10.19)	(9.24)	(8.43)	(10.09)

2.3.3.1 The Effect of Spatial Training on Spatial Scores

Does spatial training change spatial scores, and if any spatial scores do change, is this sustained over time?

There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. Spatial scores, across the three time points were normally distributed for the intervention and control groups, as assessed by Normal Q-Q Plots. There was homogeneity of variances ($p > .05$) and covariances ($p > .05$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. However Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 6.755, $p = .034$, therefore the Greenhouse-Geisser correction was subsequently used to interpret results.

An analysis of variance, with spatial training as the between-subject factor and spatial scores over three time points as the within-subjects factor revealed:

- There was no statistically significant interaction between the intervention and time on spatial scores, $F(1.825, 124.099) = 1.527$, $p = .222$, partial $\eta^2 = .022$
- The main effect of time did not show a statistically significant difference in spatial scores at the three different time points, $F(1.825, 124.099) = 2.222$, $p = .117$, partial $\eta^2 = .032$

- The main effect of group showed that there was not a statistically significant difference in spatial scores between the group that received spatial training and those who did not $F(1, 68) = 0.154$, $p = .696$, partial $\eta^2 = 0.002$

2.3.3.2 The Effect of Spatial Training on Bounded NLE Task Accuracy

Does spatial training change the percentage absolute error (PAE) rate of estimations on a bounded number line and are any changes sustained over time?

There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The PAE of the bounded NLE task, across the three time points, was normally distributed for the intervention and control groups, as assessed by Normal Q-Q Plots. There was homogeneity of variances ($p > .05$) and covariances ($p > .001$), as assessed by Levene's test of homogeneity of variances and Box's M test, respectively. However Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 75.942, $p < .001$, therefore the Greenhouse-Geisser correction was subsequently used to interpret results.

An analysis of variance, with spatial training as the between-subject factor and PAE scores from the bounded NLE task over three time points as the within-subjects factor revealed:

- There was no statistically significant interaction between the intervention and time on the PAE of estimations on the bounded number line, $F(1.192, 81.045) = .587$, p =.474, partial $\eta^2 = .009$
- The main effect of time showed a statistically significant difference in mean PAE of estimations on the bounded number line at the different time points, $F(1.192, 81.045) = 4.34$, p =.034, partial $\eta^2 = .06$. Inspection of pairwise comparisons showed that regardless of intervention group, PAE increased most significantly between time 1 and time 2.
- The main effect of group showed that there was not a statistically significant difference in PAE of estimations on the bounded number line between the group that received spatial training and those who did not $F(1, 68) = 0.913$, $p = .343$, partial $\eta^2 = 0.13$

2.3.3.3 The Effect of Spatial Training on Unbounded NLE Task Accuracy

Does spatial training change the percentage absolute error (PAE) rate of estimations on an unbounded number line and are any changes sustained over time?

There were no outliers, as assessed by examination of studentized residuals for values greater than ±3. The PAE scores from the unbounded NLE task, across the three time points, were normally distributed for the intervention and control groups, as assessed by Normal Q-Q Plots. There was homogeneity of variances ($p > .05$) as assessed by Levene's test of homogeneity of variances. Box's M test of homogeneity of covariances was violated ($p < .001$). However, this was noted but not relevant as there were no interactions to be interpreted.

Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, χ^2 (2) = 8.426, p = .015, therefore the Greenhouse-Geisser correction was subsequently used to interpret results.

An analysis of variance, with spatial training as the between-subject factor and PAE scores on the unbounded NLE task over three time points as the within-subjects factor revealed:

- There was no statistically significant interaction between the intervention and time on the PAE of estimations on the unbounded number line, $F(1.789, 121.626) = 0.151$, p =.837, partial η^2 = .002

- The main effect of time did not show a statistically significant difference in PAE of estimations on the unbounded number line at the different time points, $F(1.789, 121.626)$ = .391, p =.654, partial η^2 = .006

- The main effect of group showed that there was not a statistically significant difference in PAE of estimations on the unbounded number line between the group that received spatial training and those who did not, $F(1, 68)$ = .131, p = .719, partial η^2 = .002

2.4 Discussion

There is a wide field of ongoing interest in how cognitive individual differences relate to achievement in mathematics. Within this field, spatial skills are of particular interest as they are robustly associated with positive maths outcomes (Alloway & Passolunghi, 2011; Meyer, Salimpoor, Wu, Geary, & Menon, 2010; Delgado, & Prieto, 2004; Rasmussen & Bisanz, 2005; Carlson, Rowe, & Curby, 2013; Mix & Cheng, 2012). This study has considered the relationship between spatial skills and number magnitude understanding, which itself is thought to be an important building block of later mathematical competence, underpinning skills such as calculation (Cowan et al., 2011). Specifically, this research has focused on how 3D rotation skills might influence the precision of the developing internal representation of number magnitude, the mental number line, as measured by the NLE task.

The results confirm a strong positive correlation between SAT results and spatial scores, adding to the body of evidence that 3D rotation skill, involving intrinsic-dynamic ability, is highly influential to maths achievement (Verdine et al., 2014; Gilligan, Flouri & Farran, 2017). Previous research has used a variety of maths achievement measures to explore this association, and it is helpful to have it confirmed here using a metric standardised across the whole population of 11 year-olds in England and Wales.

The correlational results also confirmed the association between maths achievement and precision of an internal representation of number magnitude, the mental number line. Those participants whose estimates on the NLE tasks were more accurate, achieved higher SAT scores, using both the bounded and unbounded task. This finding is interesting in light of a recent analysis of the demands of the 2018 SAT papers, which suggested 'true' numerical fluency was being tested, i.e. that pupils can use and apply number facts and make connections between numbers (DfE, 2018). Researchers have suggested the mental number line as a schematic representation that not only encodes conceptual understanding, but also facilitates problem solving, including identifying numerical relationships (Case & Okamoto, 1996). This, then, may be why mental representations of number magnitude correlated so strongly with SATs scores.

Additionally, the findings describing the association between spatial scores and accuracy on the NLE tasks contributes to the debate exploring which NLE task is the purer measure of number magnitude representation. Higher spatial scores were achieved by participants who made less errors on the NLE task, and this correlation was strongest for the unbounded task. This is a surprising result as it is the unbounded task that was expected to be least influenced by spatial skill. i.e. accuracy of unbounded estimates are not reliant on proportional visuo-spatial judgements. This raises the difficulty of how to disentangle the role of spatial ability in enhancing the mental number line from any confounding influence it exerts over the metric that measures that same construct. Although this has not been resolved in this study, it does highlight the dichotomy that researchers face in this area. Also, importantly for the debate between the two types of NLE task, this study showed participants found the unbounded task more difficult than the bounded task. This was true of both the intervention and control group, across all three time points, for example the PAE for all children at the first time point was M=8.39 for the bounded task, and M=13.88 for the unbounded task. This is important to notice as it follows previous research, and when put alongside the small positive correlation between the two types of NLE task, points to evidence that the bounded NLE task **is** a valid metric for the mental number line. It does this by reason of the 'proportional judgement' argument which says the unbounded task should be easier as it requires only visual addition measures, rather than subtraction and division strategies the bounded task is thought to demand (Cohen & Sarnecka, 2014). As the current study found the unbounded task harder, the 'proportional judgement' theory is not supported and therefore the bounded task has not been shown to be confounded (any more than the unbounded task).

When considering group differences, the results showed that the spatial performance of those who received spatial training improved between time 1 and 2, and this improvement was maintained until time 3. In contrast the control group did not make any such gains. However, these differences did not reach statistical significance, and although the trend was in the right direction, the results did not replicate those from Uttal et al., (2013) meta-analysis of 217 research studies which produced an average medium effect size (g = 0.47). However, the trend of improved spatial skill found in the current study had also been found when the 3D training tool had been used previously with undergraduates (Bokhove, & Redhead, 2017). Again a level of significance was not reached and it may be possible that the tool itself, with a focus on problem solving through viewing the plane rotate and being able to visually check various configurations, limited the opportunity to practise mental visualization. Another possible explanation is the retesting effect which is suspected to be very strong in this domain (Uttal et al., 2013). The control group engaged with 72 spatial test items in total over the 15 day period, and repeated testing may have reduced the differential between any group differences.

The main hypothesis, that accuracy of estimates on the NLE tasks would improve with spatial training was not supported by the results within this study. This was the case for both the bounded and unbounded NLE task, with the most unexpected finding showing PAE on the bounded NLE task increased across time points 1, 2 and 3, regardless of whether spatial training had been received or not. Although it is unclear exactly how to interpret these findings, there are possibilities which can be considered in light of research previously discussed. Returning to the 2 x 2 classification of visuo-spatial skills, 3D mental rotation was identified as combining intrinsic (within one object,) and dynamic (moving) dimensions. It may be that developing precision in the mental number line relies more on extrinsic and / or static components, particularly as extrinsic information refers to the relation among objects in a group, relative to one another or within a framework (Uttal et al., 2013). The mental number line fits this description and so 3D rotation skills, although closely linked with maths achievement, may not be the optimum visuo-spatial skill that can transfer benefits to an internal representation of number magnitude (via improved number estimation performance). Secondly, the increase in errors over time on the bounded NLE task occurred regardless of training, which leads to the possibility that there may be other factors at work that are not directly related to spatial skill. It was noticeable that during the assessment sessions the children were more interested and excited by the computer measure for spatial skill than they were about the paper and pen NLE task. This may have resulted in task boredom for the NLE activities, with less careful and accurate estimates being made. Another factor to consider is a possible ceiling effect within the NLE task measure. Looking at the bounded PAE scores across the groups shows they are relatively small at time 1 (approx. 8%) and only go up by a small percentage by time 3 (up to approx. 12%). This may reflect a ceiling effect where gains were difficult to measure. The pilot study produced higher initial bounded PAE scores (approx. 21% compared to 8%), which led to the choice of metric (1to100 scale). In hindsight a wider scale may have been more appropriate to effectively demonstrate gains within the NLE tasks.

2.4.1 Strengths and Limitations

A main strength of this research is that it studied the relationships between number magnitude representations and spatial abilities through a randomised controlled design, attempting to explore causation within a body of research that previously primarily focused on relationships of association.

This research included a large pool of mixed ability participants recruited from different settings. This variety ensured the sample were more likely to reflect results based on a broad spectrum of skill and experience. Additionally, the use of nationally administered mathematics SAT scores, in the main externally moderated, allowed an analysis of maths achievement associations to be robust, helpfully supporting the evidence base for links between spatial ability, NLE task performance and positive maths outcomes.

The key limiting factor within this work is the ceiling effect generated by the choice of scale for the NLE task. Although the metric was chosen based on results from the pilot study, gains were difficult to evidence and this may have prevented useful findings coming to light.

Although the choice of 3D mental rotation, as the visuo-spatial component identified by the researcher as the most relevant to the focus of this study, proved not to be significant in generating expected improvement in number line estimations, this should not be viewed as a limitation. Rather, the findings point to further research examining the role of other visuo-spatial components may play in supporting number magnitude representations.

2.5 Recommendations

2.5.1 Future Research

It would be helpful to revisit the design of this study using a wider scale bounded NLE task. The ceiling effect may be masking further findings and until this is clear it would not be helpful to dismiss any role 3D mental rotation skills may have in developing the precision of the mental number line. Future research could also consider other components of visuo-spatial skills, employing the Uttal et al.,(2013) categorisations to establish areas of influence. These investigations would benefit from having a casual focus, trying to pinpoint why spatial skill has such relevance for positive maths outcomes.

In terms of number magnitude representations it would be helpful if research could establish any ongoing role as number skills develop. Although some researchers have surmised the mental number line has far reaching implications across age and mathematical complexity, the evidence so far mainly points to an influence on small integer calculations. There are some investigations pursuing the relevance of a mental number line to fractions and this line of inquiry could be replicated across other areas of mathematics involving number.

It is important to reflect on the inclusion of mainstream participants within this study and whether future research needs to explore children who may have additional needs.

Certainly the exploration of this work has been within the context of how to support children who may not have followed a developmental path of number magnitude representations that is helpful to maths achievement, but there are other SEND needs that may also interact with these difficulties. The intervention used here posits an external visual model as supportive of pupils acquiring useful internal frameworks, but future research might consider those with visual impairments who cannot access such a visual model. Concrete materials and resources might be helpful here, as size, shape and distance might be represented as useful metrics that can be felt as well as seen.

2.5.2 Conclusions and Implications for Educational Psychologists

This research into the contribution of 3D mental rotation skills within the relationship between children's number line estimations and maths achievement has not found a clear role for this particular component of visuo-spatial skills. As this may be due to the limitation of the 1to100 scale number line task as a metric used in the study, there may still be connections yet to be brought to light. With this in mind, alongside the robust, strong associations between maths achievement and spatial skill confirmed in this study, three areas of focus have been identified for Educational Psychologists who are working to support school staff help children overcome barriers to acquiring skills for numeracy. Firstly, sharing information with teachers and support staff on the importance of number magnitude understanding as a building block for approximate estimations and exact calculations; secondly, how to effectively use number line tasks to target weak conceptualisations of number magnitude; and finally, ensuring those whose maths progress is poor can access opportunities for spatial tasks within the curriculum.

The findings presented here have made a case for the bounded number line as an acceptable metric for capturing the precision of the mental number line. Correlations between accuracy and maths achievement were strong and together this supports research which suggests number magnitude is an important concept which begins as an innate more approximate non-symbolic skill and then transfers to a more exact symbolic form. It may be helpful if Educational Psychologists help staff to appreciate this transfer process and to more fully understand that number magnitude is not only about being able to count in order, or one-to-one correspondence leading to the cardinal principle, but concerned with relative size in relation to zero.

Number line tasks used within the classroom are often employed as a visual framework for stepped calculations, where jumps to 10 (or the nearest multiple) are a promoted strategy for problem solving within addition or subtraction. For children whose symbolic magnitude representation is poor, they may benefit from number lines where zero is clearly visible, from making ratio relationships explicit through distance comparisons, and considering subtraction not as a right to left move backwards along the number line, but as the positive left to right distance between two numbers. This is in line with the findings from chapter one, where evidence points to the benefits of counting on in a visually linear format, verbalising the position of a number in relation to zero during simple integer addition. A further recommendation is to ensure visual materials externally model the desired mental representations and concrete manipulatives have the flexibility to do the same e.g. Cuisenaire rods. Similarly giving children opportunities to externalise their internal representation of number, through demonstrating problem solving through manipulatives may also be helpful, as it gives the teacher a window into possible misconceptions that would determine next steps for teaching.

The spatial-numerical mapping of the mental number line, where position in space is an essential element of how number magnitude is conceptualized, led to the initial research question within this research. Despite the results in this study not clearly linking 3D rotation skills to performance on NLE tasks, robust correlations confirmed within this work continue to point to the role spatial skills play in maths achievement. Until this relationship becomes more plain, it may be helpful to encourage a variety of spatial mapping activities, especially for those children who experience difficulty with number.

In summary, alongside the six clear recommendations from chapter one, Educational Psychologists should also adapt their practice to ensure

- They share information that number magnitude skills are not only about being able to count in order, or one-to-one correspondence leading to the cardinal principle, but concerned with relative size in relation to zero.
- They promote the use of number lines where zero is clearly visible, and to encourage children experiencing problems with subtraction to replace backwards right to left moves along the number line with positive left to right counting on activities.
- They suggest visual materials and concrete manipulatives externally model the desired mental representations e.g. Cuisenaire rods.
- Teachers are encouraged to make a variety of spatial mapping activities available, especially for those children who experience difficulty with number.

Appendix A Data Extraction Table

Author Year Country	Participants / group	Design and Analysis	Intervention Control Activity Duration	Covariate	Relevant measures	Relevant findings (including effect sizes)
Siegler and Booth 2004 USA 2004	85 girls and boys 21 kindergarten *M*=5.8 years 33 first grade *M*= 6.9 years 31 second grade *M*=7.8 years	One way ANOVA with the 3 level factor of year group. Correlation / regression	N/A	Age within grade	NLE task (bounded, PAE & R^2_{LIN}) Stanford Achievement Test score for mathematics (SAT-9)	PAE was significantly different between grades. Accuracy on NLE task significantly correlated with maths achievement in all 3 grades Kindergarten *pr*=-.45 p<.05 First grade *pr*= -.66 p<.01 Second grade *pr*=-.37 p<.05 Also linearity of estimates increased as age increased. η^2 =0.35
Booth and Siegler 2008 USA	105 first graders *M*=7.2 years	Randomised controlled 2 x 2 x 2 mixed design, with the two between participants factors (computer generate, child generate) having yes / no levels and the within participants factor as measurement time points (start and end). Multiple dependant variables. ANCOVA Correlations / regressions	**Training:** All groups were given 4 addition questions to answer (based on least answered of 13 questions in a pre-test) **Computer generate** group were shown horizontal 0 to 100 number line with coloured bars that represented addends and total. **Child generate** group chose the points on the line where bars should **Control group** had no access to mapping bars on the line. **Computer and child generate** group's attempt to do the bars	Short term memory Maths level (maths section of the Wide Range Achievement test; WRAT)	**Pre and Post** NLE task (bounded, PAE) Arithmetic problems set by researchers, guided by pre-test results	**Effects on maths achievement (novel tasks)** The intervention that increased learning most was the computer generate condition (32% difference between pre and post-test, p<0.01 Cohen's d= 1.51 **Effects on NLE task** Also computer generate condition reduced PAE the most (p< .05, d= 0.66)

Appendix A

Author Year Country	**Participants / group**	**Design and Analysis**	**Intervention Control Activity Duration**	**Covariate**	**Relevant measures**	**Relevant findings (including effect sizes)**
			themselves was followed by a computer correct version. **Duration:** three 10-15 minute sessions during one week.			
Siegler and Ramani 2009 USA	88 pre-schoolers *M*=4y 8m	Randomised, controlled mixed design, with a between participants factor (linear game board, circular game board and numerical control group) and a within participants factor (pre and post- test). Multiple dependant variables. ANCOVA	**Independent variable** Board game with either a linear or circular track of play. The control condition participated in numerical activities such as counting, identifying numerals and counting objects. **Duration** Five 15-20 minute sessions		Bounded NLE –PAE and R^2 A range of numerical knowledge tasks (counting, numerical magnitude comparison, numeral identification, arithmetic).	**Effects of intervention on NLE task (PAE)** Accuracy improved most for linear group with PAE decreasing significantly $p<0.001$ $d=1.01$ **Effects of intervention on NLE task (R^2)** Linearity on the NLE task increased most for the linear group $p<0.001$ $d=1.03$ **Effects on maths achievement (numerical tasks at post-test).** No significant effects of group on counting. Performance in magnitude comparison improved more than the other 2 groups $p<0.01$, $d=0.75$ Analysis of performance on the arithmetic problems revealed a significant difference in the linear group's accuracy (number correct) but not for their rate of absolute error.
Kucian, Grond, Rotzer, Henzi, Schonmann, Plangger, Galli, Martin and von Aster 2011 Switzerland	36 boys and girls in total. 22 diagnosed with dyscalculia *M*= 9.6 16 controls *M*=9.5	Mixed 2 x 2 design with dyscalculia as the between subject factor and pre/ post training as within subject factor. ANOVA run for each dependant variable	**Training for all participants:** Intervention carried out at home using "Rescue Calcularis" computer game (training automated access to the internal mental number line) for 15 mins, 5 days a week over 5 weeks		**Pre and Post:** NLE task (bounded, using PAE and R^2) Neuropsychological Test Battery for Number Processing and Calculation in Children (ZAREKI-R)	**Effects of training on NLE task (PAE):** Significant training effects, but no effect size given. The interaction between training and group was not significant. **Effects of training on NLE task (R^2):** Significant training effects and significant interaction between training and group (no effect sizes given). While both groups improved in linearity, the dyscalculia group

Author Year Country	Participants / group	Design and Analysis	Intervention Control Activity Duration	Covariate	Relevant measures	Relevant findings (including effect sizes)
						'caught up' with the control group. **Effects of training on maths achievement:** Significant training effects, effect size not given. The interaction between training and group was not significant. Although both groups improved, the dyscalculia group did not 'catch up'.
Mazzocco, Feigenson and Halberda 2011 USA	17 boys and girls. Data collected between ages of 5-8 years old	Correlation / regression explored longitudinally (2 years)	NA	Vocab. Perceptual organisation Spatial reasoning	ANS precision tasks Test of Early Mathematics Ability (TEMA-3)	Preschool ANS precision predicted maths performance r^2=.278 p=.030 but was not a significant predictor of vocabulary, perceptual organisation, or spatial reasoning.
Gunderson, Ramirez, Beilock and Levine 2012 USA	Experiment 2 42 boys and girls. Data collected between ages of 5 and 8 years	Correlation / regression explored longitudinally	N/A	Spatial Skill Vocabulary	NLE task (bounded, R^2) Approximate symbolic calculation Approximate non-symbolic calculation	**Non-symbolic calculation** Children's performance on the approximate non-symbolic calculation task was not significantly correlated with their performance on the other measures. **Symbolic calculation** Children's spatial skill at age 5 predicted number line knowledge at 6, which in turn predicted their performance on an approximate symbolic calculation at age 8 (number line knowledge mediated the relationship).
Obersteiner, Reiss and Ufer 2013	147 first graders *M*=6.91 years	Randomised controlled 2 x 2 between participants design (with 'approximate' and 'exact training as	**Training:** Adapted versions of videogame "The Number Race" to enhance approximate and exact mental number representations.	Pre-test scores as covariate	**Pre and post**: Exact number processing tasks (subitizing and conceptual subitizing)	**Effects of training on number processing:** The exact and approximate training had positive effects only on tasks relying on exact or approximate mental rep of number respectively. The effects were mostly small

Appendix A

Author Year Country	Participants / group	Design and Analysis	Intervention Control Activity Duration	Covariate	Relevant measures	Relevant findings (including effect sizes)
Germany		separate factors, with yes / no level in each). Multiple dependant variables. ANCOVA	**Control:** Computer based word spelling and reading activities **Duration:** 10 sessions of 30 mins over 4 weeks		Approximate number processing tasks (magnitude comparison, number comparison, approximate calculation). Arithmetic test from Hamburger Rechentest	with one medium sized effect of approximate training on magnitude comparison. η^2 =0.091 **Effects of training on maths achievement:** Maths achievement improved for exact and approximate training group only, however these effects were not significant and effect size very small for both η^2 =0.003
Link, Nuerk and Moeller 2014 Germany	45 Fourth graders *M*=9.8	Correlation / regression		General cognitive ability Verbal working memory Visual working memory	NLE task (bounded and unbounded, both PAE) Arithmetic -addition and subtraction problems set by the researchers Further arithmetical competencies - multiplication, completion and number comparison (Heidelburger Rechentest subtests)	Bounded NLE task correlated highly to number comparison, addition and subtraction tasks. Correlations were not found between the unbounded NLE task and any of the basic numerical and arithmetical competencies.
Sella, Tressoldi, Langangeli and Zorzi 2016 Italy	45 pre-schoolers *M*= 5.1 years	Randomised, controlled single factor between participants design with multiple dependant variables. ANCOVA	**Training:** Videogame “The Number Race” consisting of computerized symbolic and non- symbolic number comparison tasks. **Control:** Drawing skills computer program. **Duration**: 16 half-hour sessions.	Pre-test score as covariate	**Pre and post**: NLE task (bounded, PAE) Mental calculation subtest from AC-MT battery	**Effect of training on NLE task:** Hedges g_c =0.73 p=0.012 **Effect of training on mental calculation:** Hedges g_c=1.58 p=0 .002 Significant medium to large effect sizes of training on performance on the NLE task and mental calculation subtest.

Author Year Country	**Participants / group**	**Design and Analysis**	**Intervention Control Activity Duration**	**Covariate**	**Relevant measures**	**Relevant findings (including effect sizes)**
Aulet and Lourenco 2018 USA	66 girls and boys aged between 5 and 7 years old. *M*=6.22 years	Correlation / regression	NA	Verbal proficiency Verbal working memory Spatial short term memory	Magnitude comparison Where's The Number? (WTN) Multiple measures of maths competence	Left to right orientation of number representations (correlation between performance on both SNA tasks). No significant correlations between WTN accuracy and any maths task. No significant correlations between the magnitude comparison task and measures of maths competence except for a significant negative correlation with the ACA task (Approximate Cross-Modal Arithmetic)

Appendix B Quality Assurance Assessment Framework (Original)

Criterion	Score	R1	R2	Agree coeff.	Comment
Data gathering					
Clear research question or hypothesis *e.g. well-defined, measureable constituent elements*	**1 0**				
Appropriate participant sampling *e.g. fit to research question, representativeness.*	**1 0**				
Appropriate measurement instrumentation. *e.g. sensitivity; specificity*	**1 0**				
Comprehensive data gathering *e.g. multiple measures used; context of measurement recorded (e.g. when at school vs at home)*	**1 0**				
Appropriate data gathering method used *e.g. soundness of administration*	**1 0**				
Reduction of bias within participant recruitment/ instrumentation/ administration *e.g. harder-to-reach facilitation; accessibility of instrumentation*	**1 0**				
Response rate/ completion maximised *e.g. response rate specified; piloting; access options*	**1 0**				
Population subgroup data collected	**1 0**				

e.g. participant gender; age; location					
Data analysis					
Missing data analysis *e.g. Level and treatment specified*	**1 0**				
Time trends identified *e.g. year on year changes*	**1 0**				
Geographic considerations *e.g. regional or subgroup analyses*	**1 0**				
Appropriate statistical analyses (descriptive or inferential) *e.g. coherent approach specified; sample size justification.*	**1 0**				
Multi-level or inter-group analyses present *e.g. comparison between participant groups by relevant location or characteristics*	**1 0**				
Data interpretation					
Clear criteria for rating of findings *e.g. benchmarked/ justified evaluation of found quantitative facts*	**1 0**				
Limitations of the research considered in relation to initial aims *e.g. critique of method; generalizability estimate*	**1 0**				
Implications of findings linked to rationale of research question *e.g. implications for theory, practice or future research*	**1 0**				
Total	*Max 15*			Mean coeff.	

Appendix C Quality Assurance Assessment Framework (Adapted)

	Criterion	Siegler & Booth (2004)	Booth & Siegler (2008)	Siegler & Ramani (2009)	Kucian et al. (2011)	Mazzocco et al. (2011)	Gunderson et al. (2012)	Obersteiner et al. (2013)	Link, Nuerk & Moeller (2014)	Sella et al. (2016)	Aulet & Lourenco (2018)
Data gathering	Clear research question or hypothesis	1	1	1	1	1	1	1	1	1	1
	Appropriate participant sampling	1	1	1	1	1	1	1	1	1	1
	Comprehensive data gathering	1	1	1	1	1	½	1	1	1	1
	Appropriate data gathering method used	1	1	½	½	1	½	1	1	½	1
	Reduction of bias within participant recruitment/ instrumentation/ administration	1	½	1	1	1	½	½	0	½	1
	Response rate/ completion maximised	1	1	1	1	1	1	1	1	1	1
	Population subgroup data collected	½	½	1	1	1	1	1	1	1	1
Data analysis	Missing data analysis	0	0	0	1	0	0	0	1	1	1
	Appropriate statistical analyses	1	1	1	1	1	1	1	½	1	1
	Multi-level or inter-group analyses present	0	0	0	0	0	1	0	0	0	0
Data interpretation	Clear criteria for rating of findings	1	1	1	1	1	1	1	0	1	1
	Limitations of the research considered in relation to initial aims	1	½	0	1	1	1	1	0	1	1
	Implications of findings linked to rationale of research question	1	½	1	1	1	1	1	½	1	1
	Total (max 13)	**10.5**	**9**	**9.5**	**11.5**	**12**	**10.5**	**10.5**	**9**	**11**	**12**

Adapted Review Framework for Quantitative Investigation Research.

Appendix D Assessing Weight of Evidence B and C

Weight of Evidence B **Review specific appropriateness of method** (aims, selection of participants and method of data collection)	**Weight of Evidence C** **Review specific focus of study** (specifically with regard to how mental representation of number had been operationalised, and how findings map onto developmental trajectories)	
High	High	Siegler & Booth (2004)
High	High	Booth & Siegler (2008)
High	Medium This was not given a high rating as the evaluation of mathematical performance offered a limited view of developmental progress for the arithmetical element. Children received training on the 4 addition problems and no novel items were given at post-test.	Siegler & Ramani (2009)
Medium This was not given a high rating as the main focus for data collection was EEG information.	Medium This was not given a high rating as the discussion focussed on the relationship between brain activity and maths achievement.	Kucian et al. (2011)
High	High	Mazzocco et al. (2011)
Medium This was not given a high rating as only 6 items were used to generate data about mental number line proficiency. This was the lowest of all the studies and likely to be less robust.	Medium This was not given a high rating as it failed to incorporate information about the unbounded number line estimation task, even though it post-dated Cohen, Blanc and Goldhammer's (2011) suggestion that this was an important metric for operationalising the mental number line.	Gunderson et al. (2012)
High	High	Obersteiner et al. (2013)
High	High	Link, Nuerk & Moeller (2014)
Medium This was not given a high rating as the design and methodology focused on validating the adapted tool (The Number Race) rather than exploring mental number line representations explicitly.	Medium This was not given a high rating as it failed to incorporate information about the unbounded number line estimation task, even though it post-dated Cohen, Blanc and Goldhammer's (2011) suggestion that this was an important metric for operationalising the mental number line.	Sella et al. (2016)
High	High	Aulet & Lourenco (2018)

Appendix E Pilot Study Information

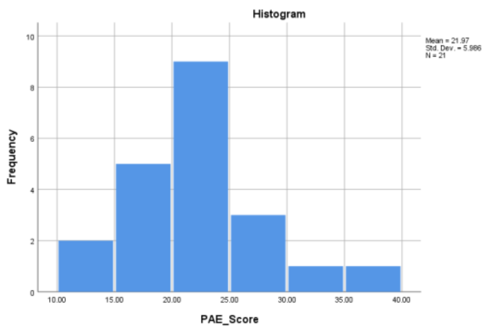
One of the schools recruited was very keen to take part in the research study, but could not commit to the full training schedule. It was agreed that they would participate in day one of data collection, where measures of spatial skill and number line estimation accuracy were completed. However, they would not take part in any training.

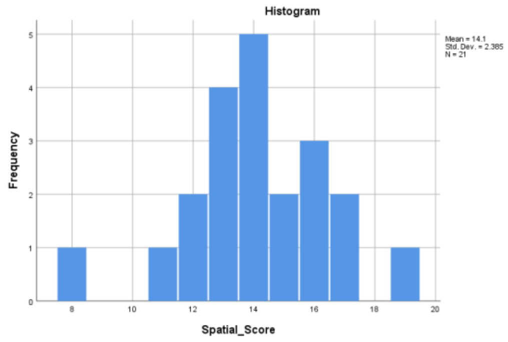
The results gathered were used to confirm the choice of materials, and to streamline the procedure, making changes in light of any encountered difficulties. All consents and permissions were obtained, including an amendment to the proposal via the Psychology Ethics Committee and University Research Governance Office.

The operational procedure went well, with the only area of change being more time needed for a clearer demonstration of the computer assessment program and renaming the modules so they were easier for the children to access.

The results were considered only in respect to the suitability of the measurement tools used, with spatial scores and PAE producing a good range of results that would allow for subsequent gains to be identified.

N=21	Mean	Standard Deviation
Spatial Scores	14.1	.52
PAE Scores	21.9	1.3





Appendix F Participant Information and Consents

UNIVERSITY OF Southampton

Parent/Guardian Information Sheet (V.1, 31.01.18)

Study Title: The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Researcher: Lesley Honour

Supervisors: Edward Redhead; Sarah Wright

ERGO Study ID number: 31704

Please read this information carefully before deciding to allow your child to take part in this research. If you are happy for your child to participate you will be asked to sign a consent form.

Who is running the research?

I am a Trainee Educational Psychologist in my third year of doctoral training at the University of Southampton and I am conducting this study as part of my course. I am very interested in how children acquire skills in maths and have focused this research around their understanding of number. I hope you find the following information helpful but if you have any further questions please contact me via the details at the end of this sheet.

What is the research about?

The purpose of this study is to explore any links between 3D shape rotation skills and placement of numbers on the number line. Previous research has shown there may be links between the two and that some areas of maths achievement can benefit from understanding these links better. The data gained from this study will further inform the evidence-base surrounding this topic and may be helpful in planning ways children can be successfully taught in the classroom.

Why has my child been chosen?

Children in year 6 are at an interesting developmental point in their understanding of number and how they represent numbers in their minds. Also the maths curriculum they experience leading up to the SATs is directed nationally which means they are likely to have had similar exposure to maths ideas which are often taught in similar ways. This helps us explore if practising a different type of task (3D rotation) makes a difference to their understanding. Additionally, your child's school is close enough to Southampton University to make it practical for the school to take part in the study.

What will happen to my child if they take part?

Children that take part will be asked to complete two different activities on a school computer; one matching 3D shapes and one placing numbers on a line. These are completed three times, once at the beginning of the study, once at a mid-point and again at the end. There will be another third activity that is completed each day at school for 2 weeks. This will last about 15 minutes and will either be more 3D shape activities or completing an activity book including word puzzles. Children will not be able to choose which activity they do as the groups will be randomised so any differences can be compared. When they log in to the computer it will be with a code so that names are not used. The researcher and head teacher will be the only ones who can match names to answers on the computer. Children will be given an assent form so that they can give their permission to taking part in the study. Also they are informed that they do not have to participate if they do not want to.

Does my child have to take part?

Your child does not have to take part if you or he/she does not wish to. Participation in the study is completely voluntary. If you would like your child to take part in the study, please sign and return the consent form to your child's teacher by (***insert date***). The study will be run soon after this date.

Are there any benefits in taking part?

Your child's primary school has agreed to participate in the study. This provides an exciting opportunity to develop a better understanding of how children gain an understanding of number and what type of activities help to improve this understanding.

Are there any risks involved?

Although the activities are similar to ones children already do within the school curriculum, some children may find it daunting to respond to questions in a computer program format. If you feel that your child will feel pressured and anxious it may be that you decide not to allow them to take part.

Will my child's participation be confidential?

Yes, all data and information collected will be held in line with the Data Protection Act 1988. All information will be coded, and stored within the university system for 10 years before it is destroyed. You and your child's information will not be identifiable in any part of the final write up.

What happens if I change my mind or my child changes his/her mind?

If you or your child decide you no longer want to participate in the study you are able to do so without facing any prejudice and without giving reason for doing so. You are able to withdraw your and your child's participation from the study at any time, up to and including 30th **April 2019**. After this date, your child's data will be included in the data analysis and subsequent final write up.

What happens if something goes wrong?

If you have any concerns or questions about this study that you would like to speak to the researcher about, please contact Lesley Honour using the email address below. If you wish to formally complain or speak to someone independent of this study, please contact the Chair of the Ethics Committee:

Chair of the Ethics Committee
School of Psychology
University of Southampton
Southampton
SO17 1BJ

Tel: 02380 594663

Where can I get more information?

If you would like any further information about the study, please contact Lesley Honour using the email address below.

Researcher contact details: Lesley Honour: l.honour@soton.ac.uk

UNIVERSITY OF Southampton

Parent / Guardian Consent Form (V.1, 31.01.18)

Study title: The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Researcher: Lesley Honour

Supervisors: Edward Redhead; Sarah Wright

ERGO Study ID number: 31704

*Please tick the box (as) if you agree with the statement(s) and return the form as soon as possible, no later than (**Insert Date**):*

I have read and understood the information sheet (V.1, 31.01.18) and have had the opportunity to ask questions about the study (via email). ☐

I give my permission for my child to take part in this study and agree for my child's data to be used for the purpose of this study. ☐

I understand my child's participation is voluntary and I/they may withdraw from the study at any time without my/their legal rights being affected. However, data must be withdrawn by ☐

30th April 2019.

Name of child (print name)…………………………………………………

Signature of parent/carer/guardian……………………………………………………………..

Date………………………………………………………………………………

UNIVERSITY OF Southampton

The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Who am I?

I am a Trainee Educational Psychologists at the University of Southampton. My name is Lesley.

What is the research about?

I would like to find out about children's skills with 3D shapes and if they are linked with how they understand how numbers are represented on a number line.

Do I have to take part?

No. It is up to you to decide whether you want to take part. You may find it helpful to talk to your parents/guardians at home about it before you make a decision.

What happens if I don't want to take part?

You do not have to take part if you do not wish to. You can tell your parents or your teacher or me that you do not wish to take part. No one will mind if you choose not to take part, it is okay and no one will be cross with you. It is completely your own choice.

What will happen if I do decide to take part? If you agree to take part you will be asked to complete two different activities on a school computer; one matching 3D shapes and one placing numbers on a line. You will do this three times (once at the beginning of the study, once near the middle and once at the end). There will be another third activity that is completed each day at school for 2 weeks. This will last about 15 minutes and will either be more 3D shape activities or completing word searches. You will not be able to choose which activity you do as the groups will be chosen by chance - a bit like picking names out of a hat. When you log on to the computer it will be with a code so that your name is not used. I will be the only one who can match names to answers on the computer. Your name will not be used in the writing about this project.

UNIVERSITY OF Southampton

Participant Assent Form (V.1, 31.01.18)

Study title: The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Researcher: Lesley Honour

Supervisors: Edward Redhead; Sarah Wright

ERGO Study ID number: 31704

Dear Year 6,

I would like you to take part in a project about some of the things you may have learned about in your school maths lessons -3D shape and the number line.

If you choose to take part you will be asked to complete two different activities on a school computer; one matching 3D shapes and one placing numbers on a line. You would do this three times in total. There will be another third activity that you have to complete everyday while you are at school for 2 weeks. This will last about 15 minutes and will either be more 3D shape activities or completing word searches. You will not be able to choose which one you do as the names will be randomised, a bit like 'pulling a name out of the hat'. It is your decision whether you would like to take part.

Even if you agree to take part but later decide that you don't want to, this is okay. You can tell the researcher on the day that you don't want to take part or afterwards you can say that you do not wish your answers to be included in the project. The researchers and your teachers will not mind.

Lastly, you will log on to the computer with a code so that you do not have to use your name. The researcher will be the only one who can match your name to the answers you give on the computer. This means your answers are confidential.

If you are happy to be a part of this project, please write your name below.

Participant Name …………………………… Date ………………………

UNIVERSITY OF Southampton

Parent / Guardian Debriefing Statement (V.1, 31.01.18)

Study Title: The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Researcher: Lesley Honour

Supervisors: Edward Redhead; Sarah Wright

ERGO Study ID number: 31704

Dear Parent/Guardian,

Thank you for allowing your child to take part in my research project. The purpose of my study was to explore any links between 3D shape rotation skills and placement of numbers on the number line. Previous research has shown there may be links between the two and that some areas of maths achievement can benefit from understanding these links better. The data gained from this study will further inform the evidence-base surrounding this topic and may be helpful in planning ways children can be successfully taught in the classroom.

Your child's data will be analysed along with other pupils' data to explore if those children who received spatial training (in this case mental rotation of 3D shapes) made more improvements in how they represent numbers on a number line, than those children who did not receive this type of training. The findings from this study will be available in a written document at the University of Southampton. I also plan to publish the results in a psychology journal and will try to find ways to share the information with parents, school staff and psychologists. This will include a summary report which I will send to your child's school to be distributed to the parents and children who participated in the study.

In the final write up of this project no names will be used as the data will be clustered together according to group. As stated in the information sheet, your child's identity and school will remain confidential, as well as any information shared by the school and the University of Southampton. If you decide that you do not wish your child's data to be included in the study, please contact me by 30th April 2019.

If you have any further questions about the project, please feel free to contact me on the following email addresses:

Lesley Honour: l.honour@soton.ac.uk

Thank you for helping me with this research.

If you have questions about your rights or your child's rights as a participant in this research, or if you feel that your child may have been placed at risk, you may contact the Chair of the Ethics Committee, Psychology, University of Southampton, Southampton, SO17 1BJ. Phone: +44 (0)23 8059 3856, email fshs-rso@soton.ac.uk

UNIVERSITY OF Southampton

Participant Debriefing Statement (V.1, 31.01.18)

Study Title: The Relationship Between Children's Number Line Estimations and Maths Achievement: The Contribution of 3D Mental Rotation Skills.

Researcher: Lesley Honour

Supervisors: Edward Redhead; Sarah Wright

ERGO Study ID number: 31704

Dear Year 6

Thank you for taking part in my research project. I hope you enjoyed the activities.

I wanted to find out if children's skills with 3D shapes are linked with how they understand how numbers are represented on a number line. When I look at the data I will try to work out if there are any links, and if having 3D training makes a difference for how they managed on the number task. This is important because if they are linked, then practising 3D activities could be a way to help children do better at their number work.

By taking part, you have contributed to helping me answer my questions about how children learn to represent number, and I really appreciate your help.

When I finish looking at the data, I will write about what I have found. You will be able to read about this as I plan to send a summary of the work to your school, and to your parents.

Thank you once again for all your help. If you have any questions, please contact me through the email address I have given to your parents.

Lesley Honour

Appendix G Bounded Number Line Protocol

Analyzing the Number-line Task
A Tutorial

John E. Opfer
Carnegie Mellon University

Introduction

The Number-line Task

The number-line task is a robust tool for characterizing representations of numerical value. Without demanding specific knowledge of measurement units (such as inches or centimeters), it taps into participants' mapping of spatial and numerical quantities across a wide range of values. The task has proven useful for characterizing subjects' representations across a wide range of ages (Siegler & Opfer, 2003). The purpose of this tutorial is to document procedures for analyzing performance on the number-line task.

The number-line task has two variants: the Number-to-Position (NP) Task and the Position-to-Number (PN) task. On the NP task (Fig. 1A), participants are shown a number and asked to estimate its position on the number line. On the PN task (Fig. 1B), participants are shown a position on a number line and asked to estimate the number that corresponds to it.

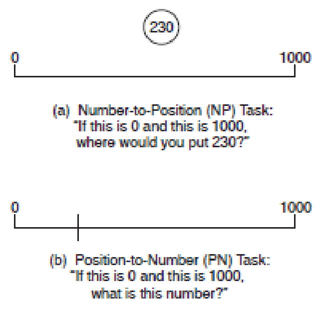
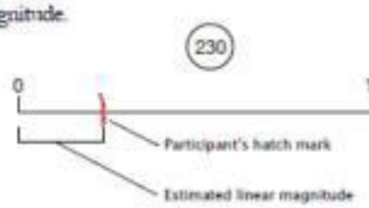


Figure 1. Two variants of the number-line task.

Recording the Data

Participants' performance on the number-line task is recorded in two ways, depending on the variant of the task.

On the NP task, participants provide a hatch mark on the number line to provide an estimate of linear magnitude.



To convert estimates of linear magnitude into a real number

1 Measure the distance from the left end point to the hatch mark (in linear units),
2 Divide that distance by the total length of the line, and
3 Multiply that value by the number given on the other endpoint.

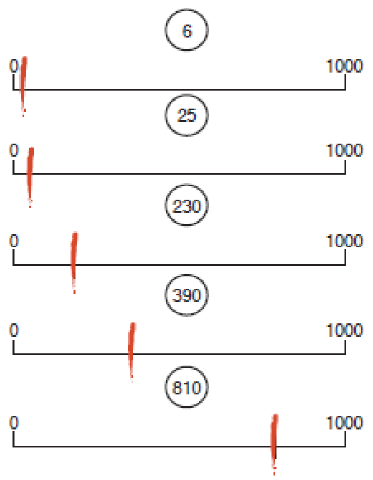
On the PN task, participants write out an integer; no conversion is needed. However, to generate stimuli on the PN task, one must convert numbers into linear magnitudes using the above procedure.

Obtaining Sufficient Data

By obtaining only a single estimate, it is not possible to discriminate among competing models of numeric representation.

Ideally, participants should be given a large range of quantities to estimate, with duplicate quantities. For example, in Siegler & Booth (2003), participants were given 24 different numerals (or linear magnitudes).

Thus, for each participant, there was a series of estimates that corresponded to a series of actual values.



Appendix H SPSS Output

Explore

Case Processing Summary

	Cases					
	Valid		Missing		Total	
	N	Percent	N	Percent	N	Percent
Spatial Score 1	73	96.1%	3	3.9%	76	100.0%
SAT Score	73	96.1%	3	3.9%	76	100.0%
Unbounded NLE 1	73	96.1%	3	3.9%	76	100.0%
Bounded NLE 1	73	96.1%	3	3.9%	76	100.0%

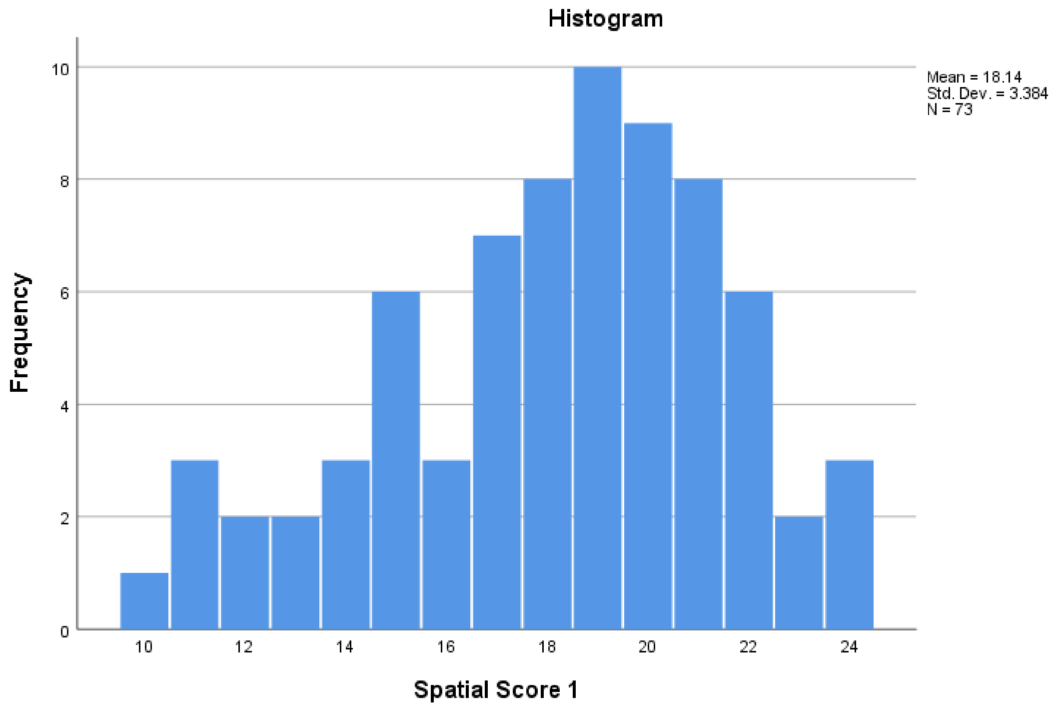
Descriptives

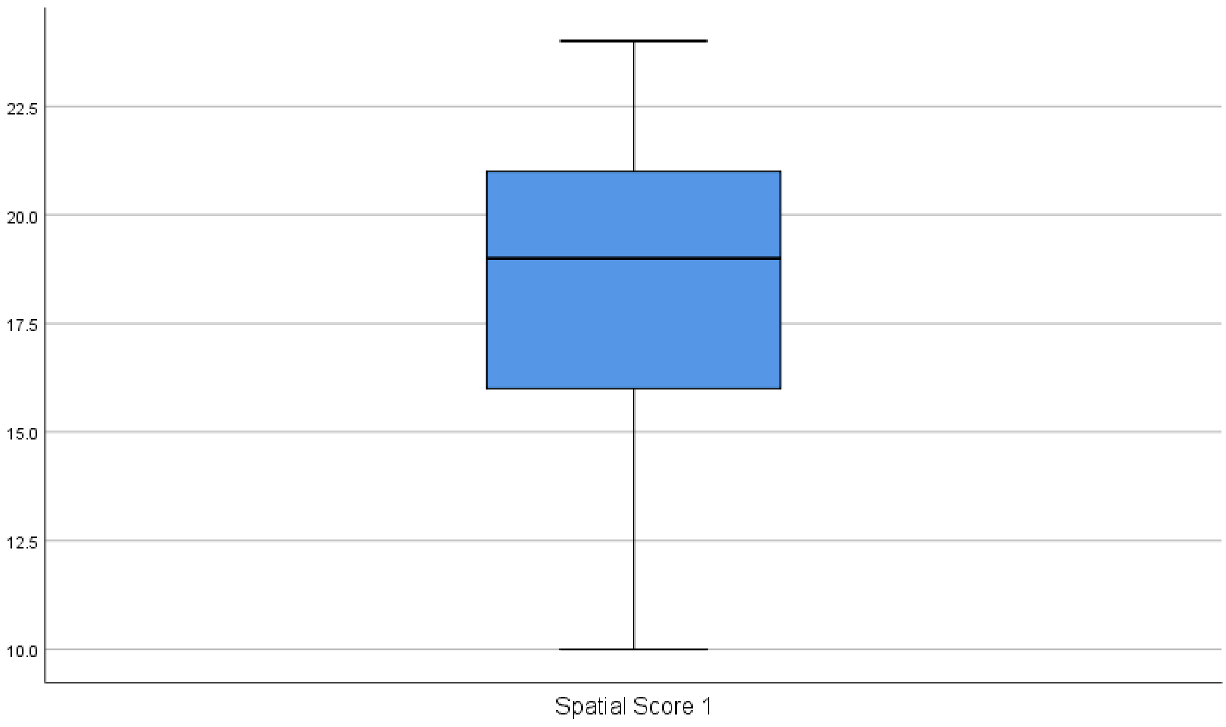
			Statistic	Std. Error
Spatial Score 1	Mean		18.14	.396
	95% Confidence Interval for Mean	Lower Bound	17.35	
		Upper Bound	18.93	
	5% Trimmed Mean		18.23	
	Median		19.00	
	Variance		11.453	
	Std. Deviation		3.384	
	Minimum		10	
	Maximum		24	
	Range		14	
	Interquartile Range		5	
	Skewness		-.504	.281

	Kurtosis		-.300	.555
SAT Score	Mean		102.32	.671
	95% Confidence Interval for Mean	Lower Bound	100.98	
		Upper Bound	103.65	
	5% Trimmed Mean		102.29	
	Median		102.00	
	Variance		32.885	
	Std. Deviation		5.735	
	Minimum		90	
	Maximum		115	
	Range		25	
	Interquartile Range		7	
	Skewness		.080	.281
	Kurtosis		-.285	.555
Unbounded NLE 1	Mean		13.988	1.1277
	95% Confidence Interval for Mean	Lower Bound	11.740	
		Upper Bound	16.236	
	5% Trimmed Mean		13.151	
	Median		11.500	
	Variance		92.833	
	Std. Deviation		9.6350	
	Minimum		2.1	
	Maximum		44.1	

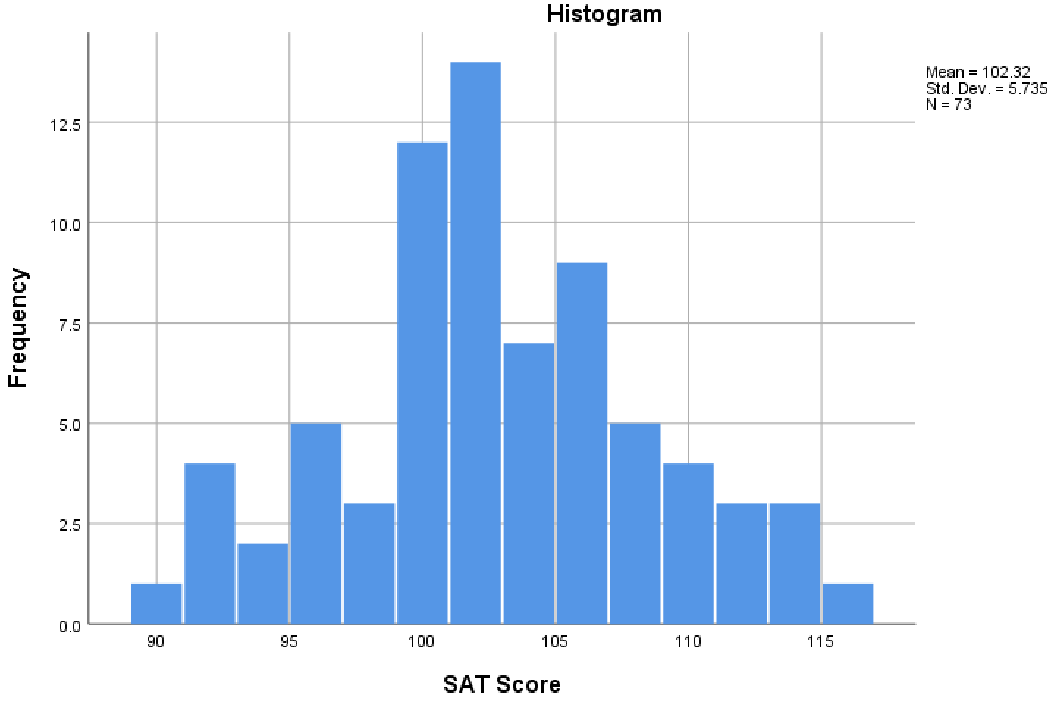
	Range		42.0	
	Interquartile Range		11.0	
	Skewness		1.351	.281
	Kurtosis		1.421	.555
Bounded NLE 1	Mean		8.542	.3971
	95% Confidence Interval for Mean	Lower Bound	7.751	
		Upper Bound	9.334	
	5% Trimmed Mean		8.380	
	Median		7.900	
	Variance		11.513	
	Std. Deviation		3.3930	
	Minimum		2.8	
	Maximum		16.6	
	Range		13.8	
	Interquartile Range		5.3	
	Skewness		.738	.281
	Kurtosis		-.229	.555

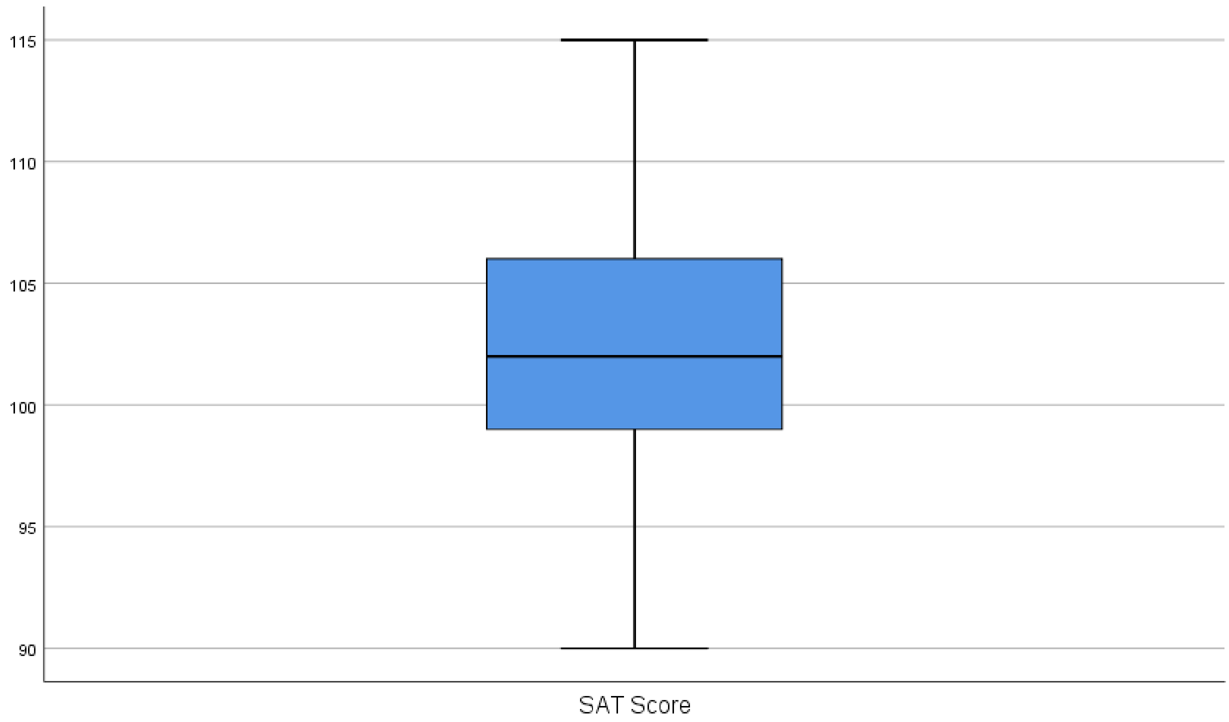
Spatial Score 1



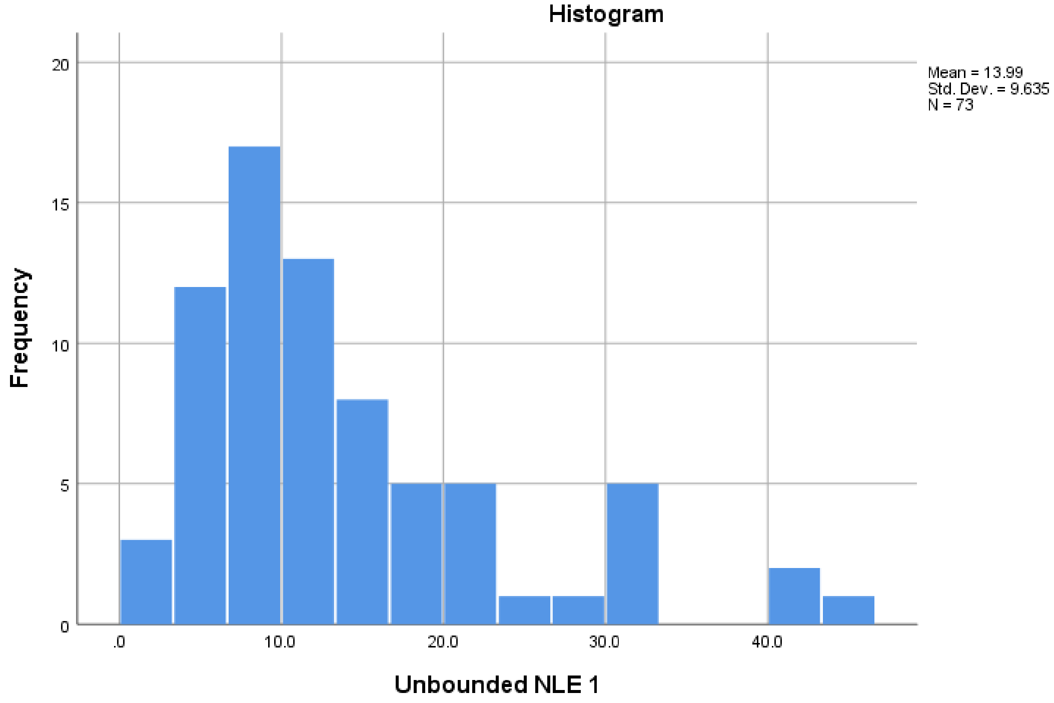


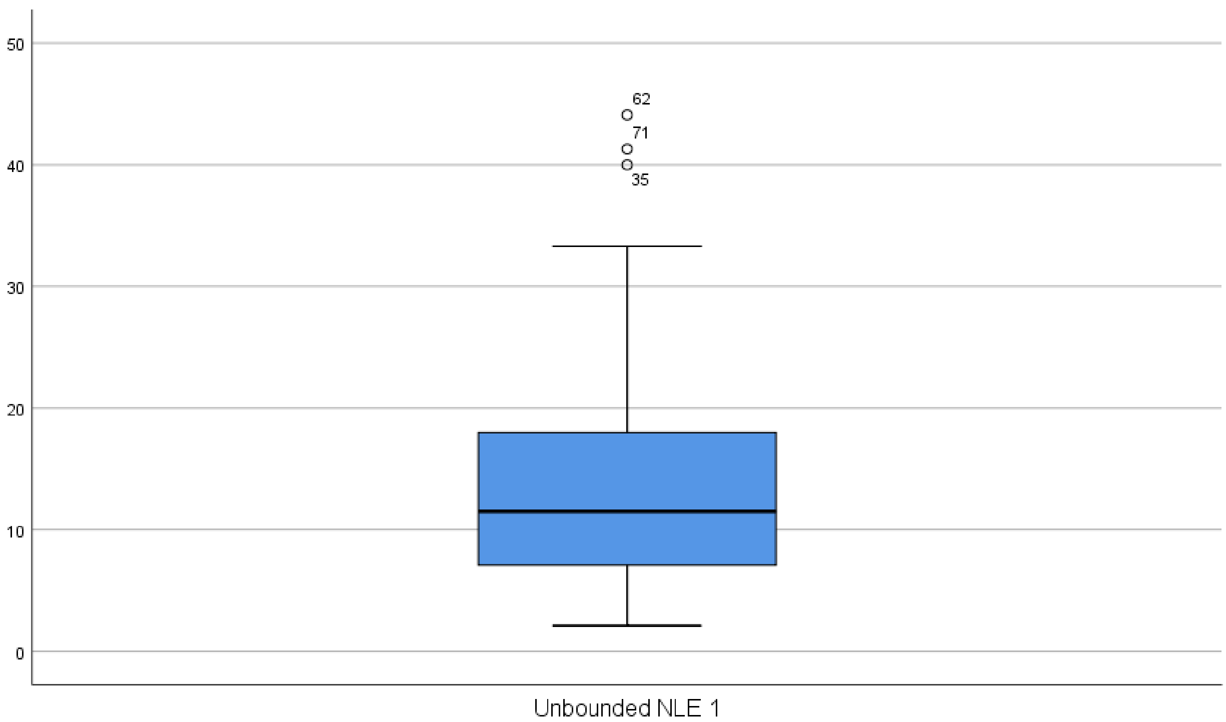
SAT Score


Histogram
Mean = 102.32
Std. Dev. = 5.735
N = 73
Frequency
12.5
10.0
7.5
5.0
2.5
0.0
90
95
100
105
110
115
SAT Score

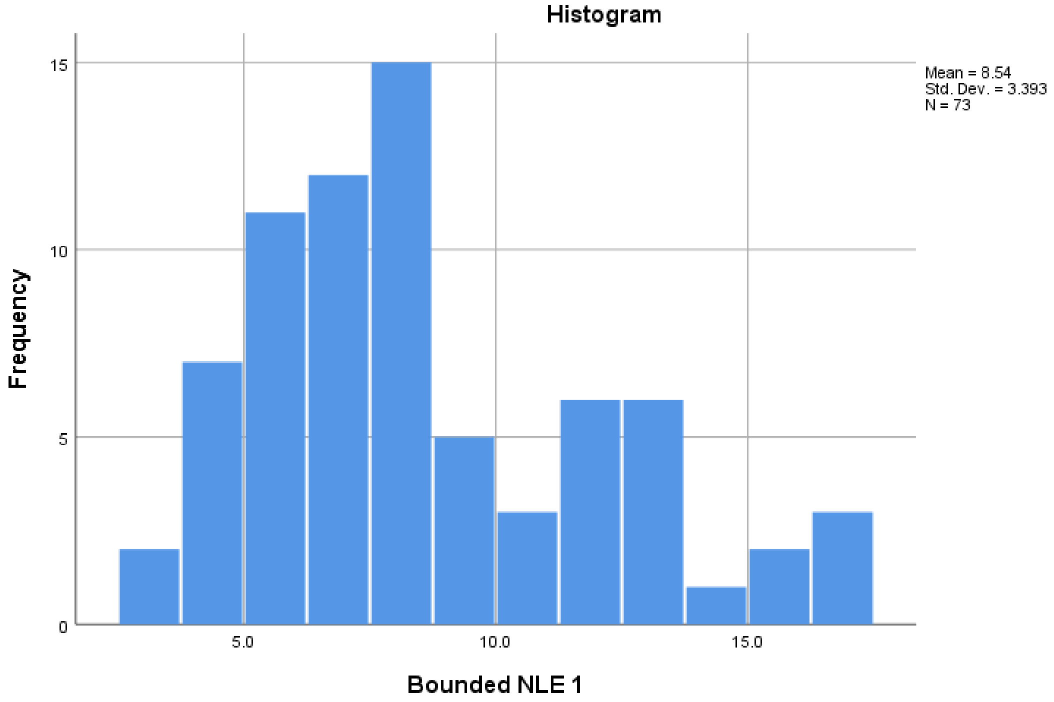

115
110
105
100
95
90
SAT Score

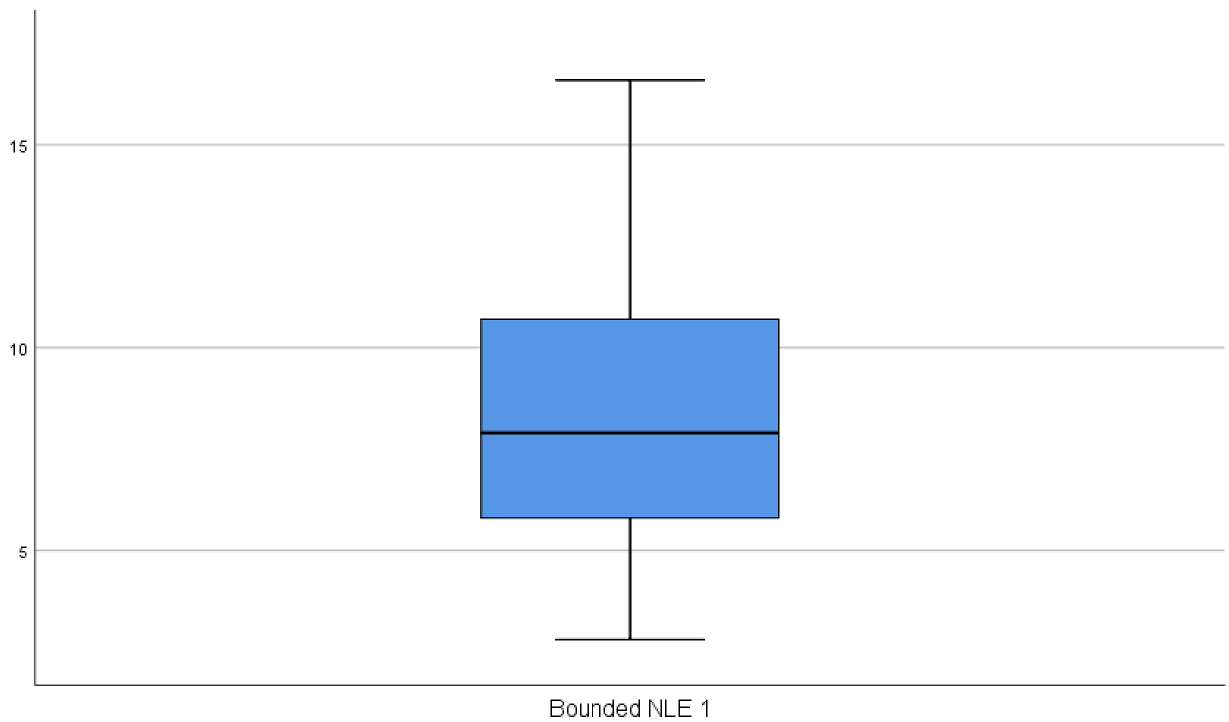
Unbounded NLE 1



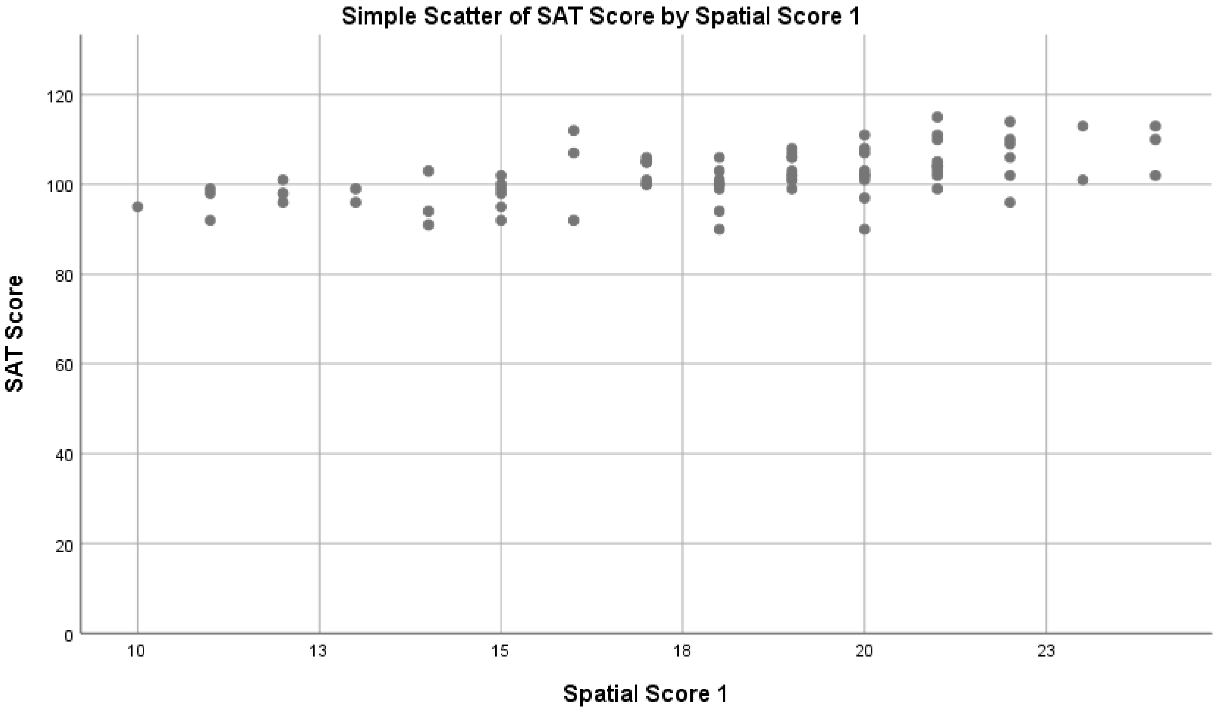


Bounded NLE 1

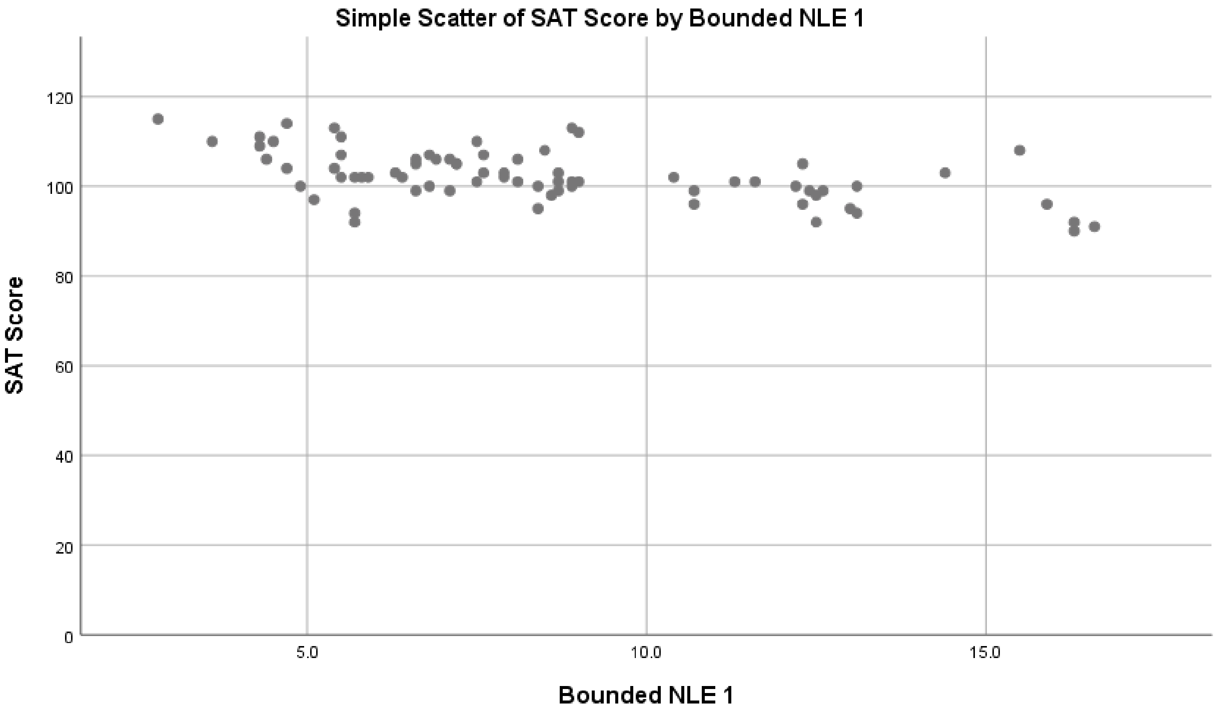




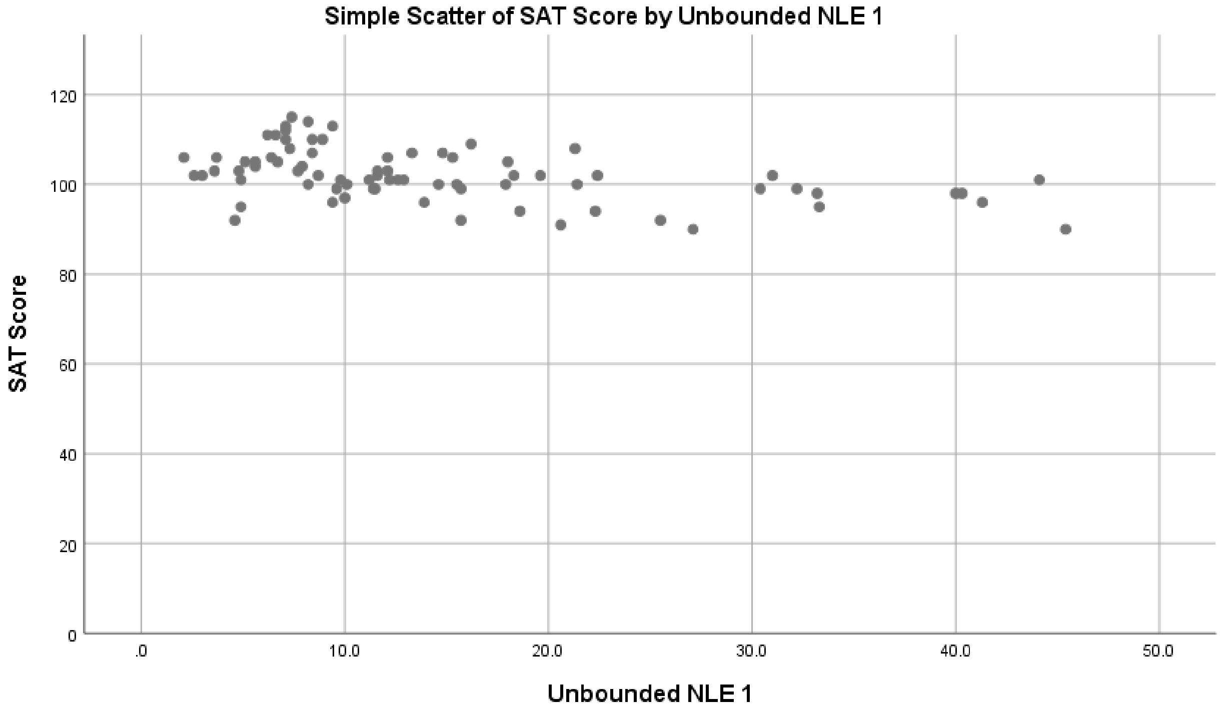
GGraph



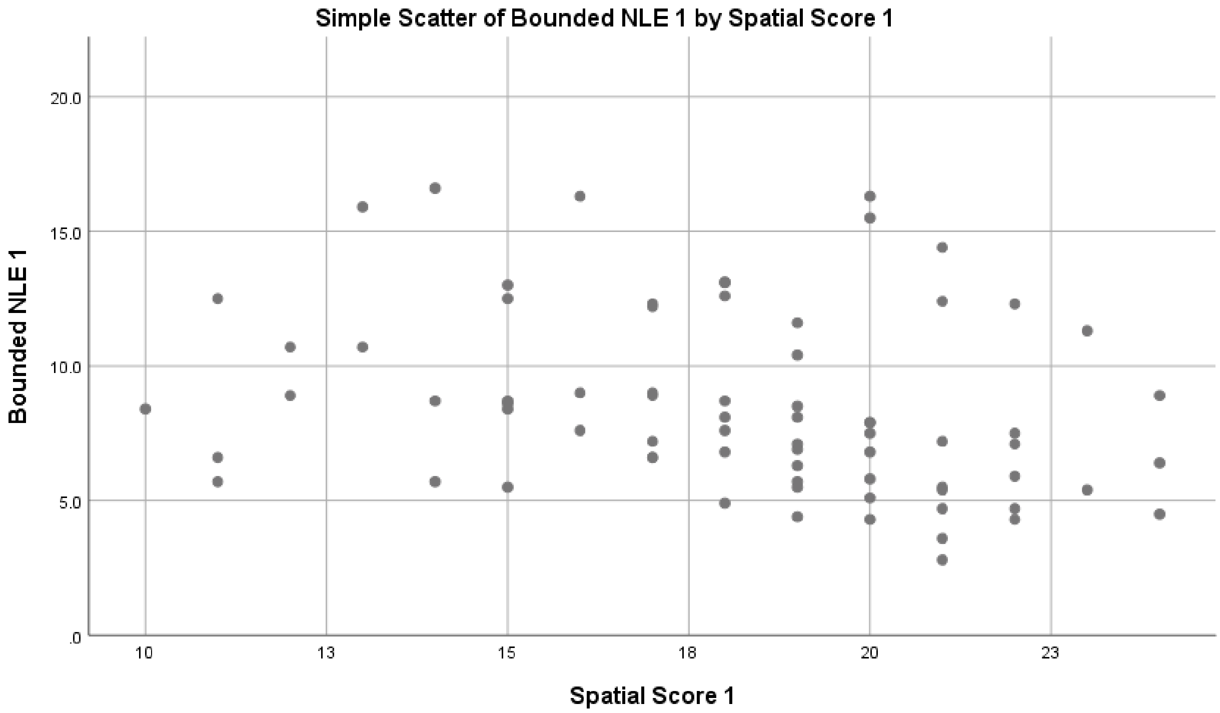
GGraph



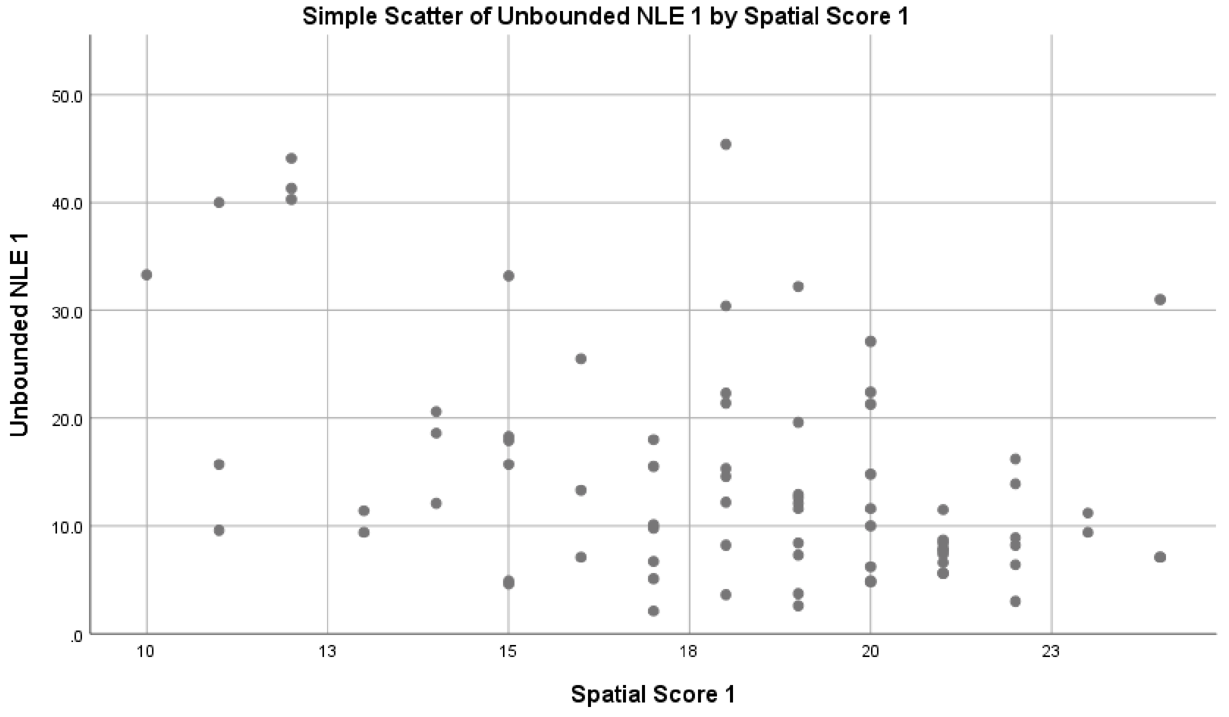
GGraph



GGraph



GGraph



Correlations

Correlations

		SAT Score	Unbounded NLE 1	Bounded NLE 1	Spatial Score 1
SAT Score	Pearson Correlation	1	-.479**	-.562**	.544**
	Sig. (2-tailed)		.000	.000	.000
	N	76	76	73	76
Unbounded NLE 1	Pearson Correlation	-.479**	1	.286*	-.428**
	Sig. (2-tailed)	.000		.014	.000
	N	76	76	73	76
Bounded NLE 1	Pearson Correlation	-.562**	.286*	1	-.282*
	Sig. (2-tailed)	.000	.014		.016
	N	73	73	73	73
Spatial Score 1	Pearson Correlation	.544**	-.428**	-.282*	1

	Sig. (2-tailed)	.000	.000	.016	
	N	76	76	73	76

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

General Linear Model

Within-Subjects Factors

Measure: Spatial_Score

Time	Dependent Variable
1	Spatial_Score_1
2	Spatial_Score_2
3	Spatial_Score_3

Between-Subjects Factors

		Value Label	N
Spatial Training	1	Intervention	34
	2	Control	36

Descriptive Statistics

	Spatial Training	Mean	Std. Deviation	N
Spatial Score 1	Intervention	17.88	3.082	34
	Control	18.44	3.557	36
	Total	18.17	3.323	70
Spatial Score 2	Intervention	19.26	3.895	34

	Control	18.58	4.771	36
	Total	18.91	4.350	70
Spatial Score 3	Intervention	19.44	4.047	34
	Control	18.58	4.759	36
	Total	19.00	4.417	70

Box's Test of Equality of Covariance Matrices[a]

Box's M	10.243
F	1.625
df1	6
df2	33194.011
Sig.	.136

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df
Time	Pillai's Trace	.071	2.572[b]	2.000	67.000
	Wilks' Lambda	.929	2.572[b]	2.000	67.000
	Hotelling's Trace	.077	2.572[b]	2.000	67.000
	Roy's Largest Root	.077	2.572[b]	2.000	67.000
Time * Spatial_Training	Pillai's Trace	.050	1.748[b]	2.000	67.000
	Wilks' Lambda	.950	1.748[b]	2.000	67.000
	Hotelling's Trace	.052	1.748[b]	2.000	67.000
	Roy's Largest Root	.052	1.748[b]	2.000	67.000

Multivariate Tests[a]

Effect		Sig.	Partial Eta Squared
Time	Pillai's Trace	.084	.071
	Wilks' Lambda	.084	.071
	Hotelling's Trace	.084	.071
	Roy's Largest Root	.084	.071
Time * Spatial_Training	Pillai's Trace	.182	.050
	Wilks' Lambda	.182	.050
	Hotelling's Trace	.182	.050
	Roy's Largest Root	.182	.050

Mauchly's Test of Sphericity[a]

Measure: Spatial_Score

					Epsilon[b]
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser
Time	.904	6.755	2	.034	.912

Mauchly's Test of Sphericity[a]

Measure: Spatial_Score

	Epsilon	
Within Subjects Effect	Huynh-Feldt	Lower-bound
Time	.950	.500

Tests of Within-Subjects Effects

Measure: Spatial_Score

Source		Type III Sum of Squares	df	Mean Square	F
Time	Sphericity Assumed	30.469	2	15.235	2.222
	Greenhouse-Geisser	30.469	1.825	16.696	2.222
	Huynh-Feldt	30.469	1.900	16.034	2.222
	Lower-bound	30.469	1.000	30.469	2.222
Time * Spatial_Training	Sphericity Assumed	20.945	2	10.473	1.527
	Greenhouse-Geisser	20.945	1.825	11.477	1.527
	Huynh-Feldt	20.945	1.900	11.022	1.527
	Lower-bound	20.945	1.000	20.945	1.527
Error(Time)	Sphericity Assumed	932.655	136	6.858	
	Greenhouse-Geisser	932.655	124.099	7.515	
	Huynh-Feldt	932.655	129.217	7.218	
	Lower-bound	932.655	68.000	13.716	

Tests of Within-Subjects Effects

Measure: Spatial_Score

Source		Sig.	Partial Eta Squared
Time	Sphericity Assumed	.112	.032
	Greenhouse-Geisser	.117	.032
	Huynh-Feldt	.115	.032

	Lower-bound	.141	.032
Time * Spatial_Training	Sphericity Assumed	.221	.022
	Greenhouse-Geisser	.222	.022
	Huynh-Feldt	.222	.022
	Lower-bound	.221	.022
Error(Time)	Sphericity Assumed		
	Greenhouse-Geisser		
	Huynh-Feldt		
	Lower-bound		

Tests of Within-Subjects Contrasts

Measure: Spatial_Score

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Linear	25.199	1	25.199	3.091	.083
	Quadratic	5.270	1	5.270	.947	.334
Time * Spatial_Training	Linear	17.627	1	17.627	2.162	.146
	Quadratic	3.318	1	3.318	.596	.443
Error(Time)	Linear	554.344	68	8.152		
	Quadratic	378.311	68	5.563		

Tests of Within-Subjects Contrasts

Measure: Spatial_Score

Source	Time	Partial Eta Squared

Time	Linear	.043
	Quadratic	.014
Time * Spatial_Training	Linear	.031
	Quadratic	.009
Error(Time)	Linear	
	Quadratic	

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Spatial Score 1	Based on Mean	.382	1	68	.539
	Based on Median	.256	1	68	.614
	Based on Median and with adjusted df	.256	1	64.130	.614
	Based on trimmed mean	.365	1	68	.548
Spatial Score 2	Based on Mean	2.818	1	68	.098
	Based on Median	2.503	1	68	.118
	Based on Median and with adjusted df	2.503	1	64.598	.118
	Based on trimmed mean	2.716	1	68	.104
Spatial Score 3	Based on Mean	2.972	1	68	.089
	Based on Median	1.779	1	68	.187
	Based on Median and with adjusted df	1.779	1	67.938	.187

Based on trimmed mean	2.957	1	68	.090

Tests of Between-Subjects Effects

Measure: Spatial_Score

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	73374.098	1	73374.098	2032.968	.000	.968
Spatial_Training	5.565	1	5.565	.154	.696	.002
Error	2454.264	68	36.092			

Estimated Marginal Means

1. Time

Estimates

Measure: Spatial_Score

Time	Mean	Std. Error	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
1	18.163	.399	17.368	18.959
2	18.924	.522	17.882	19.966
3	19.012	.529	17.956	20.069

Pairwise Comparisons

Measure: Spatial_Score

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]

					Lower Bound	Upper Bound
1	2	-.761	.369	.129	-1.666	.144
	3	-.849	.483	.250	-2.034	.336
2	1	.761	.369	.129	-.144	1.666
	3	-.088	.468	1.000	-1.238	1.061
3	1	.849	.483	.250	-.336	2.034
	2	.088	.468	1.000	-1.061	1.238

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.071	2.572[a]	2.000	67.000	.084	.071
Wilks' lambda	.929	2.572[a]	2.000	67.000	.084	.071
Hotelling's trace	.077	2.572[a]	2.000	67.000	.084	.071
Roy's largest root	.077	2.572[a]	2.000	67.000	.084	.071

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

2. Spatial Training * Time

Measure: Spatial_Score

				95% Confidence Interval	
Spatial Training	Time	Mean	Std. Error	Lower Bound	Upper Bound
Intervention	1	17.882	.572	16.741	19.024
	2	19.265	.749	17.770	20.760
	3	19.441	.759	17.926	20.956
Control	1	18.444	.556	17.335	19.554
	2	18.583	.728	17.131	20.036
	3	18.583	.738	17.111	20.056

General Linear Model

Within-Subjects Factors

Measure: BNLE

Time	Dependent Variable
1	BNLE_1
2	BNLE_2
3	BNLE_3

Between-Subjects Factors

		Value Label	N
Spatial Training	1	Intervention	34
	2	Control	36

Descriptive Statistics

	Spatial Training	Mean	Std. Deviation	N
Bounded NLE 1	Intervention	8.274	3.3242	34
	Control	8.497	3.2033	36
	Total	8.389	3.2408	70
Bounded NLE 2	Intervention	10.209	5.4825	34
	Control	8.978	4.3889	36
	Total	9.576	4.9528	70
Bounded NLE 3	Intervention	12.674	14.4205	34
	Control	10.528	4.9140	36
	Total	11.570	10.6241	70

Box's Test of Equality of Covariance Matrices[a]

Box's M	35.982
F	5.709
df1	6
df2	33194.011
Sig.	.000

Tests the null hypothesis that the observed covariance matrices of the dependent variables are equal across groups.[a]

Within Subjects Design: Time

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df
Time	Pillai's Trace	.143	5.611[b]	2.000	67.000
	Wilks' Lambda	.857	5.611[b]	2.000	67.000
	Hotelling's Trace	.167	5.611[b]	2.000	67.000
	Roy's Largest Root	.167	5.611[b]	2.000	67.000
Time * Spatial_Training	Pillai's Trace	.042	1.459[b]	2.000	67.000
	Wilks' Lambda	.958	1.459[b]	2.000	67.000
	Hotelling's Trace	.044	1.459[b]	2.000	67.000
	Roy's Largest Root	.044	1.459[b]	2.000	67.000

Multivariate Tests[a]

Effect		Sig.	Partial Eta Squared
Time	Pillai's Trace	.006	.143
	Wilks' Lambda	.006	.143
	Hotelling's Trace	.006	.143
	Roy's Largest Root	.006	.143
Time * Spatial_Training	Pillai's Trace	.240	.042
	Wilks' Lambda	.240	.042
	Hotelling's Trace	.240	.042
	Roy's Largest Root	.240	.042

a. Design: Intercept + Spatial_Training

Within Subjects Design: Time

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: BNLE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser
Time	.322	75.942	2	.000	.596

Mauchly's Test of Sphericity[a]

Measure: BNLE

Within Subjects Effect	Epsilon Huynh-Feldt	Lower-bound
Time	.609	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.[a]

a. Design: Intercept + Spatial_Training

Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: BNLE

Source		Type III Sum of Squares	df	Mean Square	F
Time	Sphericity Assumed	368.985	2	184.492	4.340
	Greenhouse-Geisser	368.985	1.192	309.594	4.340
	Huynh-Feldt	368.985	1.219	302.734	4.340
	Lower-bound	368.985	1.000	368.985	4.340
Time * Spatial_Training	Sphericity Assumed	49.935	2	24.967	.587
	Greenhouse-Geisser	49.935	1.192	41.897	.587
	Huynh-Feldt	49.935	1.219	40.969	.587
	Lower-bound	49.935	1.000	49.935	.587
Error(Time)	Sphericity Assumed	5780.873	136	42.506	
	Greenhouse-Geisser	5780.873	81.045	71.329	
	Huynh-Feldt	5780.873	82.881	69.749	
	Lower-bound	5780.873	68.000	85.013	

Tests of Within-Subjects Effects

Measure: BNLE

Source		Sig.	Partial Eta Squared
Time	Sphericity Assumed	.015	.060
	Greenhouse-Geisser	.034	.060
	Huynh-Feldt	.033	.060

	Lower-bound	.041	.060
Time * Spatial_Training	Sphericity Assumed	.557	.009
	Greenhouse-Geisser	.474	.009
	Huynh-Feldt	.477	.009
	Lower-bound	.446	.009
Error(Time)	Sphericity Assumed		
	Greenhouse-Geisser		
	Huynh-Feldt		
	Lower-bound		

Tests of Within-Subjects Contrasts

Measure: BNLE

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Linear	361.535	1	361.535	6.122	.016
	Quadratic	7.450	1	7.450	.287	.594
Time * Spatial_Training	Linear	49.085	1	49.085	.831	.365
	Quadratic	.850	1	.850	.033	.857
Error(Time)	Linear	4015.458	68	59.051		
	Quadratic	1765.414	68	25.962		

Tests of Within-Subjects Contrasts

Measure: BNLE

Source	Time	Partial Eta Squared
Time	Linear	.083
	Quadratic	.004
Time * Spatial_Training	Linear	.012
	Quadratic	.000
Error(Time)	Linear	
	Quadratic	

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Bounded NLE 1	Based on Mean	.040	1	68	.842
	Based on Median	.002	1	68	.963
	Based on Median and with adjusted df	.002	1	67.045	.963
	Based on trimmed mean	.022	1	68	.882
Bounded NLE 2	Based on Mean	1.454	1	68	.232
	Based on Median	.885	1	68	.350
	Based on Median and with adjusted df	.885	1	65.666	.350
	Based on trimmed mean	1.335	1	68	.252
Bounded NLE 3	Based on Mean	1.178	1	68	.282

	Based on Median	1.030	1	68	.314
	Based on Median and with adjusted df	1.030	1	39.016	.316
	Based on trimmed mean	.743	1	68	.392

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.[a]

a. Design: Intercept + Spatial_Training

Within Subjects Design: Time

Tests of Between-Subjects Effects

Measure: BNLE

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	20398.526	1	20398.526	321.341	.000	.825
Spatial_Training	57.948	1	57.948	.913	.343	.013
Error	4316.591	68	63.479			

Estimates

Measure: BNLE

Time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	8.385	.390	7.607	9.164
2	9.593	.592	8.412	10.774
3	11.601	1.273	9.060	14.141

Pairwise Comparisons

Measure: BNLE

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[b]	95% Confidence Interval for Difference[b]	
					Lower Bound	Upper Bound
1	2	-1.208*	.464	.034	-2.346	-.070
	3	-3.215*	1.299	.048	-6.405	-.026
2	1	1.208*	.464	.034	.070	2.346
	3	-2.007	1.320	.399	-5.248	1.233
3	1	3.215*	1.299	.048	.026	6.405
	2	2.007	1.320	.399	-1.233	5.248

Based on estimated marginal means

*. The mean difference is significant at the .05 level.

b. Adjustment for multiple comparisons: Bonferroni.

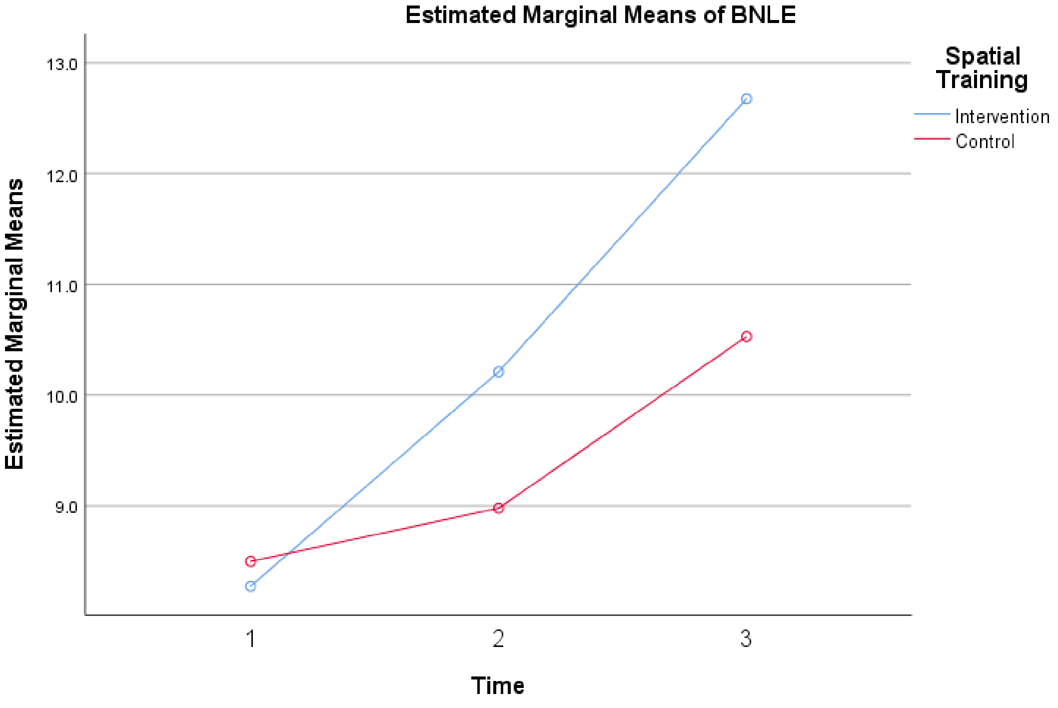
Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.143	5.611[a]	2.000	67.000	.006	.143
Wilks' lambda	.857	5.611[a]	2.000	67.000	.006	.143
Hotelling's trace	.167	5.611[a]	2.000	67.000	.006	.143
Roy's largest root	.167	5.611[a]	2.000	67.000	.006	.143

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Profile Plots



General Linear Model

Within-Subjects Factors

Measure: UBNLE

Time	Dependent Variable
1	UNLE_1
2	UNLE_2
3	UNLE_3

Between-Subjects Factors

		Value Label	N
Spatial Training	1	Intervention	34
	2	Control	36

Descriptive Statistics

	Spatial Training	Mean	Std. Deviation	N
Unbounded NLE 1	Intervention	14.171	10.2557	34
	Control	13.578	9.2450	36
	Total	13.866	9.6823	70
Unbounded NLE 2	Intervention	14.012	8.9669	34
	Control	12.711	8.4347	36
	Total	13.343	8.6586	70
Unbounded NLE 3	Intervention	14.335	10.1896	34
	Control	14.056	10.0890	36
	Total	14.191	10.0652	70

Box's Test of Equality of Covariance Matrices[a]

Box's M	7.417
F	1.177
df1	6
df2	33194.011
Sig.	.315

Multivariate Tests[a]

Effect		Value	F	Hypothesis df	Error df
Time	Pillai's Trace	.017	.586[b]	2.000	67.000
	Wilks' Lambda	.983	.586[b]	2.000	67.000
	Hotelling's Trace	.017	.586[b]	2.000	67.000
	Roy's Largest Root	.017	.586[b]	2.000	67.000
Time * Spatial_Training	Pillai's Trace	.007	.224[b]	2.000	67.000
	Wilks' Lambda	.993	.224[b]	2.000	67.000
	Hotelling's Trace	.007	.224[b]	2.000	67.000
	Roy's Largest Root	.007	.224[b]	2.000	67.000

Multivariate Tests[a]

Effect		Sig.	Partial Eta Squared
Time	Pillai's Trace	.560	.017
	Wilks' Lambda	.560	.017
	Hotelling's Trace	.560	.017
	Roy's Largest Root	.560	.017
Time * Spatial_Training	Pillai's Trace	.800	.007
	Wilks' Lambda	.800	.007
	Hotelling's Trace	.800	.007
	Roy's Largest Root	.800	.007

a. Design: Intercept + Spatial_Training

Within Subjects Design: Time

b. Exact statistic

Mauchly's Test of Sphericity[a]

Measure: UBNLE

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b] Greenhouse-Geisser
Time	.882	8.426	2	.015	.894

Mauchly's Test of Sphericity[a]

Measure: UBNLE

Within Subjects Effect	Epsilon Huynh-Feldt	Lower-bound
Time	.930	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.[a]

a. Design: Intercept + Spatial_Training

Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: UBNLE

Source		Type III Sum of Squares	df	Mean Square	F
Time	Sphericity Assumed	24.751	2	12.376	.391
	Greenhouse-Geisser	24.751	1.789	13.838	.391
	Huynh-Feldt	24.751	1.861	13.302	.391
	Lower-bound	24.751	1.000	24.751	.391
Time * Spatial_Training	Sphericity Assumed	9.567	2	4.783	.151
	Greenhouse-Geisser	9.567	1.789	5.349	.151
	Huynh-Feldt	9.567	1.861	5.141	.151
	Lower-bound	9.567	1.000	9.567	.151
Error(Time)	Sphericity Assumed	4306.317	136	31.664	
	Greenhouse-Geisser	4306.317	121.626	35.406	
	Huynh-Feldt	4306.317	126.532	34.033	
	Lower-bound	4306.317	68.000	63.328	

Tests of Within-Subjects Effects

Measure: UBNLE

Source		Sig.	Partial Eta Squared
Time	Sphericity Assumed	.677	.006
	Greenhouse-Geisser	.654	.006
	Huynh-Feldt	.662	.006

	Lower-bound	.534	.006
Time * Spatial_Training	Sphericity Assumed	.860	.002
	Greenhouse-Geisser	.837	.002
	Huynh-Feldt	.845	.002
	Lower-bound	.699	.002
Error(Time)	Sphericity Assumed		
	Greenhouse-Geisser		
	Huynh-Feldt		
	Lower-bound		

Tests of Within-Subjects Contrasts

Measure: UBNLE

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Linear	3.609	1	3.609	.094	.760
	Quadratic	21.142	1	21.142	.846	.361
Time * Spatial_Training	Linear	.857	1	.857	.022	.882
	Quadratic	8.710	1	8.710	.349	.557
Error(Time)	Linear	2607.870	68	38.351		
	Quadratic	1698.447	68	24.977		

Tests of Within-Subjects Contrasts

Measure: UBNLE

Source	Time	Partial Eta Squared

Time	Linear	.001
	Quadratic	.012
Time * Spatial_Training	Linear	.000
	Quadratic	.005
Error(Time)	Linear	
	Quadratic	

Levene's Test of Equality of Error Variances[a]

		Levene Statistic	df1	df2	Sig.
Unbounded NLE 1	Based on Mean	.553	1	68	.460
	Based on Median	.466	1	68	.497
	Based on Median and with adjusted df	.466	1	67.833	.497
	Based on trimmed mean	.574	1	68	.451
Unbounded NLE 2	Based on Mean	.034	1	68	.854
	Based on Median	.049	1	68	.825
	Based on Median and with adjusted df	.049	1	67.760	.825
	Based on trimmed mean	.031	1	68	.861
Unbounded NLE 3	Based on Mean	.061	1	68	.806
	Based on Median	.126	1	68	.723
	Based on Median and with adjusted df	.126	1	66.363	.723
	Based on trimmed mean	.087	1	68	.769

Tests of Between-Subjects Effects

Measure: UBNLE

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	40019.707	1	40019.707	190.459	.000	.737
Spatial_Training	27.527	1	27.527	.131	.719	.002
Error	14288.333	68	210.123			

Estimates

Measure: UBNLE

Time	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	13.874	1.166	11.548	16.200
2	13.361	1.040	11.286	15.437
3	14.195	1.212	11.777	16.614

Pairwise Comparisons

Measure: UBNLE

(I) Time	(J) Time	Mean Difference (I-J)	Std. Error	Sig.[a]	95% Confidence Interval for Difference[a]	
					Lower Bound	Upper Bound
1	2	.513	1.011	1.000	-1.969	2.995
	3	-.321	1.047	1.000	-2.892	2.249
2	1	-.513	1.011	1.000	-2.995	1.969
	3	-.834	.773	.853	-2.731	1.063

3	1	.321	1.047	1.000	-2.249	2.892
	2	.834	.773	.853	-1.063	2.731

Based on estimated marginal means

a. Adjustment for multiple comparisons: Bonferroni.

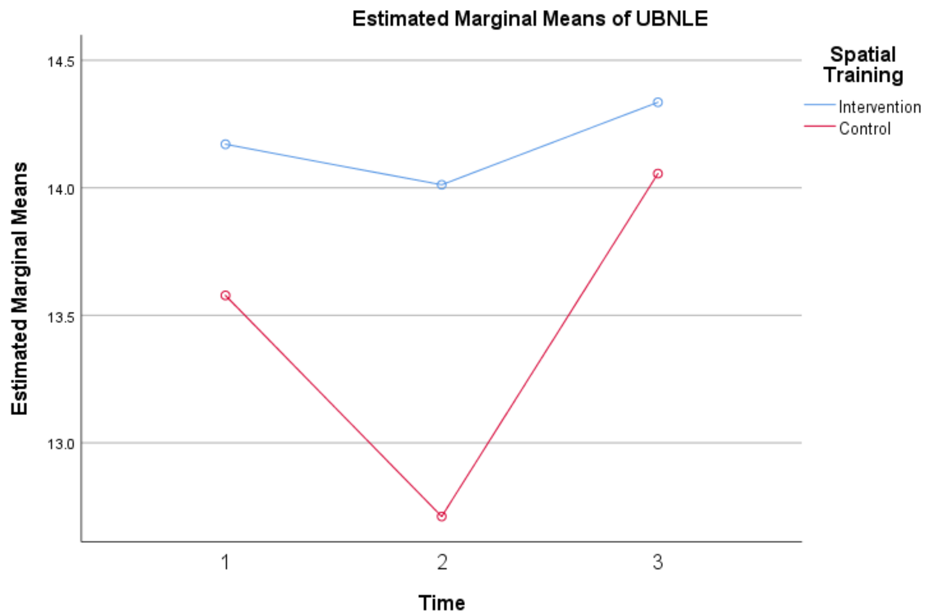
Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	.017	.586[a]	2.000	67.000	.560	.017
Wilks' lambda	.983	.586[a]	2.000	67.000	.560	.017
Hotelling's trace	.017	.586[a]	2.000	67.000	.560	.017
Roy's largest root	.017	.586[a]	2.000	67.000	.560	.017

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Profile Plots



List of References

Alloway, T. P., & Passolunghi, M. C. (2011). The relationship between working memory, IQ, and mathematical skills in children. *Learning and Individual Differences*, *21*(1), 133-137.

Aulet, L. S., & Lourenco, S. F. (2018). The Developing Mental Number Line: Does Its Directionality Relate to 5-to 7-Year-Old Children's Mathematical Abilities? *Frontiers in Psychology*, *9*, 1142.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: Evidence against a representational shift. *Developmental Science, 14*(1), 125-135.

Bokhove, C., & Redhead, E. (2017). Training mental rotation skills to improve spatial ability. *BSRLM Proceedings*, *36*(3).

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology*, *42*(1), 189.

Booth, J. L., & Siegler, R. S. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development*, *79*(4), 1016-1031.

Bruner, J. (1966). *Toward a theory of instruction*. Cambridge, MA: Belknap Press.

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

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

Carlson, A. G., Rowe, E., & Curby, T. W. (2013). Disentangling fine motor skills' relations to academic achievement: The relative contributions of visual-spatial integration and visual-motor coordination. *The Journal of Genetic Psychology*, *174*(5), 514-533.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* Cambridge: Cambridge University Press.

Case, R., & Okamoto, Y. (1996). The role of central conceptual structures in the development of children's thought. *Monographs of the Society for Research in Child Development*, *61*(1-2).

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*, *15*(1), 2-11.

Choi, B. C. (1998). Perspectives on epidemiologic surveillance in the 21st century. *Chronic Diseases in Canada*, *19,* 145-151.

Cipora, K., and Nuerk, H.-C. (2013). Is the SNARC effect related to the level of mathematics? No systematic relationship observed despite more power, more repetitions, and more direct assessment of arithmetic skill. *Quarterly Journal of Experimental Psychology 66*, 1974–1991. doi: 10.1080/17470218.2013.772215

Cohen, D. J., & Blanc-Goldhammer, D. (2011). Numerical bias in bounded and unbounded number line tasks. *Psychonomic Bulletin and Review*, *18*(2), 331-338.

Cohen, L., Manion, L., & Morrison, K. (2000). *Research methods in education 6th edition*. London, RoutledgeFalmer.

Cohen, D. J., & Sarnecka, B. W. (2014). Children's number-line estimation shows development of measurement skills (not number representations). *Developmental psychology*, *50*(6), 1640.

Cowan, R., Donlan, C., Shepherd, D.-L., Cole-Fletcher, R., Saxton, M., & Hurry, J. (2011). Basic calculation proficiency and mathematics achievement in elementary school children. *Journal of Educational Psychology, 103*(4), 786-803. http://dx.doi.org/10.1037/a0024556

De Hevia, M. D., & Spelke, E. S. (2009). Spontaneous mapping of number and space in adults and young children. *Cognition*, *110*(2), 198-207.

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. Journal of Genetic Psychology: *Research and Theory on Human Development, 163*(3), 272–282. doi: 10.1080/00221320209598683

Dehaene, S. (2011). *The number sense: How the mind creates mathematics*. OUP USA.

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, *122*(3), 371.

Dehaene, S., Molko, N., Cohen, L., & Wilson, A. J. (2004). Arithmetic and the brain. *Current Opinion in Neurobiology*, *14*(2), 218-224.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence*, *32*(1), 25-32.

DfE, (2018). *National Statistics: National curriculum assessments at key stage 2 in England, 2018 (interim.* Accessed from https://www.gov.uk/government/publications/national-curriculum-assessments-key-stage-2-2018-interim/key-stage-2-2018-interim-results-text

Edd, L. A. (2001). Mental rotations, gender, and experience (Master's thesis). *ProQuest Dissertations and Theses database.* (UMI No. 1407726)

Education and Training Foundation (2016). *Annual Review 2015/16*. Accessed from https://www.et-foundation.co.uk/wp-content/uploads/2016/12/ETF_Annual_Report_AW_v3_Pages.pdf

Eikenberry, K. D. (1988). The effects of learning to program in the computer language Logo on the visual skills and cognitive abilities of third grade children (Doctoral dissertation). *ProQuest Dissertations and Theses database*. (UMI No. 8827343)

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behaviour research methods*, *41*(4), 1149-1160.

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.

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

Flitcroft, D., & Woods, K. (2018). What does research tell high school teachers about student motivation for test performance? *Pastoral Care in Education*, *36*(2), 112-125.

Frank, M. C., & Barner, D. (2012). Representing exact number visually using mental abacus. *Journal of Experimental Psychology: General*, *141*(1), 134.

Gallistel, C. R., & Gelman, R. (2000). Non-verbal numerical cognition: From reals to integers. *Trends in Cognitive Sciences*, *4*(2), 59-65.

Ganis, G. and Kievit, R.A., 2015. A New Set of Three-Dimensional Shapes for Investigating Mental Rotation Processes: Validation Data and Stimulus Set. *Journal of Open Psychology Data*, 3(1), p.e3. DOI: http://doi.org/10.5334/jopd.ai

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology*, *47*(6), 1539.

Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PloS one*, *8*(1), e54651.

Genaidy, A. M., Lemasters, G. K., Lockey, J., Succop, P., Deddens, J., Sobeih, T., & Dunning, K. (2007). An epidemiological appraisal instrument–a tool for evaluation of epidemiological studies. *Ergonomics*, *50*(6), 920-960.

Gibbon, J., & Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology: Animal Behaviour Processes*, *7*(2), 87.

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.

Goldin, G. A., & Kaput, J. J. (1996). A joint perspective on the idea of representation in learning and doing mathematics. *Theories of Mathematical Learning*, *397*.

Gough, D. (2007). Weight of evidence: a framework for the appraisal of the quality and relevance of evidence. *Research Papers in Education*, *22*(2), 213-228.

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.

Gunderson, E. A., Spaepen, E., & Levine, S. C. (2015). Approximate number word knowledge before the cardinal principle. *Journal of Experimental Child Psychology*, *130*, 35-55.

Halberda, J., & Feigenson, L. (2008). Developmental change in the acuity of the "number sense": The approximate number Number-Sense Training 2019 system in 3-, 4-, 5-, and 6-year-olds and adults. *Developmental Psychology*, *44*, 1457–1465

Hamdan, N., & Gunderson, E. A. (2017). The number line is a critical spatial-numerical representation: Evidence from a fraction intervention. *Developmental Psychology*, *53*(3), 587.

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.

Hiebert, J. & Carpenter, T. P. (1992) Learning and teaching with understanding, in: D. A. Grouws (Ed) *Handbook of Research on Mathematics Teaching and Learning*(New York, NCTM).

Hsi, S., Linn, M. C., & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of engineering education*, *86*(2), 151-158.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience*, *6*(6), 435.

Johnson-Laird, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*(No. 6). Harvard University Press.

Kaput, J. (1991) Notations and representations as mediators of constructive processes, in E. von Glasersfeld (Ed) *Constructivism and Mathematics Education* (Dordrecht, Netherlands, Kluwer).

Kaufmann, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498-525.

Kim, D., & Opfer, J. E. (2017). A unified framework for bounde

d and unbounded numerical estimation. *Developmental psychology*, *53*(6), 1088.

Kosslyn, S. M., & Pomerantz, J. R. (1977). Imagery, propositions, and the form of internal representations. *Cognitive Psychology*, *9*(1), 52-76.

Krajewski, K., & Schneider, W. (2009). Exploring the impact of phonological awareness, visual–spatial working memory, and preschool quantity–number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of Experimental Child Psychology*, *103*(4), 516-531.

Kucian, K., Grond, U., Rotzer, S., Henzi, B., Schönmann, C., Plangger, F., Galli,M., Martin, E., & von Aster, M. (2011). Mental number line training in children with developmental dyscalculia. *NeuroImage, 57*(3), 782-795.

LeFevre, J. A., Jimenez Lira, C., Sowinski, C., Cankaya, O., Kamawar, D., & Skwarchuk, S. L. (2013). Charting the role of the number line in mathematical development. *Frontiers in Psychology*, *4*, 641.

Li, X., & Baroody, A. J. (2014). Young children's spontaneous attention to exact quantity and verbal quantification skills. *European Journal of Developmental Psychology*, *11*(5), 608-623.

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, 13, 900–906.

Link, T., Nuerk, H. C., & Moeller, K. (2014). On the relation between the mental number line and arithmetic competencies. *The Quarterly Journal of Experimental Psychology*, *67*(8), 1597-1613.

Linn, M.C., & Peterson, A. C. (1985).Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479−1498.

Mazzocco, M. M., Feigenson, L., & Halberda, J. (2011). Pre-schoolers' precision of the approximate number system predicts later school mathematics performance. *PLoS one*, *6*(9), e23749.

Meyer, M. L., Salimpoor, V. N., Wu, S. S., Geary, D. C., & Menon, V. (2010). Differential contribution of specific working memory components to mathematics achievement in 2nd and 3rd graders. *Learning and Individual Differences*, *20*(2), 101-109.

Mix, K. & Cheng, Y. (2012). *The relation between space and math: Developmental and educational implications* In J. Benson (Ed.). Advances in Child Development and Behavior, pp 197-243. Burlington, VT: Academic Press.

Moeller, K., Martignon, L., Wessolowski, S., Engel, J., & Nuerk, H. C. (2011). Effects of finger counting on numerical development–the opposing views of neurocognition and mathematics education. *Frontiers in Psychology, 2*, 328.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature*, *215*, 1519 –1520. http://dx.doi.org/ 10.1038/2151519a0

Muldoon, K., Simms, V., Towse, J., Menzies, V., & Yue, G. (2011). Cross-cultural comparisons of 5-year-olds' estimating and mathematical ability. *Journal of Cross-Cultural Psychology*, *42*(4), 669-681.

Mussolin, C., Nys, J., Leybaert, J., & Content, A. (2016). How approximate and exact number skills are related to each other across development: A review. *Developmental Review*, *39,* 1-15.

Núñez, R. E. (2011). No innate number line in the human brain. *Journal of Cross-Cultural Psychology*, *42*(4), 651-668.

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.

OECD (2014). *PISA 2012 Results in Focus; what 15-year-olds know and what they can do with what they know*. Retrieved from https://www.oecd.org/pisa/keyfindings/pisa-2012-results-overview.pdf

Ofsted. (2015). *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2014/15.* Her Majesty's Stationery Office. London. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483347/Ofsted_annual_report_education_and_skills.pdf

Opfer, J. E., & DeVries, J. M. (2008). Representational change and magnitude estimation: Why young children can make more accurate salary comparisons than adults. *Cognition*, *108*(3), 843-849.

Opfer, J. E., & Siegler, R. S. (2007). Representational change and children's numerical estimation. *Cognitive psychology*, *55*(3), 169-195.

Park, J., Bermudez, V., Roberts, R. C., & Brannon, E. M. (2016). Non-symbolic approximate arithmetic training improves math performance in pre-schoolers. *Journal of Experimental Child Psychology*, *152*, 278-293.

Piazza, M. (2011). Neurocognitive start-up tools for symbolic number representations. *Space, Time and Number in the Brain* (pp. 267-285). Academic Press.

Piazza, M., Facoetti, A., Trussardi, A. N., Berteletti, I., Conte, S., Lucangeli, D.,Dehaene, S., & Zorzi, M. (2010). Developmental trajectory of number acuity reveals a severe impairment in developmental dyscalculia. *Cognition*, *116*(1), 33-41.

Rapp, D. N., & Uttal, D. H. (2006). *Understanding and enhancing visualizations: Two models of collaboration between earth science and cognitive science*. In C. Manduca & D. Mogk (Eds.), Geological Society of America Press.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of experimental child psychology*, *91*(2), 137-157.

Reed, S. K. (1974). Structural descriptions and the limitations of visual images. *Memory and Cognition*, *2*(2), 329-336.

Rouder, J. N., & Geary, D. C. (2014). Children's cognitive representation of the mathematical number line. *Developmental Science, 17*(4), 525-536.

Schneider, M., Grabner, R. H., & Paetsch, J. (2009). Mental number line, number line estimation, and mathematical achievement: their interrelations in grades 5 and 6. *Journal of Educational Psychology, 101*(2), 359.

Sella, F., Tressoldi, P., Lucangeli, D., & Zorzi, M. (2016). Training numerical skills with the adaptive videogame “The Number Race”: A randomized controlled trial on pre=schoolers. *Trends in Neuroscience and Education, 5*(1), 20-29.

Shanley, L., Clarke, B., Doabler, C. T., Kurtz-Nelson, E., & Fien, H. (2017). Early Number Skills Gains and Mathematics Achievement: Intervening to Establish Successful Early Mathematics Trajectories. *The Journal of Special Education*, *51*(3), 177-188.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701-703.

Siegler, R. S. (2016). Magnitude knowledge: The common core of numerical development. *Developmental Science, 19*(3), 341-361.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development*, *75*(2), 428-444.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science*, *14*(3), 237-250.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear number board games—but not circular ones—improves low-income pre-schoolers’ numerical understanding. *Journal of Educational Psychology*, *101*(3), 545.

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.

Simpson, A. (2017). The misdirection of public policy: Comparing and combining standardised effect sizes. *Journal of Education Policy*, *32*(4), 450-466.

Skemp, R. R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher, 26* (3), 9-1

Slusser, E. B., Santiago, R. T., & Barth, H. C. (2013). Developmental change in numerical estimation. *Journal of Experimental Psychology: General*, *142*(1), 193.

Sorby, S. A. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, *31*(3), 459-480.

Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, *26*, 20-29.

Strauss, M. S., & Curtis, L. E. (1981). Infant perception of numerosity. *Child Development*, 1146-1152.

Thompson, C. A., & Siegler, R. S. (2010). Linear numerical-magnitude representations aid children's memory for numbers. *Psychological Science*, *21*(9), 1274-1281.

Thompson, J. M., Nuerk, H. C., Moeller, K., & Kadosh, R. C. (2013). The link between mental rotation ability and basic numerical representations. *Acta psychologica*, *144*(2), 324-331.

Tyrrell, B., & Woods, K. (2018). Gathering the views of children and young people with ASD: a systematic literature review. *British Journal of Special Education*, *45*(3), 302-328.

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*, 352– 402. doi:10.1037/a0028446

van den Bos, I., Kroesbergen, E. H., Van Luit, J. E., Xenidou-Dervou, I., Jonkman, L. M., Van der Schoot, M., & Van Lieshout, E. C. (2015). Longitudinal development of number line estimation and mathematics performance in primary school children. *Journal of experimental child psychology*, *134*, 12-29.

van Galen, M. S., & Reitsma, P. (2008). Developing access to number magnitude: A study of the SNARC effect in 7-to 9-year-olds. *Journal of experimental child psychology*, *101*(2), 99-113.

van Loosbroek, E., & Smitsman, A. W. (1990). Visual perception of numerosity in infancy. *Developmental Psychology, 26*, 916-922.

Verdine, B. N., Irwin, C. M., Golinkoff, R. M., & Hirsh-Pasek, K. (2014). Contributions of executive function and spatial skills to preschool mathematics achievement. *Journal of Experimental Child Psychology*, *126*, 37-51.

Wai, J., Lubinski, D. & Benbow, C. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. Journal of Educational Psychology, 101, 817-835. doi:10.1037/a0016127

Wallace, M., & Wray, A. (2016). *Critical reading and writing for postgraduates*. Sage.

Williams, J., Clemens, S., Oleinikova, K., & Tarvin, K. (2003). The skills for life survey. *A national needs and impact survey of literacy, numeracy and ICT skills.* London: TSO.

Xu, C., & LeFevre, J. A. (2016). Training young children on sequential relations among numbers and spatial decomposition: Differential transfer to number line and mental transformation tasks. *Developmental Psychology*, *52*(6), 854.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, *74*(1), B1-B11.

Yang, T., Chen, C., Zhou, X., Xu, J., Dong, Q., & Chen, C. (2014). Development of spatial representation of numbers: A study of the SNARC effect in Chinese children. *Journal of Experimental Child Psychology*, *117*, 1-11.