The aetiology of the number sense and its relationship with mathematics: a genetically sensitive investigation

Maria Grazia Tosto

Supervised by:

Yulia Kovas

Alice M. Gregory

Submitted in accordance with the requirements for the degree of Doctor of Philosophy in Psychology

Goldsmiths, University of London

2012

DECLARATION

I declare that the work presented in this thesis is my own. All experiments and work detailed in the text of this thesis is novel and has not been previously submitted as part of the requirements of a higher degree.

Signed ____________________________________

Date ____________________________________

ABSTRACT

Number sense is defined as the process of extracting numerical information by estimating numerosity and magnitudes of numerical symbols. Humans show great variability in estimation skills from an early age. Although little is known about the origin of individual differences in number sense, these individual variations positively correlate with mathematics. This thesis presents the first large-scale genetically sensitive investigation into the origins of number sense and into the nature of its relationship with mathematics. The research plan can be devised in two parts. In the first phase, a battery of web-tests age appropriate for 16-year olds, designed to assess number sense (as measured by numerosity and magnitude estimation), mathematics and cognitive abilities was created and validated. The battery was then administered to the large UK representative of twins of the Twins Early Development Study (TEDS). In the second phase, using data from 7,598 sixteen year-old twins from the TEDS sample, this thesis used univariate and multivariate genetic analyses to investigate the contribution of genes and environment to individual differences in number sense and to its association with mathematics. The results suggested that individual differences in number sense abilities were mostly driven by non-shared environmental factors, with modest contribution of genetic factors. No average or aetiological sex differences were found in number sense. The relationship between mathematics and number sense was largely genetically mediated. However, contrary to the predictions, the genetic relationship between number sense and mathematics was found to be mediated by *g*. The existing longitudinal data in the TEDS sample was used to investigate the retrospective relationship between number sense, measured at 16, and mathematics and a range of cognitive abilities, measured at 16 and at previous ages as far back as age 7. The results suggest that the relationship between mathematics and number sense may be uneven across development. In particular, numerosity estimation may be important only at the very early stages of mathematical learning. Overall, this investigation did not find evidence that number sense is what is "special" about mathematics. The results support the Generalist Genes Hypothesis that same genes contribute to individual differences in various aspects of cognition and learning.

STATEMENT OF AUTHORSHIP

The data for this thesis were collected in two phases: a small pilot study and as part of the large collaborative longitudinal TEDS project: Twins Early Development Study. I was actively involved in the selection and preparation of the tests for the pilot study. I then executed all aspects of the pilot study: recruited participants; administered the battery to 100 16-year old students on a one to one basis twice, to enable test-retest analyses; managed, cleaned, and analysed the pilot data; selected the best-performing and most informative tests for the final mathematical battery to be implemented online for the TEDS assessment at 16 years, as described in Chapter 4. I was involved in the testing of the newly implemented mathematical battery in the TEDS website. I analysed the data for all the studies. The GTCA analysis, described in Chapter 6, was conducted by Maciej Trzaskowski. Chapter 6 has been written up for publication (submitted) and received contribution from all the co-authors on the paper. To the best of my knowledge the remaining work presented in this thesis is original and my own.

ACKNOWLEDGMENTS

I would like to say a special thank you to my supervisor, Dr. Yulia Kovas for her expert guidance and support throughout my PhD. Her scientific insights, her passion for knowledge and understanding have been inspirational. Thank you Yulia, not only for making all of this possible, but also for being a friend. My thanks go also to Dr. Alice Gregory for being a great teacher and for the enthusiastic support and encouragement during these years.

My gratitude goes to TEDS and their families for their participation in the study; and to the funding bodies that made this research possible. TEDS has been supported by the programme grant (G9424799, now G050079) from the U.K. Medical Research Council. The work on the environment and academic achievement, including mathematics, was further supported by grants of from the U.S. National Institutes of Health (NIH; HD44454, HD46167 and HD059215), and the Government of the Russian Federation (grant 11.G34.31.0043). The pilot study has been funded by the Goldsmiths College small grant.

I would like to express my gratitude to all the people involved in the development of the mathematical battery. Particularly to Dr. Stephen Petrill, whose professional advice has been of great help throughout the process. Andy MacMillan from the TEDS team and Sean Heraghty from e-businesssystems.co.uk have been instrumental in the creation of the web test battery used in this research. Ryan Lee's assistance has been crucial for the development of our version of the Dot Task. I would also like to thank Dr. Justin Halberda for allowing the use of the version of the Dot Task programmed by his lab and for his advice. I am also very grateful to Professor Robert Plomin, for his expert guidance with the TEDS data. Thanks to all the members of InLab for creating such a stimulating research environment.

Last but not least, thank you Karim, just for being there, during all the ups and downs.

This thesis is dedicated to my Mum, Ester.

TABLE OF CONTENT:

LIST OF TABLES

LIST OF FIGURES

PUBLICATION RELEVANT TO THIS THESIS

Submitted

Maria G.Tosto, Claire M. A. Haworth, & Yulia Kovas. (in press). *Behavioural Genomics of Mathematics*. The Oxford Handbook of Numerical Cognition. Oxford University Press. (Invited Article).

In Preparation

Tosto, M. G., Petrill, S. A., Plomin R., & Kovas Y. (In preparation a). Mathematics and Number Sense (or Number Senses?) across the school years: a multivariate longitudinal investigation in 7 to 16 year old school children.

Tosto, M. G., Plomin, R., Halberda, J., Trzaskowski, M., Kovas, Y. (In preparation b). How heritable is Number Sense? A genetic investigation into numerosity estimation abilities in 16 year-old students.

Tosto, M. G., Plomin, R., Petrill, S. A., Tikhomirova, T., & Kovas, Y. (In preparation c). The myth of maths: is there anything 'special' about it?

LIST OF ABBREVIATIONS

-2LL	Minus two log-likelihood
A	Additive genetic influences
AIC	Akaike Information Criterion
ANS	Approximate Number System
BIC	Bayesian Information Criterion
C	Shared environmental influences
c^2	Shared environmental influences - proportion of variance
CRB	Criminal Record Bureau certification
DZ	Dizygotic Twins
DZf	Dizygotic females
DZ m	Dizygotic males
DZos	Dizygotic opposite sex (male-male or female-female couples)
DZ ss	Dizygotic same sex
E	Non-shared environmental influences
e^2	Non-shared environmental influences - proportion of variance
GCSE	General Certificate of Secondary Education
h^2	Narrow sense heritability - proportion of variance
ICC	Intraclass Correlation
IQ	Intelligence Quotient
MZ	Monozygotic Twins
MZf	Monozygotic females
MZm	Monozygotic males

NC	National Curriculum
nferNelson	National Foundation for Educational Research (NFER)
OECD	Organisation for Economic Co-operation and Development
Ofsted	Office for Standards in Education, Children's Services and Skills
ONS	Office for National Statistics
PIAT	Peabody Individual Achievement Test
PISA	Programme for International Student Assessment
QCA	Qualifications and Curriculum Authority
r	Correlation coefficient
ra	Genetic Correlation
rc	Shared Environmental Correlation
re	Non-shared Environmental Correlation
SNARC	Spatial Numerical Association of Response Codes
TEDS	Twins Early Development Study
TOAL	Test of Adolescent and Adult Language
TOAL - 3	Test of Adolescent and Adult Language - Third Edition
TOWRE	Test of Word Reading Efficiency
WISC	Wechsler Intelligence Scale for Children (Wechsler, 1949)
WISC-III	Wechsler Intelligence Scale for Children -Third edition (Wechsler, 1991)
WISC-III-PI	Wechsler Intelligence Scale for Children -Third edition- Primary Scale of Intelligence

Chapter 1: Theoretical Background

1.1 Overview

This introductory chapter is divided in eleven sections providing a literature review to serve as a framework for the experimental chapters of this thesis. The first 7 sections give an overview of the current efforts and progress in the understanding of the number sense and of its relationship with mathematics. The chapter provides separate sections of background research in estimation of numerical magnitude (symbolic estimation in 1.2.5) and estimation of numerosity (non-symbolic estimation in 1.2.6). The most common tasks used to detect number sense are described in section 1.2.3 together with a discussion of some of the controversies and limitations of these measurements. There are also sections giving a brief overview of the theories on sex differences in mathematics (1.2.8), theories of mathematical learning and how the number sense is related to the development of numerical concepts (1.2.2). Although the relationship between mathematics and other abilities is not the primary focus of this thesis, a synopsis of these associations is included in section 1.2.7, as the relationship between mathematics and number sense may be mediated by other cognitive mechanisms. These reviews will provide the framework for the investigation of the relationship between number sense, mathematics and cognitive abilities and the involvement of number sense in sex differences in mathematical ability and achievement.

The following sections present findings from behavioural genetic investigations into the genetic and environmental contribution to mathematical abilities (1.2.9); the aetiology of the relationship with other abilities; and the aetiology of gender differences in mathematics (1.2.10).

The final section of this chapter, 1.2.11, identifies the current gap in the understanding of the number sense structure, the aetiology of individual and sex differences in number sense and the aetiology of the relationship between number sense and mathematics.

1.2 Introduction

Number sense can be broadly defined as a basic intuition about numbers and numerical material (Dehaene, 1997). This ability is considered critical for the development of early mathematical concepts. Its definition, however, is heterogeneous with researchers and mathematical educators interpreting it in different ways. One review, for example, identified thirty definitions used in the mathematical literature that refer to a number sense construct (Berch, 2005). These definitions mostly refer to basic skills such as understanding of numerical meaning, estimation of numerical magnitudes and of quantities. The concepts of number sense also extend to mechanisms such as motivation and desire to learn mathematics, a "feel" for numbers, numerical processes developed with experience and knowledge. The review also points to a disagreement in the origin of number sense as some consider this ability as a product of experience and training, others suggest individual differences in number sense to be under strong genetic influence. Despite disagreement in the definition and aetiology of number sense, its attributes can be easily identified in people. Case (1998) describes children with good number sense as able to: conceive novel methods to perform numerical operation; represent numbers independently from a context; recognise numerical patterns; show good sense of numerical magnitude; and show confidence in translating real word quantities into numerical concepts and expressions (in Gersten & Chard, 1999). In practical terms, a good number sense defines a good grasp of numbers which is often associated with flexibility in representing basic numerical material, understanding of quantity and numerosity concepts or of relationships among numbers. These skills assist the assimilation of more advanced numerical and mathematical concepts (Gallistel & Gelman, 1992; Gersten & Chard, 1999; Butterworth, 2005).

1.2.1 Number sense and estimation

Since early age, children show various degrees of numerical understanding. For example, when entering school some children already understand that 8 is bigger than 5 by 3 units. These children may have been purposely taught this information by the family or may have acquired it accidentally through family/environmetal interaction. However, some other children lacking of this knowledge can work out the numerical magnitude with the aid of strategies such as using fingers (Gersten & Chard, 1999). These are the children showing "good number sense". Gersten & Chard draw attention to estimation abilities as the components of the number sense construct relevant to early mathematics. They propose that the ability to estimate numerical magnitudes and numerosity is at the basis of the understanding of numerical concepts. They further compare this aspect of number sense to phonemic awareness for reading, proposing estimation abilities as early precursors of mathematical learning.

Gersten & Chard (1999) consider that numerical estimation abilities in many respects have similarities with phonemic awareness. Briefly, phonemic awareness underlies the ability to identify and manipulate the smallest meaningful unit of sound. It can be considered as the awareness of the sounds that make up a word. It has been found that auditory skills positively correlate with reading skills (Kavale, 1984) and that phonemic awareness was the best predictor of early reading abilities beyond IQ, socio-economic status and readiness tests scores (Adams, 1990). Instruction in phonemic skills has been proven to enhance reading skills (Cunningham, 1990; Torgesen, Morgan & Davis, 1992), although intervention seems more beneficial to children with normal abilities as compared to children with reading disability (Smith, Simmons, & Kameenui, 1998).

The following section describes theories of development of early numerical skills (counting) and the role of estimation skills in this process. These theories agree with the role of estimation skills as fundamental processes for early mathematics, as proposed by Gersten and Chard (1999).

1.2.2 Numerosity concept and counting

Knowing about numbers is not just about recognising their symbols. When children start learning numbers they associate the symbol 1 to one unit/item and the symbol 2 to two units, for example. The symbols 1 and 2 do not just differ in their notations; their meanings are also different as they are associated with different quantities. In order to understand numbers children need to be able to mentally represent them. Conceptualization of numbers has various implications. First, children need to reason that 1 is smaller than 2, as 1 is associated with a smaller quantity than 2. This means that children need to have an appreciation of the quantity/numerosity of the set associated with the symbol. In other words, prior to being able to reason about the size or magnitude of symbols children need to perform a quantity/numerosity comparison: they need to evaluate quantities in terms of more and less. When numbers are mentally represented, they are sorted according to their magnitude and ordered, with numbers smaller in size placed before the larger. This process, described in Butterworth (2005), highlights the importance of numerosity representation as the first elemental processing behind the number system. A correct mental number representation, based on a correct numerosity representation, is fundamental for counting. Gelman & Gallistel (1978) proposed a model identifying the skills that children need to acquire in order to be able to count. First, children must understand that the number words or symbols follow always the same order: the *stable order principle*. This principle derives directly from an understanding of numerosity as described above: smaller numbers (associated with smaller numerosity) are placed before larger numbers (larger numerosity). Second, children must understand that an item can be counted only once. This is known as the *one-to-one principle* and underlies the understanding that each item to be counted is associated with one and only one number word. Third, children must be aware that the total number of items to be counted corresponds to the last counting word used. This principle, known as the *cardinal principle* encompasses knowledge of numerical magnitudes as the number sequence follows number numerical magnitudes. Fourth, it is important to understand that anything can be counted (the *abstractness* *principle*). Finally, children must understand that counting can start from any of the objects in the set – *order irrelevance principle*.

The concept of numerosity and the ability to estimate numerosity is important for counting, and in turn, counting provides the basis for arithmetic. Problems such as 3 + 2, expressed in symbolic notation, can be solved in terms of arrays of numerosity: the items of the two arrays can be combined together in a new set (a 5 items set) and then counted. Alternatively, a total can be obtained by considering the two sets separately and counting on from one of the two sets, starting from 3 for example (Carpenter & Moser, 1984). Counting is a very basic process of mathematical learning: counting skills have been found to be predictive of mathematical achievement in normally functioning children (Jordan et al., 2007), as well as children at risk for low mathematical achievement (Locuniak & Jordan, 2008). Children showing problems in mathematical learning show poor conceptual understanding of some or all of the counting principles mentioned above (Geary, 1993). For example, children with mathematical learning disability often fail to understand the order irrelevance principle. These children consider wrong counting objects not in sequence (i.e. jumping from an object to a non contiguous one) thus showing poor flexibility in mental representation of numbers (Geary, 2004).

1.2.3 Number sense tasks

The evidence above shows how numerosity and number symbols are interconnected. When referring to estimation abilities as number sense abilities, often in the literature there is no clear distinction between estimations carried out in non-symbolic (discrete number of objects) and symbolic (numerals) systems. As previously mentioned, the construct of number sense involves a variety of elements. As a consequence, in addition to estimation skills in both symbolic and non-symbolic notations, studies assessing number sense in children make use of tasks that assess a wide range of numerical abilities. Based on tests validated by research into mathematical development, Jordan, Kaplan, Oláh, & Locuniak (2006) for example, included *Counting Skills*, *Number Knowledge*, *Number Transformation*, *Estimation* and *Number Patterns* as tasks assessing number sense in kindergarten children (5/6 year-olds). In this study, the *Counting Skills* principles of one-to-one correspondence, order irrelevance and abstractness as described by Gelman & Gallistel (1978), were tested with a task adapted from Geary, Hoard, & Hamson (1999). Children were shown a puppet facing a set of alternated yellow and blue chips. They were told that the puppet was learning to countand and were asked to judge whether the puppet was counting correctly or not. The correct counting involved counting from right-to-left, left-to-right, and counting all the yellow or all the blue dots first (order irrelevance and abstractness). Incorrect counting involved counting left-to-right, but counting the first dot twice, for example. Counting the same dot twice was in violation of the one-to-one correspondence; the left-to-right counting was not a real mistake but added confusion to the real error. Enumeration and Count Sequence tasks were used to test cardinality and stable order principles. In the Enumeration task the children were asked to touch and count a set of five and a set of seven stars. The Count Sequence involved verbally counting to 10 in a sequence, not allowing the children to correct themselves in case of mistakes. In *Number Knowledge* children were shown two numerals and asked which was bigger or smaller; or asked which number comes one and two numbers after a target number. Clearly, this task assesses two different types of processing: the first involves estimation of numerical magnitudes (in the symbolic system), as it requires comparison between the magnitudes of two numbers in order to judge which one is bigger or smaller. The second involves counting and some notion of basic arithmetic, as children may derive the correct answer also by adding one or two to the target number. As part of Number Knowledge children were also asked which number was placed on top of an equilateral triangle with a number in each angle. This task assessed whether children understood the relationship among numbers rather than relying on visual perception. A task like Number Knowledge clearly assesses more developed abilities compared to basic estimation skills. In fact, Number Knowledge development relies to some extent on informal numerical instruction (e.g. numerical concepts developed in a home environment) (Saxe et al., 1987). *Number Transformation* included 3 tasks of addition and subtraction: Nonverbal Calculation, Story Problems, and Number Combination. In Nonverbal

Calculation, the children were placed in front of an array of chips and told by the experimenter how many chips there were. The experimenter then covered the chips with a lid and added or removed some of the chips under the lid (without the knowledge of the children). The children were then told exactly how many chips were added/removed and had to state how many chips remained under the lid. Story Problems presented the same arithmetical operations as the Nonverbal Calculation task but the children were not presented with physical objects on which to perform the operations. They were told stories such as "Mark has three cookies, Colleen takes away one of his cookie. How many cookies does Mark have now?”. In Number Combination, the third test of Number Transformation, the problems were verbally presented but without a story. Children were asked "How much is 2 and 1?". Number Transformation is another assessment of abilities emerging from the mastery of quantities and numerosity. As the children have to calculate the exact solution to the problem, it assesses a more conventional arithmetic that relies on some training and exposure to previous instructions. *Estimation* was assessed with a test adapted from Baroody & Gatzke (1991). The children were presented with arrays of 3, 8, 15, 25 and 35 dots randomly arranged on a card. Children had to tell how many dots there were on a particular card. Correct responses were considered answers within 25% range of the actual number; for example in the card with 25 dots, a correct answer was considered ranging from 19 to 31. Only for the card with 3 dots an exact estimation was required. This Estimation task can be considered an hybrid between symbolic and non-symbolic estimation: from a non symbolic input (dots) children have to derive a symbolic output. Although this task does not require manipulation of number or numerosities, as is the case with the previous arithmetic tasks, it does require the knowledge of numbers. Therefore, in order to be able to perform this basic task children must have undergone some numerical training. *Number Patterns* assessed the ability to combine patterns using numerical rules. Children were presented with strings of beads of 3 colours, red (R), yellow (Y) and blue (B), in sequences such as: RRBYYRRBYYRRB?Y. Children had to figure out the correct colour that completed the sequence in place of the question mark. According to Ginsburg (1997), sense of number pattern is a strategic concept for the development of early mathematics because it facilitates the development of number combination. This skill seems also sensitive to training. In fact, a correlation has been found between number patterns and social class (Starkey, Klein, & Wakely, 2004).

A more conservative approach to measuring number sense was taken in another study. Fuchs et al. (2010a) used *Number Line* and *Exact Representation of Small Quantities* to assess number sense in first graders (6 to 7 year-olds). In the *Number Line* task, children were presented with a line with the edges marked respectively 0 and 100. The task required the children to correctly place on the line a series of numerals, ranging from 0 to 100 (for example, on a line from 0 to 100 the number 50 would be correctly placed in the middle of the line). Performance on this task requires very little if not the total absence of arithmetical abilities. In order to carry out Number Line tasks, children need to be able to mentally represent numbers and to estimate on the basis of numerical magnitudes. The Number Line task will be discussed in more details in section 1.2.5.2. The task *Representation of small quantities* required children to combine together numerals and/or objects in order to match a target number. The target numbers to be represented were 5 and 9. Children were given cards containing arrays with varying numbers of objects (i.e. a group of 3 circles, another group of 5 squares etc.) and some numerals. Children had to make all the possible combinations to reach 5 or 9 which were allowed by the objects and numbers in the cards. For example they could make the target 5 by grouping together an array with 2 circles and another with 3 squares, or an array with 3 circles and the numeral 2. It is clear that in addition to estimation skills in the symbolic and non-symbolic systems, this task requires some arithmetic abilities as children should have a basic understanding of addition as combination of sets. The study assessed mathematical skills with a Word Problem task, where the children had to change, compare, and equalize relationships; as well as execute addition and subtraction.

Based on the tasks used to assess number sense in the two studies mentioned above it is possible to make the following considerations:

- Estimation in the symbolic and non-symbolic notation are used together in the same tasks (e.g. task of exact representation of small quantities in Fuchs et al., 2010a). To justify such choice, there must be the underlying assumption that symbolic and non-symbolic estimation measure the same ability or at least are supported by the same mechanism.

- Number sense abilities are measured by estimation skills, while tasks requiring arithmetical knowledge are treated as mathematical outcomes (Fuchs et al. (2010a). However, in other studies (e.g. Jordan et al., 2006) Number Combination and Story Problem measure number sense - while in Fuchs et al. they were used to measure mathematical outcomes.

Such inconsistency requires further consideration. The literature recognises that number sense is a multi-dimensional construct. Berch (2005) and Jordan et al. (2006), for example, identify low order number sense skills in comparison of numerical magnitudes, understanding of numerosity and counting. They also characterise high order number sense features in abilities resulting from conventional education such as arithmetic. One question arising from such a vague distinction between mathematics and number sense is whether some number sense tasks measure arithmetic/mathematics as an outcome, rather than number sense. Another question is whether a distinction between mathematics and number sense can be properly made.

1.2.4 Estimation and comparison

Comparison and estimation of quantities in the symbolic and non-symbolic system - two aspects of number sense - have been used to investigate the development of number concepts. Tasks using estimation processes have enabled to understand how numbers are processed, how people conceptualise and internally represent numbers, and how people access to the storage of semantic number knowledge. Studies have been conducted in adults and children using numerals in various modality (e.g. Arabic notation, verbal and auditory numbers) and numerosities.

1.2.5 Estimation and comparison in the symbolic system

The counting process is strictly related to our mental representation of numbers. Comparison of Arabic numerals activates some specific automatic responses that reveal how numbers are mentally represented.

The *SNARC effect* (Spatial Numerical Association of Response Codes) was first revealed in parity judgement tasks that require participants to decide whether the number presented is an even number. Responses to larger numbers are given faster with the right hand; conversely, smaller numbers elicit faster responses with the left hand (Dehaene, Bossini, & Giraux, 1993). Later studies showed that the SNARC effect is not specific to parity judgement, but takes place in numerical comparison tasks (e.g. Fias, Lauwereyns, & Lammertyn, 2001). This effect is a function of the relative numerical magnitude of the numerals assessed and depends on the direction of writing: in participants writing in a right-to-left direction (participants writing in Arabic language) the effect is reversed (Dehaene et al., 1993). Numerical magnitude is an important component in number representation. Magnitude characteristics have an influence on the strategies that children use to solve arithmetic (Butterworth, Zorzi, Girelli, & Jonckheere, 2001). For this reason it has been suggested that the comparison process relies on direct retrieval from the semantic store of simple arithmetic properties (Dehaene et al., 1993). This effect points to a spatial-numerical magnitude association; numerical magnitudes are spatially organised in the ascending order that follows the direction of writing. Handedness or hemispheric dominance do not affect the response. This spatial representation is also amodal, as the SNARC effect has been detected in judgement of magnitudes in visual number words and dice patterns, as well as in auditory modality (Nuerk, Wood, Willmes, 2005). The SNARC effect is a strong indicator that numerical mental representation is culturally determined (by the direction of writing); however many factors may influence the way we process numbers. Japanese language is written right-to-left in vertical columns top-to-bottom. Interestingly, in one study, Japanese writers associated small numbers with the bottom response keys and large numbers with the top response keys - which is in conflict with the writing direction top-to-bottom (Ito & Hatta, 2004). This phenomenon has been explained by the way we physically perceive "more" as something taller and therefore at the top of a scale, while "less/smaller" is placed at the bottom. Once again, this suggests that the mental organisation of numbers is culturally or environmentally affected.

Although the evidence above supports a spatial coding of numbers, research also suggests a complex relationship between numbers and space. For example negative numbers did not show a reliable spatial association in parity judgement tasks (Nuerk, Iversen, & Willmes, 2004). In fact, negative numbers larger in magnitude were associated with right responses but when the judgement was to be made between a negative and a positive number (-9 and 9 for example) with 0 in the middle as reference, the negative numbers were associated with left space suggesting a degree of automaticity in magnitudes representation and their spatial association (Fischer & Rottmann, 2004). Neuropsychological investigations have suggested that number spatial representation takes into account internal representation of numerical information. Brain damaged patients displaying hemi-neglect (with inability to attend the left side of the visual field) had selective difficulties attending the left side of space when presented with a fixed reference. The difficulty in attending the left side was observed for different references (when the number showed was either 5 or 7). This evidence suggests that the association small-numbers with the left space and large numbers with right space seems to be dynamically driven by referent frames rather than carried out on the basis of a permanent representation of numbers. Studies that have used different variants of the SNARC effect suggest that single and double digit numbers may have separate mental representations (Fias, Lammertyn, Reynvoet, Dupont, & Orban, 2003; Nuerk, Weger, & Willmes, 2001). Under these circumstances a single number line would provide a restrictive model for what in reality could be a flexible mental representation of numbers.

Another phenomenon observed during tasks of magnitude comparison (in deciding between two numbers which one is the largest) is the increase in error rate and reaction time as a linear function of the numerical distance between two numbers: performance for discriminating two numerosities declines as the distance betweenthe two decreases (Moyer & Landauer, 1967). In practice, it is easier to decide that 8 is bigger than 4 compared to 5 vs 4. This finding, known as the *numerical distance effect* suggests that when two numerals are presented, they are not seen as abstract discrete entities but they are automatically converted, from symbols to an analog format (mentally represented in a continuum, as in a continuous line). Therefore the magnitude comparison is based on the same process underlying comparison of continuous physical properties. Another phenomenon occurring when comparing numerals is that the error rates and reaction times on responses increase if the distance between the numbers is maintained constant but the absolute value of the numeral increases (e.g. it is easier to compare 10 vs 20 than 100 vs 110) (Whalen, Gallistel, & Gelman, 1999). This finding, the *number size effect*, suggests that we represent larger numbers (in magnitude) as more close to one another compared to smaller numbers. This means that mental representation of larger numerical magnitudes is much noisier compared to the smaller numerical magnitudes. As larger numbers are represented overlapping with neighbouring numbers, it is more difficult to tell them apart.

Similarly to the distance effect, numerical Stroop tasks give an indication that number processing takes into account some perceptual properties of the stimulus together with its numerical information. When comparing two written numerals, 8 vs 4 for example, the physical size of the number interferes with its magnitude. In a congruent trial, the physical size corresponds with the magnitude of the digits to be compared (e.g. 8 vs 4); an incongruent trial will have a mismatch between the size and magnitude (e.g. 8 vs 4). Incongruent trials are processed slower than congruent trials by both adults and children (e.g. Besner & Coltheart, 1979; Tzelgov, Meyer, & Henik, 1992; Rubinsten, Henik, Berger, & Shahar-Shalev, 2002). This *size-congruity effect* is a well established phenomenon and shows that physical size, although being task irrelevant, is automatically processed in numerical comparison.

1.2.5.1 Magnitude representation on a number line

Studies on magnitude comparison suggest that mental representation of numbers is a construct derived by practice with number symbols. Numbers in their symbolic system, although referring to discrete items and real/physical quantities, have the characteristics of abstractness. However, number processing is bound to physical properties due to perceptual transformations automatically made at the sight of numbers. Research also suggests that there is a one-to-one correspondence between numbers and specific locations in space that takes the trajectory of a mental number line (e.g. Dehaene, Piazza, Pinel, Cohen, 2003; Zorzi, Priftis, Umiltá, 2002).

Two major models provide an account of how people of all ages, from childhood to adulthood, represent numerical magnitudes along a number line. According to the *logarithmic ruler representation* (e.g. Dehaene, 1997, Dehaene & Mehler, 1992), the mental distance between numbers of small magnitudes (at the beginning of the line) is overestimated, in comparison to the distance between numbers with larger magnitudes. According to this model we think of the distance between 1 and 100 as greater than the distance between 700 and 800, that is to say that numbers at the end of our mental number line are represented in a "compressed" fashion. The alternative *accumulator model*, proposed by Gibbon and Church (1981) (but also Gallistel & Gelman, 1992; 2000; Whalen et al., 1999), suggests that the distance between numbers is constant, although this distance has scalar variability with the increase of the magnitude. In effect, it is difficult to assess which of the two models provides the more realistic account to mental numerical representation. The two models propose the same behavioural outcome: comparison between numbers in the high range is slower and less accurate than that between smaller numerical magnitudes. This is because magnitude representation of larger numbers, in both models, is noisier due to the closeness and overlapping with neighbouring magnitudes.

Another theory suggests that numbers are actually represented with more than one scale/model and that the scale used depends on the task demands (e.g. Siegler & Opfer, 2003; Lourenco & Longo, 2009). Siegler & Opfer reason that many studies show how children have poorer estimation skills compared to adults (e.g. Sowder & Wheeler, 1989; LeFevre, Greenham, & Naheed, 1993). Children poorer estimation skill shave been attributed to different causes, among which the inappropriate manipulation of number symbols or lack of number sense (Case & Sowder, 1990; Joram, Subrahmanyam, & Gelman, 1989). If numbers across all ages are represented according to one of the two mentioned models, the estimation process should be carried out in the same way in adults and children. Siegler & Opfer (2003) found that although even young children possess multiple numerical representations, they rely on an imprecise/inefficient logarithmic representation of numbers. With development, exposure to numerical knowledge and increased numerical skills, children are able to use the most appropriate numerical representation in various circumstances and tend to use the more accurate linear representation more often. In adulthood, estimation skills peak because estimation relies almost exclusively on the linear representation, although a logarithmic account of numbers remains present.

1.2.5.2 Estimation on a number line

To assess whether numerical mental representation follows a linear or logarithmic pattern, the most used test is the *Number Line task*. In a number-to-position task of estimation participants are presented with a line with the left edge usually marked with 0 and the right edge marked 100 or 1,000. Participants are told that the line goes from zero to one hundred (or one thousand) and are asked to place some numerals between zero and one hundred (one thousand), as accurately as possible. The level of accuracy in estimation is indexed by the discrepancy between the number to estimate and the number actually estimated. The score is calculated as the error or the absolute value of the difference between the estimated and the actual numbers divided by the scale of estimation. In a practical example if, on a line going from 0 to 100, a child positions number 20 in the place where 30 should be, the error is (20-30)/100 = .10. Figure 1.1 shows that a linear representation of numbers provides the most accurate estimation as each number on the line corresponds to the same value for the estimation and the number to be estimated (y = x). The curved line represents a logarithmic representation. From the graph it is clear that on this line the distance between numbers from 0 and ~150 is overestimated compared to the distance between numbers from 750 and 1,000 for example.

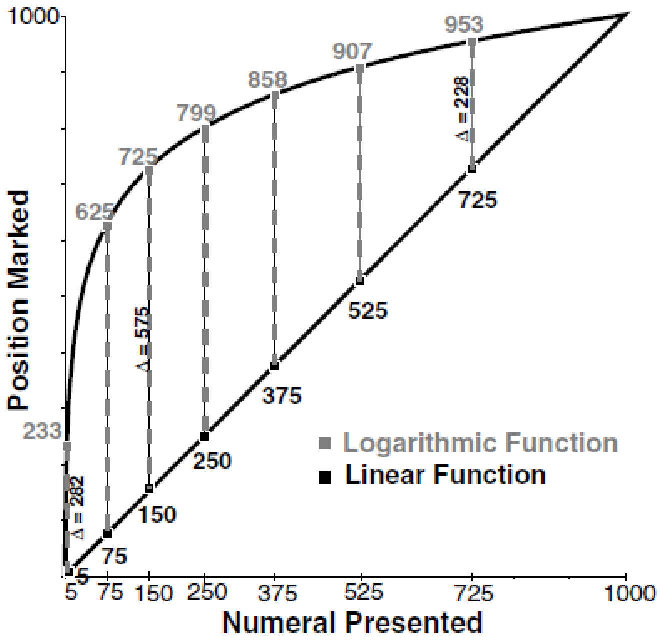


Figure 1.1: Linear vs Logarithmic representation.

The line at 45^{o}, represents a linear numerical representation. The curved line represents a logarithmic numerical representation. The greatest errors in estimations are observed for numeral 150. (from Opfer & Siegler, 2006).

Figure 1.2 shows the progression from a logarithmic to a linear representation as a consequence of development (Siegler & Booth, 2004). The children in this study were presented with a line from 0 to 100. The choice of the scale was determined by the numerical knowledge of the participants; the majority of children between kindergarten and second grade have experience with numbers within the range 0-100, while estimation on a range 0-1,000 may prove not to be age appropriate. Performance-estimation in children as young as 5 years of age fitted the logarithmic function, indicating poorer estimation skills compared to older children (7-8 years); Second Graders produced estimates consistent with the linear representation of numerical magnitudes.

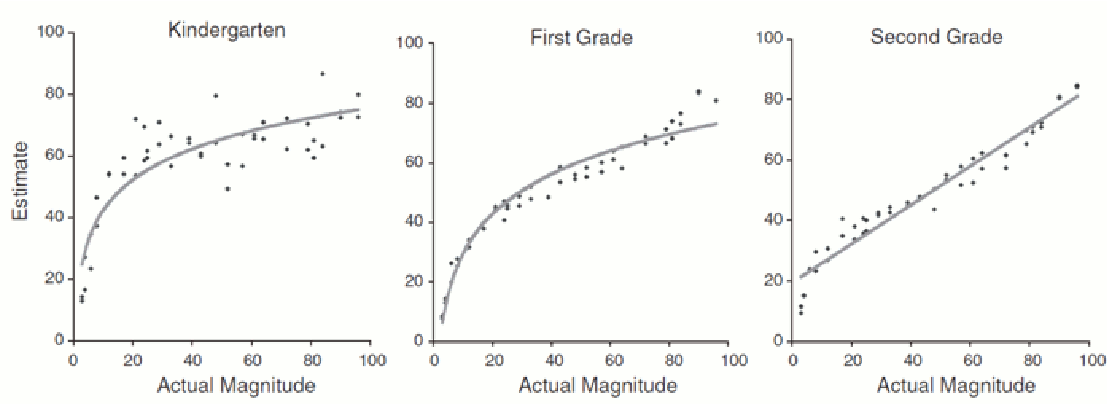


Figure 1.2: Progression from a Logarithmic to a Linear pattern of response in children.
The first panel shows performance on the number line estimation in Kindergarten children (5.8 years). The middle panel shows performance for children in the First Grade (6.9 years), the last panel shows performance for Second Graders (7.8 years). The panels display the fit function for each performance. The best fit for Kindergarten children was the logarithmic function. In the First Graders the fit of the linear and logarithmic functions did not differ, thus indicating the equal use of both representations. The liner function was the best fit for Second Graders (from Siegler & Booth, 2004).

Improvements may have occurred as a result of more mature motor control that gave children the flexibility to position the mark on the line in the wanted position; other improvements may have occurred in thinking processes. Lastly, with age, children accumulate experience in dealing with numerical material and may simply improve performance because they have more knowledge about numbers.

We use estimation in everyday life when making judgements about real quantities. For example we can approximate how much the food in our shopping trolley will cost, or, we can lift an object to judge its approximate weight. These kinds of estimations require knowledge of the measurement scale. In the aforementioned examples, in order to make the estimations we need to have some previous knowledge about money and weight. In addition, intuitively, practice with the scale of measurement may lead to more accurate estimations; for example, having experience with weighting items, people may develop sensitivity to variations in weight, leading to more accurate estimates. Some estimation processes convert measurement scales. If for example we need to estimate how long it will take to go from one place to another, we make a translation from length to time. Some other estimations may involve transforming a non numerical entity to a numeral (in weight judgement for example). All these estimations require previous knowledge of the measurement scale or familiarity with the entity assessed. Estimation on a number line, on the other hand, relies on the mental numerical representation of numbers and the knowledge of number symbols, and does not require previous knowledge or training/familiarity of measurement scales.

Despite this, estimation on a number line, to a certain extent, is driven by training. Siegler & Booth (2004) found that older children had less variability in number line scores compared to younger children. In addition, children who were given feedback on response during the number line task were more accurate compared to the children who did the task without any correction of the errors. This suggests that experience/training helps to understand numbers and their representation. This hypothesis is supported by the finding that numerical activities such as playing numerical board games improved children's estimation abilities on the number line (Siegler & Ramani, 2008; 2009).

1.2.5.3 Number line estimation and mathematics

Estimation skills are linked to children's numerical knowledge. In turn, numerical knowledge exhibited as early as in kindergarten has been shown to be predictive of mathematical achievement in later years up to high school (e.g. Duncan et al., 2007; Stevenson & Newman, 1986). Several studies have found a correlation between mathematical achievement and performance on the number line task (e.g. Siegler & Booth, 2004; Booth & Siegler, 2006; 2008; Fuchs et al., 2010a; Geary, 2011). As discussed earlier, numerical magnitude representation is linked to various cultural factors. The correlation between number line estimation and mathematics however, was found across cultures (Siegler & Mu, 2008), despite average cross-cultural differences in number line estimation. Chinese kindergarten children were significantly more accurate in a number line task compared to their American peers. This advantage was attributed to the greater exposure to numbers and to more practice using them, prior to starting schooling in China as compared to the US.

Although much evidence suggests a link between estimation abilities and mathematics, the nature and the direction of this relationship is yet to be established. Intuitively, formal mathematical education should provide some of the training responsible for accurate estimations. Research however, shows that improvement in numerical estimation does not profit from a gradual and systematic learning typical of school training. The shift from a logarithmic to a linear pattern of estimation often happens rapidly and suddenly, as a result of feedback on direct performance for example (Opfer & Siegler, 2007). It is possible that estimation drives performance in mathematics and not the other way around. In one intervention study, 7 year-old children received training on a number line task. As compared to the control group, the trained children, in addition to showing improved estimation skills, showed increased learning of novel mathematical problems that lasted several weeks after the intervention (Booth & Siegler, 2008).

1.2.5.4 Number line estimation and other cognitive abilities

The correlation between performance on number line tasks and mathematical performance may lead to the assumption that the development of estimation skills is supported by number specific systems. Although this thesis will address estimation abilities as an aspect of number sense in the general population, research on disabilities has linked estimation of numerical magnitudes to other cognitive abilities. Individuals with Williams syndrome perform poorly on tasks of magnitude comparison, showing that a general cognitive impairment is linked to impairment in magnitude representation (Patterson et al., 2006). Brain damage to the right parietal areas resulting in unilateral neglect (inability to attend to stimuli in the left side of the visual area), produces deficits in mental imagery and disrupts the ability to think of numbers in spatial terms, along a mental number line (Zorzi et al., 2002). This finding suggests an involvement of visual skills in the mental representation of numerical magnitudes. At the age of 7 and 8 years, individual differences in IQ have been found to be related to the overall accuracy in number line estimation and in the use of the logarithmic or linear pattern of responses, suggesting that logical thinking associated with intelligence assists the learning of the numerical knowledge and structure of the number line (Geary, Hoard, Nugent, Byrd-Craven, 2008). The same study found an association of the central executive component of working memory and number line in children with mathematical learning disability, thus involving central executive functions in the development of number-magnitude representation. However, some relationships between number-magnitude representation and other abilities may not be stable. An association between visuo-spatial working memory and number line was found only when the children were in the first grade, at around 7 years of age (Geary et al., 2008). When re-tested in the second grade, visuo-spatial working memory was no longer predictive of number line performance. The authors reasoned that once the children have mastered the structure of the number line, its mental representation relies on central executive attentional control rather than visual imagery.

To summarise, the way numbers are mentally represented is central to the development of early arithmetic. The most established tool to assess this mental representation is with a number line task. There is evidence of great individual variation in estimation skills, although to date the sources of these individual differences are unclear. Because of the relationship between mathematics and accuracy in number line estimation, understanding the aetiology of individual differences in this task may help to understand the origins of individual differences in mathematical achievement. Individual differences in number line estimation skills and their relationship with mathematics are investigated in this thesis using a genetically sensitive design.

1.2.6 Non-symbolic comparison and estimation

Estimation on a number line has been regarded as a "pure" process as it does not involve past exposure to the scale-measurement used or the entity assessed (Siegler & Booth, 2004). This process however requires at least the acquired knowledge of numerical symbols.

Exact representation of numerosities can be performed if the numerosities are translated into numerals. Numerosities however, can be estimated in terms of more/less or approximated in terms of greater/smaller than. This very basic type of processing is unlikely to rely on formal education. In fact the ability to discriminate more from less, as well as being present in adult humans, has been detected in human infants and in various animal species (e.g. Pica, Lemer, Izard, Dehaene, 2004; Lipton & Spelke, 2003; Xu & Spelke, 2000; Agrillo, Dadda, Serena, & Bisazza, 2009; Reznikova & Ryab, 2011).

1.2.6.1 Non-symbolic estimation in human-infants

Human infants exhibit a basic numerical knowledge in the ability to differentiate more from less. As early as 6-months, babies are able to discriminate 8 from 16 and 16 from 32 in arrays of dots or sequences of sounds. They are able to discriminate differences with a ratio of 1:2, independently from the absolute value of the items in the array. At 9 month they succeed in distinguishing between arrays with a smaller discrepancy between them, with a 2:3 ratios (Xu, Spelke & Goddard, 2005; Lipton & Spelke, 2003; Xu & Spelke, 2000; Libertus & Brannon, 2010). Studies on infants and young children show that, at the early stage of life this ability is very coarse, but improves with development. Between the age of 3 and 6 years children discriminate between 3:4 and 5:6 ratios (Halberda & Feigenson, 2008). From 14 years it is possible to discriminate ratios as fine as 9:10 (Halberda, Mazzocco, Feigenson, 2008; Pica, Lemer, Izard, & Dehaene, 2004). More recently, a study that used a dot estimation task, surveyed number sense in over 10,000 individuals between 11 and 85 years old (Halberda, Ly, Wilmer, Naiman, & Germine, 2012). The study reported individual differences and developmental changes in non-symbolic estimation skills, identifying three main transitional age-related trends in the population: a rapid increase in estimation accuracy between the age of 11 and 16 years, a steady improvement up to the age of ~30 years and a decline from 30 to 85 years.

One logical question when dealing with infants' performance is whether their response is driven by numerical information or by other continuous variables (i.e. the physical properties of the object such as the cumulative surface area, contour length or density). Studies have shown that when continuous variables are controlled for or compete for response against numerical information, infants fail to discriminate small numerosities (i.e. 1, 2 and 3) (Clearfield & Mix, 1999; Feigenson, Carey, & Spelke, 2002; Xu, 2003). For example, in work by Feigenson and colleagues, 6 month-old infants following habituation to 1 large object or 2 small objects, looked longer at the display with the objects having a novel surface area but not to the display with a novel number of objects. These findings suggest that infants cannot represent numerosities and respond on the basis of perceptual properties of the objects.

On the other hand, when the physical characteristics are controlled, infants can discriminate large numerosities (8 vs 16 and 16 vs 32) (e.g. Xu, 2003; Xu, Spelke & Goddard, 2005). These inconsistencies in babies' performance suggest that when comparing small numerosities, babies attend to the surface. However, when comparing two large numerosites, babies find the number more salient (Brannon, Abbott, Lutz, 2004). It is also possible that infants attend to the variable (numerical or continuous) depending on the context. In one experiment, 10-12 month old babies had to choose from two buckets: one containing 1 large cracker and a second with 2 smaller crackers, with a total area of half of the first cracker. The babies were allowed to keep the content of the bucket. Babies chose the container based on the surface/area of the crackers, not the number, thus maximising the quantity of food and not pieces (Feigenson, Carey, & Hauser, 2002). Although it is unclear why infants process either continuous or numerical variables in comparison of numerosities, this last study suggests that infants can represent numerosities even with small numbers.

Another recent study found that individual differences in numerosity detection at 6 month predicted the ability at 9 month beyond the babies' short-term memory (Libertus & Brannon, 2010). This suggests that individual differences in estimation of numerosities are present very early and that these differences are stable, at least in the first year of life.

1.2.6.2 Non-symbolic estimation in animals

Basic estimation skills, similar to the ones displayed in human infants, are also detected in animals. One early study on animal numerical competences found that, after training, rats were able to press a lever accordingly to a 2 or an 8 sound-sequence. In addition, the animals used the learned behaviour - responding to different numerical cues - even without reward (Meck & Church, 1983). This finding suggests that non-human animals can represent numbers using the same internal mechanism employed for timing. In more recent years, studies have provided evidence that basic numerical abilities are present in many animal species, from mammals (Beran, Evans, Leighty, Harris, & Rice, 2008) to birds (Rugani, Fontanari, Simoni, Regolin, & Vallortigara, 2009) and insects (Reznikova & Ryab, 2011).

Similarly to infants' response, with animals it is difficult to discern whether their response is driven by numerical cues or physical properties of the objects, such as the surface area or the contour. One theory suggests that animals prefer to use continuous variables in discrimination of numerosites when these are available. However, they resort to the numerical information when discrimination is not possible on the basis of other properties (Davis & Perusse, 1988; see also Davis & Memmott, 1982). There is evidence that animals spontaneously use continuous variables in numerosity discrimination (shown with cats; Pisa & Agrillo, 2009), and that they use the numerical information when they have failed the discrimination task based on other properties of the objects (e.g. shown with rats; Breukelaar & Dalrymple-Alford, 1998). This hypothesis has been challenged as there is evidence that animals automatically process numerical information irrespectively of physical properties under certain circumstances (Cantlon & Brannon, 2007; Agrillo, Dadda, Serena, & Bisazza, 2009, Agrillo, Piffer, & Bisazza, 2011, Rugani, Regolin, & Vallortigara, 2011).

In one study mosquito fish were exposed to stimuli in one modality, either numerically, continuously, or both. Fish learned to discriminate numerosities quicker when the two variables were present, as compared to when only the continuous or numerical variable was available on its own. There were no differences in learning when fish were presented with the numerical or continuous variable alone (Agrillo et al., 2011). These findings suggest that animals, even with a simple nervous system, do not find it "easier" to process continuous variables compared to numerical ones (the rate of learning was the same is the two conditions). Drawing from this evidence, it can be assumed that numerosity processing, by default, does not rely on processing of continuous variables if these are available. More importantly, these findings suggest that discrimination of numerosity is facilitated when there is another source of information in addition to the numerical one. This result is consistent with studies on infants showing that accuracy and/or reaction time improves with a multisensory stimulus (the same stimulus presented in two modalities) compared to a unisensory one (e.g. Neil et al., 2006).

Consistent with the finding above, Jordan, Suanda, & Brannon (2008) found that 6 month-old infants were able to make more accurate numerical discriminations if the same stimulus was presented simultaneously in auditory and visual modality.

In the real world, often numerical and continuous information are combined together in one stimulus. Consequently, it is more likely that processing of numerical information happens in multisensory modality and that numerical processing is aided by perceptual features of the stimuli. The studies reviewed in this section provide evidence for similarity in the way non-human animals and infants process numerosities vs continuous variables. This similarity in processing, implicitly supports a continuity between humans and animal species and suggests that numerosity estimation has been evolutionary conserved.

1.2.6.3 Non-symbolic estimation: ratio dependency and Weber Fraction

Estimation in the non-symbolic system follows similar rules to those that apply in the symbolic system. The *numerical distance effect* (it is more difficult to discriminate between 8 vs 5 compared to 6 vs 2 as their comparison is a function of their numerical distance) and *number size effect* (discrimination is harder between larger numbers compared to smaller numbers even if the numerical distance is the same, e.g. 10 vs 20 and 80 vs 90) have been observed in discrimination of numerosity sets. The phenomenon has been detected in animal studies (e.g. Gibbon, 1977; Gallistell & Gelman, 1992; Nieder & Miller, 2003), in adults (e.g. Piazza et al., 2004; Pica et al., 2004), and in infants (e.g. Libertus & Brannon, 2010).

Just as with numbers, which are more difficult to tell apart when distance between them decreases and numerical size increases, it is more difficult to discriminate between numerosities when the discrepancy between the two sets is smaller and with more items in the sets.

Discrimination of numerosities for animals, human adults and infant depends on the ratio between the sets compared. For example, comparing an array with 5 items and one with 10 (5 vs 10, ratio 0.5) should be less difficult than comparing one with 10 vs 15 (ratio 0.6) although the numerical distance between is 5 units in both examples. This pattern of response is described by the Weber Law or the law of the "just noticeable difference" (Weber, 1834). The Weber Law quantifies the minimum change we are able to perceive in a given stimulus as a constant ratio of the initial stimulus. So, if we start with an array of 100 items, we may not be able to notice 2 more items added to the set, but we may start noticing an increase after adding 20 items. 20 represents the threshold from which we start detecting the change. If we were starting with a smaller set of 50 items for example, we would start noticing the difference after adding 10 items. Intuitively, we need to add less to a smaller set to notice a variation, that is why the noticeable difference is a function of the initial stimulus and is indexed with the Weber Fraction (K) as follow: $K = \Delta I/I$, where I = initial stimulus, ΔI = incremental threshold. Therefore, the Weber Fraction measures the sensitivity of people to discern a change in the initial stimulus due an intensity increment/decrement. The Weber Fraction derived for the two examples above would be (20/100) = (10/50) = 0.2. From this ratio it is clear that smaller values from the Weber Fraction correspond to the ability to perceive smaller changes in the variation of the stimulus. This index can be applied to quantify changes in any stimulus in a variety of sensory modalities such as loudness, weight or brightness.

1.2.6.4 Non-symbolic estimation in adults and children

As previously discussed, the physical size of numbers interferes with magnitude comparison in symbolic processing tasks. What happens in non-symbolic processing and the extent to which adults process numerical information independently from continuous variables has also been investigated, mostly with non-symbolic Stroop paradigms.

Generally, Stroop tasks involve arrays of dots visually presented with congruent/incongruent trials combined with size of dots/size of the array. Studies have shown that when adults have to make judgements about the items in arrays that include both continuous (area or contours) and numerical variables available, they are unable to ignore the continuous variable (e.g. the cumulative area of the items) (Gebuis et al., 2009; Hurewitz, Gelman, & Schnitzer, 2006).

The findings are however inconsistent as other evidence suggests that continuous variables do not interfere with adults' non-symbolic numerosity judgement (e.g. Barth et al., 2006; Barth, 2008). For example, in one study adults performed numerosity comparison with cross-modal comparison (e.g. visual vs auditory) as accurately as within modality (e.g. visual vs visual) according to the Weber Law (Barth, Kanwisher, & Spelke, 2003). This finding suggests that adults are more likely to carry out judgment of non-symbolic numerosities on the basis of abstract representation of numbers rather than considering properties of the stimulus. Based on the results of the multi-modal presentation and cross-modal comparisons, the authors also concluded that the abstract representation of numerosities is achieved through multiple perceptual cues. This conclusion has a similarity with the multisensory facilitation discussed for animals and infants.

A more recent study has also shown that during numerosity comparison, adults automatically extract the continuous features of a stimulus. However, the relevant numerical information was found to be as salient as the continuous, thus holding the same chance to drive performance (Nys & Content, 2012).

As with adults, there is some controversy as to whether children process continuous or numerical variables using non-symbolic processing. Preschool children seem unable to ignore continuous task-irrelevant physical features of the stimuli in numerosity discrimination (e.g. Soltész, Szűcs, Szűcs, 2010; Gebius et al., 2009; Rousselle & Noël, 2008). Contrary to this, another study has shown that 5 year old children completed numerosity comparison tasks based on numerical variables ignoring the dot area, dot density and the dot colour (Barth et al., 2006).

To sum up, although not entirely consistent, overall evidence suggests that non-perceptual properties of the stimuli cannot be ignored during symbolic numerical processing. Adults seem to show more control over the variable features compared to children.

1.2.6.5 Non-symbolic estimation and mathematical abilities

Similarly to magnitude processing in the symbolic system, individual variations in non-symbolic numerical processing show a relationship with mathematics. Therefore, understanding the mechanisms of non-symbolic numerical processing could help to understand the mechanisms of mathematical learning. In light of the evidence examined above, it may be argued that the relationship between mathematics and non-symbolic processing may be driven by other perceptual mechanisms rather than purely numerical processing. It is important to address the issue of whether the number sense is part of some "number domain specific" construct (Fuchs et al., 2010b) or whether it is an expression of the general cognitive construct.

Regardless of the mechanism supporting this association, individual differences in non-symbolic estimation skills can be used as a predictor of mathematical skills. Mazzocco, Feigenson, & Haberda (2011) measured non-symbolic estimation skills in preschool children (age 3-6 years). The trials controlled for the size and display area of the stimuli, ensuring a response based only on the numerical information. When, two years later the same children were assessed on 7 general cognitive abilities and mathematical skills, individual differences on the non-symbolic estimation task measured earlier were found predictive only of numerical skills. Similarly, individual differences in non-symbolic estimation measured at 14 years correlated with mathematical achievement measured in the same children retrospectively all the way back to kindergarten (Halberda et al., 2008). In the same study, when controlling for 16 measures of behavioural, cognitive abilities and intelligence measured at the age of 9 years, the only correlation found was with mathematical achievement. These findings suggest a unique relationship between non-symbolic estimation and mathematics. More importantly, it means that a simple task of dot discrimination, sharing no variance with other cognitive abilities, can be used as predictor of mathematical achievement. In this respect, non-symbolic estimation can be considered an early precursor of mathematical learning, as proposed by Gersten & Chard (1999).

Studies have also shown that preschool children are able to perform non-symbolic arithmetic on numerosity sets - mentally adding the items of two different arrays and indicating whether the sum is more or less than a third item array (Barth et al., 2005). This ability correlated with symbolic mathematical achievement measured two months later, after controlling for verbal IQ (Gilmore, McCarthy, & Spelke, 2010).

In assessing the relationship between mathematical abilities and non-symbolic estimation, the nature of the tasks used needs to be carefully examined. In another longitudinal study, symbolic and non-symbolic estimation measured in kindergarten children was predictive of mathematical skills measured in grades 1 and 2 (Desoete, Ceulemans, De Weerdt, Pieters, 2010). However, it has been argued that at the time of testing the children may have had some knowledge of numbers. This, together with the fact that the display of dots did not have a fixed time, could have lead to the non-symbolic task assessing the children’s counting skills rather than numerosity estimation. In another study, tasks of non-symbolic manipulation were found to be predictive of mathematical development (Fuchs et al., 2010b). However, the tasks involved some transformation from non-symbolic to symbolic quantities and some arithmetic knowledge. These examples show that it is difficult to assess estimation skills without tapping into numerical knowledge. The observed links between non-symbolic estimation and mathematics may be reflecting the ‘mathematical aspects’ of the non-symbolic estimation measures.

Indeed, some studies have not found any relationship between pure non-symbolic estimation and mathematics in children (e.g. De Smedt & Gilmore, 2011; Soltész et al., 2010; Rousselle & Noël, 2007; Holloway & Ansari, 2009), questioning whether non-symbolic numerosity processing can be considered a precursor of early mathematics.

Some of the inconsistencies reported in the association between mathematics and non-symbolic estimation could find an explanation in the type of task used. For example, in Rousselle & Noël (2007) the non-symbolic task used only two ratios. If the relationship mathematics-estimation is of small magnitudes, perhaps individual differences in non-symbolic estimation can only be captured by a task using a greater variety of ratios. A similar suggestion was made in De Smedt & Gilmore (2011). The authors explained the absence of significant differences in non-symbolic estimation between the control and low-maths groups with the smaller numerosities used in their task compared to studies where differences were found (e.g. Landerl, Fussenegger, Moll, & Willburger, 2009). Another possibility is that the relationship between non-symbolic estimation and mathematics is uneven across development, although more research is needed to understand the developmental course of this relationship. Most studies on non-symbolic estimation to date have been conducted with children. However, some research suggests that the relationship between mathematics and non-symbolic estimation is stable across development (Nys & Content, 2012; Pica et al., 2004; Halberda et al., 2012).

1.2.7 Mathematics and cognitive abilities

Mathematical development clearly does not rely solely on estimation skills. Studies investigating mathematical learning have shown that many cognitive factors contribute to individual differences in mathematics. As previously discussed, several previous studies have failed to find a relationship between estimation of non-symbolic numerosites and other cognitive abilities, therefore suggesting a unique relationship between mathematics and estimation of numerosities (e.g. Halberda et al., 2008; 2012; Mazzocco et al., 2011). Other studies suggest a link between number sense and other abilities (e.g. Geary et al., 2008). A mediation role for cognitive abilities in the relationship between mathematics and number sense cannot be excluded. The next section will give a brief and selective overview of the main research into the relationship between mathematics and other abilities.

A significant role for memory in mathematical development has been suggested by research. Children showing poor mathematical skills perform poorly on working memory-span tasks - involving counting but not language (Siegler & Ryan, 1989). This inability to maintain and manipulate on-line numerical material suggests the contribution of central executive function to mathematical skills. In line with this hypothesis, mathematical difficulties are consistent with poor performance in tasks with high executive demand (Case, Kurland, & Golberg, 1982) especially if the task requires switching from numerical/linguistic retrieval (McLean & Hitch, 1999), thus suggesting poor retrieval of numerical information. Poor retrieval from long-term memory is manifested in children who rely on the aid of counting strategies (e.g. verbal or finger counting instead of remembering the correct answer) to solve arithmetic problems (Siegler, 1987; Geary, Bow-Thomas, & Yao, 1992).

The role of memory in mathematical learning has also been found to involve the phonological loop (e.g. Furst & Hitch, 2000; Logie, Gilhooly, & Wynn, 1994; Lee & Kang, 2002) and the visuo-spatial sketchpad (e.g. Heathcote, 1994; Lee & Kang, 2002). In particular, poor mathematical abilities were associated with poor performance in visuo-spatial working memory in the Corsi Block task (McLean & Hitch, 1999; Bull, Johnston, & Roy, 1999).

A low average IQ together with mathematical achievement test scores below the 20^{th}-25^{th} percentile identifies the population with mathematical learning disability (e.g. Geary, Hamson, & Hoard, 2000). However, this selection criterion has been criticised as many children with low average IQ do not display mathematical difficulties and vice versa (see Geary, 2004).

Speed of processing may be particularly important in mathematical learning. Studies have shown that reaction time on response is correlated with psychometric tests on intelligence (Kranzler & Jensen, 1989; Deary, Der, & Ford, 2001). Many studies have also found that children with mathematical difficulties show slow speed of number identification, visual number matching, and encoding of digits (e.g. Geary, 1993), supporting a relationship between IQ and mathematics. Alternatively, other studies have proposed an indirect association between IQ and mathematics mediated by memory. Children with poor speed of processing may have slower access to numerical material encoded in memory. Because of the slowness/difficulties in accessing the relevant numerical material and strategies from long term memory children may not develop sufficient automatic basic arithmetic facts (multiplication tables, for example) that are vital for the normal development of mathematical skills (Bull & Johnston, 1997).

Reading and language abilities have also been shown to be associated with mathematical development. Many studies have shown comorbidity between poor mathematics and poor reading skills (e.g., Geary, Brown, & Samaranayake, 1991; Hitch & McAuley, 1991; Siegel & Ryan, 1989; Lewis, Hitch, & Walker, 1994; Dirks et al., 2008; Vukovic & Siegler, 2010). Language abilities have been found to be predictive of mathematical skills (e.g. Fuchs et al., 2010).

1.2.8 Sex differences in mathematics

Sex differences in mathematical achievement, with males performing better overall, is a well known phenomenon in educational and cognitive research (e.g. Leahey & Guo, 2001, Penner & Paret, 2008). The research, however, shows a pattern of mixed results. Some studies suggest that these differences start to emerge as early as kindergarten and first grade (Penner & Paret, 2008; Rathbun, West, & Germino-Hausken, 2004). These differences may be mediated by the early male advantage in spatial abilities (as measured by the mental rotation task; Levine et al., 1999) and by the relationship between spatial abilities and mathematics (Halpern, 2000). On the other hand, it has been shown that girls outperform boys on standardised tests of early mathematical skills (Lachance & Mazzocco, 2006), problem solving and basic computation in elementary school (Hyde, Fennema, & Lamon, 1990). Despite some studies reporting early sex differences in mathematics, most research suggests that sex differences start to emerge late in the school years (e.g. Hyde et al., 1990), as between the age of 11 and 13 (middle school) no consistent differences between boys and girls in mathematical achievement have been reported (e.g. Muller, 1998; Leahery & Guo, 2001).

The reported sex differences have been attributed to biological factors, such as the different rate of maturation with the girls' faster rate of development (Tanner, 1978; Gullo & Burton, 1992). Social factors such as socialization, stereotypes and gender roles have been suggested to encourage girls to conform to gender behaviour (Sadker & Sadker, 1994; Steffens, Jelenec, & Noack, 2010). To date, the nature of sex differences in mathematical achievement remains poorly understood as research provides evidence that mathematical reasoning stems from biologically based cognitive mechanisms equally shared by males and females. A review on sex differences in mathematics and science aptitudes, presented evidence for early numerical abilities equal in males and females and the absence of infants' sex differences in numerosity processing (Spelke, 2005). In the same review it was noted that when sex differences emerge later in life, it is difficult to disentangle biological and social factors that may contribute to these differences. With evidence supporting mediation of other abilities in mathematical sex difference, one question to be addressed is whether number sense abilities are involved in the aetiology of gender gap in mathematics. This is of particular relevance given the developmental nature of number sense abilities and the developmental trend of the reported mathematical sex differences. This thesis will specifically examine the aetiology of sex differences in numerosity estimation skills and any potential association with mathematics. As research provides compelling evidence for the absence of sex differences in cognitive abilities at the basis of mathematical thinking, sex differences should not be found in estimation processing as suggested by Spelke (2005) and by Spelke and Grace (2006).

1.2.9 Behavioural Genetic findings on mathematics*

Over the last few decades, the social desirability of mathematic skills has increased as a range of advantages have been associated with high mathematical competence. A survey from the Organisation for Economic Co-operation and Development (OECD, 2010) reports that an increase of half a standard deviation in mathematical and science performance at the individual level leads to an increase of 0.87% in the country Gross Domestic Product annual growth rate. This is reflected in the PISA (Programme for International Student Assessment) report for the 2009 assessment, where the countries whose students show higher levels of mathematical performance are also the countries with the fastest rate of economic growth (OECD, 2010). Similarly, a wide range of disadvantages have been related to low numeracy. Individuals with poor mathematical skills tend to have lower level jobs and are more prone to depressive symptoms, which in turn are a cost in terms of individual suffering, health service use and loss of working days (Gross, Hudson, & Price, 2009).

These premises make the quest for understanding how people acquire and can improve mathematical skills, more important than ever. This importance is reflected in endeavours in mathematical research with an increased number of quantitative genetic studies into the aetiology of individual differences in mathematically relevant traits.

1.2.9.1 Quantitative Genetic Research

Recent behavioural genetic research leaves no doubt that individual differences in behaviour and cognition are a product of both genetic and environmental factors (e.g. Plomin et al., 2008). This research also suggests that the path leading from genes to behaviour is intertwined with the environment. While molecular genetic research aims to detect and identify specific genes involved in the variation in different aspects of behaviour and cognition, quantitative genetic

* The following sections, describing behavioural genetic findings, have been adapted from Tosto,

methodologies (e.g. twin, adoption) quantify the relative contribution of genes and environment to the variation in traits and co-variation among traits.

Genetic influences refer to the influence of multiple alleles – genetic markers that can differ in the population (rather than evolutionarily conserved invariant markers). Mostly, genetic influences are of the additive type, meaning that the variance of a trait that is attributed to genetic factors can be derived by adding the independent effects of all alleles at all loci that affect the trait. Some genetic influences may derive from interactions between genes at different loci. These epistatic processes, by which the effects of a gene on a specific trait depend on the influences of one or more other genes, remain poorly understood (e.g. Cordell, 2002).

From the behavioural genetic perspective, environmental influences are very broadly defined as effects on a trait produced by anything other than heritable DNA sequence variation. In twin and other family designs, any environmental influences contributing to differences between family-members are referred to as 'non-shared'; whereas any environmental influences contributing to the similarity between family members are referred to as 'shared' (see Plomin & Daniels, 1987; Rijsdijk & Sham, 2002; Plomin et al., 2008).

Non-shared environments defined as any events experienced or perceived differently contribute to dissimilarities among family members. These may include perinatal events, accidents, surgical procedures, and different peers. Intuitively, environments that are objectively shared among individuals within a family seem more likely to increase their similarities. We can think of nutrition, parenting practices, or socio-economic status as shared experiences that may make family members more similar in a specific trait, if these factors affect the trait in question. For example, it is reasonable to think that family eating habits could increase similarity in weight among family members. However, research shows that adult family members do not resemble each other in weight beyond genetically influenced similarity (e.g. Grilo & Pogue-Geile, 1991). Often, objectively shared environments (e.g. parenting) lead to differences rather than similarities, through differential perceptions and other poorly understood mechanisms (e.g. Plomin & Daniels, 1987; Dunn & Plomin, 1990). Parental divorce for instance, is a family event and as such shared by siblings, but research shows that divorce often impacts siblings' behaviour in different ways (Hetherington & Clingempeel, 1992; Amato, 2004). The estimate of non-shared environment in quantitative genetic methodology also includes any measurement and procedural errors, as non-systematic error can only contribute to dissimilarity in assessed traits between twins or other family members.

1.2.9.2 Genetic and Environmental aetiology of individual differences in Mathematics

In one of the first twin studies into mathematical ability, 146 MZ and 132 DZ twin pairs of the Western Reserve Twin Project, aged 6 to 12 years, completed standardised tests of school achievement in language, reading and mathematics in addition to tests of cognitive abilities (Thompson et al., 1991).The study revealed a mathematical heritability of .20 with shared and non-shared environment respectively of .71 and .10. Using a sample of 555 twin pairs with learning impairments and 570 control twin pairs with age ranging between 8 and 20 years, Alarcón, Knopick, & DeFries (2000) reported an average heritability of .90 in low mathematical abilities and control group with virtually non-significant environmental influences. The wide range of univariate estimates illustrated in these studies deserves some methodological consideration. Twins' correlations for a trait may be overestimated because of common factors unrelated to the trait. Twins in a pair are of the same age and some pairs are of the same sex. In twin analyses it is common practice to correct for age and sex in order to avoid increase in correlations because of these factors (McGue & Bouchard, 1984). In Thompson et al. (1991) this correction was not carried out, therefore age and sex may have affected the estimates. Other factors also need to be considered. Quantitative genetic investigations of complex traits suggest somewhat different patterns of genetic and environmental influences on different traits. Reading abilities for example, have shown consistently moderate genetic and shared environmental influences across ages and populations (e.g. Light et al., 1998; Stromswold, 2001; Byrne et al., 2005, 2006). Conversely, the heritability of "g" has been shown to increase consistently from early ages to middle childhood (Davis et al., 2009a; Haworth et al., 2010). Although it is unclear whether the inconsistencies in the estimates of mathematical heritability can be explained with differences in participants' age, the findings suggest that age-homogeneous twin samples should be used in behavioural genetic investigations for at least two reasons. First, the trait itself changes across the school years - what we call "mathematics" may involve very different cognitive and motivational processes at different ages, reflected in the changes in how mathematics is measured. Second, new genes and environments may become active or relevant during development, for example reflecting changes in pubertal processes or in social experiences. Estimates of genetic and environmental contributions are population and time based as they explain the sources of individual differences within a particular population within a specific time-window; these estimates may differ not only for different ages, but also for different countries or cultures. If a particular environment is uniform within a culture (e.g. national curriculum or educational standard), this specific aspect of the environment is unlikely to explain inter-individual variation. In such population, heritability of a trait may be higher. Much more research is needed to understand the sources of inconsistencies among different studies, including careful examination of cultural norms and provisions. Further, although Thompson et al. (1991) and Alarcón et al. (2000) used samples with a respectable number of twins, power calculation analyses indicate that twin studies need much larger samples to provide accurate estimates (Plomin et al., 2008).

Much recent research into the aetiology of variation in mathematical ability comes from the Twins Early Development Study (TEDS), a large-scale longitudinal study comprising three cohorts of twins recruited at birth in 1994, 1995 and 1996 in the United Kingdom. The TEDS sample is described in detail in Chapter 2.

In the assessment of over two thousand twin pairs from the TEDS sample at age 7, Oliver et al. (2004) found heritability of mathematical achievement (rated by teachers) of .66, with almost negligible shared environmental influences, and .25 non-shared environment. In the same sample the heritability and estimates of environment were highly similar for three different mathematical components: "Using and Applying Mathematics", "Numbers", and "Shapes, Space and Measures". In the assessment of TEDS at 9 years (Kovas et al., 2007a), mathematics, rated by teachers, showed genetic influences of .72, almost non-existent shared environmental influences, and very modest (.23) non-shared environment. The three mathematical components yielded estimates of genetic influences ranging between .63 (Shapes, Space and Measures) and .73 (Using and Applying Mathematics); almost non-existent shared environment; and estimates for non-shared environment with an average of .27. At 10 years, 2,674 twin pairs were assessed on three mathematical sub-tests: “Understanding Number”, “Non-Numerical Processes”, “Computation and Knowledge” (Kovas et al., 2007b). Similarly to previous estimates, shared environment had very small effects on all three measures and the non-shared environmental influences were between .42 and .48. Although “Non-Numerical processes” showed a lower heritability (.32) compared to the other two components (.42 and .45 respectively), the differences in heritability among the three measures were not significantly different. Assessment of over 5,000 TEDS twin pairs at 12 years confirmed strong genetic influences on mathematical achievement (.61), and small shared (.18) and non-shared (.21) environmental influences (Davis et al., 2009a). Overall, these results demonstrate stable genetic effects on different aspects of mathematical ability across the school years, as well as stable non-shared environmental influences.

Several other recent studies have addressed the aetiology of different aspects of mathematical development using different twin samples. The US-based Western Reserve Reading Project for Math (WRRPM) assessed 228 10 year-old twins on 4 different mathematical components: “Calculation”, “Fluency”, “Applied Problems”, and “Quantitative Concepts” (Hart et al., 2010b), reporting univariate heritability estimates of .35 and .34 for “Calculation“ and “Fluency“, and slightly higher heritability for “Applied Problems“ and “Quantitative Concepts“ (.41 and .49 respectively) . Interestingly, the shared environmental influences estimated in this study (.32-.46) were higher as compared to the TEDS estimates for mathematical sub-components at the same age (.07-.23); whereas the non-shared influences (.19-.25) were lower than in TEDS (.42-.48). The discrepancy observed in the strength of environmental influences between the two samples could be attributed to different curricula or school environments in the two countries. It is possible that the UK educational system, with its standardised curricula across schools, may lead to the smaller proportion of variance in mathematics explained by shared environment.

Alternatively, different estimates may reflect differences in the aetiology of different facets of the mathematical domain. Cross-cultural research using identical measures of mathematical ability and performance in samples of the same age are needed in order to make meaningful comparisons and to establish the true sources of differences in estimates from different studies.

1.2.9.3 The aetiology of the relationship between mathematics, cognitive traits, and motivation

Multivariate genetic methodologies allow investigation of the aetiology of the relationship between mathematics and other cognitive abilities, thus enabling understanding of the co-occurrence between traits.

One of the first studies for normal variation investigated the relationships between mathematics and English, and between mathematics and vocabulary in a sample of over 2,000 twin pairs in U.S. High Schools (Martin et al., 1984). The study reported genetic correlations of .52 and .39 respectively, suggesting that when genes associated with mathematical abilities are identified many of the same genes will be associated with English and, to a lesser extent, with vocabulary. A later study (Thompson et al., 1991) reported a large genetic correlation between mathematics and reading (.98) and mathematics and language (.98), suggesting that largely the same genes contributed to variation in each of the examined traits. The same study showed that, shared environments were also substantially the same for the three traits (shared environmental correlation was .93 on average), thus contributing to their association. Non-shared environmental factors explained most of the differences among the traits (correlations between .28 and .54). Multivariate analyses in 10 year old twins of the WRRPM study (N = 228), investigated the relationship between different components of mathematics, reading and general cognitive ability (g) (Hart, Petrill, Thompson, & Plomin, 2009). It was found that there was no significant genetic overlap between reading fluency, general cognitive ability and the mathematical subcomponent of calculation. Conversely, the shared environmental overlap with calculation and general cognitive ability was significant (~.50), suggesting that the within-families environments that are important for reading fluency and cognitive abilities, also influence the learning of calculation. The subcomponent of mathematical fluency shared genetic influences with reading fluency, but, contrarily to calculation, fluency seemed to have its own significant independent genetic effects: these explained .59 of the total genetic variance in mathematical fluency. Further, for this mathematical component there was no significant environmental overlap between reading fluency and general cognitive ability.

The assessment of TEDS at 7 years found a genetic correlation between mathematics and g of .67 and between mathematics and reading of .74 (Kovas et al., 2005). In the same sample at 10 years, the genetic correlation between mathematics and g was .68, while the genetic correlation between mathematics and reading was .73 (Davis et al., 2008). Overall, the results of several TEDS studies suggested that, to a large extent, the same genes and the same shared environments contribute to mathematics and aspects of reading and general intelligence – explaining most of the observed correlations among these traits. On the contrary, non-shared environmental overlap was very small across the measures, indicating the contribution of non-shared environment to differences among the measures (Kovas, et al., 2005; Davis et al., 2008; Haworth et al. 2008).

A number of behavioural studies also reveal significant correlations between mathematics, self-perceived ability and interest (Eccles & Wigfield, 1995; Gottfried, 1985). In more general terms, it has been shown that there is a developmental, reciprocal association between achievement (not just mathematical) and self-evaluation (Guay, Marsh, & Boivin, 2003; Marsh & Yeung, 1997). The reciprocal relationship between mathematics and self-evaluation has been investigated using longitudinal data from the TEDS sample. Applying a cross-lag design, it was found that mathematics, measured at the age of 9 years, predicted self-evaluation at the age of 12 years. Conversely, self-evaluation at 9 years predicted mathematics at age 12. The striking result was that, although these correlations were very small ($r = \sim .10$), they were almost entirely genetically mediated, meaning that the genes influencing mathematics at 9 were the same as those influencing self-evaluation of mathematics at 12, but not at 9 (Luo et al., 2011). Intuitively we may think that motivational factors strictly related to environmental conditions may help to explain mathematical learning, but this study proposes a novel perspective on the mechanisms underlying mathematical learning. These mechanisms may be more complex than previously thought.

In the TEDS sample multivariate genetic analyses have also been applied to investigate the aetiology of the links among different aspects of mathematics. At age 10 years, five different aspects of mathematics (Mathematical Application, Understanding Number, Computation and Knowledge, Mathematical Interpretation, Non-Numerical Processes) were phenotypically correlated between .45 (Computation and Knowledge and Non-Numerical Processes) and .68 (Mathematical Application and Understanding Number) (Kovas et al., 2007c). On average, the genetic correlation among the five sub-tests was .91, indicating that the same genetic influences affect these different aspects of mathematics. For example, the genetic correlation between Understanding Number and Mathematical Application was .94, meaning that the genetic influences involved in these two mathematical components were almost the same. However, their bivariate heritability was .49 indicating that only 49% in the .68 of their phenotypic correlation is genetically mediated. Similarly to bivariate heritability, the bivariate shared environment explained 29% of the phenotypic correlation, while 22% was explained by non-shared environment. Overall, these results showed that the observed covariation among different aspects of mathematics is largely explained by genetic factors.

To sum up, the patterns of genetic overlap between abilities suggest that genetic influences on individual differences in mathematics are largely the same as those on a wide range of other cognitive and learning abilities, achievement, and motivation, supporting the Generalist Genes Hypothesis (Plomin & Kovas, 2005). According to this hypothesis, genes that are involved in one learning or cognitive domain (e.g. mathematics), are also likely to be associated with other abilities, such as language and g. Conversely, most of the environmental effects on mathematics are not shared with other domains, suggesting that discrepancy in abilities largely stem from the influence of different environments (e.g. Davis et al., 2008).

1.2.9.4 High and Low mathematical abilities

A number of behavioural genetic studies have investigated the aetiology of the relationship between high/low mathematical abilities and other traits. This short section provides only a brief outline of this research to complete the picture of the genetic research in mathematical domain.

Studies conducted on different twin samples have shown that low mathematical abilities are as heritable as normal abilities (Alarcón et al., 1997; Oliver et al., 2004). In addition these studies suggest the same aetiology for low and normal abilities (Haworth et al., 2007; Kovas et al., 2007b). Similarly, the aetiology of high mathematical abilities seems to be strongly related to the aetiology of normal performance (Petrill et al., 2009; Haworth et al., 2009b). These studies also found that some genetic influences may be specific for high abilities only.

The observed similarities in the estimates of genetic and environmental contribution to high, normal and low abilities suggests that mathematical achievement across the whole range of ability is largely influenced by the same genetic factors, again providing support the Generalist Genes Hypothesis (Plomin & Kovas, 2005; Kovas et al., 2007d; Haworth et al., 2009a; Petrill et al., 2009). This also leads to the conceptualization of mathematical ability as a continuum, with one's position on this continuum not driven by 'good' or 'bad' genes, but by small additive effects of ability-increasing and ability-decreasing multiple DNA variants.

1.2.10 The aetiology of sex differences in mathematical abilities

As discussed in section 1.2.8, behavioural studies have reported mean sex differences in mathematical abilities. However, the sources of average group differences may differ from those driving group members' individual differences. Quantitive-genetic sex-limitation models examine the extent to which the same genetic and environmental factors contribute to variation within the sex-groups (testing for qualitative differences); and whether these factors have the same effect on variation in each group (testing for quantitative differences). The sex-limitation models will be described in more detail in Chapter 6 (section 6.3.3.2). Applying sex-limitation models to the data from TEDS 10 year-old twins revealed no qualitative or quantitative sex differences in the aetiology of mathematical abilities or disabilities (Kovas et al., 2007b). Furthermore, no sex differences were found in three components of mathematics (Understanding Number, Non-Numerical Processes, Computation and Knowledge), both in the low ability group and in the unselected sample (Kovas et al., 2007b). No sex differences were found in the aetiology of high mathematical abilities (Petrill et al., 2009). Similarly, no aetiological sex differences were found for parent-rated normal mathematical variation or low mathematical performance in a sample of 17-18 year-old Dutch twins (Markowitz et al., 2005). These findings suggest that genetic factors that make males better or worse at mathematics are the same genetic factors that make females better or worse at mathematics; and they exert the same amount of influence on males and females. The same studies show the equality of environmental effects on male and female variation.

1.2.11 Rationale and research questions

Quantitative genetics studies revealed a substantial genetic overlap between mathematics and other abilities. However, some genetic effects seem to be specific to mathematics (Kovas et al., 2005; Hart et al., 2009), thus indicating that some genes may be involved in processes of mathematically-relevant cognition, such as, for example, number sense. Further, given the reported average sex differences in mathematical achievement, another area of investigation is the potential involvement of number sense in the mathematical gender gap. To date, two aspects of number sense, symbolic and non-symbolic estimation have been investigated in association with mathematics. However the relationship between them remains to date unexplored. Further, many of the studies examining the relationship between mathematics and number sense are cross-sectional and the longitudinal studies cover short periods of development of mathematical skills. Finally, no studies have explored the aetiology of these relationship using genetically sensitive designs.

This thesis sets out to investigate:

1. The structure of the number sense domain by using two measures: a symbolic and a non-symbolic task of estimation.

2. The relationship of the number sense, measured at 16 years of age, and mathematical achievement spanning 10 school years (with and without controlling for general cognitive abilities measured at the time of previous mathematical assessments).

3. The aetiology of individual differences in number sense at 16 years.

4. The aetiology of sex differences in number sense at 16 years.

5. The aetiology of the relationship between number sense, mathematical abilities and general intelligence at 16 years.

Points 1 and 2 are investigated using analysis of variance (ANOVA) analyses. This is the first time that a Number Line task and a Dot Task of numerosity estimation are used together to assess number sense in a sample as old as 16 years of age. Both tasks are supposed to tap into the number sense domain; therefore a degree of association between the two measures is expected. As reviewed in section 1.2.5, a number line task may present some limitations in the assessment of symbolic estimation, as it may tap just into some aspects of number representation. However, the literature has shown a robust and consistent relationship between mathematics and number line estimation, therefore the choice of the Number Line task to assess symbolic estimation was driven on the basis of this relationship. According to previous literature, a relationship between mathematics, assessed at 16, and the two number sense measures is also expected. An association between performance in the Dot Task at 16 and all the earlier mathematical school achievement (up to the age of 7) – available in TEDS – is also hypothesised. There are no previous studies reporting on retrospective relationship between Number Line performance and mathematics. For this reason no specific predictions are made on the relationship between Number Line task at 16 and previous mathematical achievement.

Points 3, 4 and 5 are investigated using behavioural genetic methodologies (twin method). This is the first time that behavioural genetic methodologies are used to investigate number sense abilities and their relationship with mathematics and general cognitive ability. It is hypothesised that estimation of numerosity (Dot Task) has evolutionary origin, so it is possible that this ability may be highly heritable. Conversely, there is evidence that individual differences in evolutionary preserved traits are influenced very little by genetic factors. Estimation of numerosity skills have shown to have a unique association with mathematics (e.g. Halberda et al., 2008). It is possible that this ability is reflected in the specific genetic influences detected in various behavioural genetic studies of mathematics. On the other hand, Number Line estimation has shown association with other cognitive abilities (visuo-spatial working memory, general intelligence, central executive). It is possible that the multivariate genetic analyses will show a substantial genetic overlap of Number Line estimation with general cognitive ability.

The research questions were addressed following the steps listed below:

- Compilation and validation of an on-line battery of tests aimed to measure number sense and mathematical abilities in 16 year-old singletons.
- Implementation of the validated tests into the mathematical-number sense component of the on-line battery in the TEDS' assessment at 16 years.
- Administration of the battery to the first and second cohort of TEDS twins.
- Further validation of the mathematical-number sense battery (test-retest) using the first data collected from the first cohort of twins.
- Retrospective investigation of the relationship between number sense measured at 16 years, and mathematics and cognitive abilities measured at earlier ages.
- Estimation of the contribution of genetic and environmental factors to the two measures of number sense
- Investigation of the aetiology of sex differences in number sense.

- Investigation of the aetiology of the relationship between number sense, mathematics and general cognitive abilities.

Chapter 2 is a methodological chapter that describes the two samples used in this research: the sample of 16 year-old singletons recruited for the pilot study and the longitudinal TEDS (Twin Early Development Study) sample. The chapter further provides details of the experimental procedure. Description of the quantitative genetic methods used for the analyses is provided in the relevant experimental chapters. Specifically, the twin method and univariate genetic analyses are presented in Chapter 3; the sex limitation model is described in Chapter 6; while the multivariate genetic analysis is detailed in Chapter 7.

At the age of 12 years, using multivariate genetic analyses on TEDS data, it was shown that some genetic influences were specific to mathematics and independent from the genetic factors shared with reading and *g* (Kovas et al., 2005). The first study in this thesis, presented in **Chapter 3**, used the same data to estimate the genetic and environmental influences on the residuals of mathematical achievement, after removing variance explained by reading and general cognitive abilities at 12 years. The involvement of number sense in the genetic influences of the residual scores of "Pure Mathematics" is discussed. The results justify further investigations into the aetiology of number sense and its relationship with mathematics using genetically sensitive methodologies.

The investigation of number sense was conducted as part of the larger wave of testing of the TEDS sample at age 16 years. **Chapter 4** describes the piloting and validation of the on-line testing of the mathematical-number sense battery in a sample of 16 year-old singletons. This study was conducted in two stages, during which, from the initial eleven measures piloted, seven were selected for on-line implementation in the mathematical-number sense battery to be included in the TEDS on-line assessment. The seven measures comprised of three number sense measures, two tests of mathematical skills and two tests of general cognitive abilities. The chapter also presents the validation study carried out on a subset of TEDS after they completed the on-line assessment. During the pilot study some of the measures went through changes, either in the administration format (e.g. from pen and paper to computerised), or in modification of the parameters for on-line adaptation. The internal validity of the measures was good throughout the two waves of piloting, but the alterations carried out affected the test re-test reliability in the pilot study. This required an additional test re-test reliability study to be conducted at the beginning of the administration of the whole battery to the TEDS sample.

Chapter 5 investigates the relationship of the two measures of number sense (estimation of symbolic numbers and non-symbolic numerosities) assessed at 16 years with mathematics and cognitive abilities measured in previous years. The longitudinal TEDS data allowed to investigate the continuity of the relationship between mathematics and number sense, controlling for a number of cognitive abilities measures in 6 different assessments during 10 school years. The study also investigated the structure of the number sense at 16 years and the degree of the relationship among the two components: estimation of non-symbolic numerosities and estimation of symbolic numbers. This study answers the research questions posed at points 1 and 2 above.

Chapter 6 aims to estimate the contribution of genetic and environmental influences to individual and sex differences in non-symbolic estimation skills. This is the first large scale genetically sensitive investigation into this aspect of number sense. The results are discussed in relation to the evolutionary origins of this ability. This investigation answers the research questions listed at points 3 and 4 above.

The study in **Chapter 7** presents a multivariate genetic analysis investigating the nature of the relationship between the two number sense abilities (symbolic and non-symbolic estimation) and mathematics using general cognitive abilities as a covariate. The question addressed is whether number sense skills are specifically related to mathematics, beyond the contribution of general cognitive factors and answers the research question number 5.

Taken together, this thesis provides the first large-scale genetically-sensitive multivariate investigation into the origins of individual differences in number sense and the aetiological links between number sense and mathematics across school years.

Chapter 2: Methods

2.1 Overview

This chapter describes the two samples, procedures and measures used in the studies of this thesis. The models and analyses are described at later stages, in the relevant chapters. The first sample, described in section 2.2 is made up by twins from the TEDS. The section is further divided in paragraphs providing a description of the sample by age of testing, together with the procedure used for the recruitment and data collection. The section also describes the selection criteria used to recruit a subset of TEDS at 16 used for the validation study of the whole online battery administrated at 16. The second sample is made of the 16 year old students recruited specifically for the pilot study; they are described in section 2.3.

The measures are described in section 2.4. Section 2.4.1 contains brief descriptions of the measures from previous TEDS assessments at ages 7, 9, 10, 12 and 14 used for the retrospective-longitudinal analyses illustrated in Chapter 5. Chapter 3 illustrates a study using data collected from TEDS at the age of 12, therefore a detailed description of the measures at this age is further provided in Chapter 3.

The number sense measures, together with the tests of mathematical performance and numerical-relevant cognitive abilities tests used in the TEDS assessment a 16 years were developed as part of the work of this thesis. The development and description of the measures are detailed in Chapters 4. The behavioural analyses in Chapter 5 and the behavioural genetic analyses in Chapter 7 were conducted using measures from the general assessment at age 16. These measures were not-developed as part of the work for this thesis and are illustrated in details in this chapter, section 2.4.2.

2.2 TEDS sample

2.2.1 Sample description

The behavioural genetic studies and the longitudinal retrospective analyses were conducted on data collected as part of the Twins Early Development Study (TEDS). TEDS is a large longitudinal study constituted of twins born in England and Wales in 1994, 1995 and 1996. Details of 25,815 twins' families were obtained from the children's birth records from the Office for National Statistics (ONS). The families were contacted via mail. Consent and contact details for 16,810 families were obtained between 1995 and 1998 when the twins were on average one and half year old.

2.2.2 Representativeness

TEDS sample is considered a good representative of the UK population in terms of ethnicity and socio-economic status. From a survey conducted for the years 2000-2001, on average, 93% of the UK population was from a white background, 32% of the population had an A-level or higher level of education and the rate of employment was 49% for mothers and 89% for fathers (Walker et al., 2001). At the time of first contact, 91.7% of TEDS sample came from a white background, 35.5% of parents had higher education, employments for mothers was 43.1% and 91.7% for fathers' (Oliver & Plomin, 2007). Between the age of 7 and 14 the statistic figures maintain the same proportion, with 93.3 and 93.7 per cent of people from a white background, higher education between 44.4% and 48.6%, employed mother at these ages was 44.8% and 46.1% and fathers 92.8% and 93.8%. In the assessment at 16 the new survey reported an ethnicity of 92.6% white and parental higher education 46.3%. The figure of working mothers increases to 91.6%, following a decrease in childcare. This could be explained by mother returning to work after the children have grown up. Fathers' employment is 96.4%.

2.2.3 Zygosity assessment

A correct identification of the twins' zygosity is fundamental, as the twin method relies on the comparison between Monozygotic (MZ) and Dizygotic (DZ) twins. In TEDS, zygosity was assessed by means of questionnaires at the time of the first contact when the twins were one and half year old. Zygosity was further assessed through questionnaires at the age of 3 and 4. These questionnaires have been shown to be 95% accurate in determining zygosity when tested against zygosity assessed with DNA markers (Price et al., 2000). Questions were based on ratings, for example on the differences in texture of the hair, differences with ear-lobe shape or difficulty to tell the twins apart. From birth statistics, around one third of twins are MZ, one third are DZ same sex and one third are DZ opposite sex twins. The TEDS sample reflects these proportions.

2.2.4 Procedure adopted on TEDS data collection and sample statistics

During the first recruitment/contact, data was collected by means of postal booklets addressed to the parents of the twins. The questionnaires asked information about the twins, their parents and family environment. Postal booklets sent to the families at the age of 2, and 3, included further measures of language and cognitive abilities. The assessment at the age of 4 was conducted by means of parents and children booklets and home visits. This assessment contained environmental, behavioural, language and cognitive abilities measures. The data collected up to the age of 4 is not analysed in this thesis but it is relevant because it sets out the procedures and defines the sample for the future studies.

One of the aims of this thesis involves the investigation of abilities during the school years. In this period of time, TEDS have been assessed at the age of 7, 9, 10, 12 and 14. At the age of 16 the assessment included the number sense and mathematical battery developed as part of this thesis. All data collected starting from the age of 7 is used in the retrospective longitudinal study described in Chapter 5.

2.2.4.1 TEDS sample description and description of data collection at age 7

At age 7, data was collected from three sources. Firstly, parents provided information about the twins and family environment together with a consent form to the study and permission to contact the twins' teachers. Data from the parents was collected with a questionnaire and telephone interview. Secondly, the twins' teachers filled a questionnaire (sent via mail) about the children academic achievement and school behaviour. Lastly, children were assessed by phone interview. In order to carry out the telephone testing, prior to the assessment, sealed envelopes containing the booklets with questions/items of the tests were sent via post to the families. Separate envelopes with instructions explained to the parents to open the testing envelopes only at the moment of the TEDS-caller telephone call - the time and date was specified in the instruction. The telephone testing followed a set protocol; each TEDS-caller conducted the assessment following the "twin interview script". TEDS-callers also recorded twins' responses into a computerised system at the time of the interview, although each call was recorded on audio cassette tape. 14,581 families from the 16,810 of the initial TEDS sample were contacted via mail, parents returned 7,909 questionnaires, teachers 6,532. In this wave of testing only 5,421 twin-pairs born in the first and second cohort (1994 and 1995) were assessed via phone.

Prior to all analyses in this thesis (at all ages, not just at 7), children with specific medical condition such as Down syndrome, cystic fibrosis, cerebral palsy, organic brain damage, autism spectrum disorder, severe hearing loss, abnormal birth weight or gestational period and if mothers were heavy drinkers during gestation were excluded. Children whose English was not the first language were also excluded.

After language and medical exclusion the sample at age 7 provided data for 7,267 twin-pairs with mean age 7.11 (SD = .25). The twins from the first cohort received t-shirts as reward for their participation to the study.

2.2.4.2 TEDS sample description and description of data collection at age 9

In the wave of assessment at age 9 data was collected using postal questionnaires from three sources: the twins themselves, their parents and their teachers. Parents provided children behavioural information, home environment data and consents. Teachers' questionnaires informed on children school achievement, behaviour at school and motivation towards schools and achievement. The twins completed tests for cognitive abilities, and provided information about their motivation and other environmental measures.

Only children (and families) from the first and second cohort were included in this assessment. The parents and children questionnaires were sent to 7,531 families, with a return of 3,412 parents' questionnaires and 3,421 children questionnaire. 3,869 questionnaires were sent to the teachers (one questionnaire for each pair) with a return of 2,740. After medical and language exclusion, the final data for the analyses was provided from 7,162 twins (3,581 pairs) with mean age 9.03 (SD = .28). A £5 reward voucher for each twin was included together with the questionnaires sent to the families of the first cohort. The second cohort of twins received £5 reward voucher per twin after return of the questionnaire.

2.2.4.3 TEDS sample description and description of data collection at age 10

At age 10, on-line testing was introduced. Children tests of performance and achievement were administered with a web-based battery. Parents answered a short on-line questionnaire providing information about themselves. Teacher's data regarding the twin's academic achievement and school behaviour was collected with postal questionnaires. Prior to the web-testing 5,944 families (from the twins born in 1994 and 1995) were sent an information pack about the data collection. This included a consent form to the study, authorization sheet to contact the twins' teachers and the logins for parents and children to access the testing website. The tests were programmed on-line on the TEDS website at: www.teds.ac.uk. The tests were provided with online instructions and were designed to be completed without supervision; for some tasks it was possible to practice before the test trials. Internet testing has been deemed as reliable as traditional methods (e.g. Gosling et al., 2004), with the advantage of fast access to large and widespread samples. The reliability and advantages of the TEDS web-testing has been examined in Kovas et al. (2007a).

Out of the recruited 9,411 families of TEDS from the first and second cohorts, 5,944 active families received the information pack and 3,887 questionnaires (one for each twin pair) were sent to teachers. The questionnaires returned from teachers were 3,087 while 3,184 parents completed the online questionnaire. After medical and language exclusion, data available from the web-tests was from 7,198 twins (3,599 pairs) with mean age 10.09 years (SD = .28). Each twin was rewarded with £5 voucher after completion of the web-assessment, the families received £5 voucher to cover for Internet costs.

2.2.4.4 TEDS sample description and description of data collection at age 12

The assessment at the age of 12 was conducted via Internet, telephone and postal questionnaires. Parents filled questionnaires on family environment, their children behaviour and school achievement according to the ratings of the UK National Curriculum (NC). Teachers reported on the children academic achievement and school behaviour. The twins completed an on-line battery that assessed a number of cognitive abilities while a reading test (TOWRE) was administered on the telephone. In addition, the children completed a self-reported questionnaire (postal questionnaire) answering questions about themselves and their motivation towards academic achievement.

This wave of testing included the 3 cohorts of twins. Information packs containing consent forms, twins and parents questionnaires, twins log-ins to the TEDS website and permission to contact the teachers, were sent to 8,439 families (out of the 16,810 from inception). In the consent form, parents were asked to provide their phone number for the administration of the TOWRE test (if they opted in). The list of words for the reading test was in the information pack, and the test was administered to the children who had completed the on-line assessment. 6,341 questionnaires (one for each pair) were sent to teachers. After medical and language exclusion, data at age 12 was provided from 13,262 twins (6,631 pairs) with mean age 11.72 (SD =.67). Each twin was rewarded with £10 voucher after completion of the web-assessment, the families received £5 voucher to cover for Internet expenses if the children had attempted at least one test in the battery.

2.2.4.5 TEDS sample description and description of data collection at age 14

The assessment for this wave of testing was carried out with web-based tests, telephone interviews and postal questionnaires. Parents answered questionnaires, available in pen and paper format and on-line, regarding the children and family environment. They also reported on their children academic achievement, according the UK NC. Similarly, the teachers' questionnaires were also available in pen and paper and on-line format. Teachers answered questions regarding the twins' achievement and school behaviour. The twins were assessed on a range of cognitive abilities with an Internet based battery of tests. An additional Language telephone test was administered to the twins of the first cohort.

Prior to the assessment, the families received an information pack (as described in previous testing). All the active families were contacted to take part to this assessment, 11,005 families were sent log-in details to participate to the study. Teachers received questionnaires for 1,323 twin-pairs. After medical and language exclusion, data for the 14 year assessment was provided from 4,731 twin-pairs with 605 pairs language telephone interviews. The mean age was 14.10 (SD = .55).

The twins from the first cohort were rewarded with a £5 voucher for completing the web activities. All the families, for which the twins attempted at least one web test, were sent one £5 voucher as compensation for Internet expenses.

2.2.4.6 TEDS sample description and description of data collection at age 16

At 16, data was collected from the twins and their parents using web-based tests and postal questionnaires. Parents filled a pen-paper format questionnaire regarding family data and twins' behaviour. They also completed an on-line questionnaire assessing the socio-economic status. The twins completed a web-based battery assessing a range of cognitive abilities, and filled a behavioural self-report questionnaire in pen-paper format. In addition, GCSE (General Certificate of Secondary Education) and other school qualification obtained at the age of 16 were collected using an exam-result form in pen-paper format sent to the families.

GCSE results, behavioural and environmental data from the twins (pen and paper) questionnaire is being collected from the children of the three cohorts (still under way at the time of this writing with plan to be finished by August 2013).

The twins' web-data and parents' data (on-line and pen-paper questionnaires) were collected from the first and second cohorts. Prior to the web-testing the families were sent an information pack which also contained log-ins for the parents and the children. The twins' log-ins were activated after the parents logged in to complete the socio-economic status questionnaire and the on-line consent form.

At the time of the study presented in Chapter 5, data collection was still under way, therefore only data collected from the first cohort was analysed. After medical and language exclusion data from the first cohort was provided by 2,100 twins (1,050 pairs) with mean age of 16.5 (SD = .19). The behavioural genetic investigation in Chapters 6 and 7 were conducted on data from the two cohorts, on a total of 7,598 twins (3,799 pairs). After medical and language exclusion the final sample from the two cohorts was provided by 6,854 twins (3,427 pairs) with mean age 16.6 (SD = .28). For completion of the web activities each twin was rewarded with a £10 shopping voucher an entry to a prize draw.

2.2.4.7 Subset of TEDS for validation study at 16

This sample was recruited to carry out test-retest reliability and internal validity of the whole TEDS Internet battery administered at 16 years. The assessment of the first cohort of TEDS started in October 2010. By January 2011, around 600 families had completed the web testing. The twins recruited for the validation sample were selected among these families. Twenty-four twin-pairs were chosen to match the 600 families in SES (from the online parents questionnaire) and IQ (from the Raven and Mill Hill Vocabulary tests). They were invited to the Social Genetic Developmental Psychiatry Centre at King's College, University of London (Centre of the TEDS study) to repeated the whole battery in the Centre premises. The re-test took place two months after the twins completed the online assessment for the first time. Each twin received £30 voucher for participating to the validation study.

2.3 16-year old singletons

2.3.1 Recruitment of the school and of students

The pilot study took place between October 2009 and May 2010. It was conducted on a sample of 16 year-old singleton students. These students were matched on age to TEDS to make sure that number sense/mathematical tests included in the TEDS assessment at 16 were age appropriate. The 16 year-olds were recruited at year 12 (sixth-form) among students who had taken their GCSE exams in mathematics.

The students were recruited from two schools in the Greater London area. One of the schools contacted Goldsmiths College - Educational Department, asking to involve the students in experimental projects as part of their Psychology class activities. The Psychology teachers from the second school were contacted via letter and showed interest in letting their students participating to the study. In both schools, the Head teachers gave permission to carry out the testing on school premises. In order to comply with legislative requirements, a Criminal Record Bureau certification (CRB) was required for the tester (the author of this thesis). The two schools received different ratings in performance and services from the Office for Standards in Education, Children's Services and Skills (Ofsted). From its latest Ofsted inspection (11/2006) the first school received a Grade 1 rating (Outstanding) in all the fields object of the assessment. The second school was inspected in 05/2009 receiving ratings of Grades 3 and 4 (respectively Satisfactory and Inadequate), with a notice to improve.

2.3.2 Procedure

The pilot study was conducted in two sessions. During the first session, students were administered a battery of 11 tests: 2 in pen-paper format, 3 computerised and administered on a laptop off-line and 6 were web-based. These web-based tests are available on-line at http://lab.kctam.com/stroop/. Reliability and internal validity analyses were carried out on the data collect in the first session. Guided by the results of these analyses, the battery was reduced to 7 tests, all programmed on-line in the pilot website available at:

http://www.e-businesssystems.co.uk/teds/. The development of the 7 tests is detailed in Chapter 4. The second session of the pilot study took place a couple of months later, after that the 7 tests were programmed and implemented online in the pilot website. During this session, students recruited in the first session of the pilot completed the battery on-line for test-retest reliability and on-line validation purposes.

Students' participation was voluntary. They received an information sheet (Appendix - 1) together with a consent form (Appendix - 2) and if opted-in they returned the consent forms signed by their parents/guardians and themselves. The consent form was valid for the two testing sessions of the pilot. The testing was conducted in a quiet area of the school, on a one to one basis and it lasted around one hour. During this phase the results of their GCSE mathematical grade were collected (students self report). Although the details of the study were clearly explained in the information sheet, a debrief followed the testing session. The information given during this phase was the same as in the information sheet. In some cases, more details were given if requested by the students. Prior to testing, students were reminded that they were free to withdrawn from the study (and from the testing session) at any time, without giving explanations. However, no students interrupted the experiment nor reported discomfort or distress due to testing.

When the battery comprising of the 7 tests was ready and available online, individual log-in details were issued to the students to access the testing website. Teachers involved in the testing made sure that the students received the log-ins.

Students were rewarded with one £10 voucher for each session completed. As token of appreciation, 4 vouchers of £10 each were given to each teacher (3 teachers in total) involved in the study.

2.3.3 Pilot sample description

100 students were recruited during the first pilot session, with valid data for 98. Mean age of the 98 students was 16.78 (SD = .91). There was a predominance of female students (83 females vs 15 males) as in the first schools the student-population was mostly composed by females. 75 students took part to the second session of the pilot and repeated the tests on-line. 68 were females and 7 males, with mean age 16.74 (SD = .93).

2.4 Measures

Measures from TEDS assessments from the age of 7 to 14 are described at a later stage as were used in the longitudinal study presented in Chapter 5. TEDS general assessment at 16 included other cognitive measures that were not developed as part of the work of this thesis. These measures were also used in the study presented in Chapter 5, therefore are detailed in Chapter 5. Longitudinal (7 to 14 years) and contemporaneous measures (at 16 years, not developed as part of this thesis) are described in this section.

2.4.1 Longitudinal Measures

Measures from previous TEDS assessments, their validity, and administration procedures are described in detail in previous TEDS publications (e.g., Kovas et al., 2007a; Haworth et al., 2009). The following paragraphs provide only a brief outline of each measure analysed for the purposes of this thesis.

2.4.1.1 Measures at 7 years

Cognitive ability measures at 7 were collected using telephone testing. Children mathematical school achievements were derived from the teacher's questionnaires.

Verbal Ability: A composite measure was obtained from the Wechsler Intelligence Scale for Children (WISC-III) Vocabulary test and the WISC-III Similarities test (Wechsler, 1992).

Non-Verbal Ability: A composite measure was obtained from the test of Conceptual Grouping (McCarthy, 1972) and WISC Picture Completion test (Wechsler, 1992).

Reading Ability: Composite of fluency reading was obtained from scores of the two TOWRE (Test of Word Reading Efficiency) sub-tests - the Test of Sight Word Efficiency (non-words) and the Phonemic Decoding Efficiency test (words) - (Torgesen, Wagner, & Rashotte, 1999).

Mathematical Achievement: Teachers assessed children's achievement, based on the expected UK standard at Key stage 1 (QCA-Qualifications and Curriculum Authority Key stage-1 www.qca.org.uk/ca/tests) on three mathematical components: "using and applying mathematics", "numbers", "shapes, space and measures". The composite score of the three assessments was used.

2.4.1.2 Measures at 9 years

Data for this wave of testing was collected by postal booklets. The cognitive measures were derived from the child-completed booklets; the mathematical scores were derived from the teachers' questionnaires.

Verbal Ability: The verbal cognitive score was obtained from the Vocabulary Multiple Choice and General Knowledge tests taken from WISC-III-PI (Kaplan et al., 1999).

Non Verbal Ability: The non verbal cognitive score was derived from scores of two tests: the "Puzzle" and "Shapes" tests adapted from Smith, Fernandes, & Strand, (2001).

Mathematical Achievement: Teachers assessed children's achievement, based on the expected UK standard at Key stage-2 (QCA Key stage-2) on three mathematical components: "using and applying mathematics", "numbers and algebra", "shapes, space and measures". A composite of the three mathematical scales was used.

2.4.1.3 Measures at 10 years

Data at 10 years were collected using a web-based battery of cognitive tests and mathematical performance. An additional measure of mathematical school achievement was available from the teachers' questionnaire.

Verbal Ability: The verbal ability scale was derived combining the Vocabulary Multiple Choice and the General Knowledge web-tests from the WISC-III-PI (Kaplan et al., 1999).

Non Verbal Ability: The scale was obtained from two web tests: Picture Completion (Wechsler, 1992) and the Raven Standard Progressive Matrices (Raven et al., 1996).

Reading Ability: Reading ability was assessed with the web-test of reading comprehension PIAT (Peabody Individual Achievement test; Markwardt, 1997).

Mathematics Web Test: The test was based on the items of the NFER 5–14 Mathematics Series. The mathematics web score was derived combining together 3 tests assessing the mathematical sub-components: "non-numerical processes", "understanding numbers", "computation and knowledge".

Mathematical Achievement: Mathematical school achievement was assessed by teacher questionnaires based on the standards required from the UK National Curriculum at Key Stage 2 (QCA Key Stage-2) on the mathematical components of: "using and applying mathematics", "number and algebra" and "shapes, space and measures".

2.4.1.4 Measures at 12 years

Data at 12 years were collected using a web-based battery of cognitive tests and mathematical performance. An additional measure of mathematical school achievement was available from the teachers' questionnaire.

Verbal Ability: The verbal ability scale was derived from Vocabulary Multiple Choice and General Knowledge web-tests from the WISC-III- PI (Kaplan et al., 1999).

Non Verbal Ability: The composite score of non verbal ability was obtained from two non-verbal reasoning web-tests: Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1998) and the Picture Completion Test (Wechsler, 1992).

Reading Ability: Reading ability was assessed with 3 tests of reading fluency and 2 tests of reading comprehension: TOWRE, reading fluency of words and non-words (administered on the phone) (Torgesen, Wagner, & Rashotte, 1999); Woodcock-Johnson III Reading Fluency Test (Woodcock et al., 2001); GOAL-Reading comprehension–Formative Assessment in Literacy - Key Stage 3 (GOAL plc, 2002); PIAT- Reading comprehension-Peabody Individual Achievement Test (Markwardt, 1997). The scores of the 5 tests were combined together in a composite reading score.

Language Ability: A composite score was obtained from three web-tests: Figurative Language - the test of semantic language competences (Wiig et al., 1989); Inferences – the test assessing pragmatics skills (Wiig et al., 1989); and TOAL-3 - the test of grammar (Hammill et al., 1994).

Spatial Ability: A composite score was obtained from the “Hidden Shapes” and the “Jigsaw” web-tests, both taken from the nferNelson Spatial Reasoning Series books (nferNelson, 2002).

Mathematics Web test: The test was based on the nferNelson 5–14 Mathematics Series. A composite score was created from scores on the three mathematical components: "understanding numbers", "non-numerical processes", "computation and knowledge" (nferNelson, 1994; 1999).

Mathematical Achievement: Mathematics was assessed by teacher questionnaires that rated the twins' school achievement on four mathematical components: "using and applying mathematics", "number and algebra", "shape, space and measures", "handling data". The rating was based on the requirement levels set by UK National Curriculum and taken from the booklets Mathematics 5-14 of the National Foundation for Educational Research (nferNelson, 1999).

2.4.1.4 Measures at 14 years

Data uses in the analyses at 14 years were collected using a web-based battery of cognitive tests. The measure of mathematical school achievement was available from the teachers’ questionnaire. Scores of the Language test obtained by phone were not used in the longitudinal analyses.

Verbal Ability: The verbal ability scale was derived from Vocabulary Multiple Choice and General Knowledge from the WISC-III- PI (Kaplan et al., 1999).

Non Verbal Ability: The non verbal ability was measured by the Raven’s Standard Progressive Matrices test (Raven et al., 1998).

Mathematical Achievement. Teachers rated the twins mathematical school achievement according to the levels required from the UK National Curriculum on four components: "using and applying mathematics", "number and algebra", "shape, space and measures", "handling data". Teachers' questionnaire was based on the booklets Mathematics 5-14 of the National Foundation for Educational Research (nferNelson, 1999).

2.4.2 Contemporaneous measures: Tests of General Cognitive Ability a age 16 - not developed as part of this thesis

5 tests of general cognitive ability were used to measure non-verbal intelligence (Raven's Progressive Matrices), verbal ability (Mill Hill Vocabulary test), language ability (Figurative Language) and reading skills (reading fluency and reading comprehension). These tests were adapted from previous versions of the TEDS web batteries. However, the new versions went through some alteration, therefore they are described in details in this section, as no previous TEDS publications report on these measures.

Non-Verbal Ability - Raven Progressive Matrices: This computerized test of non-verbal (fluid) intelligence was adapted from Raven, Court, & Raven (1996). The test started with a set of animated instructions and one practice trial that could be repeated at the discretion of the participant. Participants were presented with a matrix of patterns with one piece missing from each pattern. The task required to select the missing pattern from a choice of 8 by clicking on it with a mouse. The test consisted of 30 items organised in 3 levels with 6 items each, and a 4^{th} level with 12 items. The first 3 items of the first level were presented sequentially. Twins progressed within the same level if a correct response was given to at least one of the 3 items. If the first 3 items of the level were answered incorrectly, the following 3 items were skipped and the test advanced to the next level. One point was assigned for each correct answer; the skipped items received no points. The maximum score for this test was 30. If no answer was given within 5 minutes the program returned to the main page of the website. When resuming the session the same question was presented. After each response, the next question followed immediately. Participants could, however, take a break at any point in the test. Accuracy (measured as the total number of correct answers in the test) and response reaction time (time measured from the appearing of the stimulus on the screen to the time of response) was recorded by the program.

Verbal Ability: The test was programmed based on the Mill Hill Vocabulary test (Raven, Raven, and Court, 1998). It consisted of 33 questions, where a single word was displayed on the screen with 6 other words below it. Words constituting the target and the 6 responses were taken from the published measure. This is an example of target word: "fascinated" and the 6 choices: "ill-treated, poisoned, frightened, modelled, charmed, copied". The task required to click with a mouse on one of the words with the meaning closest to the target word presented on top. Only one among the 6 choices was the correct answer and one point score was assigned for each correct response. Maximum score for the test was 33. A set of instructions started the test followed by one practice trial. A button to start the test appeared after the practice and the items were presented in the same order for all participants. After a response was given, the next question was presented immediately and it was possible to pause the test and resume it later. There was no time out for response. The test was discontinued after 7 consecutive incorrect answers. Accuracy and response reaction time were recorded by the program.

Language Ability: Programmed according to Wiig, Secord, and Sabers (1989), the test was administered with the additional auditory modality to ensure that children with reading disability would not be disadvantaged. The stimuli consisted of 15 target expressions or figures of speech, referred to a situation and were displayed one at a time with 4 other expressions having similar meaning. For example, the situation was described as “Two boys talking at a dog show”; the expression referred to this situation was: “He is crazy about that pet”; the 4 possible responses were: 1) “The pet makes him angry”, 2) “He is up in arms about the pet”, 3) “The pet is really wild”, 4) “He is wild about the pet”. The task required to match one of the 4 choices with the target expression. One point was assigned for each correct answer. A tutorial played at the beginning of the test advised to switch on the sound on the computer. There was one practice trial that provided feedback on response. During the practice and the test trials, the situation, target expression and the 4 choices were audio-played while the text was displayed on screen. It was possible to respond by clicking on the choice before the end of the audio recording. The next item was displayed immediately following the response. The choice to pause the test and resume it later was displayed following each item. Response time out was 60 sec. If during this time no answer was given, the question was recorded as incorrect and the next item was displayed. Accuracy and response reaction time were recorded by the program.

Reading Comprehension: The test was developed by Hayiou-Thomas & Dale, it is not published and is available from the authors. The test was based on two passages of written text. The task entailed reading the text and answering multiple choice comprehension-questions based on the passage. Response was given by clicking on 1 of the 4 choices, of which, only one was correct. One point score was assigned for each correct answer. Thirteen questions were asked for each of the two passages, with maximum total score of 26. The test started with an introduction tutorial and there was no practice. After the tutorial the first passage was presented on the screen, with an option to click on the "next" button to proceed to the first question and multiple choice answers. The text remained on the screen together with the questions. The next question appeared immediately following the response. It was possible to pause the test at any time. Time-out for each response was 5 minutes. If during this time no answer was given, the question was recorded as wrong and the next question followed. The reading time was recorded from the appearance of the passage to the click of the "next" button to see the first question. The program recorded accuracy and response reaction time.

Reading Fluency: The test was programmed according to Woodcock-Johnson III (Woodcock, McGrew, & Mather, 2001). This was a timed test with a limit of 2 minutes and 30 seconds and consisted of 98 statements which required a yes/no answer (this is an example of a statement: "A jug may be used to pour water"). Response was given by clicking with a mouse on the "Yes" or "No" buttons displayed together with each statement. The next item was displayed immediately following the response.

One point was awarded for each correct answer; no points were assigned for timed out or incorrect responses. Maximum time for a response was 40 seconds, if no response was given during this time the question was recorded as incorrect and the next question was presented. The test was provided with a set of instructions and one practice trial with feedback. The program recorded accuracy and response reaction time. The total time of completion was also recorded.

Chapter 3: Sources of Individual Differences in Maths beyond IQ: insights from 12-year old twins

3.1 Abstract

Behavioural studies have shown that individual differences in mathematical abilities are related to a variety of cognitive and environmental factors. These studies also suggest that about half of the variance in mathematics is related to DNA variation. Environmental factors have also been shown to be important, especially those, that are not shared by family members. Previous research found substantial genetic overlap between factors influencing variation in mathematics and other cognitive domains at the age of 7 and 10. This study set out to investigate individual differences in mathematics and the nature of the relationship of mathematics with reading and *g* at the age of 12. More than 13,000 12 year-old twins (part of the Twins Early Development Study, TEDS) were assessed on 11 measures of mathematics, reading, and general cognitive abilities. The results showed that removing overlapping variance with reading and *g* from mathematics scores did not reduce mathematics variability. The variable obtained removing the common variance with reading and *g* from mathematical scores (Pure Mathematics), was still moderately heritable. Beyond genetic influences, only non-shared environment explained individual differences in Pure Mathematics. These results suggest that objectively shared environments such as growing up in the same family, going to the same school, and attending the same teacher's class, may not make two children more similar in their mathematical development beyond genetic similarity.

3.2 Introduction

Several mechanisms contribute to a successful acquisition of mathematics. Environmental factors associated with individual differences in mathematics include educational resources, such as access to books and parental education (reports from the Third International Math and Science Study -TIMMS, Mullis et al., 2001), content of the curriculum and structure of textbooks (Carnine, 1991), teachers' professional development (Saxe et al., 2001). It is important to note that not one single environmental factor has been definitively demonstrated to have a large effect on individual differences. For example, being in the same class and being taught by the same teacher contributes very little to making children more similar to each other in achievement (Kovas, Haworth, Dale, & Plomin, 2007a; Byrne et al., 2010). Moreover, some factors, considered to be 'environments', such as parental education, may actually reflect some gene-environment correlations, and therefore not be pure measures of environments.

Beyond environmental factors, mathematical learning relies on a number of general cognitive abilities (e.g.: Geary, 2004; Fuchs et al., 2010). Impairments in any of these supporting cognitive mechanisms may give rise to mathematical difficulties. For example, reading difficulties and low mathematical performance often co-occur (Vukovic, Lesaux & Siegel, 2010; Rubinsten, 2009; Dirks, Spyers & van Lieshout, 2008; Badian, 1999). Children with poor mathematical skills, however, can show no impairment in non-verbal intelligence and language abilities (Landerl, Bevan, & Butterworth, 2004).

Low scores in mathematical achievement combined with low-average IQ identify Mathematical Learning Disability (Geary, Hamson, & Hoard, 2000). However, the relationship between mathematics and IQ is complex as children with average IQ can also have low mathematical skills. It is possible that this relationship is mediated by other abilities. For example, low mathematical achieving children with an IQ similar to typically achieving children, scored as low as mathematics disabled children (with significantly lower IQ) on mathematical tests (Geary, Hoard, Byrd-Craven, & Numtee, 2007). Conversely, no differences in IQ (Raven's Coloured Progressive

Matrices test) were detected between children with Dyscalculia only, reading deficits only, reading and Dyscalculia and control (Landerl, et al., 2004).

Studies have also shown that children with poor mathematical abilities are slower when performing mathematical operations requiring retrieval from long term memory (Geary et al., 2000; Geary, 1993). Because speed of processing has been associated with fluid intelligence (Deary, Der, & Ford, 2001), it has been suggested that the association between mathematics and memory may be mediated through speed of processing, thus partially explaining the link of mathematics with IQ (Bull & Johnston, 1997). Indeed, in children with learning disabilities, full scale intelligence scores can account for variance in achievement scores in all academic domains (Hale, Fiorello, Kavanagh, Hoeppner, & Gaither, 2001).

Such inconsistent findings have encouraged researches to use different approaches in exploring the aetiology of mathematical skills. Recent genetically sensitive research has highlighted the importance of genetic factors. By using the twin design, behavioural genetic studies can estimate the portion of variance in a trait that is attributed to genetic and environmental influences. Using representative age-homogeneous samples, heritability of mathematical ability and achievement has been found to be moderate to substantial and consistent across development (Kovas et al., 2005; 2007a; Hart et al., 2009). Findings reported from studies on the large longitudinal sample of Twins Early Development Study (TEDS), estimated heritability at .68, with non-significant shared environment (Haworth, Kovas, Petrill, & Plomin, 2007). Moderate genetic (between .32 and .45) and non-shared environmental influences (.42 - .48) were found in the TEDS sample at 10 years of age (Kovas, et al., 2007c). This study also investigated different mathematical components addressing the mathematical heterogeneity. Non-numerical processing, that requires the understanding of concepts such as mental rotation and spatial operations, was found to be slightly less heritable (.32) than other mathematical processing (between .42 and .45). However, the genetic correlations (indexing common genetic factors) between the 5 components of mathematics were extremely high (~.90), indicating that mostly the same genes affect these aspects of mathematics, with only a small degree of aetiological specificity.

Similar estimates have been reported in studies conducted on different samples of twins. Heritability of mathematical fluency (the accuracy and speed in retrieving mathematical information and strategies from long term memory) was estimated respectively at .63 and .47 in 8.5 and 10 year old twins from the Western Reserve Reading Project Math (WRRPM) (Hart et al., 2009). Furthermore, the mathematical component of Applied Problems indicated a moderate heritability of .54 only at 10 years of age, suggesting developmental changes in genetic effects.

The comorbidity of reading and mathematical difficulties with low/normal IQ described earlier is reflected in genetic correlations between abilities. TEDS studies have reported the genetic correlation between mathematics and reading of .74; and between mathematics and *g* of .67 (Kovas, Harlaar, Petrill, & Plomin, 2005). These correlations suggest that if a gene is involved in the development of mathematics, there is 74% of probability that the same gene is also involved in reading ability and 67% of chance that is involved in *g.* Similarly, other studies have found genetic correlations between mathematics and reading ranging from .47 to .61 (Knopik & DeFries, 1999) to .98 (Thompson, Detterman, & Plomin, 1991). Genetic correlations have also been estimated between measures of reading and *g* (e.g. Harlaar, Hayiou-Thomas, & Plomin, 2005; Tiu, Wadsworth, Olson, & DeFries, 2004; Thompson et al., 1991). These studies suggest that most of the genes influencing mathematics also influence reading and/or *g*. This is referred as the "Generalist Genes Hypothesis" (Plomin & Kovas, 2005), according to which different cognitive domains are, to a large extent, influenced by the same pool of genes. According to this theory, different cognitive abilities share much of the same aetiology. To further support this theory a recent report from TEDS at the age of 12 confirmed a strong genetic correlation among different abilities: .62 between reading and *g*; .58 between reading and mathematics; .75 between *g* and mathematics (Haworth et al., 2009a).

Although research suggests that the genes involved in learning are, to some extent, shared across abilities, there is evidence that some genetic influences are trait-specific. An investigation in TEDS at the age of 7 showed that despite the great genetic overlap between mathematics, reading and *g*, approximately 25% of the genetic influences in mathematics are independent from reading and *g* (Kovas et al., 2005). At the age of 10, around 20% of the genetic variance in mathematical abilities was explained by genetic influence not shared with reading, science and *g* (Kovas et al., 2007a). Studies that used other samples to investigate the relationship between mathematics and other abilities have also confirmed genetic effects specific to mathematics. In a sample of 10 year old US children, it was found that 59% of the genetic variance in mathematical component of fluency was not shared with reading fluency and *g* (Hart et al., 2009).

These mathematics-specific genetic influences seem to be stable, at least between the ages of 7 and 10. One important question is whether this independence from other abilities is maintained at later stages, when the mathematical concepts become more complex and abstract. It is possible that different genetic influences get involved as a consequence of the change in mathematics, or new genetic influences arise with development. Better understanding of the aetiology of these mathematics-specific genetic influences may help to improve mathematical learning.

3.2.1 Research question

1. The present study reports the results from 13,262 of 12 year-old twins, part of the Twins' Early Development Study. This large representative sample allowed us to address the following research questions: Do the relative contributions of genetic and environmental factors to variation in mathematics remain similar to those, previously reported at 10 years of age in the same sample? This question is not trivial for three reasons. First, at least for some educationally-relevant traits, heritability has been shown to increase with age. Second, many biological changes occur in children over the two year period. Third, under the UK National Curriculum set-up, at 10 years of age, children are still in primary school, where the same teacher is likely to teach most subjects, including mathematics and reading. At 12, all children are in the secondary school, where different subjects are taught by a specialist teacher.

These factors may lead to an abrupt change in the relative contribution of genes and environments to the variation in mathematics.

2. How variable is mathematical ability at this age after controlling for variability in reading and *g*? This question is of major practical and scientific interest. As described above, a large portion of the mathematical variance has been shown to be shared with reading and *g* at previous developmental stages described in this sample. Do reading and general cognitive intelligence become even more closely associated with mathematics at 12, or does a lot of independent variance in this trait emerge with the increased complexity of the subject?

3. What is the aetiology of this “Pure Mathematics” variance? One hypothesis would be that individual differences in mathematics, that are not associated with reading or *g* (Pure Mathematics), would be highly genetic. On the other hand, with a specialist mathematics teacher providing mathematical education for the first time, shared environmental influences might become particularly important. As many children move to a different school between the ages of 10 and 12, an additional source of environmental, potentially non-shared influence may be new peers.

The aetiology of “Pure Mathematics” was explored at 12 years of age. Univariate genetic analysis was used to estimate genetic and environmental influences on mathematical scores after controlling for the effects of reading and *g*. It is hypothesized that by removing the overlapping variance of reading and *g*, will lead to a change in the estimates of genetic and environmental influences on mathematical variation at this developmental stage.

3.3 Method

3.3.1 Participants

Participants in this study were the twins of the TEDS sample at age 12. The sample is described in the section 2.2.4.4. Prior to any analysis, all the twins with severe medical problems such as cerebral palsy, Down syndrome or autism were excluded. For the purpose of this analysis, twins for which English was not the first language were also excluded from the analysis. The final sample constituted 13,262 twins (6,631 pairs), of which 4,636 were MZ (2,318 pairs), 8,626 (4,313 pairs) were DZ, with a mean age of 11.71

3.3.2 Measures

Measures of reading, mathematical and general cognitive ability were obtained using an internet web-based battery and telephone testing. Internet based assessments give the opportunity to collect data from large samples quickly and at low cost. Furthermore, web data collection requires little human manipulation in terms of extraction and data entry, reducing data errors. TEDS internet assessments have been previously validated and successfully carried out when the twins were 10 year of age (Kovas et al., 2007a). In addition to the mathematical scores obtained from the web assessment, measures of mathematical school achievement were collected with questionnaires sent to the twins' teachers. Each measure used in this study is described below, together with means and standard deviation on unstandardised scores. Further analyses were conducted on standardized variables of the accuracy scores.

READING was assessed on 2 tests of reading comprehension and 2 tests of reading fluency.

PIAT – Peabody Individual Achievement Test (Markwardt, 1997). The 3 practice trials, followed by 82 test trials were sentences testing the twins' reading comprehension. A written sentence was presented on the screen. The text was replaced by four pictures. Answers were given by clicking with the mouse on one of pictures that referred to the sentence (M=58.03, SD=10.64).

GOAL – Formative Assessment in Literacy - Key Stage 3 (GOAL plc 2002). The test is based on academic reading achievement required for the UK National Curriculum Key stage 3. Reading comprehension was assessed on 36 questions with 4-choice answer. (M=23.53, SD=6.37).

Woodcock-Johnson III Reading Fluency Test – (Woodcock, McGrew, & Mather, 2001). Reading fluency was measured with a list of up to 98 statements, either true or false, that could be answered with "yes" or "no" within 3 minutes time limit (M=58.41, SD=13.38).

TOWRE – Test of Word Reading Efficiency (Torgesen, Wagner, & Rashotte, 1999).The test measured reading fluency and accuracy on two subtests. The test of Sight Word Efficiency required the twins to read aloud, as quickly as possible a list of 54 words; the Phonemic Decoding Efficiency test required reading a list of 85 non-words. For each list there was a time limit of 45 seconds. The testing was carried out via telephone. Twins received the lists via mail, in sealed envelopes with separate instructions to open the envelopes only at the time of the testing session. (TOWRE Word: M=71.71, SD=10.87; TOWRE non-word M=41.95, SD=11.41).

A reading composite score was computed by averaging the standardized means of the five reading tests (M=.005, SD=.997).

GENERAL COGNITIVE ABILITY (g) was measured with two verbal and two non-verbal ability tests.

General Knowledge. Adaptation of the WISC-III- PI (Kaplan, Fein, Kramer, Delis, & Morris, 1999). This test of verbal reasoning consisted of 30 general knowledge questions. Response was given by choosing one of the 4 proposed answers (M=21.37, SD=4.22).

Vocabulary Test. WISC-III- PI (Kaplan, Fein, Kramer, Delis, & Morris, 1999). This verbal test consisted of 30 vocabulary questions. Out of the 4 proposed answers, 3 or 4 were correct solutions. 2 points were assigned for the best correct answer, 1 point for the others (M=39.77, SD=10.12).

Picture Completion Test. (Wechsler, 1992). The trials of this test of non verbal reasoning consisted of 30 pictures in which one recognisable part was missing. The task was to point the screen in the place of the missing detail (M=19.99, SD=3.88).

Raven's Standard Progressive Matrices. (Raven, Court, & Raven (1998). Stimuli consisted of 24 incomplete patterns (matrices). The task was to choose the correct pattern to complete the matrix among the choice of 8 (M=10.81, SD=3.50).

Standardised means of the 4 tests were averaged to create a composite score of *g* (M=-.00, SD=.99).

MATHEMATICS was measured with a web-based battery of tests and with Teacher Assessment.

Web-Testing. Test items for this task were 95 questions selected from the nferNelson booklets levels 1 to 8 (NFER-Nelson, 1994, 1999, 2001). nferNelson independent assessment of mathematic skills reflects the learning achievement levels required by the UK National Curriculum. 33 questions were taken from the category "understanding numbers". Answer to the questions of this test required understanding of the relationship between algebraic and mathematical operations (for example, "Type the correct number in each box: 123 +__= 123; 123 -__= 123; 123 x__= 123; 123 ÷__= 123"). 25 questions belonged to the "non-numerical processes" category and their solution required processing concepts of shapes and space without the aid of numerical information (for example: "Which is the longest drinking straw? Click on it"). 37 questions were taken from "computation and knowledge" which required the recall and application of well rehearsed mathematical facts (for example ''Work out the answer to this sum and type it in the box: 64 + 905 = __"). The task required to click on one of the multiple choice solution proposed or to type the correct answer in the box on the screen. For a complete description of scoring systems and administration for the mathematical test, see Kovas et al. (2007a).

Questions of the three categories were administered in a single mathematical test and for the purpose of this analysis the mathematical score was calculated on the whole 95 questions, therefore there was no need to compute a composite score (M=67.11, SD=14.78 on unstandardized scores, M=.004, SD=.996 on standardized scores).

Teacher Ratings of the National Curriculum Mathematical Achievement. Teachers filled in a questionnaire that rated children on their

mathematical achievement level according to the UK N.C. The children were assessed in 4 sub-components of mathematics: using and applying mathematics; number and algebra; shape, space and measures, handling data. Ratings were given on a 1 to 8 scale, with 9 given for exceptional performance. The correlations among the four sub-components ranged between .93 and .96 ($p<.01$, 2-tailed). For analysis purposes we derived a single score that best summarised the overall performance according to the Teacher rating of achievement on the National Curriculum for each child. The 4 scores were combined together into the Teacher rating score using the principal component analysis, with the unrotated first principal component explaining 95.7% of the variance (M=0.44, SD=0.93).

Teacher ratings have been previously shown to be a reliable measure of school achievement (Hoge & Coladarci, 1989; Kovas, et al., 2007a). In the present sample there was a good correlation ($r=.55$; N=5,367; $p<.01$; 2-tailed) between teacher ratings of the twins' academic achievement and the web-testing scores.

3.3.3 Twin method

This thesis utilises the twin method as part of the quantitative genetic methodology to conduct the genetic sensitive investigation.

In any population, measurements of any traits reveal a degree of variability around the means. However, the variability observed among family members is less than that across unrelated individuals. The twin method takes advantage of the twins' genetic relatedness to make inferences regarding the origins of individual differences in traits. Monozygotic (MZ) twins result from the division of a single fertilised egg - the zygote, therefore they inherit the same genetic information. It is generally assumed that MZ twins are genetically identical, although recent research suggests some genotypic differences (e.g. Bruder et al., 2008; Kaminsky et al., 2009). Because of their genetic relatedness, MZ twin genetic correlations (described in Chapter 1, section 1.2.9.2) is assumed to be 100%. Dizygotic (DZ) twins occur when two eggs are fertilised at the same time. DZ twins, like any other pair of siblings, share on average 50 percent of the segregating genes; for this reason their genetic correlation is assumed to be 50%. As described in Chapter 1, section 1.2.9.2, genetic effects arise also as results of more complex non-additive effects, these are referred to as "dominance". As MZ twins share the same genetic information their correlation due to dominance is assumed to be 100%. Dizygotic twins' alleles are made up from a random combination of one allele from each parent. In receiving just one allele from one of the parents the interaction among genes (causing dominance) will not be transmitted from parents to offspring. The chance that offspring inherit the parents' genetic interaction is 25%, therefore, DZ twins are expected to correlate on average .25 for dominance. Dominance effects are inferred if the correlation of MZ twins is more than two times higher than the correlations of DZ twins.

Behavioural genetic research suggests that variations in observed traits occur because of genetic and environmental reasons (Plomin et al., 2008). In quantitative genetic analyses, the total variance in a trait is decomposed into sources that contribute to the variation: additive and non-additive genetic influences; and shared and non-shared environmental influences. As it is not possible to estimate additive and non-additive genetic effects in the same model (non-additive genetic effects assume different ratio correlation for DZ twin compared to the additive effects), it is common practice to model non-additive effects only when they are indicated by the patterns of the twin correlations (if the MZ twin correlation is double the value of the DZ twin correlation). Heritability estimates provided by the additive models are referred to as "narrow-sense heritability", as opposed to the "broad-sense heritability" that pertains to additive and non-additive genetic influences. Dominance effects will not be further discussed in this report as no dominance was indicated by the twin correlations in mathematics or any other variables examined in this thesis.

Twins brought up in the same family may be similar to each other because of the influences of their common environment, as well as because of the influences of their shared genes. If genes play an important role in a trait, identical (MZ) twins must be more similar on that trait, compared to fraternal (DZ) twins. The influence of genetic factors (heritability) is calculated as twice the difference between the MZ and DZ twin correlations. Shared environmental factors are implicated if the DZ twin correlation is greater than half of the MZ twin correlation, and can be calculated as the difference between the MZ twin correlation and the heritability. Non-shared environmental influences are indicated by the extent to which the correlation between MZ twins is not 100%. In a practical example, if the MZ twin correlation for a trait is 0.8 and the DZ correlation is 0.5, heritability for that trait would be 60% [2 * (0.8 - 0.5) = 0.6], shared environment would be 20% (0.8 – 0.6 = 0.2) and non-shared environment would be 20% (1 – 0.8 = 0.2).

3.3.4 Assumptions of the twin method

As in many statistical methods, the twin model relies on some assumptions. These assumptions are made on equal environment, zygosity determination, assortative mating, and generalisability. Violation of these assumptions may lead to incorrect conclusions on the causes of variations for that trait.

3.3.4.1 Equal environment assumption

One important question about twins is whether the degree of genetic similarity is reflected in their environmental experience. In other words, do MZ twins go through more similar experience compared to DZ twins because they are more similar genetically? If this was true, the environmental influences would be incorrectly included in the genetic estimates (Plomin et al., 2008). In the classic twin method it is assumed that the same shared environmental influences will equally affect MZ and DZ twins that is to say that shared environment is the same for both MZ and DZ twins. Although MZ twins are more likely to be treated alike (Loehlin & Nichols, 1976), research that has investigated the equal environments assumption has shown that this similar treatment is a reflection of their increased genetic similarities (e.g. Evans & Martin, 2000). Using simulation analyses and real data on aggression and spatial abilities, it was found no violation of this assumption (Derks, Dolan, & Boomsma, 2006). Equality of environment was also tested in relation to the perceived twins' zygosity - whether they were perceived more similar/different rather than being classified according to zygosity, showing that perception of zygosity had no influence on estimates obtained using the twin method (Scarr, & Carter-Saltzman, 1979). The zygosity perception was treated as a specific family environment showing that zygosity perception had no significant influence on any of 5 psychiatric disorders investigated (Kendler et al., 1993). Another argument posed against the equal environment is that MZ twins have more similar prenatal experience compared to DZ twins, as they often share the same amniotic sac (chorion). However, the evidence is mixed. For example, no differences in heritability estimates of blood pressure between MZ twins have been found as effect of sharing the same chorion (Fagard et al., 2003). Conversely, MZ twins sharing the same chorion (monochorionic) were more similar in almost all subscales assessing IQ compared to both, the MZ twins who did not share chorions (dichorionic) and to DZ same-sex twins (Jacobs et al., 2001). Similarly, monochorionic MZ twins were found to be more similar than dichorionic MZ twins in a range of behavioural traits (Sokol et al., 1995). However, it is also likely that MZ twins experience differences in prenatal environments as they have to compete for nutrients during gestation, in fact monochorionic MZ twins show greater weight differences at birth, as well as in dental characteristics, compared to dichorionic MZ twins (Race, Townsend, Hughes, 2006). The equality of environment has been one of the most debated assumptions of the twin method; the contrasting evidence comes mostly from the pre-natal environmental conditions, calling for more research in the area (e.g. Rijsdijk & Sham, 2002).

3.3.4.2 Zygosity determination

Determining zygosity correctly is a crucial point of the twin method as it relies on the MZ - DZ comparison. In the case of opposite sex twins, there are no doubts about their zygosity, they can be only dizygotic. MZ twins are always of the same sex, but same sex DZ twins occur in around one third of the twin-births. Zygosity can be assessed with an accuracy of around 99% using DNA markers (Chen et al., 2010). However, costs and time are involved in carrying out such procedures. An alternative method to determining zygosity is through questionnaires asking about twins' similarity on a variety of traits and features (as described in TEDS zygosity questionnaires, Chapter 2, section 2.2.3). The method yields 95% accuracy against validation of zygosity with DNA markers (Price et al., 2000).

3.3.4.3 Assortative mating

Assortative mating refers to a type of gene-environment relationship, by which partners chose each other on the basis of specific characteristics rather than randomly. For example, individuals may choose partners similar to themselves on a particular trait (positive assortative mating) or dissimilar (negative assortative mating). A choice based on phenotypic traits, indirectly may reflect the partners' genetic and environmental similarity. Their offspring may receive a non-random transmission of genetic influences, with positive assortative mating increasing the similarity between parents and offspring. Positive assortative mating (that increases similarity among family members) leads to the increase of DZ twin correlations. The correlations of the MZ would be unaffected as they are genetically identical. This translates into an overestimation of the shared environmental influences at the expense of genetic influences (Rijsdijk & Sham, 2002; Neal & Maes, 2002). The presence of non-random assortment can be detected in family studies, by looking at the parents' phenotypic correlations. For example, there is evidence that partners have preferences for some physical characteristics such as weight and height, although correlations between spouses are modest, .20 and .25, respectively (Spuhler, 1968). Assortative mating is more substantial in behavioural and cognitive traits such as alcohol abuse (e.g. correlation .38 between parents, Agrawal et al., 2006) and *g* (correlation .40, Jensen, 1978). Although heritability estimates in twin studies may be biased by the presence of assortative mating, these effects are modest (Loehlin, Harden, & Turkheimer, 2009) suggesting that in many cases violation of this assumption is not crucial.

3.3.4.4 Generalisability

Another important question is whether findings from twin studies can be generalised to non-twins. Twins do not constitute a random sample and their representativeness has been questioned. For example, newborn twins on average weight less than singletons (Rutter & Redshaw, 1991); however by middle childhood the weight differences disappear (MacGillivray et al., 1988). Other differences between twins and singletons are alleged on physical characteristics such as increased congenital malformations or perinatal mortality, linked to pre-natal condition (Rutter & Redshaw, 1991). Twins also often have shorter gestation periods and tend to be born three-four weeks earlier (Phillips, 1993). Although these early differences may point to the possibility of different developmental paths between twins and singletons, studies found that they do not differ for brain volume measures (Hulshoff Pol et al., 2001). Similarly, twins' achievement in adolescence has been shown to be the same as singletons' (Christensen et al., 2006), despite the slightly lower average IQ displayed by twins in early childhood (Ronalds et al., 2005). Using genome-wide scans, the heritability of height has been estimated from sibling pairs finding estimates similar to twins studies (Visscher et al., 2006). Many studies tested the twin method and concluded that findings from twin studies are valid and are applicable to the general population (e.g. Kovas et al., 2007a; Plomin et al., 2008; Hart et al., 2010a).

Overall, the twin method has been found reliable and valid by many studies. However, as with any other method, it is important to test its assumptions in relation to each trait of interest, as well as to find convergent support for its findings using other methods. Adoption studies, experimental methods, animal studies, and molecular genetic approaches, can all be used to replicate and build on the findings from the twin studies.

3.3.5 Univariate genetic analysis

3.3.5.1 Twin analyses

As previously described, the sources of the observed variation in traits can be attributed to genetic and environmental influences. Behavioural genetic methods examine phenotypic similarity in relation to genetic similarity. In the case of twins, it is possible to make accurate inferences on their genetic similarity. The degree of their phenotypic similarity is indexed by intraclass correlations (ICC; Shrout & Fleiss, 1979). This type of correlation is appropriate when the data is structured in groups; in the case of twin data, groups of two (the twin pair). The ICC is a ratio of the variance between the twins in the pair (within-variance) over the observed total variance in a trait. Therefore, this index can be taken as the proportion of total variance that is explained by the within twin-pairs variation. In the univariate genetic analysis, the comparison of ICCs between MZ and DZ twins allows to estimate the proportion of variance in a trait that can be attributed to additive genetic (A), shared (C) and non-shared (E) environmental influences (e.g. Plomin et al. 2008). These proportions of variance can be estimated using the Falconer's formula (Falconer & MacKay, 1996), which has already been used in the example in section 3.3.1. The MZ twin correlation on a trait is due to the influences of their shared environments and the total of the genetic influences: rMZ = A + C, where "r" is the MZ ICC coefficient. The DZ correlation due to additive genetics can be half of the correlation displayed by MZ twins (DZ share around 50% of the segregating genes): A/2, therefore the overall correlation on a trait for DZ twins will be: rDZ = (A/2) + C. Subtracting the rDZ equation from the rMZ it is possible to derive the heritability as: A = 2(rMZ - rDZ). In the model, genetic and environmental influences explain 100% of the variance in the trait, that is to say that their sum must be 1. From A + C + E = 1 (total variance), it is possible to obtain the shared environmental influences: C = rMZ – A; and the non-shared environmental influences: E = 1 - rMZ. In the formula the A, C, and E represent the genetic, shared environmental, and non-shared environmental components of variance explained. The notations h^2, c^2 and e^2 correspond to the standardised values of the ACE parameters and represent the proportion of variance explained. Using the proportion of variance, the Falconer's formula can then be expressed as:

$h^2 = 2(rMZ - rDZ)$	proportion of variance explained by narrow sense heritability	(3.1)
$c^2 = rMZ - h^2$	proportion of variance explained by shared environment	(3.2)
$e^2 = 1 - rMZ$	proportion of variance explained by non-shared environment	(3.3)

Analyses of twin data are carried out on the residuals of standardised scores corrected for average effects of age and sex (McGue & Bouchard, 1984). This is because twins' age across pairs is completely correlated, which could inflate twin correlations and could be wrongly attributed to shared environmental influences. The same applies to sex because MZ co-twins are all of the same sex, as are half of DZ pairs.

3.3.5.5 Univariate model fitting

Although the results of twin analyses can be easily ball parked from the simple twin correlations, the effects of genetic, shared and non-shared environmental influences are more accurately estimated using structural equation modelling. In twin model fitting the A, C, E parameters are estimated as the latent variables that more closely reproduce the observed MZ and DZ variance-covariance.

The first step of model fitting is to create a model, the saturated model, that makes use of the maximum number of parameters to reproduce the observed data (the MZ and DZ variance-covariance). The modeled data is then compared to the set of observations in order to obtain the residuals. The comparison between the modeled and observed data is iterated until the set of parameters minimising the residuals is found. The optimisation process is usually performed with specific software; the behavioural genetic analyses in this thesis were conducted using OpenMx software (Boker et al., 2011) running in the R environment (http://www.R-project.org).

The process of parameter estimation is carried out using maximum likelihood. This method provides an index, -2LL (minus two log-likelihood) indicating how well the modeled data fits the observed data. It has to be noted that the saturated model provides the best fit possible to the data because it uses the maximum numbers of parameters to model the data. The -2LL of the saturated model does not offer a valid assessment of the quality of the fit, but it is used to compare different models. Other parameters providing information for models comparison are the Akaike Information Criterion (AIC; Akaike, 1987) and the Bayesian Information Criterion (BIC).

The next step is to model the additive genetic variance (A), the shared environmental variance (C) and the non-shared environmental variance (E). The random error in the model is modeled in the E parameter estimate. The goodness of fit of this last model (ACE model) and the saturated model are compared. If the differences in -2LL between the two models is non-significant (p-value > 0.05), it means that the ACE model is not a worse fit than the saturated model. The AIC and BIC are also used to compare the fit: the lowest AIC and BIC refer to the most parsimonious (preferred) model.

The better fit is determined following the principle of parsimony, the Occam's razor principle, by which the better solution is reached using the fewest resources (parameters). Once the ACE parameters are estimated, nested (simpler) models are run dropping parameters and comparing the fit to the ACE model. If the difference in -2LL between the nested models and the ACE model is non-significant, it means that two estimated parameters describe the observed data as efficiently as three, therefore the nested model is preferred. The AE and CE nested models are obtained constraining C and A parameters (respectively) to zero. It is not possible to constraint E to zero as this would assume no measurement error. Model fitting allows to calculate confidence intervals around the estimates, which give an indication of their significance (see Plomin et al., 2008).

3.4 Results

3.4.1 Derived variables

Correlational analysis conducted on unstandardised scores showed a moderate correlation between the Mathematical web scores and the 9 measures of general cognitive ability and reading. Correlation coefficients ranged between .33 and .55 ($p<.01$, 2-tailed). The magnitudes of these correlations justified the creation of composite scores. First, all the scores were standardised, then a variable of general cognitive ability, *g*, was obtained by averaging the standardised means of the 4 general cognitive tests (General Knowledge, Vocabulary, Picture Completion, Raven). Similarly, a single reading score was created by averaging the standardised means of the 5 reading scales (GOAL, PIAT, TOWRE-words test, TOWRE-non words test, Reading Fluency). The correlation between the Mathematical web-test with the reading and *g* composites was respectively .55 and .61 while between composites of reading and *g* the correlation coefficient was .54 ($p<.01$, 2-tailed for all correlations). Twins are perfectly correlated in age and sex and this correlation can inflate estimates of the shared environment (McGue and Bouchard, 1984). For this reason variables and the composite scores were corrected for age and sex.

The correlation between Mathematical web scores and mathematical achievement rated by teachers was .55 ($p<.01$, 2-tailed). Therefore, a composite score was created averaging the standardised means of the mathematical web test scores and the teachers' ratings scores. From this variable and from the Mathematical web score alone the variance in common with reading and *g* was removed, - creating two ‘Pure Mathematics’ variables: web + teacher and web score only . Scores outside +/-3 standard deviations were considered outliers and removed from the analyses.

3.4.2 Descriptive statistics

Table 3.1 shows means, standard deviations, and the effect of sex and zygosity on the standardised composite scores of *g*, reading, Mathematics web test, and the two variables of Pure Mathematics. No effects of sex or interaction with zygosity were significant in any of the measures. The effects of zygosity were significant on both the reading and *g* composites, with MZ twins showing lower average performance than DZ twins. However, the effect sizes were very small, with zygosity accounting for less than 1% of the variance (η^2=.00).

Table 3.1

Means, Standard Deviations and ANOVA results by sex and zygosity

Mathematics web-test (N)	***N=4,655***		***N=1,685***		***N=2,970***		***N=2,101***		***N=2,554***							
Reading (N)	***N=3,720***		***N=1,355***		***N=2,365***		***N=1,638***		***N=2,082***							
"g" (N)	***N=3,922***		***N=1,444***		***N=2,478***		***N=1,711***		***N=2,211***							
Pure Mathematics web	***N=3,363***		***N=1,231***		***N=2,132***		***N=1,461***		***N=1,902***							
Pure Mathematics web+teach	***N=2,045***		***N= 748***		***N=1,297***		***N= 886***		***N=1,159***							
											ANOVA					
	All		***MZ***		***DZ***		***All Males***		***All Females***		***Zyg.***		***Sex***		***Zyg.*Sex***	***Tot.***
Measures	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***M***	***SD***	***p***	η^2	***p***	η^2	***p***	η^2
Mathematics web test	.03	.95	.00	.95	.04	.96	.03	.94	.03	.96	.15	.00	.84	.00	.75	.00
Reading	.02	.97	-.05	.97	.06	.97	.03	.97	.02	.97	.00	.00	.53	.00	.07	.00
"g"	.01	.98	-.07	.98	.06	.97	-.01	.97	.03	.98	.00	.00	.39	.00	.24	.00
Pure Mathematics web	.02	.95	.06	.94	.00	.95	.01	.98	.03	.94	.09	.00	.58	.00	.83	.00
Pure Mathematics web+teach	.01	.96	.02	.93	.01	.97	.00	.97	.02	.94	.74	.00	.83	.00	.95	.00

Measures	Tot. R^2
Mathematics web test	.001
Reading	.004
"g"	.005
Pure Mathematics web	.001
Pure Mathematics web+teach	.000

Mathematics web test = Mathematics web scores containing the overlapping variance with reading and "g". In the variables "Pure Mathematics" the variance shared between mathematics, reading and *g* has been removed. Pure Mathematics web = Mathematics web scores removed from the overlapping variance with reading and *g.;* Pure Mathematics web+teach = Composite of Mathematics web scores and mathematics teachers rating removed from overlapping variance with reading an *g*. All the measure are standardised with mean of zero and standard deviation of 1

3.4.3 Univariate genetic analyses

The intraclass correlations, indexing twins' similarity on Mathematics and the two Pure Mathematics variables, are shown in Table 3.2.

Table 3.2

Intraclass correlation for the Mathematical scores and the measures of Pure Mathematics, parameter estimates and 95% Confidence Intervals

Measure	***MZ intraclass correl. (C.I)***	***DZ intraclass correl.(C.I)***	***Variance of A (C.I)***	***Variance of C (C.I)***	***Variance of E (C.I)***
Mathematics (web test)	.66 (.64 - .69)	.42 (.39 – 45)	.52 (.44 - .59)	.16 (.10 - .22)	.32 (.30 - .35)
Pure Mathematics (web test)	.44 (.39 - .49)	.21 (.17 – 25)	.44 (.40 - .48)	N/A	.56 (.52 - .60)
Pure Mathematics (web+teachers)	.59 (.54 - .64)	.29 (.24 - .35)	.62 (.57 - .66)	N/A	.38 (.34 - .43)

C.I. = 95% confidence intervals; A=variance explained by genetic factors, C=variance explained by shared environmental factors, E=variance explained by non shared factors plus the error model.

MZ correlations were higher than DZ correlations for all variables, suggesting genetic influences on all variables. The significance of the genetic influences was indicated from the lack of overlap between the confidence intervals between MZ and DZ intraclass correlations.

The genetic, shared and non-shared environmental influences on the 3 measures were estimates with a univariate genetic analysis. Summary of the model fit comparison is presented in Table 3.3.

Table 3.3

Model Fitting univariate genetic analysis Mathematics and Pure Mathematics at age 12

		Model Fit					
		-2 log likelihood	degrees of freedom				Par. Est.
Measure	Model	-2LL	df	(Δ-2LL)	AIC	BIC	ep
	Sat.Model	23781.27	9269		5243.27	-57781.40	10
Mathematics Web	**ACE**	**23787.29**	**9275**	**-6.02**	**5237.27**	**-57828.18**	**4**
test	AE	23812.62	9276	-31.35	5260.63	-57811.64	3
	Sat. Model	18083.38	6764		4555.38	-41436.51	10
Pure Mathematics	ACE	18087.66	6670	-4.28	4547.66	-41485.03	4
web test	**AE**	**18087.66**	**6671**	**-4.28**	**4545.66**	**-41493.83**	**3**
	Sat. Model	10902.61	4087		2728.61	-25060.99	10
Pure Mathematics	ACE	10904.82	4093	2.21	2718.83	-25111.57	4
Web+teacher rating	**AE**	**10904.82**	**4094**	**2.21**	**2716.83**	**-25120.37**	**3**

Sat. Model = Saturated Model; Par. Est. = Parameters Estimated; AIC = Akaike's information criterion; BIC = Bayesian information criterion. Lower values of the two indices indicate better fit of the model; Δ-2LL= difference between the Saturated Model and the nested models likelihoods. The better fitting models are indicated in bold. For the Mathematics web scores, the best fit is provided by the ACE model. For the two Pure Mathematics variables, the best fit is provided by the AE models: with no decrease in likelihood two estimated parameters fit the data as well as three estimated parameters. The better fit is confirmed by the smaller AIC and BIC indices in the AE models.

The estimates, reported in Table 3.2, were consistent with the pattern of the twin correlations. Individual differences in Mathematics web test scores (including variance with reading and *g*) were explained mostly by genetic influences (.52); with non-shared influences of .32; and very modest but significant influence of shared environment (.16). The shared environmental influences on the Pure Mathematical scores (both web test, and web+teacher composite) were non-significant. Individual differences in Pure Mathematical web scores were explained by genetic and non-shared environmental influence in almost equal measure (.44 and .56 respectively). In the Pure Mathematics web+teacher composite, there was a stronger genetic contribution (.62) compared to non-shared environmental influences (.38)

3.5 Discussion

The aim of this study was to investigate the aetiology of individual differences in Mathematics and in Pure Mathematics (a variable obtained by removing the phenotypic variance shared by mathematics with reading and *g)* at 12 years of age, using quantitative genetic methodologies.

The results revealed that at age 12, mathematics is a heritable trait (.52) with environmental influences of the non-shared type (.32), with very modest shared environmental influences. These findings are consistent with other TEDS studies, suggesting stable genetic influences in mathematics across development.

Between the age of 7 and 12, the concepts involved in mathematics become more abstract and more complex. However, the results of this investigation show that genetic influences contributing to mathematical differences among children have highly similar strength at ages 12 and 7. Stable genetic influences indicate that environmental influences do not become more or less important in mathematics, despite the changes in the trait of study. Starting at the age of 12, in the UK school mathematical education is delivered by specialist mathematical teachers. This change in the objective-mathematical environment does not seem to contribute to a change in the amount of subjectively relevant environments (shared environments) to individual differences in mathematics. It is possible that the structured UK National Curriculum and the homogeneous training of mathematical teachers contribute to the stability of the trait, whereas genetic influences contribute to individual differences. The experiences unique to each individual (non-shared environments) also contribute to children's differences, almost in the same amount, at ages of 7 and 12.

Given the genetic and phenotypic overlap of mathematics with other abilities, it is also important to understand whether the observed stability in heritability estimates is mediated by other abilities. In other words, it is possible that later mathematical ability is supported more by reading skills, reasoning, and general intelligence than earlier mathematics. Our results suggest that this is not the case. After removing the shared variance between mathematics, reading and *g*, what is left in terms of mathematical variation, is still heritable. This means that similarly to what was previously found for mathematics at the age of 7 and 10 (Kovas et al., 2005; Kovas et al., 2007a), some genetic influences are specific to mathematics at the age of 12. The mathematics-specific genetic influences in this study were higher compared to the unique genetic influences reported at the age 7 and 10. It is possible that later mathematics (at 12) shares less genetic influences with other domains than the earlier mathematics.

Alternatively, the different estimates may be partially due to the different mathematical components assessed at the different ages. Some evidence indeed exists for different heritability for mathematical components (e.g. Hart et al., 2009; Kovas et al. 2007c).

Pure Mathematics of web scores was less heritable than when teacher-rated school achievement was included in the composite. This difference in heritability is of interest, as school achievement may differ from the web assessment, as they assess a wider range of skills over time. Alternatively, a rating bias of teachers may inflate MZ correlation - resulting in higher heritability estimates.

Despite some discrepancy in genetic estimates, the measures of "Pure Mathematics"showed the absence of shared environmental influences. Only individual- specific, rather than family- or school-wide environmental experiences showed influences on unique variance in mathematics

3.5.1 Conclusion

The existence of mathematical variance independent from other cognitive abilities calls for the search for cognitive mechanisms underlying such variance. Recently, cognitive literature has identified some processes that appear to be specific to the mathematical domain. These processes, referred to as "number sense" underpin the ability to estimate and approximate numerosites (Dehahene, 1997; Halberda, Mazzocco, & Feigenson, 2008). Approximation abilities have been found to correlate with mathematical test scores and achievement, even after controlling for other cognitive abilities (e.g. Halberda et al., 2008; Booth & Siegler, 2006). One of the possibilities is that causes of variation in estimation abilities overlap with those of variation in Pure Mathematics. If number sense contributes to mathematical skills, independently from other abilities, the understanding of its aetiology could clarify the mechanisms that facilitate a successful mathematical acquisition. In turn, this information could guide changes in the way we conceptualise and teach mathematics. It could also lead to efficient screening for predicting individual variation in mathematical skills and apply suitable individualized environmental interventions. The following chapters (Chapter 5, 6 and 7) utilises the unique large longitudinal TEDS sample to carry out the first genetically sensitive investigation into individual differences in number sense and its links with mathematics.

Chapter 4: Development and validation of a mathematical-number sense web-based battery: a pilot study

4.1 Abstract

The aim of this study was to create and validate a web-based battery of tests, age appropriate for 16 year olds, and designed to assess mathematical skills, general cognitive abilities and number sense. The purpose was to add this new tool of assessment to a larger Internet battery of tests to be administered to the 16 year-old twins, as part of the Twin Early Development Study (TEDS). The study was conducted in two stages. The first stage involved the selection of 11 measures: 4 arithmetic tasks, 2 numerical Stroop tasks, 3 number sense tasks and 2 general cognitive tasks. The 11 test were administered to a sample of 100 16-year-old students, either in pen and paper format, or on computers. In the second stage, 6 of these tests were selected based on reliability analyses, conducted on the data collected in the first phase. During this stage, a new mathematical task was added. The 7 measures were programmed, implemented online and administered to the same sample of 16-year old students for test-retest reliability and online validation purposes. In order to be implemented in the final TEDS Internet battery, all the measures were further modified, therefore a second validation study was conducted on a subset of 48 TEDS twins (24 twin-pairs) soon after the first wave of TEDS testing. The new battery revealed to be a reliable tool of assessment showing good internal validity and reliability for all the measures.

4.2 Introduction

4.2.1 Definition of the research problem

The acquisition of mathematical competences is a gradual process that requires many cognitive abilities and skills working in synergy. The attainment of mathematics relies on the support of general cognitive processes including memory, reasoning and IQ (Butterworth, 2005). Evidence that acquired skills, such as reading and writing, may support mathematical development is provided by the fact that dysfunctions in any of these abilities can lead to different manifestations of mathematical impairments (e.g. Vukovic et al., 2010; Dirks et al., 2008, Geary, 1993).

Overall, less research has been dedicated to mathematics compared to reading or language, leading to less information available about mathematical cognition compared to these other domains. For example, while a number of established tests are used as valid predictors of reading difficulties (Gersten, Jordan & Flojo, 2005), a valid instrument for screening measures of mathematical difficulties is yet to be determined (Dowker, 2005). The task of predicting and screening for mathematical difficulties starts with identifying performance that falls below "normal" levels. These tests aim to assess performance into the different mathematical components. However, these instruments tag cognitive mechanisms that may not necessarily be the ones underlying mathematical difficulty per se. Mathematical difficulties are diagnosed on the basis of scores that fall below an arbitrary cut-off point of performance (see Dowker, 2005). This kind of assessment raises a number of issues. Firstly, these instruments may actually measure a number of cognitive traits involved in the mathematical domain, not mathematical abilities. Secondly, reaching a diagnostic threshold on one scale may not necessarily mean reaching a diagnostic threshold on another scale. Lastly, the sensitivity of the measures is determined a priori by the cut off points along a continuous trait. By means of this kind of assessment mathematical problems and disabilities are placed at the low end of a continuously distributed trait. It remains unclear whether mathematical disability is a qualitatively different dimension from mathematical difficulty. This is evident in existing literature where the term “disability” and “difficulty” are often used interchangeably (Dowker, 2005).

Although the work reported in this thesis is based on normal abilities, this brief overview on mathematical difficulties emphasizes the current understanding of the multifaceted nature of the mathematical domain. The heterogeneity of mathematical disabilities hinders our understanding of the relationship between cognitive abilities and mathematical performance. Fundamental questions, such as the definition of mathematical ability and disability are still being debated.

Two competing (or complementary) views on the development of mathematical ability and disability have been proposed. The first attributes mathematical difficulties to some dysfunction of the general cognitive mechanisms that support the mathematical domain and are shared with other abilities (e.g. Geary et al., 2007). The second view attributes the deficits to some numerical processes that are specific only to the mathematical domain (e.g. Butterworth, 2005). Genetically sensitive designs can be employed to provide deeper insights into the nature of the links between mathematics and other cognitive domains.

4.2.2 Domain General Abilities

Individual differences in mathematical performance may be driven by the same cognitive mechanisms that influence the learning of any skill. According to this view, mathematical performance can be predicted from general cognitive mechanisms such speed of processing and working memory.

4.2.2.1 Memory

Working memory plays a central role in learning as it is considered a temporary storage where information is maintained and manipulated before being transferred to the long-term memory system (Baddeley, 1983).The role of working memory in the development of mathematical skills has been extensively investigated. Mathematical impairments have been attributed to dysfunctions of the three working memory components: the central executive, phonological loop and visuo-spatial sketchpad. According to Baddeley's working memory model (Baddeley, 1983), the central executive role is to coordinate and mediate information flow and functions of the two "slave" systems: phonological loop and visuo-spatial sketchpad. It also performs more complex functions, such as selecting and switching between different strategies and providing temporary activation of information from long term memory (Baddeley, 1996; Miyake et al., 2000).The phonological loop encodes and maintains phonological information with the process of subvocal rehearsal; the visuo-spatial sketchpad maintains and manipulates visual and spatial material (Baddeley, 1992).

A number of studies have demonstrated that the execution of arithmetical processes is disrupted when the central executive, phonological loop and visuo-spatial sketchpad, are engaged by another task, such as in dual task paradigms (e.g. Logie, Gilhooloy & Wynn, 1994; Lemaire et al., 1996; Lee & Kang, 2002). Furthermore, children with arithmetical difficulties show short term memory deficits. Evidence is provided by the fact that children who are slower at counting make greater use of fingers to aid their memory and keep track of digits: counting relies on memory span and fingers are used as an external support for memory (Geary, 1993). Siegler & Ryan (1989) argued that children with impairments in central executive functions showed arithmetical difficulties only in material involving numbers. They suggested that central executive dysfunction prevented the ability to maintain and manipulate numerical information. Another study suggests that impairments in the central executive prevent access and retrieval of information from the numerical lexicon (D'Amico & Guarnera, 2005).

Poor mathematical abilities are also associated with poor performance in visuo-spatial working memory in the Corsi Block task (McLean & Hitch, 1999; Bull et al., 1999). However, this association may not be significant beyond the common variance with reading and IQ (Bull et al., 1999), suggesting that visuo-spatial abilities are only indirectly related to mathematical skills. Phonological loop impairments have been linked with problems in recalling numerical material. Children with poor mathematical skills perform worse than controls in digit span tasks (Swanson & Sachse-Lee, 2001). However, other studies have assessed normal performance in digit span tasks in low mathematics performing children (e.g. Bull & Johnston, 1997; Passolunghi & Siegler, 2001). These studies show correlation between memory components and mathematical disabilities. However, our limited understanding of the underlying mechanisms does not allow us to suggest that this relationship is causal.

4.2.2.2 Speed of processing

The examples provided above show that there are some inconsistencies in the mathematical literature regarding the role of working memory in mathematical difficulties. It is still debated in the literature whether memory deficits directly affect numerical processing. Individual differences in speed of processing (speed of number identification, visual number matching, encoding of digits) have been found to affect retrieval of information from long term memory, leading to poor mathematical performance. This suggests an indirect involvement of memory in mathematical abilities, mediated by speed of processing (Bull & Johnston, 1997; Geary & Wiley, 1991). Speed of processing is associated with physiological factors of the human nervous system at the neuronal level (Neubauer, 1997). Perceptual, motor and cognitive tasks that require rapid response show slower reaction times in children, compared to adults (Kail, 1991). These differences have been attributed to children's developing information processing system, and this supports the theory that speed of processing is an index of intellectual development (Kail, 1992). Reaction times are related to the individuals' abilities to gain information and have been associated to psychometric intelligence tests scores (Kail, 1992; Deary, Der, Ford, 2001).

Individual differences in speed of processing appear to be important in mathematical skills. Children with poor mathematical abilities were found to be slower when performing mathematical operations that require retrieval from long term memory (Geary, 1993). One of the explanations given is that poor mathematicians simply have a slower information processing. Alternatively, it is possible that children with mathematical disability fail to develop the automated processes of the arithmetical operations because of slow processing of numerical information. In support of this theory, it was found that children with arithmetic learning disability had no evident impairments in working memory capacity (Hitch & McAuley, 1991). However, their poor arithmetical performance indicated that they had slower access to numerical material in long term memory due to the lack of familiarity with mathematical material. Bull & Johnston (1997) reported lack of automaticity in the retrieval of mathematical facts in children with low mathematical performance. In this study, difficulties with accessing mathematical material from long term memory were attributed to slow speed of processing. Their results were evaluated against Hitch and McAuley's theory, hypothesizing that children with poor familiarity in a particular area are not motivated to use that material. The infrequent activation creates weaker links with that subject - preventing the creation of automaticity of the material in question.

To summarise, speed of processing is an indicator of general cognitive abilities that is also involved in mathematical skills, linking mathematical abilities with other general cognitive functions. This relationship adds an additional layer of complexity to the understanding of mathematical abilities in relation to general cognitive functions.

4.2.3 Number Specific Abilities

Recent literature suggests that some cognitive mechanisms may be uniquely associated with mathematics (e.g. Halberda et al., 2008). Abilities underpinning these mechanisms are referred to as "number sense" (Dehaene, 1997). Among these abilities, estimation processes have been singled out as predictive of mathematical skills (e.g. Gersten & Chard, 1999; Butterworth, 2005).

4.2.3.1 Numerosity estimation

Discrimination of numerosity sets (the number of items in a set) is based on comparison of sets in terms of more or less. Estimation of numerosities reproduces inexact quantities, and in this respect is different from exact counting. There is evidence that a specific mechanism, the Approximate Number System (ANS), is the quantification system responsible for estimation functions (Feigenson, Dehaene & Spelke, 2004). Estimations of numerosity are ratio dependent (see Halberda & Feigenson, 2008) rather than modality dependent (Barth, Kanwisher & Spelke, 2003). Estimation of numerosity follows the Weber Law (Weber, 1834) and is indexed by the Weber Fraction (e.g. Pica et al., 2004). Weber Fraction is an amodal index of the ability to perceive changes in the appraised measure. For example, given two arrays, one with 4 items and one with 6, (ratio 2:3) the Weber Fraction is derived by dividing the difference of the two ratios by the smallest number in the ratio: [(3-2)/2] =0.5. Understanding numerosities is of particular relevance as it has been suggested that mathematical disabilities and difficulties can arise from deficits in numerosity processing (Butterworth, 2005). The aetiology, of estimation remains poorly understood.

Empirical evidence points to an evolutionary origin of the ANS. Estimation abilities are found in animal species (Meck & Church, 1983), as well as in humans at pre-verbal age. Infants as young as four days seem to discriminate two and three syllable words (Bijeljac-Babic, Bertoncini, & Mehler, 1993). Six month old infants respond to numerosity (Xu & Spelke, 2000) and have expectations on the outcome of simple arithmetic calculations like adding up to 5 items (McKrink & Wynn, 2004). Furthermore discrimination abilities improve with development (Lipton & Spelke, 2003). It has also been demonstrated that people differ significantly in estimation abilities and that these differences positively correlate with mathematical performance (Halberda et al., 2008).

4.2.3.2 Estimation of numerical magnitudes

Another specific numerical process is the estimation of numerical magnitudes. Estimation of numerical magnitude is a specific type of estimation that is based exclusively on numerical knowledge. If, for example, we were to estimate the weight of an object (without using a weighting instrument), we should have some experience with weight. Booth & Siegler (2006) propose that estimation on a number line (where the estimation occurs between numbers) is based on numerical knowledge only. Such estimation requires just the pre-existence knowledge of a numerical construct, whereas estimation of numerosity does not require understanding numbers. Estimation skills improve with development and experience (Dowker, 2003). For example, number line estimation seems to improve with age (Siegler & Opfer, 2003), and with improving of number knowledge (Siegler & Ramani, 2008). Children's inaccurate estimation compared to adults' has been attributed to the use of an immature logarithmic mental representation of numbers. In a logarithmic numerical representation the distance between numerical magnitudes at the end of the range are underestimated compared to magnitudes at the beginning and the middle. For example, the distance between 10 and 20 is mentally represented as greater than the distance between 90 and 100. As a result, in a logarithmic representation the numbers at the end of the range are "compressed". Adults produce more accurate estimates, as they rely on the more efficient linear representation of numbers, although a logarithmic representation is still present (Siegler & Opfer, 2003; Siegler & Booth, 2004; Booth & Siegler, 2006). Estimation abilities vary across individuals, with these individual differences positively correlating with mathematical test scores (e.g. Booth & Siegler, 2006).

4.2.4 Genetic Studies

Quantitative genetics can provide clearer picture of the aetiology of abilities and of the nature of relationships among abilities. Behavioural genetic methodologies (e.g. twin studies) allow for estimation of the portion of the variance of a trait that is attributed to genetic and environmental influences. These methodologies further allow for estimation of the genetic overlap between different measures (i.e. the extent to which genetic factors influencing one measure also influence another measure).

Quantitative Genetic research has shown the contribution of both genetic and environmental influences to mathematical abilities, with wide variability of the estimates across studies. Heritability of mathematics has been estimated from a low .20 (Thompson, Detterman, & Plomin, 1991) to a high .90 (Alarcón, Knopik, & DeFries, 2000). The same variability has been found for environmental estimates. Shared environmental estimates ranged from zero (Oliver et al., 2004) to low.07 to .16; Kovas, Haworth, Petrill, and Plomin, 2007b), to very high .73 (Thompson et al., 1991). These discrepancies have been attributed to differences in samples, age range, and measures used. Quantitative genetic also allows to investigate the stability and change in mathematical performance. For example, one study estimated heritability of mathematical *fluency* as .47 and .63 respectively in 8.5 and 10 year old twins (Hart, Petrill, Thompson, & Plomin, 2009). The same study found significant heritability (.54) for *applied mathematical problems* only at 10 years, suggesting potential developmental changes in genetic effects. These findings highlight the need of longitudinal designs as well as paying particular attention to the ages of the compared samples.

One large study that has extensive longitudinal data on a wide range of cognitive traits is the Twin Early Developmental Studies (TEDS). Mathematical assessment conducted on this sample at 9 years of age, estimated heritability at .68, with non-significant shared environment (Haworth, Kovas, Petrill, & Plomin, 2007). Moderate genetic influences (between .29 and .46) and moderate non shared environmental influences were found in the TEDS sample at 10 year of age (Kovas, Petrill, & Plomin, 2007c).

Multivariate genetic methods have also been used to investigate the relationship between mathematical and other cognitive abilities and achievement measures. Genetic and environmental correlations have been estimated between mathematics and reading (e.g. Thompson et al., 1991; Knopik & DeFries, 1999);

mathematics and language abilities (Plomin & Kovas, 2005); and between mathematics, reading and *g* (Kovas et al., 2005). This suggests that mathematical skills share some of the same aetiology with reading, language and *g*. These findings support the "Generalist Genes Hypothesis" (Kovas & Plomin, 2006), which suggests that different cognitive domains are, to a large extent, influenced by the same genes. Under this assumption, most genes influencing mathematics also influence reading and/or language. These results are also consistent with behavioural literature that argues that general cognitive abilities are shared across domains.

Although a large number of genetic factors influence both mathematical skills and other general cognitive abilities, it appears that there might be a set of genes that is specific only to mathematics (Kovas et al., 2005). It is possible that this specificity relates to some specific aspects of numerical or mathematical cognition. In support of this hypothesis, a recent study found that mathematical fluency (the ease and accuracy with which maths facts are retrieved and strategies to solve arithmetic problems are adopted), unlike other mathematics-related measures in this study, showed some unique variance (Hart et al., 2009).

4.2.5 Research question

From the evidence discussed so far, it emerges that research has not yet satisfactorily answered the fundamental questions of what mechanisms are driving mathematical acquisition and the aetiology of mathematical abilities. Quantitative genetic research suggests that many genetic factors that affect reading, language abilities, and *g* are involved in the mathematical domain. This is consistent with behavioural findings where a number of general cognitive abilities have been shown to be involved in the acquisition of mathematical skills. Number specific processes of estimation of numerosity and of numerical magnitudes may be the skills associated with the mathematics-specific genetic influences, independent from other abilities reported at the age of 7, 10 and 12.

A question that still needs to be addressed is to what extent genetic and environmental factors contribute to individual differences in specific number abilities. Another question, yet to be addressed, is the extent of the aetiological overlap between specific number abilities, mathematics and other abilities. In order to investigate these issues, the first step is to identify and assess the abilities of interest.

As shown by genetic studies, heritability may change across development, therefore age-homogeneous samples are preferred to control for developmental bias. The genetically sensitive investigations in this thesis are conducted on the twins of the TEDS sample at the age of 16, therefore the tests needed to be appropriate for this age. The aim of the study reported in this chapter was to design and validate a battery of tests assessing mathematical abilities, as well as number specific and general abilities involved in mathematical acquisition as discussed earlier.

This pilot study was devised in two phases of testing and validation on a sample of singleton students, with an additional third phase of validation in a subset of the TEDS sample.

- Phase 1. The first phase involved the identification of measures of specific numerical abilities of estimation, approximation of numerosities, general cognitive abilities, and mathematical performance. This newly developed instrument was administered to a sample of 16 year-old singletons.

- Phase 2. The second phase involved the online implementation of the selected measure and their administration to the same sample of 16 year-old students for test-retest reliability and online validation purposes.

The ultimate aim was to use this new instrument as part of the TEDS assessment at the age of 16.

- Phase 3. The newly developed instrument was administered to the TEDS sample. From the first cohort of assessment, a small subset of twins was retested on the same tests two months after they completed the online assessment at 16.

4.3 Methods

4.3.1 Participants

A sample of one hundred 16 year olds was recruited (as described in section 2.3.3) from secondary schools, at the beginning of the first term, after completion of the GCSE exams. The age of the sample was chosen to match that of TEDS. The GCSE scores in mathematics provided a further assessment of mathematical achievement.

4.3.1.1 Participants in Phase 1

The final sample of the Phase 1 was comprised of ninety-eight 16 year old students (mean age = 16.78, SD = .91). All participants had successfully taken the General Certificate of Secondary Education (GCSE) exam in mathematics. Testing took place at the beginning of Autumn Term after completion of their GCSEs.

4.3.1.2 Participants Phase 2

Out of 98 initial participants, 75 students (mean age = 16.74, SD = 0.93) repeated the tests online for test-retest reliability and online validation purposes during Spring Term (the year after completion of their GCSE).

4.3.1.3 Participants in the final validation on the TEDS sample.

The assessment of the first cohort of TEDS started in October 2010. By January 2011, around 600 families had completed the web testing. 48 twins recruited for the validation sample were selected from these families. The selection criteria and sample description are provided in section 2.2.4.7.

4.3.2 Procedures and material

4.3.2.1 Procedures and materials in Phase 1

The tasks chosen for this first phase of assessment were selected among the tests that from the behavioural literature were considered to tap into the abilities of interest for this investigation. Nine of these tasks, assessing number specific abilities, visuo-spatial working memory, speed of processing and mathematical skills, were administered online, using a laptop. Two tasks (Number Line and Corsi Block) were in pen and paper format. The assessment was carried out, on a one-to-one basis, on school premises with the permission of the school. The whole battery lasted 45-60 minutes. The students were given a £10 voucher for their participation. Detailed procedure is described in section 2.3.2.

4.3.2.2 Procedures and materials in Phase 2

Analysis on the data from the first wave of testing informed the choice of the tests to include in the second wave of assessment. Six out of the eleven tasks that were used in the first wave of testing, and one new mathematical task, were programmed and assembled in a web-based battery. The tests were implemented online at: http://e-businesssystems.co.uk/teds. This testing session lasted approximately 35-40 minutes and was designed to be carried out without supervision. Detailed procedure is also described in section 2.3.2.

4.3.2.3 Procedure in Phase 3

Twenty four twin-pairs were invited (via mail) to the Social Genetic Developmental Psychiatry Centre at King's College, University of London (Centre of the TEDS study) to repeat the whole battery on the Centre premises, using the Centre's equipment. The re-test took place two months after the twins completed the online assessment for the first time (description of the sample and testing procedures are described in section 2.2.4.7).

4.3.3 Measures

4.3.3.1 Measures Phase 1

Students' number specific abilities were assessed with three tests: Dot Task, Dot Matching, and Number Line. As part of general cognitive abilities, speed of processing was assessed with Reaction Time task; and visuo-spatial working memory was assessed with the Corsi Tapping block task. Mathematical skills were assessed with a timed task of arithmetic (Problem Verification Task); as well as a series of tests containing addition, multiplication, and subtraction. The battery also contained two mathematical Stroop tasks.

1. Dot Task (Halberda et al., 2008). This computerised task administered via laptop assessed the specific numerical ability to estimate large numerosity. Stimuli and parameters have been provided by the author of the task, Dr. Justin Halberda. The trials consisted of arrays of yellow and blue dots mixed together flashing on the screen for 400 ms (see Appendix 3 for a sample trial). There were 55 trials with ratios ranging between 0.33 and 0.66; 115 trials had ratios ranging between 0.66 and 0.83 and in 80 trials ratios ranged between 0.83 and 0.89.The trials were administered in random order. The task required to judge which of the two arrays had more dots. Responses were given by pressing the "F" key for more yellow, "J" for more blue. To facilitate response, the keys on the keyboard were colour coded. After response, the next trial was shown by pressing the space-bar. The 250 trials were divided in 5 blocks of 50 trials. At the end of each block it was possible to pause the test and take a break from it. If the students wanted to continue they had to press the space-bar to start a new block of trials. The Weber Fraction measure derived from this assessment was the acuity of the ANS (Approximate Number System) (see Halberda et al., 2008 method section). Cronbach alpha in this test was .89 with split half of .74.

2. Number Line (Opfer & Siegler, 2003). The task assessed estimation of numerical magnitudes and was in pen and paper format. Participants were shown an A4 paper with a line 25 cm long drawn in the middle with the extremes marked 0 and 100 (sample in Appendix - 4). The task required to estimate the position of a series of integers along the line. The 20 numbers to be estimate were presented with same order to all participants as follow: 25, 75, 50, 20, 5, 11, 17, 22, 29, 33, 40, 44, 95, 57, 60, 65, 71, 78, 83, 86. The choice of the range 0-100 was driven to make sure that the task was suitable for 16 year-olds and that students were able to perform the task within this range. Administration mode and stimuli were the same used in the assessment of the US twins of the Western Reserve Reading Project Math. This would have given comparable data for the Number Line in different twin samples. In the instructions given prior to the test trials, students were shown an A4 sheet displaying the line divided in 10 portions (Appendix - 5) and did one practice by estimating the number 50. Scores were the mean absolute difference between the correct location of the number on the line and the point of estimation made by participants. Cronbach alpha in this assessment was .75 with split half .29. The low split half correlation was probably due to the presentation order. The second half of the trials contained more numbers with larger magnitudes compared to the first half. Magnitude comparison in the higher range can be more sensitive to fluctuation in estimation as the numbers are represented "compressed". The presentation order was modified in the final version of the task, with large and small magnitudes randomly and evenly allocated. This improved the split half of the final version to .87 (section 4.3.3.2).

3. Dot Matching task. This computerised task assessed estimation of small and large numerosities. The task was available online at the website of our collaborators in Hong Kong: http://lab.kctam.com/stroop. The 36 stimuli consisted of arrays of dots presented with a number on the side (see Appendix - 6 for sample trial). Numbers and dots ranged between 1 and 9. Half of the trials were congruent (the number matched the dots) and half incongruent. Half of the trials had the numbers of the right-side and the dots on the left-side. The program also varied the spread and layout of the dots across participants. Students had to judge if the number matched the dots in the array. Responses were given by pressing the "J" for a match, "F" if there was no match between dots and numbers. For left-handed participants the keys

were reversed. To avoid confusion with the responding keys, an A4 sheet with a reminder of the response keys was placed next to the screen. After response the next trial followed without delay. Time out for response was 8 seconds, but students were encouraged to answer as quickly as possible. To start the test students had to provide their date of birth and handedness, after which the program generated a unique log-in. There was a set of on screen instruction and 4 practice trials with feedback. The practice could be repeated if needed. One point was awarded for each correct response, incorrect answers received zero points. The first 4 test trials allowed to adjust to the task therefore they did not concur to the calculation of the final score. Maximum score for this task was 32. Reaction time and accuracy on response were recorded. However, due to the slow internet connection and slow download times experienced during the administration of the task, reaction time was not used in the analyses. For the purpose of the analyses, proportion of correct answers computed as the total of correct answers divided by 32, was derived for each participant. Cronbach alpha in this assessment was .93 with split half .82.

4. Reaction Time (Deary et al., 2001). This task was a measure of general cognitive ability - speed of processing. It was programmed in E-Prime 2 following the procedure described in Deary et al. (2001) and was administered via laptop. The trials consisted of a single number presented in the middle of the screen. The computer program generated random sequences of the numbers 1, 2, 3 and 4 repeating them 10 times each (for a total of 40 trials) with a random interval of 1, 2 and 3 seconds; the 10 trials displaying the same number had an equal number of 1 second, 2 seconds and 3 seconds intervals. The task required to press as fast and accurately as possible the key corresponding to the number appearing on the screen. The responding keys were: X = 1, C = 2, B = 3, N = 4. In order to facilitate response the responding keys had stickers with the corresponding numbers. Reaction time and accuracy were recorded. Each correct answer was given one point, incorrect answers scored zero. To correct for speed-accuracy trade off effects in the analysis we used Efficiency (median Response time/proportion of correct response). Cronbach alpha in this assessment was .96 with split half .94.

5. Corsi Tapping Block (Corsi, 1972; Pagulayan et al., 2006). This test assessed non-verbal visuo-spatial short term memory. During phase 1 the test was administered using the Corsi Block apparatus. This consisted of a 25cm X 20cm wooden block with nine 3.1cm cubes placed as in Figure 4.1. The procedure and apparatus for this task are described in Pagulayan et al. (2006).

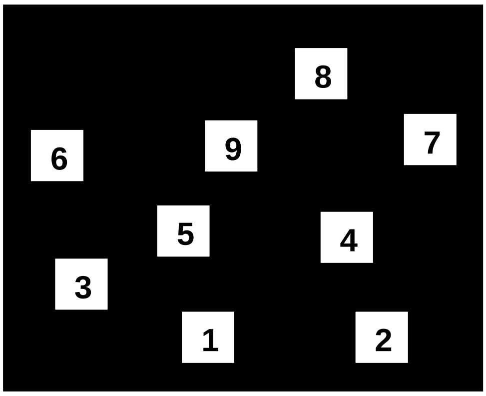


Figure 4.1: Corsi Block apparatus

The numbers are not visible to participants. They can be seen only by the experimenter and are used to determine the pattern of tapping.

Briefly, the experimenter tapped the cubes in predetermined patterns. The patterns were the same for all participants. Students had to reproduce the patterns or sequences by tapping the cubes back in the same order. The list of sequences and administration test rules are shown in Table 4.1. The number of cubes tapped increased up to a maximum of nine or until participants failed to correctly reproduce 5 sequences with the same number of tapping. Each level had a maximum of 5 sequences. One point score was assigned for each correct sequence. Maximum score for this test was 30. Cronbach alpha in this assessment was .75 with split half .40.

Table 4.1
Corsi Block tapping sequence - Phase 1 of testing

Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
4 sequences	5 sequences	6 sequences	7 sequences	8 sequences	9 sequences
3 2 4 9	3 4 8 7 5	5 6 8 1 9 4	5 4 1 6 3 9 7	1 4 7 2 8 3 6 9	3 7 4 1 9 2 5 6 8
1 7 6 5	8 1 5 3 6	4 1 6 7 8 2	7 1 9 3 4 6 2	5 9 3 7 2 1 8 4	7 5 2 8 6 4 1 9 3
3 5 7 1	6 4 5 2 9	2 5 9 6 3 1	9 5 3 6 8 1 7	7 6 9 3 1 5 4 8	9 5 3 7 1 2 6 4 8
6 8 4 9	3 1 7 2 4	7 9 2 5 1 3	2 6 8 1 4 9 5	4 2 1 3 9 5 7 8	6 3 7 4 2 5 1 9 8
(5 2 6 4)	(4 9 3 6 2)	(8 6 1 4 3 9)	(2 9 1 7 4 8 3)	(9 3 1 4 2 6 5 7)	(5 8 6 2 9 1 4 3 7)

The first level starts with a sequence of 4 numbers/items. This is an appropriate level of difficulty for 16 year-olds. Each level is successfully completed if 4 sequences are reproduced correctly. Five points score are awarded for each level successfully completed. The sequences in brackets are administered only if one of the first 4 is reproduced incorrectly. If all the first 4 sequences are successfully reproduced, one point for the sequence is bracket is credited. The test is discontinued after failure to reproduce all the 5 sequences in one level.

6. Problem Verification Task: This test was a timed task assessing mathematical fluency. It was programmed in E-Prime 2 following the procedure described in Murphy & Mazzocco (2008) and it was administered via laptop. The stimuli consisted of 88 arithmetic problems as follow: 14 additions (7 small double-digit and 7 large-three or four digits operands); 14 subtractions (10 small single and double-digit operands, 4 large three or four-digit); 14 multiplication (7 small single digit, 7 large-two and three-digit operands); 14 divisions (7 small one or two-digits, and 7 large two and three-digits); 32 fraction (12 additions and 20 equations). The problems were presented on screen in a fixed sequence together with a proposed answer. The task required to judge whether the answer was correct or not; the proposed answers were correct half of the times. Response was given by pressing the keys "T" for correct, "Y" incorrect, "U", don't know. In addition, participants used the same keys to rate how confident they were about their response: "T" confident, "Y" not confident, "U", don't know. The responding keys were colour coded to facilitate response: T was green, Y was red and U yellow. An A4 sheet with the coloured letters T, Y, U and their meaning for response was placed next to screen as reminder. After response was given the next trial followed without delay. One point was given for correct responses, the incorrect and "don't know" answers received zero point. The ratings on the confidence of response were coded as 1 = confident, 2 = not confident, 3 =

don't know. The task had on screen instructions and two practice trials. The program recorded accuracy and reaction time on correct responses. Accuracy was used for the analyses. Cronbach alpha in this assessment was .91 with split half .84.

7. Arithmetic Problems. This test assessed mathematical fluency. It was composed of 3 separate subtests: 20 single digits addition and multiplication problems, 20 subtractions (10 single digits and 10 with 1 double digit operand). The tasks were available online on the website: http://lab.kctam.com/stroop. The arithmetic problems were presented sequentially on screen together with a proposed answer. Students had to judge whether the answer was correct or not, within 8 seconds time limit. Right handed participants responded by pressing "J" for correct, "F" for incorrect. The keys were reversed for left-handed participants to allow correct responses with the dominant hand. Each sub-test started with its own on-line instructions followed by 4 practice trials. To avoid confusion with the keys to be pressed, a reminder with the responding keys was placed next to the screen. Accuracy and reaction time on response were recorded. The analyses were carried out on accuracy scores.

8. Numerical Stroop tasks - Physical and Numerical Comparison tests were taken from the website: http://lab.kctam.com/stroop. The stimuli consisted of two double digit numbers presented next to each other in the middle of the screen. The font-size of one of the numbers in the pair was always bigger than the other. The task involved identifying a larger number in the pair, according to its size in the physical comparison; or in numerical value (regardless the size) in the numerical comparison. Response was given by pressing "J" if the numbers larger in magnitude or larger in physical size were on the right side of the screen. Participants pressed "F" if the numbers larger in size or in magnitudes were on the left side of the screen. The keys were reversed for left-handed participants. Time out for response was 8 seconds. Both tests started with their own instructions followed by 4 practice trials; participants completed 52 numerical and 52 physical comparisons. Accuracy and reaction time on response were recorded.

4.3.3.2 Measures in Phase 2

The analysis of the data collected from Phase 1 informed the choice of tests to be used in Phase 2. The online tests of addition, subtraction, multiplication and the two Stroop tasks revealed a ceiling effect in participants' response. The small variation in the distribution of the above measures did not add new information to the analysis. It was decided to limit the duration of the whole battery to 30 minutes in order to maximise participation in the main TEDS data collection. With these time constraints, it was decided to keep only Problem Verification Task as a test of mathematical fluency, as it showed the best distribution and validity. For this reason, the 3 tests of arithmetic problems were dropped from the battery, together with the two Stroop tasks.

The six remaining tasks and a new test were programmed and assembled in a web-based battery found online at: http://e-businesssystems.co.uk/teds. This website was used only for Phase 2 of the pilot testing. In order to be programmed online, the six tests used in Phase 1 were modified from the description given in section 4.3.3. Therefore, the following description pertains only to the changes made to the six tests from Phase 1 and to the new test 'Understanding Numbers'.

1. Understanding Numbers. This test was introduced in the on-line battery as further assessment of mathematical skills according to the standards of the UK National Curriculum. Test items were word problems selected from the nferNelson booklets (level 1 to 8) (NFER-Nelson, 1994, 1999, 2001), from the mathematical component "understanding numbers". The solution of the problems required understanding of the relationship between numerical expressions and patterns of numbers, understanding of mathematical operations, as well as of relationships among mathematical operations (e.g., subtraction is the inverse of addition). One example of a trial: "Work out the value of x: 6x + 9 = 8x. Click on your answer" - 5 options were given as possible responses. A graphic sample of a test trial is provided in Appendix 7. The test comprised 27 items arranged in increasing level of difficulty. Difficulty was assessed by the percentage correct on the National Curriculum standardisation sample (reported in the Group Record Sheets; NFER-Nelson). 20 items were already

been used in the web assessment of TEDS at age 12 and for these the level of difficulty was assessed taking also into account the percentage of accuracy in response received during the previous TEDS web testing. The 27 questions were organized in 3 levels of 9 items each. Each level was further divided into 3 sub-levels of items with increasing difficulty. All participants started with the same question of medium difficulty. The subsequent presentation order was determined by participants' answers: answering correctly to the problems of one level advanced the test progressively to the more difficult questions; and items from the easier levels were credited as correct. If the problems within a level were answered incorrectly the test branched down to easier levels. The task stopped when three questions in a row were answered incorrectly. The test started with a set of instructions and there was no practice trial. For some problems, the answers needed to be typed in; others had multiple choice answers and response required to click on the correct answers. For some problems, a simple calculator appeared on the screen alongside the question. After response was given, participants submitted their answers by clicking on the provided "ok" button. A new screen presented the option to progress to the next question or take a break and resume the test later. Maximum response time was 5 minutes and prompts encouraged participants to answer during this time. If no answer was given during the 5 minutes, participants were given the choice to go to the next question or take a break from the test. One point was awarded for each correct/credited answer; no points were given for timed out or incorrect answers, therefore maximum score on this test was 27. The program recorded accuracy and time of response computed from the onset of the problem on screen to the submission of the answer. Cronbach alpha in this assessment was .82 with split half .61.

2. Dot Task. The length was shortened from 250 to 150 trials. The reduction was carried out by the author of the task on the basis of analyses conducted on the data from Phase 1. The 150 stimuli were chosen from the original 250, their ratios maximised performance of this shorter version. The length of the original test (250 trials) was split in 3 and each third was correlated with the length of the whole test. The subset of stimuli chosen yielded a correlation of .91 with the full length test. It was also decided to administer the test in a fixed sequence to reduce eventual fatigue

and disadvantages deriving from random sequences where more difficult trials were presented towards the end of the test. The number of dots ranged between 5 and 21 for each colour with ratios organised in 8 bins with the lowest ratio of each bin serving as the top boundary of the following bin. The bins' ratios were organised as follow: 11 trials with a ratio randomly chosen between 8/7 and 7/6; 26 trials between 7/6 and 6/5; 28 trials between 6/5 and 5/4; 29 trials between 5/4 and 4/3; 26 trials between 4/3 and 3/2; 18 trials between 3/2 and 2; 8 trials between 2 and 3; 4 trials between 3 and 4. In all trials the average size of yellow dots was equal to the average size of blue dots. With this display, the array with more dots also occupied more area on the screen. Such change in the design of the task may have induced responses to the stimuli on the basis of the area rather than numerosity. However, studies have shown that, if required, adults can suppress response on the basis of continuous properties of a stimulus (area) and respond to numerosity (Nys & Content, 2012). The decision to control for size only and not for area, was taken on the basis of previous unpublished studies conducted by the author of the Dot Task. These studies showed no changes in individuals' Weber Fractions if calculated on trials controlled for area only, size only or both provided that the task was able to measure fine grain individual variation (personal communication from the author of the task). A greater sensitivity of the task could have been reached with a greater number of ratios. After the reduction of the number of trials, controlling only for one parameter gave the possibility to increase the number of ratios (the previous version of the task used three ratio-bins). The exposure time remained the same (400 ms) but the responding keys were changed to "Y" for more yellow and "B" for more blue dots. Maximum time allowed for response was 8 seconds. If no answer was given during this time, the answer was recorded as wrong and a message appeared on the screen to encourage pressing the space bar to see the next trial. The message disappeared after 20 seconds and the next trial was displayed only after a press of the space-bar. There was a two item practice trial, with feedback and an option to repeat the practice. At the end of the practice trial it was made clear that the task measured speed as well as accuracy and invited participants to respond as quickly as possible. The task was divided in three blocks of 50 trials. At the end of each block it was possible to take a break and resume the test later. The test recorded accuracy and reaction time for each trial. A Weber Fraction score for each individual was derived using the method described in the supplementary information of Halberda et al. (2008). In addition, reaction time on response was used for a further correction. The Weber Fraction for each participant was derived only on trials not considered outliers according to the Jolicoeur method (Van Selts & Jolicoeur, 1994). On average, 3.9 trials were removed from each performance, with a minimum of 0 and a maximum of 10. Cronbach alpha in this assessment was .88 with split half .80.

3. Number Line. The test was an online adaptation of the pen and paper version administered during the first wave of testing. As documented in many studies (e.g., Siegler & Booth, 2004; Booth & Siegler, 2006), at 16 years most children rely on linear representations of numerical magnitudes on a mental number line from 0 to 1,000. Consistent with this finding, the analysis of the pen and paper data from Phase 1 of testing revealed a linear trend in the pattern of estimation. Moreover the data collected using estimations of numerical values from 0 to 100 did not add any predictive value to the analysis. For this reason, it was decided to change the estimations range from 0 to 1,000 for the on-line battery. In this new version, a line, with the left edge marked with “0” and the right edge marked with “1000” was presented in the middle of the screen with a numeral above the line. The task required participants to indicate where they thought the numeral should be, by dragging and releasing a red cursor along the line. The twenty-two numbers to be estimated were taken from Opfer and Sigler (2007) and were arranged in random order with low and high magnitudes evenly spread. They were presented in the same order to all participants as follow: 246, 179, 818, 78, 722, 150, 366, 122, 738, 5, 147, 938, 18, 606, 2, 34, 754, 100, 56, 163, 486, and 725. The choice of a fixed order was driven to minimise advantages - favouring accuracy - that may have been produced by easier random presentations (where numbers in the same range were presented in close trials). Total length of the line was 500 pixels with each unit 0.5 pixels long, therefore accuracy in response recorded to the nearest 0.5 units. The marks on the line were converted into numbers based on number of units (pixels); the scores were calculated as the mean of the absolute deviations from the correct position of the numbers on the line. This test allowed only one practice trial to reduce the effects of training/learning as this has been shown to positively affect estimation accuracy. At

each screen participants were given the option to continue with the task or to resume it later. The program recorded the scores as described as well as response reaction time. Cronbach alpha in this assessment was .87 with split half .87.

4. Dot-Number Matching. The changes implemented in this task involved the presentation and exposure of the stimuli. The stimuli consisted in static pictures resembling the same arrays presented in the first wave of testing. Differently from the first wave of testing, the spread and arrangement of the dots was also the same for all participants. To prevent participants from counting the dots, especially in the estimation of large numerosity, exposure of the stimuli was reduced to 2 seconds; the instruction however, invited participants to respond as quickly and accurately as possible. There was no option to pause the test. Cronbach alpha in this assessment was .85 with split half .78.

5. Reaction Time. The only change implemented in the task was in the presentation and the change of the responding key to adapt the test for online testing. The numbers 1, 2, 3 and 4 appeared 10 times each in the same randomised order for all participants. The interval between numbers was also maintained the same: the interval of 1 second between presentations was repeated 14 times and the interval of 2 and 3 seconds was repeated 13 times each. Instructions reminded participants to respond as quickly and accurately as possible. The responding keys were changed from letters to the numbers on the keyboard due to technical constraints. Cronbach alpha in this assessment was .94 with split half .91.

6. Corsi Tapping. The test was an online adaptation of the pen and paper version administered during the first wave of testing. The number of levels remained the same. The total number of trials was reduced from 30 to 18 based on an internal validity analysis conducted on data from Phase 1. The list of sequence and the administration rules are shown in Table 4.2.

Table 4.2
Corsi Block tapping sequence - Phase 2 of testing

Level 4	Level 5	Level 6	Level 7	Level 8	Level 9
4 sequences	5 sequences	6 sequences	7 sequences	8 sequences	9 sequences
1 7 6 5	3 4 8 7 5	5 6 8 1 9 4	5 4 1 6 3 9 7	1 4 7 2 8 3 6 9	3 7 4 1 9 2 5 6 8
3 5 7 1	8 1 5 3 6	4 1 6 7 8 2	7 1 9 3 4 6 2	5 9 3 7 2 1 8 4	7 5 2 8 6 4 1 9 3
5 2 6 4	6 4 5 2 9	7 9 2 5 1 3	2 6 8 1 4 9 5	7 6 9 3 1 5 4 8	5 8 6 2 9 1 4 3 7

The number of items per level has been reduced based on reliability analysis. The difference in loss of reliability items deleted was very small for all items. We retained the items that generated the least of correct responses, although for some item these differences were of one or two responses. Each level is successfully completed to the correct sequence of 3 trials. The test is discontinued after failure to reproduce all the 3 items in one level.

The trials of the online version showed on the screen an image, depicting a black rectangle similar to the original Corsi apparatus, with 9 small cubes-blocks arranged inside as shown in Figure 4.2.

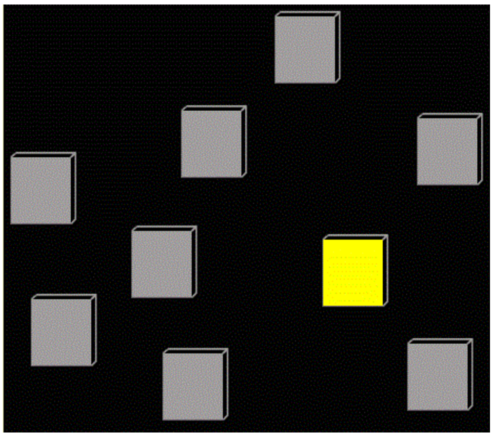


Figure 4.2: Online image Corsi Block display

The cubes lit up turning yellow for 1 sec in a patterned sequence, with a 1 sec interval between each cube. Participants had to reproduce the pattern by clicking on the cubes with a mouse. Each cube had a number from 1 to 9 associated with it, so that each sequence could be identified with the numerical string shown in Table 4.2

(the numbers were not showing on the screen). 6 difficulty levels were administered, with three sequences within each level. In hardest level, participants had to reproduce patterns of 9 numbers. Immediately after the last item of the sequence was presented, a text prompt appeared on the screen inviting participants to start reproducing the sequence; they responded by clicking on the blocks in turn, using a mouse. As participants clicked on the block, this turned yellow and remained yellow until the next block was clicked. Clicks in the black areas between blocks were not registered. Responses were irreversible. After each response participants were presented with a screen with two buttons to choose either to continue with the test, or to come back to it later. The test had audio-visual instructions and 3 item practice trial which could be repeated until the participant was familiar with the task. If students correctly completed at least one sequences in a level, they progressed to the first item of the next level. The test was terminated when all the sequences in the same level were reproduced incorrectly. One point was assigned for each sequence correctly reproduced, with maximum score of 18. There was no time limit for response. The program recorded accuracy and reaction time for each trial. Cronbach alpha in this assessment was .75 with split half .63

7. Problem Verification Task. The off-line version of the task was implemented online with the only change of dropping the confidence rating of response. Besides to time constraints, this decision was driven by the fact that this rating did not add useful information to the analysis: the ratings did not produce any variability as ~90% of the students answered 1 = confident. At the bottom of the screen the letters T, Y, U, and their meaning were displayed as reminder for response. Cronbach alpha in this assessment was .92 with split half .87.

4.3.3.3 Final version of the Battery

The tests described in the above section, were administered to the same sample of students for on-line validation purposes. Following this new wave of assessment some tests were modified and optimised for web testing. For some of the test the total duration needed to be reduced. The description that follows is limited to the modification implemented.

1. Understanding Numbers. The total number of items was reduced from 27 to 18. An internal reliability analysis conducted on the data of Phase 2 of testing, revealed that the differences in reliability from dropping any of the items was minimal. It was decided to drop the first item of each level. Each level of difficulty comprised 2 items and the test branched up or down after two questions. The discontinue rules accommodated this change: the test was discontinued after two consecutive items answered incorrectly.

2. Corsi Block. The only change implemented in this test was the reduction of the total number of items from 18 to 12. The internal validity analysis on the data of the Phase 2 of testing showed almost no change in reliability by dropping any of the items. For this reason it was decided to remove the first sequence of each level. As a consequence of the reduced numbers of items on each level, the discontinue rule applied after two incorrect answers at the same level.

3. Problem Verification Task. The total number of items was reduced from 88 to 48. The final version of this task consisted of 24 fraction problems and 6 of each: multiplication, division, subtraction and addition problems. The selection of the items was based on correlation and reliability analysis conducted on the data of Phase 2 of testing. The test composed of the final 48 items correlated .97 with the test composed of 88. Cronbach alpha on the 48 items was .86 with split half of .79. In order to unify the responding keys across the battery, responding keys were modified to "F", "J" or "K" respectively for “correct”, “incorrect”, and “don’t know”. The reminder of which keys to press was shown at the bottom of the screen. A sample of a test trial is provided in Appendix 8. Instructions reminded participants to respond as quickly and accurately as possible. Maximum time for response was changed to 10 sec and a time bar added on the top-left corner of the screen reminded participants of the elapsing time. If no answer was given during this time the next trial followed. The next item was presented immediately following a response. After the 24th trial participants were presented with a screen with two buttons that gave the option either to continue with the test or to take a break.

4.4 Results and discussion of Phases 1 and 2

The main aim of this study was to create a valid battery with the power and sensitivity to predict mathematical performance and achievement from measures of number sense abilities in the presence of other general cognitive abilities. The analysis presented is not aimed at addressing a specific hypothesis, but rather to evaluate the predictive validity and reliability of the measures employed in the newly developed battery.

It was important that in the process of implementation on the web and length reduction, the measures would not lose any of their reliability. The internal validity of the measures was accurately monitored throughout the process; overall the measures were accurately translated from different formats into a bespoke online tool of assessment. The second point of attention was that the measures employed were consistent with existing literature. In other words, we needed to make sure that the measures tapped into the cognitive domains of interest for future analyses.

4.4.1 Identification of outliers and transformation of the variables

Comparison of the students' performance between the online and one-to-one administration of the computerised tests was crucial in understanding whether the variation in performance was due to the change of administration modality, or due to other confounding variables. For the 75 students who completed both waves of testing, new variables were derived computing the difference between Phase 1 and Phase 2 testing scores as follow: for Reaction Time task this variable was created on both Accuracy and Efficiency scores; for all the remaining variables the variable was created only on the accuracy scores. As Corsi Block and Number Line were in pen and paper format during Phase 1 of assessment, the comparison of the tests scores between Phase 1 and Phase 2 could have been affected by the change of format. For this reason no variable was created for these two tasks. For Number Line and Corsi Block, the outlier scores were identified if laying outside +/- 3 standard deviations. The new computed variables were plotted and observations falling outside 3 standard deviations were used to identify participants with inconsistent performance between the two waves of testing. Outlier cases were identified in the frequency tables for each variable and removed on both on-line and one-to-one tests, for that variable. A total of 25 cases (on different variables) were excluded from the analyses. Two more cases were excluded from Number Line and Dot Task analysis as their z-scores were above 3 standard deviations

The analyses were conducted on standardised scores with the exception of Weber Fraction and Reaction Time which is reported as Reaction Time Efficiency (see section 4.3.3.1). The GCSE were provided in grades ranging from A* to E. These grades were scored from 6 to 1 assigning 6 to the highest score (A*) to 1 for the lowest (E). The GCSE were therefore transformed in scores as follow; 6 = A*; 5 = A; 4 = B, 3 = C; 2 = D; 1 = E. These scores were standardised to carry out the analyses.

4.4.2 Results summary from Phase1

Table 4.3 shows the correlation between measures of general cognitive abilities and specific number abilities with mathematical scores. For the purposes of the analysis, Fraction problems were analysed separately from the total score in the task, as the solution of fractions requires conceptual knowledge (Hecht, 1998), as opposed to the procedural knowledge required for the rest of problems in the task. Mathematical performance indexed by the GCSE significantly correlated with the general cognitive measures: positively with Corsi and negatively with Reaction time efficiency (lower efficiency scores index better performance, hence the negative correlation with GCSEs). However, mathematical performance did not correlate with the two number sense measures. Weber Fraction negatively correlated with mathematical fluency, assessed by the Problem Verification Task (smaller Weber Fraction scores indicate more accurate performance on the Dot Task). Number Line significantly correlated only with Weber Fraction, and not with any of the mathematical scores. This correlation was positive as for both measures smaller scores indicate better performance.

Table 4.3

Descriptive statistics and correlations among measures of general cognitive ability, specific numerical ability, and mathematical performance – Phase 1 of testing.

	Measure	Mean (SD)	1	2	3	4	5	6	7	8
1	GCSE scores	4.5 (.82)	--							
2	Fraction	.62 (.20)	.62**	--						
3	Problem Verif.	.73 (.15)	.55**	.91**	--					
4	Weber Fraction	.25 (.09)	-.17	-.34**	-.33**	--				
5	Dot Matching	.90 (.07)	.14	.18	.22*	-.17	--			
6	Number Line	3.37 (1.23)	-.21	-.05	-.04	.23*	-.02	--		
7	Corsi	.67 (.08)	.30**	.14	.12	-.16	.08	-.10	--	
8	Reaction Time Eff.	596.82 (84.30)	-.27**	-.24*	-.21*	.09	-.13	.11	-.14	--

*p<.05; **p<.01 (2-tailed test)

SD = Standard deviation; Reaction Time Eff. = Reaction Time Efficiency (computed as median reaction time of correct answers/proportion of correct answers); Problem Verif. = Problem Verification Task total score. Means and standard deviations for Fraction, Problem Verification, Corsi and Dot-Matching are reported for unstandardised proportion of correct answer. For GCSEs, Weber Fraction, Number Line and Reaction time efficiency means and standard deviations are reported on the unstandardised accuracy scores.

4.4.3 Test re-tests Phase 1 and 2 of testing

A reduction in test re-test correlation between Phase 1 and 2 of testing was expected since all the tasks' format, administration parameters and length changed. during the two phases However, despite all the modifications, the measures showed significant correlations between the two waves of testing. The tasks with the least changes in administration parameters showed the highest correlations. A summary of these correlations is shown in Table 4.4.

Table 4.4
Test- retest coefficients between Phase 1 and 2 of testing

Measures	r test re-test	N
Fraction	.71**	63
Problem Verificat. (tot.)	.60**	66
Weber Fraction	.43**	67
Dot Matching	.12	67
Number Line	.30*	64
Corsi	.57**	61
Reaction Time Efficiency	.41**	67

***p <.000; *p <.05*
r = correlation coefficient; N = number of cases; Problem Verificat. (tot.) = Total score for the Problem verification task; Fraction = Fraction problems from the Problem verification task. The correlation between the Phase 1 and 2 of testing of the Dot Matching task is non-significant.

Only the Dot Matching task had a non-significant re-test coefficient. It has to be noted that during the first phase of the pilot, when the Dot Matching task was downloaded from the website in Hong Kong, students experienced problems with the stimuli-images. This was probably due to internet bandwidth issues and the image size of the stimuli. Other tasks downloaded from the same website were not affected as the stimuli were only text based. This technical problem was completely resolved in the second wave of assessment.

4.4.4 Results summary from Phase 2 of testing

Means and standard deviations of the measures used in Phase 2 of testing are shown in Table 4.5. It can be noted that estimations on the Number Line in the range 0-1000 showed an average mean greater than in the range 0-100 (43.85 vs 3.35), the standard deviation was however smaller in Phase 2 compared to Phase 1 (the SD = 1.23 of Phase 1 is 1/3 of the average mean, while the SD = 17.30 of Phase 2 is less that 1/3 of the average mean). This shows that by small amount, the error in estimation was less spread when students did the test online. Perhaps this indicates that the online task generated a more consistent performance-less variability of the estimation error. However, the two versions of the Number Line differed in both administration mode and range of the magnitudes of estimation, so inferences on the less variability of the errors cannot be made in this occasion. This point is also supported by the modest test re-test correlation (.30). On the other hand, consistent with existent literature, this version of the Number Line task showed a significant correlation with mathematical school achievement measured by the GCSEs (-.35).

Table 4.5

Descriptive statistics and correlation among measures of general cognitive ability, specific numerical ability and mathematical performance – Phase 2 of testing

	Measures	Means (SD)	1	2	3	4	5	6	7	8	9
1	GCSE scores	4.6 (.80)	-								
2	Underst. Numbers	.77(.09)	.48**	-							
3	Fractions	.58 (.21)	.46**	.32**	-						
4	Problem Verif.	.69 (.16)	.44**	.29**	.91**	-					
5	Weber Fraction	.36 (.22)	-.04	-.16	-.16	-.31*	-				
6	Dot Matching	.77(.09)	-.04	-.10	.11	.11	-.22*	-			
7	Number line	43.85 (17.30)	-.35**	-.04	-.26	-.21	.12	-.21	-		
8	Corsi	.50 (.14)	.26	.15	23*	.28*	-.17	.08	-.21	-	
9	React. Time Eff.	645.93 (121.54)	-.37**	-.22	-.26*	-.26*	.30*	-.04	.24	.03	-

***p<.01; *p<.05 (2-tailed test)*

SD = Standard deviation; Underst. Numbers = Understanding Numbers total score; Problem Verif. = Problem Verification Task total score; React. Time Eff. = Reaction Time Efficiency (calculated as median reaction time of correct responses/proportion of correct responses). Means and standard deviations for Understanding Numbers, Fraction, Problem Verification, Corsi and Dot-Matching are reported for unstandardised proportion of correct answer. For GCSEs, Weber Fraction, Number Line and Reaction time efficiency means and standard deviations are reported on the unstandardised accuracy scores. The GSCE scores mean is calculated assigning values from 1=E to 6=A*.

The mean of the Weber Fraction scores also showed some difference between the two phases of testing. It has to be noted that only in the second phase, reaction time was used to correct outlier-trials according to the Van Selts & Jolicoeur (1994) method. Similarly to the Phase 1, in this second phase Weber Fraction showed a significant correlation with mathematical fluency (with Problem Verification, r= -.31). It also showed significant correlation with Reaction Time efficiency (.30) and Dot Matching (-.22). It is not excluded that these correlations may have been partially driven by the perceptual and timed nature of the tasks.

The GCSE scores showed significant correlations with Reaction Time efficiency (-.37) but not with the other measure of general cognitive ability Corsi. As expected GCSE scores correlated with all the mathematical tests: Understanding Numbers (.48), Problem Verification (.44) and the Fractions (.46).

Besides GCSEs and mathematical fluency, the new test, Understanding Numbers, showed no significant correlations with any of the measures. It is possible that the small sample used in the pilot was not suited for this kind of analysis. It is of value to note that this test was adapted from a previous version used in TEDS where it successfully assessed mathematical skills.

Although the mean and standard deviation of Reaction Time efficiency was higher than in Phase 1, the measure still correlated with mathematical achievement (-.37 as mentioned earlier) and fluency (r= -.26 for both Problem Verification and Fraction). Corsi showed significant correlation with mathematical fluency: .23 with Problem verification and .28 with Fractions.

Overall, this pilot study was successful in selecting and validating measures for a new bespoke tool of assessment of mathematics, general abilities and number sense skills. The measures implemented online showed reliability and the ability to capture different aspects of the mathematical construct.

4.5 Methods and results of Phase 3

The pilot study described above dealt with the selection and implementation on line of the tests in the mathematical battery. The tests showed consistent internal validity throughout this process. However, the tests already piloted needed to go through further changes, mainly aimed to reduce the time of the whole battery. These changes took place prior to implementation in the final battery just before being administered to the twins. The first cohort of TEDS started the online assessment with the modified tests as described in section 4.3.2.2.

Although the modification of the tests was guided by the pilot data, internal validity needed to be re-examined to assess the effects of these changes to the tests. Test re-test reliability analysis was conducted during the pilot study, however, the modification of the tests between these two phases could have compromised the correlation, for this reason a second reliability analysis was conducted with this validation study.

4.5.1 Data collection and analyses

Data for the complete TEDS' battery at 16, which also included tests of language, reading, and general intelligence, were collected from the sample of 48 twins, selected as described in section 4.3.1.3.

Reliability: Intraclass correlations were used to assess the relationship between the twins' performance (24 pairs) on the first web assessment and performance on the same tests after two months. As the re-test was administered online, this analysis allowed investigating stability of the online measurements over time. It was also possible to assess if the result of the tests were affected or not by changes in testing environment.

Internal validity: Cronbach alpha coefficients were calculated for the items within each test. For the purposes of these analyses only one randomly selected twin in each pair was used. The analyses were conducted on all the data collected from the testing of the first cohort, over 2,000 twin-pairs.

4.5.2 Results and discussion

4.5.2.1 Test re-tests reliability

The mathematical battery was created as a bespoke tool of assessment. We therefore have no reliability data from previous studies to which to compare the test re-test reliability of our measures. However, the test re-retests correlation of the battery ranged between .44 and .78, as shown in Table 4.6. The magnitude of these correlations is within the range of reliability of many cognitive tests

The major changes occurring in the process of online implementation of Problem Verification involved the reduction of the items. This task was the least affected by changes in format and procedures of administration. Its reliability correlation during the pilot was higher compared to other tasks (.60). The re-test correlation in the validation study was the highest (.78) and reliable as shown by the narrow confidence intervals (Table 4.6). Overall, all other tests showed an improvement in test re-test correlation from the pilot. The only test for which it was possible to compare reliability was Understanding Numbers. As mentioned in section 3.3.5, this test was modified from previous versions of the online TEDS assessments, to make it age appropriate to 16 year-olds. In previous assessments (at age 12), its test re-rest reliability was .92. This validation study reported a strong, although lower reliability correlation of .67.

Table 4.6

Test re-test reliability correlations with 95% Confidence Intervals. Cronbach alpha is calculated on the data from the whole web-assessment at 16 years.

Measures	***Test re-test from validation study***			***Cronbach alpha on all TEDS data from the web-testing***	
	r	**(95% C.I.)**	**N**	**α**	**N**
Corsi Visuo-Spatial Memory	.64	(.01 - .90)	44	.67	742
Number Line	.50	(.25 - .69)	48	.72	822
Problem Verification	.78	(.64 - .87)	48	.86	776
Reaction Time (on accuracy)	.68	(.48 - .80)	48	--	--
Reaction Time (on time of response)	.58	(.35 - .73)	48	.96	728
Weber Fraction (on accuracy)	.62	(.40 - .76)	48	.79	824
Dot Matching	.44	(.47 - .80)	46	--	--
Number Game	.67	(.18 - .65)	48	.91	713
Non Verbal -Raven	.72	(.53 - .82)	48	.79	756
Verbal - Vocabulary	.65	(.44 - .78)	48	.81	891
Language	.74	(.54 - .83)	48	.69	818
Reading Fluency	.81	(.69 - .89)	48	.96	836
Reading Comprehension	.67	(.43 - .78)	48	.72	645

The 24 twin-pairs repeated the test for the second time two months after the first web-assessment. Cronbach alpha is calculated on the data from the first wave of web-assessment at age 16. Test re-test on Reaction Time test is calculated on accuracy of response and on reaction time of response. Cronbach alpha on Reaction Time test is calculated only on time of response because of the speed-accuracy trade off effects on this task. Weber Fraction validation is conducted on the scores of the Dot Task.

Dot Matching was the only test with a moderate reliability coefficient (.44). Although the test re-test correlation improved from the pilot, it is possible that this test is particularly affected by extreme scores as shown by the plots of the distributions on the test and re-test scores (Figure 4.3). Larger samples may provide more accurate estimates for this measure. The other non-mathematical tests showed good test re-test reliability with correlations ranging between .65 and .81.

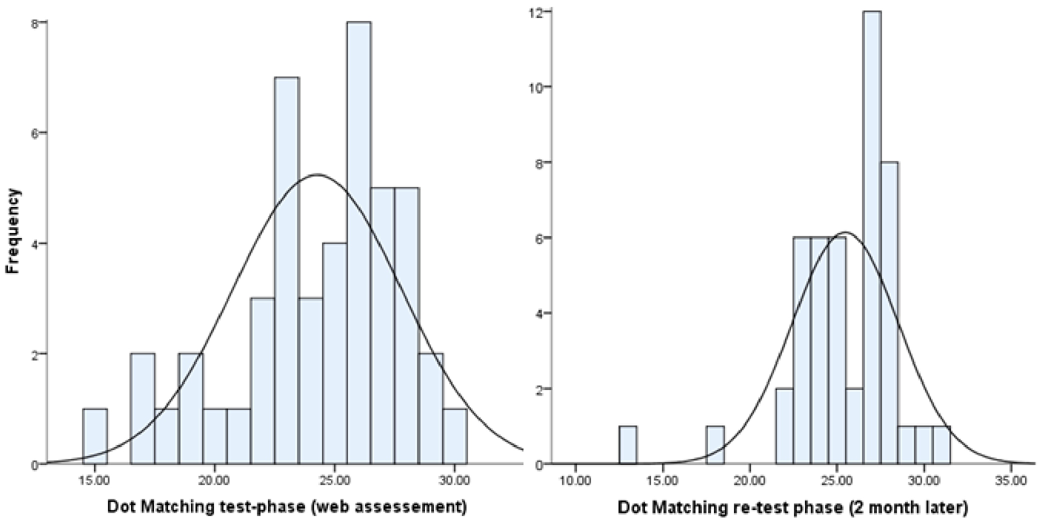


Figure 4.3: Plots distribution of Dot Matching during the tests phase (1st graph) and re-test (2nd graph). The plot of the re-test shows the same trend of the test phase, this plot shows also presence of extreme scores.

4.5.2.2 Internal validity

The internal consistency of the measures was examined with the Cronbach alpha coefficients, shown in Table 4.6. The average alpha coefficient for the mathematical battery was high, .78, ranging between .54 and .96. The average Cronbach alpha for the non-mathematical tests was .79. The test Understanding Numbers showed an internal validity of .91. This result was consistent with the internal validity of the previous assessment at age 12 for which the same test yielded a value of .92.

4.5.2.3 Conclusion

The results of this validation study showed good reliability and validity of the measures in the mathematical battery and in the whole TEDS online battery. Overall, the measures had high Cronbach alpha values before being reduced in length. It is acknowledged that a high number of items may inflate the value of alpha (Cortina, 1993), however, after reducing the number of items on all the test, we did not observe a drastic decrease in alpha coefficients. This suggests that the modifications performed on the tests have not altered the original internal consistency. Also, the strong tests re-test reliability correlations suggest stability of the data obtained with these instruments.

However, Cronbach alpha may not be the best parameter to assess validity. In fact, internal consistency can be inflated in tests with branching or discontinue rules (Understanding Numbers and Corsi Block) because participants may answer only a few questions before being discontinued, for example. Another consideration concerns the change of the environment in the re-test phase. Although the twins were re-tested on the same battery, online, using a computer, we do not know if using a different computer or being in an unfamiliar place has affected performance. Also, the whole battery is quite long and breaks are allowed to avoid fatigue; it is possible that some children may have felt pressure to finish the battery quickly and not have taken breaks or rushed through the tests. This may explain some of the extreme scores of some test (Dot Matching).

Overall, it can be concluded that the mathematical battery was successfully implemented online. This instrument can be considered a reliable tool to gather data via Internet.

Chapter 5: Mathematics and Number Sense (or Number Senses?) across the school years: a multivariate longitudinal investigation in 7 to 16 year old school children*

5.1 Abstract

Variation in number sense (awareness about numerosity, numbers and their relationships) correlates with mathematics. We used Weber Fraction scores (from a non-symbolic numerosity discrimination task) and Symbolic magnitudes estimation scores (from the Number Line task) to assess number sense in a sample of 2,382 16-year-olds, who were also assessed on measures of mathematics, verbal and non-verbal ability at ages 7, 9, 10, 12, 14, and 16. Although Weber Fraction and Number Line scores correlated 0.22 with each other, they correlated 0.33 and 0.47, respectively, with mathematical performance. We also found significant correlations between both number sense measures at age 16 and mathematics as early as age 7. However, after controlling for the contemporaneous mathematical performance (age 16), only mathematics at age 7 predicted Weber Fraction; and none of the earlier measures of mathematics predicted Number Line scores. After controlling for mathematical ability, reading and non-verbal ability at earlier ages were not significant predictors of number sense at 16. Our results suggest that the two number sense abilities may be largely supported by distinct processes and have a different relationship with mathematics across development. For example, better ability to approximate numerosity (measured by a Dot task) may bootstrap early mathematical development by facilitating the association between numerosity and its symbolic meaning. This ability may be bootstrapped by mathematical training later in development. No sex differences were found in the means or variances in number sense, suggesting that any observed mean differences in mathematical achievement do not relate to number sense.

---

* This Chapter has been adapted from Tosto et al. (in preparation-a)

5.2 Introduction

In day to day situations, approximate solutions to problems are often satisfactory. For example, determining whether the train is more crowded in the morning than in the afternoon, does not require counting all the people in; a judgment in terms of more or less suffices. Similarly, we may spot an error in our bill if the total is different from the ball-park figure we had in mind. These are two examples of estimation. The first case is a process of non-symbolic estimation as it is carried out on discrete items; this process does not require any knowledge of numbers and their symbols. The second process is a symbolic estimation as involves numerical comparison and requires numerical knowledge.

Recent literature on mathematical development suggests that non-symbolic quantity estimation and discrimination abilities may provide the basis for the development of mathematical ability. These "number sense" abilities (Dehaene, 1997) may have evolutionary origins. For example, basic numerical perceptions and discriminations have been reported for various species of animals (e.g.: Stevens, Wood, Hauser, 2007; Wood, Hauser, Glynn, Barner, 2008; Al Aïn, Giret, Grand, Kreutzer, Bovet, 2009; Agrillo, Piffer, Bisazza, 2011; Reznikova & Ryab, 2011) and for human infants (Feigenson, Carey, Spelke, 2002; Xu, 2003; Xu &Spelke, 2005; Jordan, Suanda, Brannon, 2008; Libertus & Brannon, 2010). Despite this evolutionarily preserved ability to perceive quantity, a large amount of individual variation in judgment of numerosity (i.e. number of items grouped together) seems to exist in humans. For example, some adult individuals are able to discriminate without counting between displays of discrete visual or auditory stimuli, even when the ratio between the smaller and the larger numerosity is very small (9:10) (Halberda & Feigenson, 2008). Estimation skills are shown to progressively improve with development, leading to the ability to discriminate larger numerosities and smaller ratios (Xu, 2003; Xu & Arriaga, 2007, Halberda & Feigenson, 2008).

The same ratio dependency that is observed in comparisons of non-symbolic numerosities, is present in comparisons of numerals (which requires mapping between symbolic number representations and their meaning in terms of numerosity). This ability is also regarded as another measure of number sense (e.g. Berch, 2005). For example, adults and children are faster and more accurate in judging the difference between two numerical magnitudes when the numerical distance between the numerals is larger (1 vs 9) than when it is smaller (6 vs 8) (e.g., Moyer & Landauer, 1967; Dehaene, Dupoux & Meheler, 1990). This distance effect has been taken as indirect evidence that the symbolic representation of numbers builds on the approximate representation of non-symbolic numersosity (Feigenson, Dehaene, & Spelke, 2004) and that numbers are mentally represented along a mental "number line" (Siegler & Opfer, 2003). It is hypothesized that numbers on the mental number line are represented logarithmically compressed, so that the mental distance between numbers at the low end of the line are over-estimated compared to that at the high end (for example, the representation of the distance between 1 and 20 is greater than the mental distance between 130 and 150). Logarithmic representations are therefore less accurate than linear ones. Over the course of development a gradual shift occurs from a less accurate logarithmic mental number representation to the more precise linear representation. Although both systems seem to operate at the same time, from the age of 6 to 8 years a linear representation becomes dominant as evidenced from improved performance on estimation on the number line task (Siegler & Booth, 2004).

There is evidence that individual differences in estimation of symbolic numerical magnitudes on a number line positively correlate with mathematical achievement (Siegler and Opfer, 2003; Siegler & Booth, 2004; Siegler & Mu, 2008). This relationship may be at least partially mediated by visuo-spatial working memory (Geary et al., 2008). Similarly, several recent studies suggested that individual differences in non-symbolic estimation, assessed by numerosity comparison tasks, also correlate with concurrent and past mathematical achievement all the way back to kindergarten (Halberda et al., 2008) and predict mathematical skills in 6 year olds (Mazzocco et al., 2011). However, at least one study did not find any relationship between numerosity estimation and mathematical ability in kindergarten children after controlling for verbal abilities, verbal and numerical short-term memory, knowledge of numerals, and mathematical fact retrieval (Soltész, et al., 2010). The sources of these inconsistencies remain unclear and more research is required to clarify whether the variation in non-symbolic and symbolic number sense is consistently and causally related to variation in mathematical ability across development.

It is possible that the relationship between number sense and mathematics is uneven across development, so that number sense bootstraps mathematical development at one age, and the relationship is reversed at other ages. It is also possible that the relationship differs in different cultures, as suggested by a recent cross-cultural study with 5-7 year old children (Rodic, et al., in press). It is not excluded that the cross-cultural differences are at least partially explained by differences in exposure or training to numerical material. Both symbolic (on number line task) and non-symbolic (numerosity) estimation abilities are sensitive to training (e.g., Jordan, Suanda & Brannon, 2008; Siegler & Ramani, 2008). It has been proposed that the superior number line estimation of Chinese kindergarten children compared to their American peers stems from the greater exposure to numerical activities of the Chinese children before entering school (Siegler & Mu, 2008). However, currently there is little evidence of long lasting effects of training in estimation skills. One intervention study (Booth & Siegler, 2008) suggested a causal relationship between numerical estimations and mathematical skills. In this study, 7 year old children were presented with a number line on a computer screen. One of the experimental manipulations consisted in the exposure to the correct visual information of numerical magnitudes: the number line in this condition showed the correct position of the numbers constituting addition-problems (the two addends and the total). Children who received this type of intervention/feedback showed improved estimation abilities on a number line task compared to children in other experimental conditions. Moreover, two weeks later, these children showed increased learning of novel mathematical problems compared to the other groups. The authors argued that exposures to accurate visual representation of numerical magnitudes contribute to the structure of a more accurate mental number line. This in turn influences learning by providing a solid framework in which organising numerical knowledge. Much more research on the links between mathematical ability and numerical abilities of estimation is necessary in order to establish the developmental course and the direction of effect of these associations.

Another important question that is yet to be fully addressed is the issue of specificity of association between estimation skills and mathematics. Recent longitudinal studies suggest that the association between non-symbolic estimation and mathematics is unique. For example, one study found no correlations between estimation of numerosities on a dot estimation task at age 14 and sixteen cognitive measures, including visuo-spatial reasoning, working memory, reading, word knowledge and object perception at age 9 - after controlling for mathematical ability (Halberda et al., 2008). Similarly, non-symbolic estimation skills measured prior entering school, were found to be uniquely associated with mathematical abilities measured in primary school (Mazzocco et al., 2011). On the other hand, mathematical ability consistently co-varies with reading, language, and general cognitive factors (Lewis, Hitch, & Walker, 1994; Dirks et al., 2008; Fuchs et al., 2010b; Kovas et al., 2005; Kovas et al., 2007d). Working memory, including phonological loop, visuo-spatial sketchpad and central executive, have been found to be related to mathematical achievement and performance (e.g.: Swanson and Sachse-Lee, 2001; McLean and Hitch, 1999; D'Amico and Guarnera, 2005; Siegler & Ryan, 1989; McLean and Hitch, 1999; Bull, Johnston, & Roy, 1999; Geary, 2011). However, the association between mathematics and memory may be mediated by speed of processing (Case, Kurland & Goldberg, 1982; Bull & Johnston, 1997; Bull et al., 1999).

Because of the relationship between estimation and mathematics, investigating sex difference in estimation abilities may help to understand the reported sex differences in mathematical achievement and performance. Similarly to mathematics, the issue of sex differences in symbolic and non symbolic estimation is debated. One study that measured non-symbolic numerosity estimation using a computerised task reported a small male advantage in 4 year olds (Soltész et al., 2010). However, the authors attributed this finding to the different use/familiarity of computers between boys and girls. This explanation is in line with literature on non-symbolic estimation suggesting that the same biological mechanisms provide the basis of mathematical learning in boys and girls (Spelke, 2005; Spelke & Grace, 2006). For symbolic estimation, sex differences have been reported between the ages of 7 and 9 years, with better performance of boys in number line tasks (LeFevre et al., 2010; Thomson & Opfer, 2008). However, Thomson and Opfer also noted some developmental changes, as sex differences were significant at the age of 7, while decreased when the children were around the age of 9 (boys performed better in number line estimating integer numerals, while girls performed better than boys in estimating fractions). Sex differences in number line estimation suggest that numerical magnitude representation may be involved in the observed sex differences in mathematical performance. Besides providing some evidence for sex differences, these studies also suggest developmental changes; therefore the investigation of sex differences at later ages is of particular importance.

5.2.1 Research question

The present study is the first large-scale longitudinal multivariate investigation into specificity and stability of the relationship between two number sense measures (symbolic estimation with Number Line task and non-symbolic comparison with a Dot Estimation task) and mathematics across school years (age 7 to 16). The study had three major aims: (1) to examine whether the two measures of number sense - symbolic estimation and non-symbolic numerosity comparison - are closely related to each other; (2) to examine the stability of the relationship between mathematical abilities and number sense measures across development; (3) to examine the specificity of the longitudinal relationships between number sense and mathematics, by including in the analyses data on language, reading, verbal and non verbal abilities available from the same children at 7, 9, 10, 12, and 14 years of age. The large sample employed in this study also allowed us to investigate any potential gender differences in number sense at 16 years of age.

5.3 Methods

5.3.1 Participants and procedure

This study used data collected from the TEDS sample (described in section 2.2) at various school ages. At age 16 the analyses were conducted using data from the first cohort as described in section 2.2.4.6. The final sample at this age consisted of 2,100 twins (1,050 pairs) with mean age 16.5 (SD = .19). For the purpose of the longitudinal analyses, all the data collected at ages of 14, 12, 10, 9 and 7 was used. At age 14, there were 6,209 twins (3,105 pairs) with mean age 14.10 (SD = .55). At age 12 the final sample was of 10,878 twins (5,439 pairs) with mean age 11.72 (SD =.67). At age 10 there were 5,634 twin (2,817 pairs) with mean age 10.09 years (SD = .28). At the age of 9 there were 6,329 twins (3,164 pairs) with mean age 9.03 (SD = .28) and at age 7 there were 9,824 twins (4,912 pairs) with mean age 7.11 (SD = .25). The procedure of data collection at each age has been described in details in section 2.2.4.

5.3.2 Measures

The measures used in the analyses are described in details in sections 2.4 and 4.3.3.2; section 1.2.11 provides justifications for the choice of measures used. Here we provide a brief summary of the variables grouped by age of testing.

5.3.2.1 Contemporaneous measures age 16

1. Number sense ability - *Number Line* task, assessing estimation of numerical magnitude; *Dot Task*, assessing estimation of large numerosity. The analyses used *Weber Fraction* scores derived from the Dot Task as described in section 4.3.3.2.

2. Tests of Mathematical performance – This was the composite score of following tests: *Problem Verification Task*, a measure of mathematical fluency and *Understanding Numbers*, a measure of mathematical school achievement according the UK National Curriculum.

3. Tests of General Cognitive ability - Corsi *Tapping Block*, measuring visuo-spatial working memory; *Reaction Time* task, measuring speed of processing; *Raven's Progressive Matrices*, a measure of non-verbal intelligence; *Mill Hill Vocabulary* test, measuring verbal ability; *Figurative Language*, assessing language ability; *Reading*, assessing reading fluency and reading comprehension.

5.3.2.2 Longitudinal measures

14 years

1. Mathematics - *Mathematical achievement* as assessed by teachers.

2. Verbal ability - *Vocabulary* test.

3. Non Verbal ability - *Raven's Standard Progressive Matrices* test.

12 years

1. Mathematics - *Mathematical achievement* as assessed by teachers; *Mathematical web test*. The two measures are analysed separately.

2. Verbal ability - *General Knowledge* test

3. Non Verbal ability - : Composite score of *Raven's Standard Progressive Matrices* test and *Picture Completion* Test.

4. Spatial ability - Composite score of *Hidden Shape* and *Jigsaw* tests.

5. Reading ability - Composite score of reading fluency (*TOWRE, reading words*, *TOWRE, reading non-words*; Woodcock-Johnson-III Reading Fluency Test) and reading comprehension (*GOAL* and *PIAT*).

6. Language ability - Composite of *Figurative language*, *Inferences* and *Grammar* tests

10 years

1. Mathematics - *Mathematical achievement* as assessed by teachers; *Mathematical web test*. The two measures are analysed separately.

2. Verbal ability - Composite of *Vocabulary* and *General Knowledge* tests.

3. Non Verbal ability - : Composite score of *Raven's Standard Progressive Matrices* test and *Picture Completion* Test.

4. Reading ability - *PIAT*, test of reading comprehension.

9 years

1. Mathematics - *Mathematical achievement* as assessed by teachers.

2. Verbal ability - Composite of *Vocabulary* and *General Knowledge* tests.

3. Non Verbal ability - Composite score of *Conceptual Grouping* and *Picture Completion* tests.

7 years

1. Mathematics - *Mathematical achievement* as assessed by teachers.

2. Verbal ability - Composite of *Vocabulary* and *Similarity* tests.

3. Non Verbal ability - Composite score of *Puzzle* and *Shapes* tests.

4. Reading ability - Composite score of *TOWRE, reading words* and *TOWRE, reading non-words*. Both tests assessed reading fluency.

5.4 Results

All the variables were standardised with mean of zero and standard deviation of one. The analyses were conducted on the sample constituted by one randomly selected twin from each pair. The analyses were replicated on the sample constituted by the co-twins and only results holding statistical significance in both samples were considered significant. Table 5.1 shows the correlations among all the measures. Consistent with the analyses conducted during the pilot study (previous sections 4.4.2 and 4.4.4) both number sense measures had significant negative correlation with contemporaneous mathemtics, at age 16. Their correlation with mathematics was also significant at all the previous ages. The two measures had also significant negative relationship with contemporaneous and previous measures of: reading, language, verbal, non-verbal ability, spatial ability and visuo-spatial working memory. The sign of the negative correlation reflects the fact that smaller values of Weber Fraction and Number Line scores indicate better performance in the two number sense tasks - and are associated with higher performance in the mentioned abilities. Both Number Line and Weber Fraction held a positive relationship only with Reaction Time (efficiency) as for the three measures, the smaller the score, the better the performance on the tasks. In addition to mathematics, the two number sense measures correlated with all the other cognitive measures, contemporaneous and longitudinal, suggesting a non-unique relationship mathematics-number sense.

Table 5.1
Correlations all variables across all ages

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
1	*Weber Fraction*	-																															
2	*Number Line*	.22**																															
3	*Teacher Maths composite at 7*	-.28**	-.34**																														
4	*VERBAL composite at 7*	-.12**	-.17**	.33**																													
5	*NON-VERBAL at 7*	-.10**	-.14**	.28**	31**																												
6	*Reading at 7*	-.22**	-.25**	.53**	.40**	.23**																											
7	*Teacher Maths composite at 9*	-.23**	-.35**	.58**	.35**	.31**	.55**																										
8	*Verbal composite at 9*	-.11**	-.21**	.31**	.30**	.22**	.35**	.31**																									
9	*NON_Verbal composite at 9*	-.24**	-.25**	.37**	.22**	.24**	.33**	.39**	.44**																								
10	*Teacher Maths composite at 10*	-.22**	-.38**	.58**	.37**	.29**	.52**	.69**	.35**	.41**																							
11	*Maths web test at 10*	-.35**	-.40**	.44**	.27**	.25**	.36**	.46**	.38**	.51**	.50**																						
12	*Verbal composite at 10*	-.19**	-.22**	.37**	.36**	.25**	.36**	.36**	.47**	.40**	.40**	.52**																					
13	*NON_Verbal composite at 10*	-.27**	-.29**	.30**	.24**	.27**	.21**	.33**	.30**	.47**	.36**	.56**	.51**																				
14	*Reading at 10*	-.18**	-.25**	.29**	.36**	.23**	.42**	.34**	.45**	.37**	.36**	.47**	.54**	.48**																			
15	*Maths web tests at 12*	-.35**	-.41**	.54**	.34**	.33**	.48**	.55**	.40**	.53**	.62**	.64**	.44**	.50**	.47**																		
16	*Teacher Maths composite at 12*	-.17**	-.42**	.47**	.25**	.18**	.48**	.48**	.26**	.34**	.50**	.49**	.33**	.34**	.32**	.58**																	
17	*Verbal composite at 12*	-.14**	-.20**	.41**	.40**	.25**	.40**	.35**	.48**	.40**	.41**	.45**	.57**	.41**	.52**	.51**	.42**																
18	*NON_Verbal composite at 12*	-.19**	-.23**	.36**	.29**	.28**	.28**	.22**	.34**	.44**	.35**	.38**	.35**	.54**	.37**	.49**	.34**	.46**															
19	*Reading composite at 12*	-.26**	-.29**	.44**	.42**	.26**	.70**	.50**	.43**	.36**	.49**	.41**	.47**	.33**	.58**	.59**	.45**	.53**	.38**														
20	*Language composite at 12*	-.19**	-.23**	.44**	.44**	.32**	.45**	.41**	.37**	.40**	.44**	.42**	.43**	.38**	.46**	.61**	.51**	.58**	.47**	.58**													
21	*Spatial composite at 12 yrs*	-.19**	-.29**	.29**	.22**	.23**	.26**	.30**	.20**	.33**	.34**	.34**	.28**	.42**	.31**	.47**	.36**	.32**	.48**	.32**	.37**												
22	*Teacher Maths composite at 14*	-.24**	-.39**	.56**	.33**	.31**	.45**	.65**	.37**	.45**	.71**	.56**	.42**	.47**	.36**	.74**	.63**	.54**	.42**	.54**	.61**	.44**											
23	*Verbal composite at 14*	-.21**	-.26**	.36**	.35**	.24**	.41**	.40**	.34**	.34**	.38**	.41**	.47**	.34**	.38**	.47**	.38**	.53**	.40**	.52**	.51**	.29**	.44**										
24	*NON_Verbal composite at 14*	-.27**	-.27**	.41**	.28**	.24**	.34**	.40**	.21**	.40**	.45**	.39**	.31**	.43**	.33**	.52**	.38**	.36**	.52**	.37**	.42**	.44**	.48**	.42**									
25	*Maths -web at 16 composite*	-.33**	-.47**	.54**	.32**	.27**	.42**	.58**	.31**	.44**	.61**	.59**	.43**	.43**	.35**	.71**	.53**	.49**	.36**	.52**	.53**	.41**	.73**	.49**	.53**								
26	*Reaction Time at 16*	.25**	.21**	-.25**	-.18**	-.12**	-.19**	-.28**	-.16**	-.25**	-.26**	-.22**	-.13**	-.22**	-.16**	-.33**	-.30**	-.17**	-.23**	-.26**	-.25**	-.23**	-.34**	-.26**	-.23**	-.36**							
27	*Corsi at 16*	-.22**	-.23**	.28**	.14**	.17**	.19**	.29**	.08	.29**	.30**	.27**	.16**	.25**	.10*	.34**	.25**	.14**	.22**	.21**	.18**	.26**	.26**	.25**	.33**	.37**	-.23**						
28	*VERBAL at 16*	-.19**	-.23**	.28**	.34**	.22**	.41**	.32**	.32**	.28**	.34**	.29**	.41**	.28**	.40**	.45**	.33**	.45**	.31**	.49**	.44**	.23**	.43**	.40**	.35**	.46**	-.16**	.22**					
29	*NON_VERBAL at 16*	-.27**	-.34**	.37**	.26**	.24**	.26**	.31**	.21**	.40**	.42**	.45**	.32**	.44**	.31**	.56**	.37**	.32**	.49**	.34**	.43**	.35**	.50**	.38**	.58**	.58**	-.25**	.29**	.35**				
30	*READING composite at 16*	-.16**	-.25**	.38**	.36**	.24**	.53**	.40**	.31**	.37**	.43**	.38**	.45**	.30**	.49**	.48**	.42**	.44**	.33**	.68**	.56**	.22**	.48**	.48**	.38**	.47**	-.33**	.23**	.52**	.36**			
31	*LANGUAGE at 16*	-.21**	-.28**	.33**	.39**	.24**	.36**	.35**	.28**	.30**	.37**	.34**	.41**	.30**	.39**	.47**	.30**	.51**	.37**	.53**	.53**	.21**	.46**	.51**	.37**	.52**	-.23**	.22**	.49**	.37**	.54**		
32	*Sex*	-.02	-.02	.02	-.00	-.05	-.01	.03	.06	.05	.06	.12**	.14**	.08	.04	.06	.04	.13**	.01	-.01	-.05	.05	.037	-.04	.01	.20**	-.02	.10*	-.02	.03	-.04	-.04	

*** p< .01 ; * p<.05 (2-tailed)*

5.4.1 Examining the number sense construct

The correlation (Table 5.1) between Number Line and Weber Fraction scores was modest (.22) and did not justify the creation of a composite score of number sense. We further explored the Number Line and Weber Fraction measures by including them in a Factor Analysis (FA) together with the mathematics and reading composite scores, verbal, non-verbal scores, language, reaction time and visuo-spatial working memory. We used FA rather than Principal Component Analysis (PCA) to carry out this investigation and used the extraction method of Principal Axis Factoring on the 9 variables obtained from the web testing at 16. Although FA and PCA often yield similar results in reducing large set of variables they deal differently with the variables' variances-covariances. PCA calculates the factors on the basis of the total variance in the variables belonging to the structure. FA extracts factors that explain the variables' covariance. FA is suited to investigate constructs underlying the data.

The adequacy of the sample size (N = 490) was confirmed with Kaiser Meyer Olkin test which returned a good KMO value of .83 with the Bartlett's sphericity test χ^2 (36) = 989.56, $p < .001$. The eigen values were first obtained for all components, only two components with an eigen value greater than 1 were retained. A comparison with the scree plot confirmed the retention of these two factors. The analysis was performed with orthogonal and oblique rotation (Table 5.2). The two rotations yielded similar results in terms of loadings although the oblique rotation provided a cleaner output as mathematics and non-verbal abilities did not cross-load on the two factors. This suggested a relationship between the two dimensions. Non-Verbal abilities clustered on Factor 1, while Factor 2 represented a verbal and language dimension. Both Weber Fraction and Number Line loaded on Factor 1. As Weber Fraction, Number Line and Reaction Time correlated negatively with mathematics (and positively among them) (Table 5.1), their correlation with Factor 1 was in the opposite direction (negative) than the correlation of the Mathematics composite with Factor 1 (positive). Similarly, Non-Verbal ability and Visuo-Spatial memory had a positive relationship with Factor 1.

Table 5.2

Summary of the Exploratory Factor Analysis- Analysis carried out with orthogonal and oblique rotation. Extraction method Principal Axis Factoring. The analysis is conducted on the measures obtained from the web testing at 16 years.

	Orthogonal rotation 2 factors extraction (Varimax)		Oblique rotation 2 factors extraction (Oblimin)		Oblique rotation 3 factors extraction (Oblimin)		
	Total Variance explained after rotation 36.5%		Total Variance explained after rotation 36.5%		Total Variance explained after rotation 40.5%		
N=490	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 1**	**Factor 2**	**Factor 3**
Language		.67		.68		.66	
Reading Composite		.73		.74		.75	
Verbal ability		.64		.69		.68	
Mathematics composite	.73	.42	.70		.72	.21	
Non-Verbal Ability (Raven)	.50	.34	.46		.46	.21	
Visuo-Spatial Memory (Corsi)	.42		.45		.38		
Reaction Time	-.32		-.30				.60
Number Line	-.47		-.51		-.55		
Weber Fraction	-.32		-.35		-.22		.21
Eigen values	*3.3*	*1.1*	*3.3*	*1.2*	*3.3*	*1.2*	*.94*
α	*.41*	*.80*	*.41*	*.76*	*.25*	*.80*	*.40*

In the extraction of two factors, both orthogonal and oblique, only factors with loading greater than 3 were retained. In the extraction of three factors the threshold was lowered at factors with loading greater than 2.

Although there was a concordance between the scree plot and the Kaiser criterion based on the eigen values in extracting only two factors, a third factor yielded an eigen value of .94 in the initial extraction. As the factor analysis had the purpose to understand the underlying structure within the data, a second analysis was performed to extract three factors. These results are shown in Table 5.2. In this analysis a Factor 1 representing a non verbal dimension and Factor 2 a verbal-general dimension could still be identified. The third factor could be mapped to speed. Reaction Time loaded only on Factor 3 while Weber Fraction loaded on both Factor 3 and Factor 1 together with Number Line and mathematics. The load of Weber Fraction was small on both factors, and only present when retaining variables with loadings greater than .2. In fact, Weber fraction did not significantly load on Factor 3 in the replication sample.

To summarise, Number Line, Weber Fraction, mathematics and non verbal abilities clustered on Factor 1 with a Cronbach alpha of .25. Language, reading and verbal abilities loaded on Factor 2 (α = .80). Reaction Time and Weber Fraction loaded on Factor 3 (α = .40). In addition it was observed that deleting the non-verbal ability scores from Factor 1 (in the 3 factors extraction), its internal validity increased to .72. This may indicate a weaker relationship of non-verbal abilities with the mathematics-estimation construct. These results further suggested that Weber Fraction, Number Line and mathematics belong to the same construct, but perhaps estimation of numerosities (Weber Fraction) has a weaker relationship with the mathematical/non-verbal dimension compared to estimation of numerical magnitudes (Number Line). More importantly these analyses showed that even forcing the model to three factors we could not detect a singular "number sense dimension". For this reason, all further analyses were conducted separately for Weber Fraction and Number Line measures – representing different hypothesised aspects of number sense.

5.4.2 Examining Sex Differences in number sense

Table 5.3 shows descriptive statistics of the standardised scores for the number sense measures and mathematics. Means and standard deviation on raw data were as follow: Problem Verification, Mean = 35.86, SD = 6.97; Understanding Numbers Mean = 11.63, SD = 4.39; Number Line Mean = 36.41, SD = 15.32; Weber Fraction Mean = .28, SD = .13. The divergence of the means from zero indicates that Mathematics composite and Weber Fraction were not normally distributed. Weber fraction was negatively skewed; Mathematics composite had a small positive skewness. However, the mathematical composite, was less kurtotic as its standard deviation was closer to 1 compared to the Weber Fraction's. Independent sample t-tests with equal variance assumed showed no significant differences between the mean scores of males and females in Number Line or Weber Fraction. Further ANOVAs on Number Line, Weber Fraction and mathematics at 16 confirmed no sex effects in number sense, and significant but very modest (η^2 =.04) effect on mathematics (Table 5.3). In the following multiple regressions sex was not found to be a significant predictor of either measure of number sense.

Table5.3

Mean, Standard deviation and ANOVA results by Sex

Mathematics composite 16	*N=644*		*N=242*		*N=402*				
Weber Fraction at 16	*N=729*		*N=272*		*N=457*				
Number Line at 16	*N=857*		*N=329*		*N=528*				
							Effect Size		
	All		***Males***		***Females***		***Sex***		**R^2**
Measures	***M***	***SD***	***M***	***SD***	***M***	***SD***	***p***	***η^2***	
Mathematics composite 16	.02	.99	.05	.97	.01	1.0	.00	.04	.04
Weber Fraction at 16	-.12	.77	-.11	.81	-.13	.75	.83	.00	.00
Number Line at 16	.00	.98	.03	.99	-.02	.98	.55	.00	.00

M = Mean; SD = Standard Deviation; p = p-value; η^2 = effect size; R^2 = variance explained by sex in the model.

5.4.3 Predicting Mathematics from number sense

As shown in the correlational matrix (Table 5.1), Number Line and Weber Fraction, assessed at 16 years of age, correlated significantly with mathematics and other cognitive measures at all ages. Smaller values of Number Line scores indicate more accurate estimations, while smaller values of Weber Fraction refer to greater accuracy in discriminating finer ratios – reflected in negative correlations of these parameters with other cognitive measures.

First, we investigated whether any relationship exists between mathematics at different ages and number sense at age 16. We conducted a series of regressions, entering both Number Line and Weber Fraction scores simultaneously as predictor variables, and mathematics at different ages as criterion. Separate regressions were run for teacher rated mathematics and for the web-assessed mathematics. The results, summarised in Table 5.4, suggested an overall weaker and less consistent relationship between mathematics and Weber Fraction than that between mathematics and Number Line. Moreover, both number sense measures explained more variance in web-assessed mathematics than in teacher-rated mathematics.

Table 5.4

Summary regression method forced entry of Mathematics Teachers' Ratings and Mathematics Web tests at all ages. Mathematics has been entered as DV, Weber Fraction and Number Line were entered as IV.

	Mathematics Teacher's Ratings				**Mathematics Web Tests**			
	B	SE-B	Beta	CI (95%) for B	B	SE-B	Beta	CI (95%) for B
Age 7								
Number Line @ 16	-.26	.04	-.28***	(-.34, -.19)				
Weber Fraction @ 16	-.25	.05	-.20***	(-.35, -.15)				
	*R^2=.149; F(2,525)=45.89****							
Age 9								
Number Line @ 16	-.30	.05	-.31***	(-.39, -.21)				
Weber Fraction @ 16	-.20	.06	-.15**	(-.33, -.08)▲				
	*R^2=.143; F(2,410)=34.25****							
Age 10								
Number Line @ 16	-.33	.04	-.34***	(-.41, -.24)	-.25	.04	-.30***	(-.32, -.18)
Weber Fraction @ 16	-.20	.06	-.15**	(-.31, -.08)▲	-.32	.05	-.29***	(-.41, -.23)▲
	*R^2=.162; F(2,412)=39.91****				*R^2=.213; F(2,445)=60.10****			
Age 12								
Number Line @ 16	-.30	.04	-.37***	(-.39, -.22)	-.34	.04	-.35***	(-.42, -.26)
Weber Fraction @ 16	-.11	.06	-.09	(-.22, .01)	-.34	.05	-.27***	(-.44, -.24)▲
	*R^2=.160; F(2,364)=34.60****				*R^2=.241; F(2,474)=75.27****			
Age 14								
Number Line @ 16	-.29	.04	-.33***	(-.37, -.21)				
Weber Fraction @ 16	-.21	.06	-.17***	(-.32, -.09)				
	*R^2=.167; F(2,412)=41.22****							
Age 16								
Number Line @ 16					-.41	.04	-.41***	(-.48, -.34)
Weber Fraction @ 16					-.32	.05	-.25***	(-.41, -.23)
					*R^2=.271; F(2,577)=107.43****			

*p<.05; **p<.01; ***p<.001

▲ indicates discrepancy of significance on the analysis conducted on the replication sample

5.4.4 Predicting number sense from contemporaneous and earlier measures of mathematics

The next series of analyses explored whether early mathematical achievement explained additional variance in individual differences in estimation abilities at 16, beyond contemporaneous mathematical achievement - addressing the causal links between estimation abilities and mathematics. Number Line and Weber Fraction were entered as dependent variables in two separate stepwise regressions. Mathematics at 16 was entered as predictor in the first step, while mathematics scores at the earlier ages were added in the second step. The results are summarised in Table 5.5. The first regression showed that the contribution of early mathematics to the variance in Weber Fraction was not significant, with the possible exception of age 7. This result suggests that if the causal relationship between mathematics and Weber Fraction exists, it may not be consistent across development. The results of the second regression showed that none of the early mathematics measures significantly predicted Number Line scores, after controlling for association with mathematics at 16.

Table 5.5

Regression stepwise method: Weber fraction and Number Line scores predicted by mathematics at all ages. Mathematics web-scores at 16 have been entered in the first step, mathematics at earlier years have been entered in the second step.

Weber fraction predicted by Mathematics					Number Line predicted by Mathematics				
	B	SE -B	Beta	CI (95%) for B		B	SE - B	Beta	CI (95%) for B
1st step					***1st step***				
Maths 16-web	-.23	.07	-.36**	(-.38, -.08)	Maths 16-web	-.39	.09	-.44***	(-.57, -.21)
R^2=.130; F(1,65)=9.67, p<.01					*R^2=.194; F(1,75)=18.10, p<.001*				
2nd step					***2nd step***				
Maths 16-web	-.27	.13	-.42*	(-.52, -.01)	Maths 16-web	-.43	.17	-.49**	(-.77, -.09)
Maths 14-teach	.21	.17	.32	(-.12, .55)	Maths 14-teach	-.00	.23	-.00	(-.46, .45)
Maths 12-teach	.10	.14	.12	(-.18, .38)	Maths 12-teach	-.31	.19	-.26	(-.68, .07)
Maths 12-web	-.37	.16	-.05	(-.35, .28)	Maths 12-web	.14	.20	.13	(-.25, .52)
Maths 10-teach	.02	.12	.02	(-.23, .26)	Maths 10-teach	.06	.15	.06	(-.24, .35)
Maths 10-web	-.22	.15	-.27	(-.53, .08)	Maths 10-web	.09	.20	.08	(-.31, .49)
Maths 9 -teach	.09	.13	.14	(-.17, .35)	Maths 9 -teach	.14	.16	.14	(-.18, .46)
Maths 7-teach	-.22	.11	-.31*	(-.43, -.01) ▲	Maths 7-teach	-.16	.15	-.16	(-.45, .13)
ΔR=.107; Sig. F-change=.34					*ΔR=.064; Sig. F-change=.56*				

****p<.001, ** p<.01, *p=.05*

▲ Indicates discrepancy of significance in the analysis conducted on the replication sample.; Maths web = Mathematics assessed with the web tests; Maths teach. = Mathematics assessed by teachers.

5.4.5 Predicting Mathematics from number sense and other cognitive abilities

Next, we examined the specificity of the relationship between mathematics and number sense. In a series of multiple regressions, Number Line and Weber Fraction were entered into regressions as predictors of mathematics (separately for Web Tests and Teacher's Ratings), together with the general cognitive abilities (separately at ages 7, 9, 10, 12, 14, and 16). The results are summarised in Table 5.6. In the presence of other cognitive abilities, Number Line maintained a significant relationship with mathematics (both Teachers' Ratings and Web Tests) across all ages. On the other hand, Weber Fraction did not maintain a consistent significant relationship with mathematics, with the exception of the contemporaneous mathematics measure. As evident from Table 6.6, controlling for 6 other cognitive measures, variance in mathematics at 16 was significantly explained by Weber Fraction (6%), Number Line (22%), Reaction Time (11%), spatial memory (Corsi) (12%), verbal intelligence (14%), non-verbal intelligence (23%), and language (20%). These results speak against the specificity of the relationship between number sense and mathematics at this age. In other words, it is likely that the relationship between number sense and mathematics is largely explained by the variance overlapping with other cognitive measures. In fact, as shown in Table 6.1, Number Line and Weber Fraction presented significant correlation with all general cognitive abilities in the study.

Table 5.6

Summary multiple regressions, method forced entry, Mathematics Web Tests and Mathematics Teachers' Ratings were entered as DV in the models at all ages, the General Cognitive measure at the age of 7,9,10,12, 14 and Weber Fraction and Number Line at 16 were entered as IV.

	Mathematics Teacher's Ratings				**Mathematics Web Tests**			
	B	SE - B	Beta	CI (95%) for B	B	SE - B	Beta	CI (95%) for B
Age 7								
Number Line -16	-.16	.04	-.17***	(-.23, -.09)				
Weber Fraction -16	-.14	.05	-.11**	(-.23, -.05) ▲				
Verbal Comp. - 7	.11	.04	.12**	(.04, .18)				
Non-Verbal Comp. - 7	.12	.04	.12**	(.04, .19)				
Reading -7	.40	.04	.39***	(.32, .48)				
Sex	.13	.07	.07	(.00, .26)				
*R^2=.373; F(6,494)=49.06****								
Age 9								
Number Line- 16	-.21	.05	-.22***	(-.30, -.12)				
Weber Fraction- 16	-.16	.06	-.12*	(-.28, -.04) ▲				
Verbal Comp. - 9	.15	.05	.14**	(.05, .25)				
Non-Verbal Comp. - 9	.25	.05	.25***	(.15, .25)				
Sex	.12	.09	.06	(-.05, .29)				
*R^2=.236; F(5,380)=23.49****								
Age 10								
Number Line - 16	-.22	.05	-.23***	(-.31, -.13)	-.15	.03	-.18***	(-.20, -.09)
Weber Fraction - 16	-.11	.06	-.08	(-.23, .02)	-.14	.04	-.13***	(-.22, -.07) ▲
Verbal Comp. - 10	.24	.06	.23***	(.11, .36)	.19	.04	.22***	(.11, .26)
Non-Verbal Comp. -10	.12	.06	.11	(-.01, .24) ▲	.29	.04	.31***	(.21, .38)
Reading -10	.12	.06	.11	(-.01, .24) ▲	.12	.04	.14**	(.04, .19)
Sex	.17	.09	.09	(-.02, .35) ▲	.11	.06	.66	(-.01, .22)
*R^2=.282; F(6,309)=20.27****					*R^2=.472; F(6,421)=62.71****			
Age 12								
Number Line -16	-.19	.05	-.23***	(-.29, -.09)	-.15	.04	-.16***	(-.23, -.08)
Weber Fraction -16	-.02	.07	-.01	(-.16, .13)	-.14	.05	-.12**	(-.24, -.05) ▲
Verbal Comp. -12	.09	.07	.09	(-.06, .23)	.12	.05	.12*	(.02, .21)
Non-Verbal Comp.-12	-.03	.07	-.03	(-.16, .10)	.08	.04	.08	(-.00, .17) ▲
Reading Comp.-12	.12	.06	.14	(-.01, .24) ▲	.27	.05	.28***	(.17, .35)
Language Comp.- 12	.20	.07	.23**	(.06, .33)	.19	.04	.21***	(.10, .28)
Spatial Comp. -12	.10	.05	.12	(-.01, .21)	.13	.03	.15***	(.06, .19)
Sex	-.02	.10	-.01	(-022, .18)	.10	.07	.06	(-.03, .24)
*R^2=.309,F(8,213)=11.93****					*R^2=.574; F(8,319)=53.65****			
Age 14								
Number Line -16	-.19	.04	-.22***	(-.26, -.11)				
Weber Fraction -16	-.04	.05	-.04	(-.15, .06)				
Verbal Comp. -14	.24	.04	.24***	(.05, .25)				
Non-Verbal Comp. - 14	.33	.04	.37***	(.25, .41)				
Sex	.20	.07	.11**	(.06, .33)				
*R^2=.387; F(5,373)=47.11****								
Age 16								
Number Line - 16					-.21	.03	-.22***	(-.27, -14)
Weber Fraction - 16					-.08	.04	-.06*	(-.17, .00)
Reaction Time -16					-.13	.04	-.11**	(-.21, -.06)
Corsi -16					.12	.03	.12**	(.05, .18)
Verbal Comp.- 16					.14	.04	.14***	(.07, .22)
Non-Verbal Comp. -16					.23	.04	.23***	(.16, .30)
Reading Comp.-16					.07	.04	.08	(-.00, .15) ▲
Language Comp. - 16					.21	.04	.20***	(.13, .28)
Sex					.36	.06	.19***	(.25, .48)
					*R^2=.537; F(9,480)=61.79****			

*p<.05; **p<.01; ***p<.001* ▲ indicates discrepancy of significance on the analysis conducted on the other half of the sample

5.4.6 Predicting number sense from Mathematics and cognitive skills

The observed relationship between number sense and mathematics may reflect the influence of formal mathematical activities. To examine whether number sense at 16 is best predicted by mathematics, as opposed to other cognitive skills, Number Line and Weber Fraction were entered as dependent variables in two separate blocks of multiple regressions, where general cognitive abilities and mathematics at ages 7, 9, 10, 12, 14, and 16 were the independent predictors (Table 5.7).

At age 16, Number Line estimation was significantly associated only with mathematics, after controlling for speed of processing (Reaction Time), visuo-spatial working memory (Corsi), verbal, non-verbal, reading and language abilities at the same age. In the replication sample, non verbal ability also showed a significant correlation with Number Line. All the longitudinal measures of mathematics also significantly predicted individual differences in Number Line estimation at 16, after controlling for other cognitive abilities at the same ages. Although reading abilities and non-verbal ability also predicted Number Line performance at some ages, these relationships were weaker than those with mathematics and were not replicated in both samples.

At age 16, Weber Fraction was associated only with mathematics and speed of processing (Reaction Time), after controlling for contemporaneous measures of cognitive ability. The relationship with visuo-spatial working memory (Corsi) and non-verbal scores were only significant in one sample. Similarly to Number Line, Weber Fraction scores at 16 were significantly predicted by mathematics in the presence of other general cognitive abilities at earlier ages. However, the relationship between mathematics and Weber Fraction was overall weaker and less consistent across ages. Weber Fraction significantly correlated with the earliest (age 7) and the contemporaneous (age 16) mathematics in the presence of other cognitive abilities. Besides mathematics, the only significant correlations were with the non-verbal ability at 9 and 10

Table 5.7

Summary of multiple regressions, method forced entry, Weber Fraction and Number Line measures at 16 were entered as DV. - , The general cognitive measure and mathematics at the age of 7,9,10,12 and 16 were entered as IV. At the ages of 10 and 12 Number Line and Weber Fraction were estimated in two different models having alternatively Teacher's rating and Web assessment. Sex was entered in all the regression but was non-significant at all ages for both DV and it is not reported in this table. The degrees of freedom reflect Sex entered in the regression

	Weber Fraction				Number Line			
	B	SE - B	Beta	CI (95%) for B	B	SE - B	Beta	CI (95%) for B
Teachers-Maths -7	-.17	.04	-.22***	(-.25, -.09)	-.29	.05	-.27***	(-.39, -.19)
Verbal Composite -7	.01	.04	.01	(-.06, .08)	-.03	.04	-.03	(-.12, .05)
Non-Verbal Comp. - 7	.00	.04	.00	(-.07, .07)	-.03	.04	-.03	(-.12, .05)
Reading Ability - 7	-.10	.04	-.13*	(-.18, -.02) ▲	-.10	.05	-.09*	(-.20, -.00) ▲
	*R^2=.093; F(5,500)=10.25****				*R^2=.127; F(5,585)=17.08****			
Teachers-Maths - 9	-.15	.04	-.19***	(-.23, -.07) ▲	-.28	.05	-.28***	(-.30, -.18)
Verbal Composite - 9	.06	.04	.08	(-.03, .15)	-.02	.05	-.02	(-.13, .08)
Non-Verbal Comp. - 9	-.12	.04	-.16**	(-.21, -.04)	-.08	.05	-.08	(-.19, .02) ▲
	*R^2=.072; F(4,385)=7.42****				*R^2=.109; F(4,449)=13.74****			
Teachers-Maths @ 10	-.10	.05	-.13*	(-.20, -.01) ▲	-.30	.06	-.28***	(-.42, -.18)
Verbal Composite - 10	.03	.06	.03	(-.09, .14)	.04	.07	.04	(-.10, .19)
Non-Verbal Com. - 10	-.23	.06	-.26***	(-.34, -.12) ▲	-.17	.07	-.15*	(-.31, -.04) ▲
Reading Ability - 10	-.02	.06	-.02	(-.13, .09)	-.07	.07	-.06	(-.20, .06) ▲
	*R^2=.109; F(5,313)=7.64****				*R^2=.140; F(5,352)=11.45****			
WEB-Maths - 10	-.22	.06	-.24***	(-.33, -.11) ▲	-.44	.07	-.36***	(-.57, -.31)
Verbal Composite - 10	.02	.05	.03	(-.07, .12)	.05	.06	.04	(-.07, .16)
Non-Verbal Com - 10	-.14	.05	-.15**	(-.24, -.03)	-.07	.06	-.06	(-.19, .05)
Reading Ability - 10	-.01	.05	-.02	(-.11, .08)	-.07	.06	-.06	(-.18, .04)
	*R^2=.116; F(5,427)=11.26****				*R^2=.164; F(5,488)=19.11****			
Teachers-Maths -12	-.04	.06	-.05	(-.16, .09)	-.36	.08	-.30***	(-.52, -.20)
Verbal Ability - 12	-.02	.07	-.03	(-.16, .11)	.08	.09	.07	(-.09, .26)
Non-Verbal Comp -12	-.02	.06	-.03	(-.15, .10)	-.18	.08	-.16*	(-.34, -.03) ▲
Reading]Comp.- 12	-.08	.06	-.11	(-.20, .04) ▲	-.17	.08	-.16*	(-.32, -.02) ▲
Language Comp.- 12	-.02	.07	-.02	(-.14, .11)	-.02	.08	-.02	(-.18, .15)
Spatial Composite- 12	-.12	.05	-.18*	(-.22, -.02) ▲	-.02	.07	-.02	(-.15, .11) ▲
	*R^2=.094; F(7,216)=3.21***				*R^2=.201; F(7,244)=8.77****			
WEB- Maths - 12	-.21	.06	-.26**	(-.34, -.09) ▲	-.34	.07	-.32***	(-.48, -.19)
Verbal Ability - 12	.07	.05	.09	(-.03, .18)	.08	.07	.08	(-.05, .21)
Non-Verb. Comp. -12	-.06	.05	-.08	(-.16, .04)	-.05	.06	-.05	(-.17, .07)
Reading Comp. - 12	-.12	.06	-.15*	(-.23, -.00) ▲	-.13	.07	-.13*	(-.26, .00) ▲
Language Comp. - 12	.04	.05	.05	(-.07, .14)	.03	.06	.03	(-.09, .16)
Spatial Comp. - 12	-.02	.04	-.03	(-.10, .06)	-.02	.05	-.02	(-.12, .08) ▲
	*R^2=.124; F(7,323)=6.55****				*R^2=.150; F(7,378)=9.54****			
Teachers-Maths - 14	-.07	.05	-.08	(-.17, .03)	-.40	.06	-.34***	(-.52, -.27)
Verbal Comp. - 14	-.10	.05	-.12*	(-.19, -.01) ▲	-.04	.06	-.04	(-.15, .07) ▲
Non-Verbal Comp- 14	-.11	.04	-.14*	(-.19, -.13)	-.05	.06	-.04	(-.16, .06) ▲
	*R^2=.076; F(4,378)=7.79****				*R^2=.151; F(4,431)=19.22****			
WEB-Maths 16	-.11	.05	-.15*	(-.20, -.02)	-.41	.06	-.40***	(-.52, -.30)
Reaction Time -16	.11	.04	.13**	(.04, .19)	-.01	.05	.00	(-.10, .09)
Corsi -16	-.07	.04	-.10*	(-.14, -.01) ▲	-.04	.04	-.04	(-.12, .05)
Verbal Ability - 16	-.02	.04	-.02	(-.10, .06)	.06	.05	.06	(-.04, .16)
Non-Verb. Ability -16	-.03	.04	-.05	(-.11, .04) ▲	-.08	.05	-.08	(-.18, .01) ▲
Reading Ability.- 16	.01	.04	.02	(-.07, .09)	-.05	.05	-.06	(-.15, .04)
Language Ability- 16	.01	.04	.02	(-.07, .10)	.02	.05	.02	(-.09, .12)
	*R^2=.084; F(8,483)=5.55****				*R^2=.204; F(8,514)=16.52****			

**p<.05; **p<.01; ***p<.001*

▲ Indicates discrepancy of significance on the analysis conducted on the replication sample.

It is important to note, that much of the variance in number sense measures remained unexplained. Even for contemporaneous analyses at 16, only 8.4% and 20.4% of the variance was explained by all available measures in Weber Fraction and Number Line respectively, of which most of the variance in both regressions was attributed to mathematics.

5.5 Discussion

The study reported here is the first large-scale investigation into the relationship between mathematics and two different measures of number sense, administered to the representative longitudinal TEDS sample. The longitudinal design allowed us to explore the continuity of this relationship. Specifically, we tested number sense at age 16 and looked at its retrospective relationship with mathematical achievement (rated by teachers) and performance (on web tests) measured at 7, 9, 10, 12 and 14 years. The strength and uniqueness of the relationship between mathematics and the two measures was also assessed, controlling for a number of cognitive abilities measured across ages. Lastly, sex differences in number sense were explored. A particular strength of this study was the employment of a discovery-replication approach, by using one twin from each pair - generating two matching samples.

In order to thoroughly assess individuals' number sense, this study utilized two tasks of estimation abilities. As expected, the two measures were significantly related to mathematics at 16, collectively explaining 27.1% of the total variance in mathematics. Despite their correlation with mathematics, the two number sense measures correlated only modestly (.22) with each other, seemingly tapping into different aspects of estimating ability. Therefore, no "number sense" construct could be derived from the two measures, as reflected in the results of the exploratory Factor Analysis carried out on the measures at 16. Although Number Line and Weber Fraction loaded on a common (non-verbal ability) factor, Weber Fraction also loaded partially onto another factor, revealing a certain degree of autonomy between the two measures. It is likely that this autonomy reflects the fact that Number Line requires knowledge of the symbolic numerosity, whereas Dot Task relies purely on non-symbolic estimation. Moreover, the results suggest that both tasks may simply reflect two aspects of the general cognitive domain, as both loaded on the non-verbal ability factor. These findings suggest that the concept of number sense may be misleading, and that different estimation abilities may instead be viewed as separate measures of general intelligence. For these reasons, we discuss the relationship between number sense and mathematics separately.

5.5.1 Weber Fraction

The results confirmed the existence of the relationship between mathematics and Weber Fraction. However, this relationship was weak and was not constant across development. Controlling for Number Line, Weber Fraction showed a significant correlation with mathematics at ages 7, 14 and 16, but not at 10 and 12. When other cognitive abilities were also controlled for, only the relationship with contemporaneous mathematics (at 16) survived significance. The absence of the stable relationship between early mathematics and Weber fraction at 16 suggests the absence of a causal link between the two abilities.

The result of the stepwise regression suggested that only mathematics at 7 predicted variance in non-symbolic estimation skills beyond current mathematical proficiency. It is possible, that engaging in mathematical activities at an early age improves non-symbolic estimation skills, or that early mathematics is built on this aspect of number sense. In the current study we could not assess the relationship between early number sense and mathematics. However, the suggested relationship between mathematics at 7 and Weber Fraction at 16 may be the product of this early association. The absence of the relationship between mathematics at other ages (9, 10, 12, 14) and Weber Fraction at 16 may indicate that mathematical skills learned at these ages rely more on other cognitive abilities than on estimation of numerosity. For example, as mathematics becomes more complex and abstract, a basic process such non-symbolic estimation is no longer sufficient or even required for mathematical acquisition, while abilities like spatial memory, speed of processing, and language become more relevant.

At 16, after controlling for reading, language, verbal and non verbal abilities, Weber Fraction maintained a significant relationship only with speed of processing, mathematics, and visuo-spatial short term memory (although this last association was not significant in the replication sample). This pattern of correlations may indicate a mediation of speed of processing in the relationship between mathematics and Weber Fraction. This is also suggested by the fact that the association with teacher rated mathematics was overall weaker than with the web-assessed mathematics, which included the component of fluency - the ease with which numerical material and strategies are retrieved from memory. Previous research suggests that speed of processing mediates the relationship between mathematic and memory in: (1) the creation of automaticity for the basic arithmetic that arise when practicing numerical material and (2) in retrieval for mathematics (Geary & Wiley, 1991; Hitch and McAuley, 1991; Bull and Johnston, 1997). It is possible that non-symbolic estimation skills are involved in acquiring automaticity in basic arithmetic but no longer necessary to mathematical acquisition once such automaticity has been achieved; this is supported by the association found between Weber Fraction and mathematics at 7, when children start to learn the symbols and rule of mathematics. This explanation reconciles with findings of a recent study which found that in adults, higher mathematical skills were associated with faster or more automatic access to non-symbolic numerosity information rather that with the precision of non-symbolic discrimination. (Nys & Content, 2012).

There is to consider that in addition to speed of processing, the Reaction Time task may have assessed processes involving number and key identification and number-key mapping. In Bull and Johnston (1997), speed of processing was measured with a task of simple motor skill and another that used numerical stimuli. Children with poor mathematical abilities performed poorly in both tasks (compare to normal achieving children) suggesting that lower speed is a feature in poor mathematical ability irrespectively of the stimuli used. In this study we did not have a task measuring pure perceptual speed, however the element of speed was captured by the Factor Analysis. In a three dimension construct Reaction Time and Weber Fraction loaded on a single factor. If Reaction Time would have measured some numerical or mapping processing, we would have seen a cross-loading of Reaction Time on Factor 1 (non-verbal ability factor) even in a 3 dimension construct. Weber Fraction on the other hand cross-loaded in the non-verbal dimension as the task required both speed and non-symbolic numerical processing.

Weber Fraction was also associated with reading and non-verbal abilities at some ages. For reading, this association was true when word recognition/decoding was assessed (age 7, 12), but not when only comprehension was assessed (age 10). Word recognition relies on pattern recognition processing (so do the non-verbal tests in this study: Raven, Puzzle and Shapes), and it is possible that the same mechanism may be involved in the estimating of numerosities. The process of pattern recognition may be at the core of the reading-mathematics-estimation association. Previous research suggests, that word recognition deficits in children are associated with problems in retrieving numerical material from memory (e.g. Geary, 1993). It is possible that pattern processing is at the core of the development of the correspondence between numerosity of a set and its symbolic representation of number, underlying mathematical learning (Gelman & Galistel, 1978; Butterworth, 2005).

5.5.2 Number Line

Individual differences in Number Line accuracy were significantly related to teacher-rated and web-assessed mathematics at all ages, even after controlling for general cognitive abilities. However, after controlling for mathematics at 16, early mathematics was not associated with Number Line scores. This could indicate that the process of symbolic estimation is dynamic, where more efficient numerical representations take place with more advanced mathematics. Therefore the earlier mathematics did add any variance to the Number Line scores to the variance already explained by the contemporaneous - more updated number representation. This explanation fits with previous findings reporting improvements in Number Line tasks after training (e.g. Siegler & Booth, 2004; Siegler & Mu, 2008). These results also suggest that, contrarily to non-symbolic estimation skills, being able to mentally represent numbers in their symbolic system is a process continuously required in mathematical acquisition. The directionality of this relationship remains still unclear.

The nature of the association of Number Line and mathematics seemed to be specific. Out of 7 cognitive measures, assessed at age 16, Number Line estimation was significantly associated only with mathematics. Similarly, when all cognitive measures at earlier ages were entered in the regressions as predictors of Number Line scores, only mathematics (both teacher and web assessed) was significantly associated with Number Line in both samples.

In line with previous findings (Geary et al., 2008), we did not find a relationship between Number Line estimation and speed of processing at 16. However, some weak correlations were also observed between Number Line and spatial abilities, reading, verbal abilities, and non-verbal abilities. These results suggest that, representation of numerical magnitude is part of a general cognitive mechanism captured by number line tasks. As argued by Geary et al. (2008) a certain degree of intelligence is required in order to organise numbers in a logical structure (e.g. a mental number line) that facilitate mathematical learning. These results are consistent with previous research which shows that children with relatively low IQ have impaired representation of numerical magnitudes on number line (Bachot, Gevers, Fias, Roeyers, 2005).

In summary, the results of these analyses suggest that symbolic and non-symbolic estimation may represent distinct processes. These distinct processes do not belong to the same construct just because they correlate with mathematics and because involve the processing of numerical information of different type. With these results in mind, the use of "number sense" referring to estimation as a global process, can be misleading. Further, the relationship with mathematics and the two processes seems quite different; non-symbolic estimation processes may bootstrap mathematics just at some stages of mathematical development, while symbolic estimation seems a pervasive process in mathematics.

5.5.3 Sex Differences

Our large sample was well suited to explore sex differences in number sense. No sex differences were observed in either Weber Fraction or Number Line estimation. Previous studies that have reported sex differences on number line tasks have used samples of young children (between 4.5 and 9 years; LeFevre et al., 2010; Thomson & Opfer, 2008). These studies have also suggested a decrease of sex differences with development. It is possible that the absence of sex differences in symbolic estimation of numerical magnitudes at the age 16 is the product of the development of magnitude estimation skills. Conversely, consistently with previous literature (e.g. Kovas et al., 2007a), we found small, but significant sex differences in mathematics at age 16. These results are in favour of the theory suggesting that cognitive mechanisms providing the basis for mathematical learning are the same for boys and girls (Spelke, 2005). As suggested by Spelke, perhaps the roots of sex differences in mathematics are to be searched beyond cognitive abilities or biological factors.

5.5.4 Conclusion

The results of this study suggest that estimation of numerosities (Weber Fraction) and estimation of numerical magnitude (Number Line) rely on independent mechanisms to a large extent. We propose that "number sense" should not be used as an umbrella term, as different measures seem to be characterised by different processes and differentially relate to other cognitive traits. In the presence of the Number Line task, which showed a consistent strong association with mathematical ability, Weber Fraction task was overall not related to mathematics. Contrary to the suggestions in the literature of the importance of the number sense for mathematical development, both measures explained only a small fraction of variation in mathematical ability at all ages.

Chapter 6: How heritable is number sense? A genetic investigation into numerosity estimation abilities in 16 year-old students*

6.1 Abstract

Basic abilities of quantity and numerosity estimation have been detected across animal species. Such skills are referred to as "number sense". For human species, individual differences in number sense are detectable early in life, and persist in later development. The origins of these individual differences are unknown. To address this question, we conducted the first large-scale genetically sensitive investigation of number sense, assessing estimation abilities in 1,241 pairs of monozygotic and 2,348 pairs of dizygotic 16-year-old twin pairs. Univariate genetic analysis of the twin data revealed that number sense is only modestly heritable (32%), with individual differences being largely explained by non-shared environmental influences (68%), with no contribution from shared environmental factors. Sex-limitation model fitting revealed the same aetiology individual differences in estimation skills in males and females. We also carried out the *Genome-wide Complex Trait Analysis* (GCTA) that estimates the population variance explained by DNA differences among unrelated individuals. For 1118 unrelated individuals in our sample with genotyping information on 1.7 million DNA markers, GCTA-estimated non significant, zero heritability of number sense , unlike results for other cognitive abilities from this study where the GCTA heritability estimates were about 25%. Because directional selection reduces genetic variance, the low heritability of number sense suggests the evolutionary importance of this trait.

* This Chapter has been adapted from Tosto et al. (in preparation b)

6.2 Introduction

Numbers, in their symbolic notation, form a basic tally system to answer the questions of "how much" or "how many". Numerals are an efficient way to keep track of discrete quantities and numerosities. This is particularly useful if the numerosities to be represented are relatively large. An alternative way to represent quantities and numerosities is to evaluate them in terms of "more" or "less"; this approach does not require the use of symbols or any learned system and is based on approximation. The mechanism supporting such approximations, the approximate number system, is also often referred to as "number sense" (see Dehaene, 1997 for a review). The exact definition and measurement of number sense are often debated (see Berch, 2005). This paper will refer to number sense as the skill allowing us to represent, estimate and manipulate non-symbolic quantities/numerosities. A practical example of when we use our number sense is when, without counting and after a quick glance, we join the queue with the least people.

One of the theories underlying mathematical learning is that numeracy skills partially originate from the non-symbolic/numerosity ability interfacing with the taught numerical system (e.g. Dehaene, 1997; Feigenson, Dehaene, & Spelke, 2004; Izard Pica, Spelke, & Dehaene, 2008). For example, it has been proposed, that deficits in manipulating numerosity are one of the signatures of mathematical difficulties (Butterworth, 1999; 2010; Landerl et al., 2004; Mazzocco, Feigenson & Halberda, 2011a). There is evidence that the symbolic and non-symbolic number systems contribute interactively to the development of normal arithmetic skills. For example, the native language of a small Amazonian tribe, the Mundurukú , has words for numbers only up to five. Although Mundurukú participants can approximate quantities well above their naming range, they fail to manipulate exact numbers. This indicates that the approximate number system is independent from the verbal encoding of numbers that produces exact numerical representations. Further, if the non-symbolic quantities fail to map onto their symbolic correspondence, the emergence of exact arithmetic may not typically develop (Pica et al., 2004).

Some studies, however, challenge the view of a significant relationship between symbolic and non-symbolic representation of numbers. In one study, mathematical achievement in 6- to 7-year-old children correlated with Numerical Distance Effect (speed and accuracy in number comparison are greater when the numerical distance separating two numbers is relatively large, i.e. 3 and 9 vs 3 and 5) in symbolic, but not in non-symbolic comparisons (Holloway & Ansari, 2009). Similarly, Rousselle & Noël (2007) found that children with mathematical disabilities show impairments in comparisons of number symbols, but not in the processing of non-symbolic numerical magnitudes.

The approximate number system is not unique to humans; many animal species can approximate numerosities and to some extent can remember discrete number of objects and events. Basic numerical competences have been reported in ants (Reznikova & Ryabko, 2011). Rats can distinguish between arrays with different numbers of auditory signals (Meck & Church, 1983); mosquito fish discriminate quantities using numerical cues and can be trained to recognise a set of two items from another with three (Agrillo, Dadda, Serena, & Bisazza, 2009; Agrillo, Piffer, & Bisazza, 2011). In addition to estimation abilities, rudimentary arithmetic skills performed on numerosity sets (i.e. collection of discrete items) have been reported by studies that used attachment paradigms with newborn chicks (Rugani et al., 2009; 2011).

Animal evidence suggests that basic numerical competences are independent from language and are present at birth. Studies of human infants also show that this ability is preverbal. Using habituation paradigms it has been shown that babies as old as 6 months are able to distinguish between arrays of items or sequences of sounds of 4 from 8, and 8 from 16 (ratio 1:2) (Lipton & Spelke, 2003; Xu & Spelke, 2000). Older babies can discriminate between finer ratios. At 9 months for example, babies can discriminate between displays of 8 and 12 items (ratio 2:3) (Lipton & Spelke, 2003) and between the age of 3 and 6 years, children can distinguish between ratios of 3:4 and 5:6 (Halberda & Feigenson, 2008). In adulthood, estimation skills peak, allowing to discriminate between arrays with ratios of 9:10 (Pica, Lemer, Izard, Dehaene, 2004; Halberda, Mazzocco, & Feigenson, 2008). Such evidence from animal and infant studies indicates that basic estimation skills involved in number sense are unlearned and present across species, which suggests that number sense may have been evolutionarily preserved. However, if number sense has evolved, this does not imply that the origins of individual differences in number sense are genetic in origin. An example from another field is attachment which appears to have evolved in mammalian species yet individual differences in attachment in the human species show little genetic influence (Plomin, DeFries, Knopik, & Neiderhiser, 2013).

One fundamental parameter in estimation skills, used to assess an individual's number sense acuity, is the ratio of the items in the arrays that are being compared. Discrimination of numerosities in animals, infant humans and adult humans follows the Weber Law (Nieder & Miller, 2003; 2004; Libertus & Brannon, 2009; 2010; Pica et al. 2004; Halberda et al., 2008). The Weber Law (Weber, 1834) describes the relationship between the magnitude of the stimulus appraised and the ability to detect "the just noticeable change" in such magnitude. Judging whether a set has more items than another is difficult when the discrepancy between the two displays is small. According to the Weber Law, the threshold of the minimum difference that can be detected is equal to the difference between the numbers of items in the two sets (the increment in quantity) divided by the number of items in the smallest of the two sets. This threshold is indexed by the Weber Fraction. In a practical example, if one can tell, without counting, which is the larger set between a display with 5 items and one with 7 (ratio 5:7), the Weber Fraction associated to the number sense acuity for that person is 0.4, [(7-5)/5]. As previously mentioned, estimation abilities improve with development, therefore Weber Fraction values have been measured in the range from 1, on performance at 6 months, to 0.11 in adulthood.

Individual differences in the ability to approximate and compare numerosities emerge early in life. At 6 months, infants already show stable individual differences in numerical discrimination (Libertus & Brannon, 2010). Individual differences in the ability to detect changes in numerosity at 6 months have been shown to predict this ability at 9 months beyond short-term memory skills. Individual differences in estimation abilities were also detected in 3- to 4 year-olds, as well as, 6, 14 and 16-year-olds (Halberda & Feigenson, 2008; Mazzocco et al., 2011b; Halberda et al. 2008;

study in Chapter 5). Investigation of number sense have been carried out mainly on young children, however, a recent study surveyed number sense in over 10,000 individuals between 11 and 85 years old (Halberda, Ly, Wilmer, Naiman & Germine, 2012). The study reported individual differences and developmental changes in non-symbolic estimation skills, identifying three main transitional age-related trends in the population: a rapid increase in estimation accuracy between the age of 11 and 16, a steady improvement up the age of ~30 and a decline from 30 to 85.

It is possible that individual differences in estimation abilities in children are driven by differences in the processing of perceptual characteristics of the stimuli rather than of the numerical information per se. Pre-school children have difficulties in ignoring continuous, non-numerical irrelevant information (e.g., the area occupied by the dots in display) in non-symbolic numerical comparisons (Rousselle & Noël, 2008). For example, when the perceptual information was in conflict with the numerical information (e.g., when arrays with more number of dots had smaller physical dot size than the array with less dots), 4 year-olds were unable to discriminate between numerosities independently from the physical appearance of the stimuli (Soltesz, et al., 2010). Adults also seem to automatically process irrelevant non-numerical information (the area occupied by the dots for example) in numerosity discrimination, (e.g. Gebuis, Cohen Kadosh, de Haan, & Henik, 2009; Barth, La Mont, Lipton, Dehaene, Kanwisher, & Spelke, 2006). However, in adulthood the numerical information is as salient as the non-numerical (area) information, allowing to respond to numerosity (discrete) rather than continuous properties of the stimulus (Nys & Content, 2012).

Whether individual differences in the processing of numerosity stem from perceptual processing of continuous or discrete information, accuracy in a simple task of judging which of two arrays has more items, has been associated with mathematical abilities (e.g. Harberda et al., 2008; Mazzocco et al., 2011b; Nys & Content, 2012, study in Chapter 5). Further, this association has been shown to persist across the life span (Halberda et al., 2012). Understanding the aetiology of individual differences in number sense skills can contribute to our understanding of processes involved in mathematical learning.

6.2.1 Research question

The present study is the first large scale genetic investigation into the aetiology of individual differences in number sense. We assessed number sense in 16-year-old twins and conducted univariate genetic analysis in order to estimate the relative contribution of genetic and environmental factors to variation in number sense. The large representative sample, which included both same-sex and opposite-sex twin pairs, also allowed the investigation of any sex differences in the aetiology of the variation in number sense. We hypothesised, that the ability to judge more from less might be a product of "directional" evolutionary selection due to importance of this ability for survival (for example, increasing success in foraging) ,Such "directional natural selection" should reduce individual variability through de-selection of "unhelpful" genetic variants, leading to reduced trait-relevant genetic population variation (Plomin, DeFries, McClearn, & McGuffin, 2008). Therefore, we did not expect a large genetic contribution to individual differences in number sense.

In addition to estimating heritability of number sense using the twin method, we used the Genome-wide Complex Trait Analysis (GCTA), to estimate heritability directly from DNA, using 1.7 million DNA markers genotyped in 1118 unrelated individuals in our sample.

6.3 Methods

6.3.1 Participants and procedure

Participants of this study were the twins of TEDS (Twin Early Developments Study) of the first and second cohort. These were the twins born between January 1994 and August 1995 tested when they were 16 years old. Detailed description of the sample and procedure of recruitment can be found in section 2.2.4.6. For the purpose of this study, twins with specific medical problems and whose English was not the first language were excluded from the analyses. The final sample consisted of

4,518 twins [2,259 pairs of which 838 monozygotic (MZ), 733 dizygotic same-sex (DZss) and 689 dizygotic opposite-sex (DZos)]. The mean age for the sample was 16.6 (SD = .28)

6.3.2 Measures

The number sense measure used in this study was the Weber Fraction. Scores for this variable were derived variable from the web-based Dot Task - assessing discrimination of large numerosity, Full description of the Dot task and Weber Fraction scores is found in section 4.3.3.2.

6.3.3 Analyses

Standard quantitative methodologies were used to estimate genetic and environmental contribution on individual differences in number sense (Plomin et al., 2013). The twin method used here is described in sections 3.3.1 and 3.3.2; the univariate genetic analysis and model fitting are described in section. 3.3.3.

6.3.3.1 Sex-limitation model

Sex limitation model fitting was used to investigate the aetiology of gender differences in number sense. Sex-limitation models are an extension of the basic univariate model. They are used to uncover differences between males and females A, C, E parameters (Neale & Maes, 2003). Quantitative sex differences refer to differences in the magnitude of the A, C, E estimates of males and females. Qualitative sex differences rest on comparisons between same-sex and opposite-sex DZ twins, which indicates the extent to which the same genetic or environmental factors affect individual differences for males and females. It should be noted that quantitative and qualitative sex differences in the aetiology of individual differences are unrelated to any observed mean gender difference. Scalar differences refer to differences in variance of the measured trait between males and females.

The model relies on comparison of DZos and DZss genetic relatedness coefficient (r*g*). This genetic relatedness is assumed to be 0.5 because DZ twin share on average 50% of their segregating genes. If different genes affect males and females the genetic relatedness of the DZos will be less than 0.5, this is the case of qualitative differences. If sex differences are quantitative, the same genetic factors influence males and females, therefore the r*g* in the DZos will be 0.5, but the A, C, E estimates for males and females will be significantly different. The same logic applies to the coefficient indicating relatedness due to shared environmental factors (r*c*), this coeefficient should be equal to 1 as twins in the same family share the same environments. It is not possible to estimate r*g* and r*c* at the same time (the model is not statistically defined), so qualitative and quantitative differences in genetic influences have to be modeled separately from shared environmental influences.

6.3.3.2 Sex-limitation model fitting

The ACE parameters and their 95% confidence intervals are estimated separately for males and females in a full sex-limitation model. The data is entered dividing the groups in MZ male (MZm), MZ female (MZf), DZ male (DZm), DZ female (DZF), and DZ opposite-sex (DZos) twin pairs. In this model the *rg* (or *rc*) are also free to be estimated. Subsequently, three nested models are compared to the full model in order to determine which model best describe the observed data.

In the Common Effects Sex-Limitation model, the A, C, E parameters are estimated separately for males and females but the r*g* of the DZos, is constrained to 0.5. The model allows for quantitative and variance differences to occur. If the fit of this model is not worse compared to the full model, quantitative but not qualitative differences between males and females would be indicated. The goodness of fit is evaluated following the same criteria described in the univariate model fitting (section 3.3.3). The Scalar Effects Sex-Limitation model tests for variance differences between males and females. The A, C, E parameters of the females are constrained to be the same as males parameters, while r*g* is constrained to 0.5 in the DZos. The model allows for phenotypic variance by constraining the variance of one sex to be a multiple scalar of the other sex. If this model is not a worse fit compared to the full model, this indicated differences in variances between males and females. In the Null model, all the parameters are constrained to be the same for males and females, thus testing the null hypothesis: non-worse fit of this compared to the full model would indicate no aetiological or phenotypic variance differences between males and females.

6.3.3.3 Genome-wide Complex Trait Analysis (GCTA) Genome-wide Complex Trait

Analysis (GCTA), can be used to estimate genetic variance accounted for by all the SNPs genotyped in samples free from constraints of family relationships such as twins or relatives. (Lee, Wray, Goddard, & Visscher, 2011; Yang, Lee, Goddard, & Visscher, 2011a; Yang, Manolio et al., 2011b). GCTA requires large samples in which each individual has been genotyped for hundreds of thousands of DNA markers, typically SNPs. Large samples and extensive genotyping are also needed in genome-wide association studies (GWAS), thus data collected in GWAS are suitable to conduct GCTA analyses. GWA genotyping data of the TEDS sample has been used to conduct the first GWAS of general cognitive abilities, mathematics and reading (e.g. Docherty et al., 2010; Haworth et al., 2007) and the first GCTA studies of cognitive abilities at the age of 12, estimating heritability between 20% to 35% for diverse cognitive abilities (Plomin et al., 2012). GCTA has been used to estimate heritability as captured by genotyping arrays for height (Yang et al., 2010), weight (Yang, et al., 2011b), psychiatric and other medical disorders (Lee et al., 2011, 2012; Lubke et al., 2012), and personality (Vinkhuyzen et al., 2012a). GCTA has also been applied to general cognitive ability in studies of adults (Chabris et al., in press; Davies et al., 2011) and children (Deary et al., 2012).

At age 16, GWA genotyping TEDS data was available on around 1,000 individuals. Because GWA analysis needs to correct for multiple testing of hundreds of thousands of genotyping tests, this data is not suitable for GWAS but can be used in GCTA analyses to estimate of genetic influence as a check on the heritability estimate based on the twin method.

GWAS design associate a SNP with a trait using statistical tests, GCTA instead, uses chance similarity across hundreds of thousands of SNPs to predict phenotypic similarity pair by pair in a large sample of unrelated individuals. The essence of GCTA is to estimate genetic influence on a trait by predicting phenotypic similarity for each pair of individuals in the sample from their total SNP similarity. In contrast to the twin method, which estimates heritability by comparing phenotypic similarity of identical and fraternal twin pairs whose genetic similarity is roughly 100 percent and 50 percent respectively, GCTA relies on comparisons of pairs of individuals whose genetic similarity varies from 0 to 2 percent. GCTA extracts this tiny genetic signal from the noise of hundreds of thousands of DNA markers (single nucleotide polymorphisms, SNPs) using the massive information available from a large sample of individuals, each compared pair by pair with every other individual in the sample.

GCTA genetic similarity is not limited to the genotyped SNPs themselves but also includes unknown causal variants to the extent that they are correlated with the SNPs. Mendel's second law of inheritance is that genes (as we now call them) are inherited independently (now called linkage equilibrium), but Mendel did not know that genes can be on the same chromosome, in which case they are not inherited independently (linkage disequilibrium). This violation of Mendel's second law is complicated by the fact that during meiosis, chromosomes from the mother and father recombine on average once per meiosis, which means that, in the population, genes on the same chromosome are separated by this process of recombination to the extent that the genes are not close together on the chromosome. GCTA provides a lower-limit estimate of heritability because it misses genetic influence due to causal variants that are not highly correlated with the common SNPs on genotyping arrays.

The genetic effect on a trait may not just derive from the simple sum of independent genetic actions, they may stem from more complex gene-gene interactions. One of the assumptions of the twin method is that the variance explained by genetic influences is attributed to additive genetic effects. In practice, the method captures both additive and non-additive genetic effects because the DNA sequence of identical twins is virtually identical and thus they share all genetic effects including non-additive effects (see Plomin et al. 2012, for details). Conversely GCTA adds up the effect of each SNP, therefore it does not include gene-gene interaction effects; this is why the method provides lower-limit estimates of heritability.

Genotyping on the Affymetrix 6.0 GeneChip and subsequent quality control was carried out as part of the WTCCC2 project (The UK IBD Genetics Consortium & the Wellcome Trust Case Control Consortium, 2009) for 1118 individuals (one member of a twin pair) for whom number sense data at age 16 were also available. In addition to nearly 700,000 genotyped single-nucleotide polymorphisms (SNPs), more than one million other SNPs were imputed using IMPUTE v.2 software (Howie, Donnelly, & Marchini, 2009). GCTA estimates were obtained using the GCTA software package (Yang et al., 2011a). In GCTA, any pairs whose genetic similarity exceeded +/- 0.025 (i.e., greater genetic relatedness than fourth-degree relatives) are removed so that genetic similarity is random and can be treated in a random effects model. By this criterion, no individuals were excluded.

6.4 Results

The analyses were conducted using Weber Fraction and accuracy scores on the Dot Task. Prior to analyses, accuracy scores were squared and a square root transformation was applied to Weber Fraction scores. The variables were then standardised (mean of zero and standard deviation of one); corrected for age and sex; and scores outside +/- 3 standard deviations were considered as outliers and excluded. The transformation improved normality of both variables. However, even after transformation, Weber Fraction scores did not fully meet assumptions of normality as skewness was 1.09 (SE = .05) and kurtosis 1.27 (SE = .10). Maximum Likelihood estimation provides efficient parameter estimates under the assumption of normality of the data, but the method has also been shown to be reliable when assumptions of normality are violated (c.f. Boomsma & Hoogland, 2001). We therefore report also the results of the analyses conducted on the accuracy as on this variable skweness was -.43 (SE = .05) and kurtosis -.02 (SE = .10). The Number Sense measure (accuracy scores) showed good internal validity (alpha = .79) and test-retest (.62) (see Table 4.6 in section 4.5.2.1 for internal validity and test re-test correlations). Descriptives of the data collected on the TEDS sample are also consistent with estimation abilities reported for 16 year-olds in Halberda et al. (2012) (Figure 6.1).

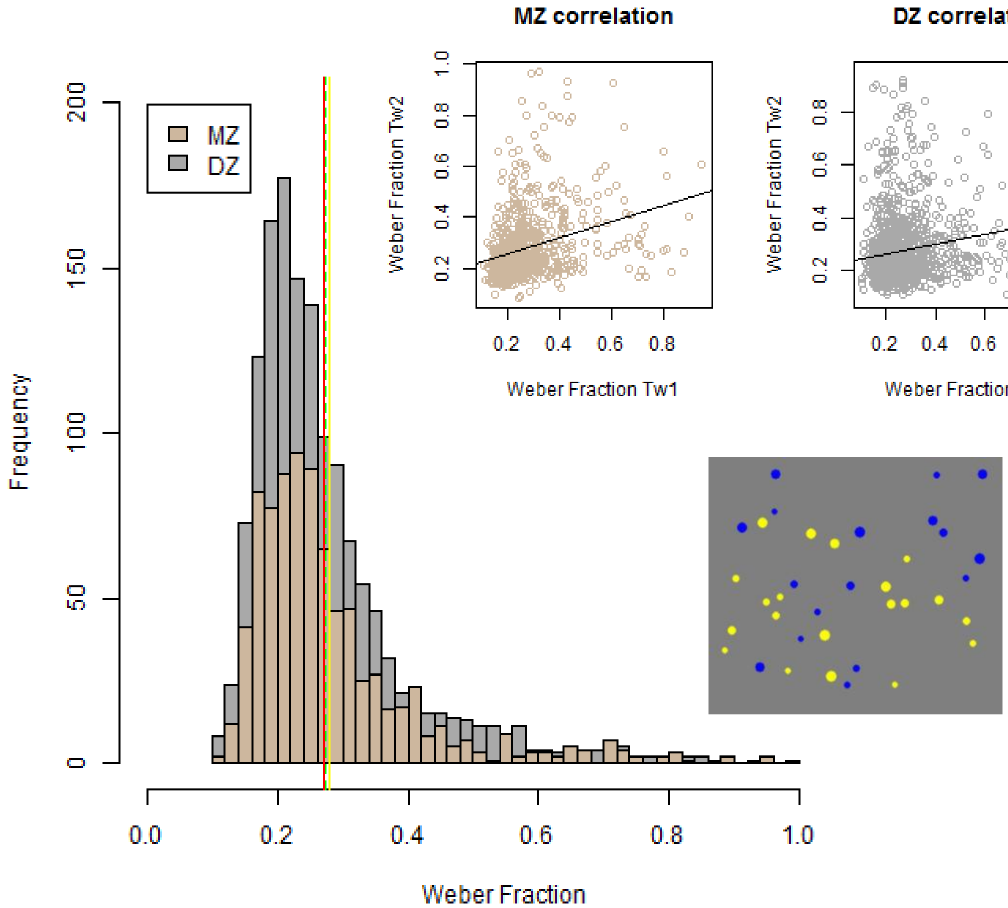


*Figure 6.1: S*catter plots correlations MZ (Monozygotic, in brown) and DZ (Dizygotic, in grey) twins with their co-twins in Weber Fraction. The Weber Fraction scores were derived from accuracy in the Dot Task. The display of yellow and blue dots is an example of a test trial. The twins had to judge whether there were more yellow or blue dot following an exposure of 400 milliseconds.
The overlapping distributions of the Weber Fraction scores of the MZ (brown) and DZ (grey) show the means: MZ= .28 (yellow line); DZ= .27 (red line). These are compared with the 16-year olds means reported in Halberda et al. (2012) = .275 (green dashed line).

Table 6.1 shows means, standard deviations and ANOVA result by sex and zygosity for Number Sense accuracy and Weber Fraction scores. One twin was chosen at random from each pair (N = 2,472). Mean accuracy score on the Number Sense accuracy was 115.82 (SD = 9.57; range = 79 – 140, out of a possible 150). Mean Weber Fraction score was 0.28 (SD = .13; range .10 - .99). No significant mean sex differences were found, nor were there zygosity differences. Descriptive analyses run on the second half of the sample yielded highly similar results (available from the authors).

Table 6.1

Means, Standard Deviations and ANOVA results by sex and zygosity

																							ANOVA						
	All		*MZ*		*DZ*		*Female*		*Male*		*MZm*		*DZm*		*MZf*		*DZf*		*DZo*		*DZss*		*Zyg.*		*Sex*		*Zyg.*Sex*		*Tot.*
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	η^2	*p*	η^2	*p*	η^2	R^2
Number Sense accuracy	*N=2258*		*N=836*		*N=1422*		*N=1315*		*N=943*		*N=317*		*N=626*		*N=519*		*N=796*		*N=689*		*N=733*								
	.03	.98	.00	.99	.04	.98	.05	.95	.00	1.0	-.07	1.1	.04	1.0	.04	.93	.05	.97	.04	.99	.05	.98	.17	.00	.16	.00	.29	.00	.000
	N=2214		*N=817*		*N=1397*		*N=1298*		*N=916*		*N=301*		*N=615*		*N=516*		*N=782*		*N=677*		*N=720*								
Weber Fraction	-.10	.85	-.07	.85	-.11	.84	-.10	.83	-.10	.87	-.08	.86	-.10	.88	-.07	.85	-.12	.82	-.10	.85	-.12	.83	.34	.00	.95	.00	.79	.00	.001

Number Sense accuracy = accuracy scores on Dot Task - variable squared transformed; Weber Fraction = Weber Fraction score variable square root transformed; M = mean; SD = standard deviation; MZ = monzygotic twins; DZ = dizygotic twins; MZm = monzygotic males; MZf = monzygotic females; DZo = dizygotic opposite sex; DZss = dizygotic same sex; p = p-value associated with the effect size of sex, zygosity and the interaction of the two on the means of all groups; η^2 = magnitude of the effect of sex, zygosity and the interaction of the two on the means of all groups; R^2 = proportion of variance explained by sex and zygosity in Number Sense variables; N = number of the twins: one randomly selected from each pair.

Table 6.2 shows the intraclass correlations (indexing the similarity of co-twins) of Number Sense accuracy and Weber Fraction scores with 95% confidence intervals. The most striking finding is that, despite good validity of the measure, the intraclass correlations for number sense were modest, even for MZ twins, suggesting that twins differ markedly in their number sense ability and pointing to a significant contribution of non-shared environmental influences. Nonetheless, MZ twin correlations were greater than DZ correlations, suggesting the presence of genetic influences on number sense as well.

Table 6.2

Intraclass correlations on number sense variables and model fitting parameters estimates

	r MZ	*N*	*r DZ*	*N*	**Variance of A (95%CI)**	**Variance of C (95%CI)**	**Variance of E (95%CI)**
Number Sense acc.	0.35	730	0.18	1175	0.35	N/A	0.65
95% CI	*(.28 - .41)*		*(.13 - .24)*		*(.30 - .41)*		*(.59 - .70)*
Weber Fraction	0.31	700	0.15	1140	0.32	N/A	0.68
95% CI	*(.24 - .38)*		*(.09 - .20)*		*(.26- .37)*		*(.63 - .74)*

rMZ = intraclass correlation for monozygotic twins; rDZ = intraclass correlation for dizygotic twins; N = number of complete pairs; Variance components A,C,E = estimates respectively of genetic influences, shared environment, non shared environment . 95% CI = 95% Confidence Intervals. The best fitting model did not include estimates for shared environment.

Indeed, the result of the univariate genetic analyses (Table 6.3) showed that individual differences in Weber Fraction scores were explained by modest genetic influences (.32) and largely by non-shared environmental influences (.68). Shared environmental influences were non-significant. The parameter estimates for the accuracy scores were similar to the estimates of the Weber Fraction scores, .35 and .65 respectively for genetic and non-shared environmental influences. The most parsimonious models explaining the observed data were the AE models, as indicated in Table 6.3.

Table 6.3

Model Fitting univariate genetic analysis: model-fit statistics

Measure	Model	Model Fit						
		-2LL	df	(Δ -2LL)	AIC	(Δ-AIC)	BIC	p-value
Number Sense accuracy	*Saturated*	-12649.05	4505		3639.05		-12009.57	-
	ACE	-12658.33	4511	-9.28	3636.33	2.72	-12029.34	.10
	AE	-12658.58	4512	-.25	3634.58	1.75	-12033.29	.60
	E	-12791.82	4513	-133.49	3765.82	-129.49	-11970.74	.00
Weber Fraction	*Saturated*	-11170.54	4415		2340.54		-12382.55	-
	ACE	-11185.37	4421	-14.83	2343.37	-2.83	-12399.55	.02
	AE	-11185.37	4422	.00	2341.37	2.00	-12403.62	1.00
	E	-11282.29	4423	-96.92	2436.29	-92.92	-12359.23	.00

-2LL = minus log-likelihood; df = degrees of freedom; Δ-2LL = difference in likelihood; AIC = Akaike's Information Criterion; Δ-AIC = difference in AIC, this is calculated between the Saturated and full ACE model, and between the full ACE model and the AE, E nested models. BIC = Bayesian Information Criterion; p-value = associated with the differences in likelihood ratio between the Saturated and the full ACE model, and between the full ACE model and the AE, E nested models. Ep = parameters estimated.
The p-value shows no significant differences in likelihood between the Saturated and the full ACE model for accuracy in the Number Sense task. AIC shows good fit of the ACE model compared to the Saturated model in Number Sense scores (lower AIC of full ACE). The same parameter shows the better fit of the AE model. The goodness of fit for the Weber Fraction model is demonstrated to a lesser extent by the AIC. The BIC however shows a good fit of the full ACE model to the observed data and confirms the best fit of the AE model for the Weber fraction variables.

The results of the sex-limitation model fitting are shown in Table 6.4. In summary, no quantitative or qualitative sex differences were found for Number Sense accuracy or Weber Fraction. Genetic and environmental influences were estimated separately for males and females by fitting a Full Sex-Limitation model. The parameters for the accuracy and the Weber Fraction scores with their 95% confidence intervals are also shown in Table 6.4. The models testing for qualitative and quantitative differences in both number sense variables (respectively the Common Effects and Scalar Effects models), did not differ significantly from the Full Sex-Limitation model. The AIC and BIC parameters confirmed that the best fit was provided by the Null Model, indicating that there are no qualitative or quantitative differences in the aetiology of number sense between males and females.

Table 6.4
Sex-Limitation Model Fitting for Weber Fraction and Number Sense Accuracy scores.

		Sex Limitation Model Fit						Parameter estimates from the sex-limitation model fitting, for males and females		
								Variance of A (95%CI) *Males*	**Variance of C (95%CI)** *Males*	**Variance of E (95%CI)** *Males*
Measure	**Model**	**-2LL**	**df**	**(Δ-2LL)**	**p-value**	**AIC**	**BIC**	***Females***	***Females***	***Females***
Number Sense accuracy	Full sex-limitation model	10791.89	3839	-	-	3113.89	-9898.58			
	Common.Eff. (Qualit. diff.)	10793.7	3840	1.81	.18	3113.70	-9901.66	.35 (.07 - .44)	.00 (.00 - .24)	.65 (.56 - .75)
	Scalar.Eff. (Quantit. diff.)	10792.21	3842	.32	1.00	3108.21	-9910.37	.34 (.13 - .41)	.00 (.00 - .17)	.66 (.59 - .74)
	Null Model	**10794.07**	**3843**	**2.18**	**0.70**	**3108.07**	**-9913.43**			
Weber Fraction	Full sex-limitation model	9533.15	3761	-	-	2011.15	-20434.4			
	Common.Eff. (Qualit. diff.)	9533.3	3762	.15	.70	2009.30	-20442.21	.34 (.13 - .43)	.00 (.00 - .16)	.67 (.56 - .77)
	Scalar.Eff. (Quantit. diff.)	9533.79	3764	.49	.89	2005.79	-20457.66	.29 (.06 - .37)	.00 (.00 - .19)	.71 (.63 - .80)
	Null Model	**9533.99**	**3765**	**.84**	**.88**	**2003.99**	**-20465.43**			

-2LL= minus log-likelihood; df = degrees of freedom; Δ-2LL = difference in likelihood; df = degrees of freedom; p-value = associated with the differences in likelihood ratio between each of the nested models and the full sex limitation model. AIC = Akaike's Information Criterion; BIC = Bayesian Information Criterion. The models testing for qualitative and quantitative differences show no significant difference in fit compared to the full model (p-values non significant). The Null Model shows no significant difference in fit with the full sex-limitation model suggesting no qualitative or quantitative differences, or variance differences between males and females in Number Sense accuracy and Weber Fraction.The bold characters indicate the best fitting model.

A,C,E = estimates respectively of genetic influences, shared environment, non shared environment . 95% CI = 95% Confidence Intervals. Estimates separate for males (above) and female (below) for the accuracy scores and the Weber Fraction scores.

Genome-wide Complex Trait Analysis (GCTA) yielded a non-significant estimate of zero heritability for number sense. Because the GCTA heritability estimate was so low and the sample size was relatively small for GCTA analysis, the standard error of estimate was large (0.29). Nonetheless, this GCTA analysis provides strong support for the relatively low twin study estimate of heritability for number sense.

6.5 Discussion

In this study we performed the first large-scale genetic analysis on number sense; we found that individual differences in this ability at 16, as indexed by the Weber Fraction, were largely influenced by non-shared environment (.68), with the rest of the variance explained by genetic influences (.32). This modest estimate of heritability from the twin study was underlined by a GCTA estimate of zero heritability. As number sense is an evolutionarily preserved trait (many animal species are equipped to deal with quantities and numerosities), the low heritability estimate may come as a surprise. However, heritability estimates refer to genetic influences on individual differences, which may be different from those affecting the presence of a trait in evolutionary terms. As discussed earlier, evolutionarily useful traits are not necessarily heritable, in fact there is evidence that they are to some extent learned. Acquisition, habituation and extinction of fear of snakes and spiders for example, are moderately heritable, with most of variance in individual differences explained by the environmental influences (Hettema, Annas, Neale, Kendler, & Fredrikson, 2003). In terms of genetic influences on such an evolutionarily preserved trait as number sense, one set of genes may provide a blueprint for the development of this ability across many species; whereas a different set of genes may contribute to variation in it within any population. Such "individual differences" genes may work through various mechanisms, affecting for example perceptual processes, speed of processing, and other functions relevant to perform estimation of numerosities.

Heritability estimates are time and population specific as the genetic influences captured by the parameters depend on how the genes express in the environmental condition at the time of the assessment. For this reason, we cannot extend the heritability of number sense at 16 to other ages or other cultures/populations. For example, reading abilities show consistent genetic and environmental estimates across ages and across populations (Byrne et al., 2005, 2006; Stromswold, 2001), while the heritability of "g" increases from early age to young adulthood (Davis et al., 2009; Haworth et al., 2010). Similarly, we cannot exclude the possibility of developmental changes in the heritability of number sense. It is possible that the marked individual differences in number sense acuity observed in infancy (Libertus & Brannon, 2010) may be under stronger influence of genes. This could explain why during infancy babies already show individual differences in estimation of numerosities. However, later in development, factors such as exposure to numerical stimuli, individual's interest and amount of practice in number-related activities, may all contribute to the development of this ability. With number sense becoming more and more precise, individual differences in this precision may be under stronger and stronger environmental influences. Since this study is the first large scale genetically sensitive investigation on number sense, further research needs to be conducted using longitudinal twin samples assessing aetiological change and continuity of influences on number sense. In addition, the strong non-shared environmental influences indicated in this study call for cross-cultural genetically-sensitive investigations to examine the relative contributions of genes and environments to the number sense in different cultures, where different educational, linguistic, and social practices are in operation.

One of the implications of the large environmental influence is that higher levels of accuracy in estimation of numerosity may be achieved through training. One study involving 6 month-old infants showed that when babies were simultaneously presented with a congruent visual and auditory stimulus they were able to discriminate numerosities with a ratio usually present in 9 month-old infants (Jordan, Suanda, Brannon, 2008). One explanation given by the authors was that the greater amount of numerical information received in two rather than one sensory modality increases infants' arousal leading to increased sensitivity to numerical changes. The number sense in animals seems also to be influenced by external cues in the same way as in humans. In one study, fish learned to discriminate numerosity faster if the numerical information was available in more than one sensory source, suggesting that multisensory numerical information facilitate discrimination learning (Agrillo, et al., 2011).

Although we need to understand in more depth the mechanisms through which the environment acts upon estimation skills, there are some studies showing how estimation of numerosity skills can be manipulated through exposure to numerical material. It has been suggested that playing numerical board games gives children familiarity about numbers and improves their estimation of numerical magnitudes (Siegler & Ramani, 2008). It is not clear why such influences should be non-shared by twins in the same family. It is possible that active and evocative gene-environment correlations, by which children choose specific activities or receive specific environmental inputs partly based on their genetic predispositions, play a role. Future studies should examine the similarity in twin and non-twin siblings in the willingness and frequency of engagement in the relevant activities – to evaluate whether they can explain some of the non-shared environmental influences on number sense development. Ultimately, understanding which environments influence number sense may help to understand how people learn mathematics.

Studies on artificial learning provide further evidence that individual differences in numerosity skills similar to number sense can emerge from differences in training. Neural network models can be modeled to detect numerosity from visual inputs (Domijan, 2004), with the quality of detection depending on the quality (e.g., frequency) of the inputs One study has shown that models not programmed a priori in numerosity recognition can learn to discriminate numerosities according to the Weber Law through "unsupervised learning" (Stoianov & Zorzi, 2012). The model in the study was also able to simulate response to numerosities similarly to the neurons in the areas responsible for numerosity representation (later intraparietal area, LIP) of the human (Santens et al., 2010) and monkey brain (Roitman, Brannon, Platt, 2007). As it is possible for models to develop different levels of number sense just by being exposed to different quality of visual stimuli, humans could develop differences in number sense through different exposure to numerical material - as opposed to genetic influences setting individual differences (programs in the case of the models).

Our results add a novel perspective on a current debate in the mathematical literature. One theory proposes that mathematical disability, more precisely Developmental Dyscalculia, emerges from difficulties in understanding numerosity concepts and dealing with numerosity. This occurs even in the absence of general cognitive impairments (Butterworth, 2005; Landerl et al., 2004). It has been suggested that this problem with basic numerosity manipulation may be genetic in origin (Butterworth, 2005). Indeed, although multivariate genetic research suggests that individual differences in mathematical ability and disability are largely influenced by the same genetic factors as those that affect other learning and cognitive traits, some unique genetic effects also exist (Kovas et al., 2007a). These unique genetic effects could be those shared between number sense and mathematics.

Evidence shows that, variation in number sense may also arise under the influences of general cognitive development (e.g. Geary et al., 2008, 2011; Swanson and Sachse-Lee, 2001, Cowan et al., in press). It is possible that children with poor reading or poor memory engage in less effective or insufficient numerical practices (e.g. less games with numerical content during pre-school age,), compared to children with non-impaired general abilities. In the long term, these discrepancies in numerical environments may lead to the poorer numerosity estimation in children with poor mathematical abilities. Alternatively, the same aetiological factors could affect both traits without any reciprocal contributions between the two. Under these circumstances and with the large environmental estimates found in this study for estimation ability, it is possible that this skill may be a product, rather than a cause of mathematical variation.

As previously mentioned, earlier quantitative genetic investigations have found no sex differences in the aetiology of different aspects of mathematical abilities, disabilities, or high abilities. This indicates that same genetic and environmental factors affect individual differences in mathematics equally in males and females (e.g. Kovas et al., 2007a, 2007b; Markowitz, Willemsen, Trumbetta, van Beijsterveldt, & Boomsma, 2005; Petrill, Kovas, Hart, Thompson, Plomin, 2009). For number sense, no mean sex differences have been reported, with the exception of one study that reported marginal male advantage in estimation abilities in 4 year-old children (Soltesz et al., 2010). In our study we did not find any mean sex differences in number sense. This result suggests that any observed average sex differences in mathematics are not mediated by estimation of numerosity skills.

As mentioned earlier, the aetiology of average differences may be independent of the aetiology of the variation. The present study was the first to examine whether the aetiology of individual differences in number sense was the same for males and females. We found no quantitative or qualitative differences between the two sexes. In other words, factors that make males differ from one another in number sense are the same that make females differ from one another.

6.5.1 Conclusion

The two methods employed in this study, the twin method and the GCTA analysis, showed that individual differences in numerosity estimation skill are modestly influenced by genetic factors. In fact, the heritability of number sense in this study seems significantly lower than that of other cognitive abilities in this sample at the same age (study in Chapter 7). Similarly to many other naturally selected traits the dis-advantageous "genes" may have been successfully de-selected in population, leading to reduced within trait variability due to genetic influences. The same genetic factors influence estimation skills in males and females equally. In order to take advantage of the positive relationship between number sense and mathematics, our efforts should proceed in identifying the successful environments contributing to individual differences in number sense.

Chapter 7: The myth of maths: is there anything 'special' about it?*

7.1 Abstract

Symbolic (Arabic numbers) and non symbolic (arrays of dots) estimation tasks have been recently used as predictors of mathematical achievement. These tasks are thought to tap into a cognitive mechanism known as number sense. This study employed a web-based task of symbolic estimation (Number Line) and a non-symbolic one (Dot Task to assess number sense in a sample of 7,598 twins (3,799 pairs) of the TEDS sample at the age of 16. The twins' mathematical abilities were measured with two web-based tasks assessing fluency and mathematical skills according to the UK National Curriculum. A measure of *g* was derived from two tests of verbal and non-verbal abilities. Multivariate and univariate genetic analyses revealed that individual differences in number sense are largely driven by non-shared environmental influences (~ .70), very small genetic influences (~.26) with negligible influences from shared environment (.04). There was an almost complete genetic overlap between symbolic and non-symbolic estimation skills (.90), with phenotypic correlation between the Dot Task and Number Line (.21) entirely mediated by the common genetic factors. The results also showed no genetic influences shared between the two number sense measures and Mathematics independently from *g*. These results suggest that the number sense measured at 16 years of age can be identified as a component of a *g* factor. No qualitative or quantitative sex differences were detected in the aetiology of individual differences in the four measures, suggesting that the same genetic and environmental factors equally influence number sense, Mathematics and *g* in boys and girls.

* This Chapter has been adapted from Tosto et al. (in preparation c)

7.2 Introduction

Mathematics is a complex domain drawing resources from different cognitive mechanisms. The links between mathematics and other cognitive abilities have been explored to gain further understanding into mathematical disabilities and the mechanisms of normal mathematical acquisition. Several studies suggest that impairment in other general cognitive abilities, such as reading, memory, or intelligence is associated with lower mathematical performance (e.g. Jordan, Bull & Johnston, 1997; McLean & Hitch, 1999; Passolunghi & Siegler, 2001, Swanson & Sachse-Lee, 2001; Hanich & Kaplan, 2003; Fuchs, Fuchs, & Prentice, 2004). However the extent to which low mathematical performance is a consequence of other impairments, or co-occurs independently remains unclear. With such variability in mathematical difficulties and outcomes, diagnostic criteria of mathematical disability, or the distinction between disability and difficulties, are object of debate (e.g. Shalev, Manor, & Gross- Tsur, 2005; Murphy, Mazzocco, Hanich, & Early, 2007). Moreover, the complex relationships between mathematics and other abilities make it difficult to predict mathematical proficiency (e.g. Gersten, Jordan, Flojo, 2005). On the other hand, the most prevalent approach in assessing and predicting mathematical abilities involves general intelligence and number sense ability (e.g. Geary et al., 2009; Geary, 2011). The focus of this study is the investigation of the relationship between mathematics, general intelligence (*g*) and number sense using a genetically sensitive design.

7.2.1 Mathematics and intelligence

Intelligence can be conceptualised as a general mental capability indicating the ease with which novel things are assimilated; it also includes broader capabilities such understanding our surroundings, planning and reasoning (Gottfredson, 1997). Overall, intelligence (*g*) is considered the best single predictor of school achievement and other cognitive outcomes (Jensen, 1998). Not surprisingly, individual differences in *g* have been found to be associated with many mathematical outcomes (Hoard, Geary, & Hamson, 1999). The first formal definition of learning disability was provided by Kirk & Bateman (1962) as a retardation, disorder or delay in the development of cognitive abilities, such as mathematics. In line with this tradition, mathematical difficulties are often identified on the basis of scores that fall below arbitrary cut offs on the normal distribution of performance. Severe mathematical problems are identified by restrictive cut offs (usually below the 5th or 10th percentile); performance falling below the 30th percentile diagnoses milder mathematical disabilities. IQ scores are often taken into account, so that specific mathematical disability is defined as low mathematical performance in the presence of average or above average IQ scores; or simply as a significant discrepancy between mathematical and IQ performance (e.g., Geary, Hamson, & Hoard, 2000; Gross-Tsur, Manor, & Shalev, 1996). However, such IQ discrepancy criterion has been criticised by many researchers as lower than expected mathematical performance does not always identify a mathematical disability (see Kavale & Forness, 2000; Geary, 2004).

Recent quantitative genetic research found a common genetic aetiology for learning ability and disability, leading to the formulation of the so called "Generalist Genes Hypothesis" (Plomin & Kovas, 2005). According to this hypothesis, largely the same genes influence all learning. This means that despite the uneven relationship between mathematics and *g* observed in behavioural research, the Generalist Genes theory predicts common genetic influences largely contributing to their covariation, with discrepancies between them largely explained by environmental factors (Plomin & Kovas, 2005).

One of the early twin studies investigated the relationship between mathematics and IQ, in a sample where at least one of the twins had reading disability (Light, DeFries, & Olson, 1998). In agreement with the Generalist Genes Hypothesis, it was found that around 82% of the phenotypic correlation between mathematics and reading was mediated by the same genetic factors influencing verbal IQ and phonological decoding. Following studies found a genetic correlation of .95 between *g* and mathematics in a sample of normally achieving twins aged 8 to 20 years (Alarcón, Knopik, & DeFries, 2000). The same study showed a similarly high genetic correlation between mathematics and *g* in a sample of twins where at least one of the twins showed reading, mathematical problems, or both. Another study used the WRRPM US sample of 6 to 10 year old normally developing twins to investigate the relationship between a general factor of intelligence and the mathematical components of *fluency*, *applied problems*, *quantitative concepts,* and *calculation* (Hart et al., 2009). The results showed common genetic influences between the mathematical components and *g* ranging between .20 and .51. In normally developing twins of the TEDS sample, between the ages 7 to 10, the genetic correlation between mathematics and *g* ranged from .54 to .67 (Kovas et al., 2005; Kovas et al., 2007a). At the age of 12 the correlation was .86 (Davis, Haworth, & Plomin, 2009b).

To sum up, quantitative genetic studies across different samples, different levels of mathematical abilities, and different ages suggest a substantial and stable genetic overlap between mathematics and *g*. However, this genetic overlap is not complete; the studies also indicate some genetic effects that influence mathematics independently from *g* (Kovas et al., 2005; Hart et al., 2009; Chapter 3 of this thesis). Understanding which cognitive mechanisms may be associated with these mathematics-specific genetic factors could aid better understanding of mathematical learning. One candidate for a cognitive process that may be uniquely associated with mathematics is number sense.

7.2.2 Mathematics, number sense and cognitive ability

Recent research endeavours have successfully identified some early predictors of mathematical abilities. Tools such as the *Number Sets Test* (Geary, Bailey, & Hoard, 2009), have been developed and used to predict mathematical achievement and to identify children at risks of low mathematical ability (e.g. Fuchs et al., 2010a). Tests such as *Estimation* (Baroody & Gatzke, 1991) have been found to correlate with mathematical achievement (Jordan et al., 2006). Such tests seem to tap into a cognitive mechanism pivotal to mathematical acquisition - the number sense (Dehaene, 1997). There is disagreement in the definition of number sense and what this ability exactly entails. However, it is generally accepted that number sense abilities allow the appreciation of the relative and absolute magnitudes of numerosity sets (i.e. discrimination of which of two arrays has more items); and of numbers in their symbolic system (i.e. estimation of numerical magnitudes of number symbols).

In estimation of numerosities, quantities are appreciated in terms of more and less, for this reason this process does not require counting and can be performed without any formal mathematical training. Numerosity discrimination can be performed by infant humans (Xu & Spelke, 2000; Libertus & Brannon, 2009), children (Halberda et al., 2008), and adults (Pica et al., 2004; Halberda et al., 2012). The ability to discriminate numerosities without counting has also been detected in various animal species (Uller, Jaeger, Guidry, & Martin, 2003; Pisa & Agrillo, 2009), suggesting a continuity of this ability across species. For this reason it has been proposed that number sense may have evolutionary origin. Although infants show discrimination abilities as early as 6 months, this ability improves from infancy to childhood (Xu & Spelke, 2000; Lipton & Spelke, 2003; Halberda & Feigenson, 2008; Mazzocco et al., 2011). A recent study that assessed estimation of numerosity skills in a sample of over 10,000 people, from the age of 11 to 85 year was able to map the developmental trajectory of this ability, with a rapid improvement from the age of 11 to 16, a peak at 30, and a decline between the ages of 30 and 85 (Halberda et al., 2012). This study also suggested a specific relationship between estimation of numerosity skills and mathematical achievement across ages, after controlling for age, sex, science ability, writing, and computer skills. Similarly to Halberda et al. (2012), another study found non-symbolic estimation exclusively associated with mathematics after controlling for general intelligence (Lourenco, Bonny, Fernandez, & Rao, 2012). Estimation of numerosity measured in a sample of 3 to 6 year old children, predicted only numerical skills measures in the same children two years later (Mazzocco et al., 2011). Non-symbolic estimation measured in preschool children, predicted mathematical outcomes two months later, after controlling for Verbal IQ (Gilmore, McCarthy, & Spelke, 2010). Another study showed a significant relationship between estimation of numerosity measured at 14 years and school mathematical achievement in previous years after controlling for 16 cognitive abilities (Halberda et al., 2008). Although these studies suggest a unique relationship between mathematics and non-symbolic estimation, some other studies have shown that this relationship may be mediated by IQ (Soltész, Szűcs, Szűcs, 2010) or other cognitive abilities (Chapter 5 of this thesis for a relationship of non-symbolic estimation and early reading). Some of the studies are cross-sectional therefore some of the inconsistencies in the association between non-symbolic estimations and other abilities may stem by different developmental trajectories of other abilities. For example, some longitudinal studies have shown that mean IQ scores in teenage years highly correlate, suggesting a large degree of stability but also some changes (Neisser et al., 1996). In light of these variations in abilities it is possible that non-symbolic estimation links with other abilities only in particular stages of development. Although more longitudinal studies are needed to understand the relationship between non-symbolic estimation and IQ, the literature suggests that this aspect of number sense may have a unique and stable relationship with mathematics across ages, therefore making a good candidate for the mechanism behind the mathematic-specific genetic influences discussed above.

Similarly to estimation of numerosity, numerical estimation (in Arabic number symbols) does not require exact calculation. This process is driven by reasoning as opposed to instinct, and relies on numerical concepts and knowledge about the characteristics of the magnitudes. For example, we do not need to compute 4 x 5 to decide that the magnitude of this product is greater than 4 x 3. This process however, involves some knowledge and training as it requires understanding of numbers and their relationships. A process that does not require any previous experience or training in mathematical processes is estimation of numerical magnitudes on a number line. To perform on this task the only skill required is number knowledge. Based on these properties, Booth and Siegler (2006) refer to number line estimation as a process of "pure numerical estimation". In a number line task participants are presented with a line with the edges marked, for example, 0 and 100; they are then required to place on the line some numerals ranging from 0 and 100.

Estimation skills improve with development and experience (Dowker, 2003), and number line estimation also seems to improve with age (Siegler & Opfer, 2003), and with exposure to number knowledge (Siegler & Ramani, 2008). Children's difficulties in estimation compared to adults are due to the use of an immature logarithmic mental representation of numbers. In a logarithmic numerical representation the distance between numerical magnitudes at the end of the range are underestimated compared to magnitudes at the beginning and the middle. For example the distance between 10 and 20 is mentally represented greater than the distance between 90 and 100. As a result, in a logarithmic representation the numbers at the end of the range are “compressed”. Adults produce more accurate estimates, as they rely on the more efficient linear representation of numbers, although a logarithmic representation is still present (Siegler & Opfer, 2003; Siegler & Booth, 2004; Booth & Siegler, 2006). The accuracy of the mental numerical magnitudes representation varies across individuals. These individual differences positively correlate with mathematical test scores and achievement (Booth & Siegler, 2006).

Numerical estimation has shown association with other abilities, besides mathematics. For example, patients with unilateral neglect (attention deficit as a consequence of damage to one brain-hemisphere) show poor performance on the number line task (Zorzi et al., 2002). Number Line estimation accuracy was associated with individual differences in IQ in a sample of 7 to 8 years old children (Geary et al., 2008). These studies suggest that mental representation of numbers, irrespectively of their magnitudes, relies on visual-imagery and logical thinking associated with general intelligence. It is also possible that the association between numerical estimation and other abilities is subject to developmental changes. The association between the Number Line task and visuo-spatial working memory was found significant at age 7, while in the same children at the age of 8 executive functions, but not visuo-spatial working memory predicted performance on the Number Line task (Geary et al., 2008). This suggests that visual imagery is necessary just at the early stages of mental number representation. Once the Number Line structure is in place, numerical representation may rely on attentional control driven by the central executive.

7.2.3 Research question

The present study is the first genetically sensitive investigation into the aetiology of number sense and its relationship with mathematics and *g*, using the large and age homogeneous TEDS sample. The study uses multivariate genetic analyses to investigate:

- the aetiology of the relationship between two aspects of number sense at age 16: estimation of numerosity, assessed with a Dot Task; and estimation of numerical magnitudes, assessed with a Number Line task.
- the aetiology of the contemporaneous relationship between mathematics, two aspects of number sense; and *g,* assessed with verbal and a non-verbal ability tasks at 16 years of age.

According to the Generalist Genes Hypothesis a certain degree of genetic overlap among all the measures is expected. A substantial genetic overlap is predicted between the two measures of Number sense (Number Line and Dot Task). On the other hand, if number sense abilities are specific to mathematics, as suggested by the behavioral literature, it is possible that they may share genetic aetiology only with mathematics. However, given the relationship between Number Line with IQ shown in behavioral studies, it is reasonable to expect a degree of genetic overlap between Number Line and *g*.

Although the main questions of this study are multivariate (examining the aetiology of the covariance among the measures), this research also provides the first estimates of genetic and environmental contributions to the variance in Number Line, mathematics and *g* at 16 years of age in the TEDS sample. The study employs the sex-limitation model to investigate for the first time the aetiology of sex differences in Number Line estimation, as well as the aetiology of sex differences in mathematics and *g* at age 16.

7.3 Methods

7.3.1 Participants

At the age of 16, data were collected on the web from 7,598 twins (3,799 pairs) of the first two cohorts (twins born between January 1994 and August 1995). As described in section 2.2.4.1, twins with medical problems and for who English was not the first language were removed from the analyses. Table 7.1 summarises the number of twins and their age by sex and zygosity

Table 7.1
Number of twins and their age in each zygosity group at 16

Groups	N (pairs)	AGE (SD)
All	6,854 (3,427)	16.6 (.26)
All Males	3,097	16.6 (.28)
All Females	3,757	16.6 (.27)
MZ	2,482 (1,241)	16.6 (.27)
DZ	4,372 (2,186)	16.6 (.28)
MZm	1,034 (517)	16.6 (.27)
DZm	2,063 (1,031)	16.6 (.28)
MZf	1,448 (724)	16.6 (.27)
DZf	2,309 (1,154)	16.6 (.27)
DZss	2,214 (1,107)	16.6 (.28)
Dzos	2,158 (1,079)	16.6 (.28)

MZ = Monozygotic, DZ = Dizygotic, MZm = Monozygotic males, DZm = Dizygotic males, MZf = Monozygotic females, DZf = Dizygotic females, DZss = Dizygotic same sex (male-male couples or female-females couples), DZos = Dizygotic opposite sex (male-female couple) SD = standard deviation

7.3.2 Measures

The number sense measures included in these analyses were the Weber Fraction (obtained from the Dot Task scores) and Number Line scores, derived as described in section 4.3.3.2. Weber Fraction scores were corrected for normality using a square root transformation; Number Line scores were log10 transformed. The mathematical measure was obtained averaging the standardised accuracy scores of the two web tests, Problem Verification and Understanding Numbers. A measure of general intelligence *g* was obtained averaging the standardised means of the two web tests of non verbal ability (Raven progressive matrices) and verbal ability (Vocabulary test), these two measures are described in section 2.4.2. The analyses were conducted on the standardised scores, corrected for the effects of age and sex (McGue & Bouchard, 1984). Scores outside +/-3standard deviation were removed as outliers.

7.3.3 Genetic analyses

This study makes use of multivariate genetic analysis to investigate the aetiology of the covariation between *g*, mathematics, Number Line and Weber Fraction. Sex-Limitation models are used to investigate the aetiology of sex differences in the four measures.

7.3.3.1 Multivariate genetic analyses

Univariate genetic analysis is primarily designed to estimate the relative contribution of genetic and environmental factors in variation in a trait. In this analysis, the cross-twin correlation on the same trait is modeled. One of the aims of multivariate genetic analysis is to estimate the genetic and environmental sources of covariation among different traits (Martin & Eaves, 1977). The comparison between MZ and DZ twins is conducted on the cross-trait twin correlations (for example, the correlation between performance of one of the twins in the pair in mathematics, and the co-twin's performance in number sense). When MZ twin cross-trait correlation is greater than the DZ twin cross-trait correlation, this indicates that the covariation between the two traits is due to common genetic factors. When the DZ twin correlation is more than half the MZ twin correlation, this indicates shared environmental influences. If the cross-trait twin correlations are non-significant, this implies that the co-occurrence of the two traits is due to non-shared environmental factors.

In multivariate genetic analyses, the sources of covariation among traits are attributed to 3 latent factor responsible for: genetic influences (A), shared environmental (C) and non-shared (E) environmental influences. These latent variables are estimated using structural equation modeling of the cross-trait and within-trait twin variances. There are several models that can be used to model the variance and covariance among traits. Each model is based on different sets of assumptions. For example, in the Independent pathway it is assumed that the traits are directly influenced by one common set of A, C, E, latent variables; the magnitude of the influences can be different for each variable. The residual variance in each variable is then explained by a set of trait specific A, C, E factors. In the Common pathway model, a single latent common factor mediates genetic and environmental factors across all traits; the residual variance, not explained by the common factor is further decomposed in A, C, E variance of trait specific variables.

One of the methods most used to conduct multivariate genetic analyses is the Cholesky decomposition (Loehlin, 1996). This approach assumes genetic and environmental factors specific for each trait with genetic and environmental factors on each trait correlating to some extent. In a classic Cholesky decomposition, the order of the variables is defined by a priori hypotheses. In the diagram in Figure 7.1 the order of Trait 1 and Trait 2 has been established on the basis of some logical assumptions. For example, in longitudinal analyses the sequence of the variables is established by the temporal order, in multivariate cases, the first variable should be the most general and more likely to share variance with the others.

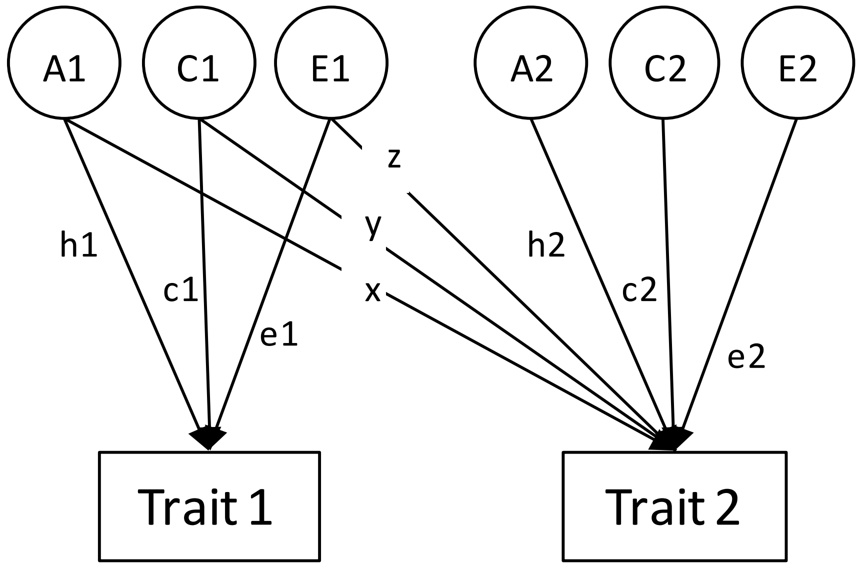


Figure 7.1: Path diagram Cholesky decomposition.

The latent variables represented by the circles A1, C1, E1 explain all the variance in Trait 1 and some of the variance in Trait 2 (traits are represented by squares). The residual variance in Trait 2 is explained by the latent variables A2, C2, E2 specific to Trait 2, independent from Trait 1. The vectors x, y and z represent the genetic (x), shared environmental influences (y) and non-shared environmental influence (z) in common between the two traits or variables.

The path diagram shown in Figure 7.1 illustrates a Cholesky decomposition. The variables, Trait 1 and Trait 2 are influenced by genetic, shared and non-shared environmental latent variables, respectively A1, C1, E1, and A2, C2, E2. The influences from a latent variable to the trait are represented by the arrows or paths. The first set of variables influences both traits. Genetic influences on Trait 1 are represented by the path h1. The influence of A1 on Trait 2 are the genetic influences in common between the two traits and are represented by the path x. The second set of latent variables A2, C2, E2 influences Trait 2 only. The Cholesky procedure is conceptually similar to a hierarchical regression where the trait specific variance is explained by trait-specific factors after the common variance has been accounted for by common factors. In such model the genetic variance explained by A1 in Trait 1 corresponds to its heritability estimate as this is its total genetic variance for Trait 1. The total genetic variance in Trait 2 is explained by trait-specific variance, (h2) and the genetic variance in common with Trait 1 (x). The heritability of Trait 2 is the sum of h2 + x. The same logic is applied to shared and non-shared environmental influences. Multivariate analyses may be used to investigate the covariation of more than two variables. In these cases rather than establishing the order of the variables a priori, it is useful to transform the initial Cholesky decomposition in a model that is free from hierarchical constraints.

The Correlated Factor solution is one of the possible transformations of the initial Cholesky decomposition. In this model a set of A, C, E variables explains genetic, shared and non-shared environmental components in each trait. The degree of covariation among the variables is indexed by the correlations of the latent variable across traits. Figure 7.2 illustrates the diagram of the Correlated Factor Model.

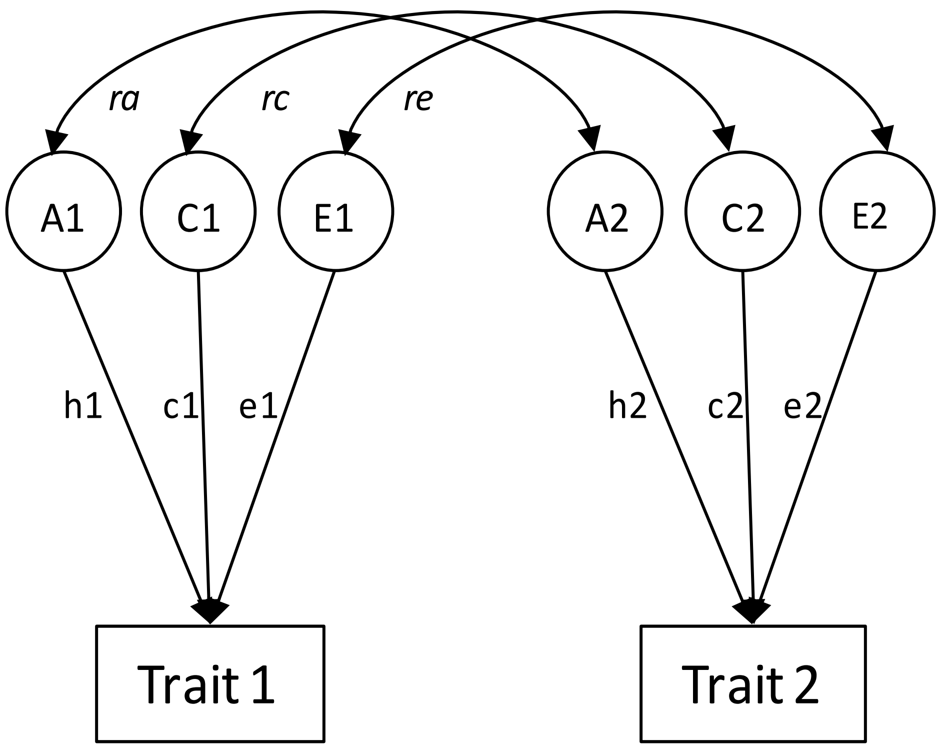


Figure 7.2: Correlated Factor Model

The total variance in Trait 1 is the sum of the path coefficients h1, c1, e1. The squared path-values (h^21, c^21, e^21) represent the univariate heritability estimates for Trait 1. The coefficient *ra* indexes the correlation between the genetic factors A1 and A2. Similarly, the shared (*rc*) and non shared (*re*) environmental correlations index the correlation between the latent shared environmental factors C1 and C2, and between E1 and E2.

Like in the Cholesky path diagram, in the Correlated Factor solution a set of latent variables explains the contribution of genetic and environmental influence on each trait. For Trait 1 the latent variables A1, C1, E1, explain the relative contribution of genetic (h1), shared (c1) and non-shared (e1) influences in Trait 1. The variance of Trait 2 is decomposed in the same way. The parameter *ra* indexes the genetic correlation between Trait 1 and Trait 2 and represents the amount of genetic overlap between the two traits. In other words, the genetic correlation is the extent to which the genetic effects of one trait are in common with another trait. The shared (*rc*) and non-shared (*re*) environmental correlations index the extent to which these environmental influences affect two traits. The absence of common paths between latent variables releases from the necessity to establish the order of variables a priory. It is of value to note that the term non-shared environmental correlation refers to influences non-shared between family members (not between traits). Non-shared environmental correlation exists between two traits as effects of environments that family members do not share but that influence the two traits.

Genetic correlation is independent from heritability. For example, if the genetic correlation between two traits is 1, all the genes involved in Trait 1 are also involved in Trait 2. These two traits may have low heritability (meaning that individual differences in that trait are only modestly influenced by genetic factors) but their high genetic correlation implies that all the genes involved in the individual differences in Trait 1 are also involved in Trait 2. The same logic applies to environmental correlations. A low shared environmental correlation indicates that factors that make the twins or family members more similar in Trait 1 are independent from the environmental factors that make the twins or family members more similar in Trait 2.

Another important statistic that is derived from multivariate analysis is the bivariate heritability. This represents the contribution of genetic factors to the phenotypic correlation between traits. The bivariate heritability is obtained with the formula:

$$Bivariate\ Heritability = \frac{\sqrt{\text{ATrait 1}}\ \text{x}\ ra\ \text{x}\ \sqrt{\text{ATrait 2}}}{\text{Phentotypic Correlation}} \qquad (7.1)$$

where $\sqrt{\mathrm{ATrait}}$ is the square root of the univariate heritability of the two traits and *ra* is the genetic correlation between Trait 1 and Trait 2. Similarly, it is possible to derive the bivariate shared and non-shared environmental heritability that index how much of the phenotypic correlation is mediated respectively by shared and non shared environmental factors.

7.3.3.2 Sex-Limitation models

The aetiology of gender differences in individual differences in the four measures was investigated using sex-limitation model fitting. The procedure has been described in section 6.3.3.1.

7.3.4 Model fitting

The genetic and environmental variance components were estimated using structural equation model fitting. As described in section 3.3.3.3 for the univariate model fitting, the multivariate and sex-limitation model fitting were performed using OpenMx software (Boker et al, 2011) running in the R environment (www.R-project.org). The criteria of the model fitting are the same as described in section 3.3.3.3.

7.4 Results

7.4.1 ANOVA results

Descriptive statistics and analyses of variance (ANOVA) by sex and zygosity are presented in Table 7.2. These analyses show that there are no significant effects of sex and zygosity on the means of the four variables. Overall, sex and zygosity effects were non-significant in all the four variables.

Table 7.2

Means, Standard Deviations and ANOVA results by sex and zygosity for: g, Mathematics, Number Line, Weber Fraction

																							ANOVA						
g	*N=2,197*		*N=831*		*N=1,366*		*N=1,299*		*N=898*		*N=305*		*N=593*		*N=526*		*N=773*		*N=650*		*N=716*								
Mathematics	*N=2,145*		*N=814*		*N=1,331*		*N=1,267*		*N=878*		*N=299*		*N=579*		*N=515*		*N=752*		*N=636*		*N=695*								
Number Line	*N=2,526*		*N=938*		*N=1,588*		*N=1,470*		*N=1,056*		*N=357*		*N=699*		*N=581*		*N=889*		*N=774*		*N=814*								
Weber Fraction	*N=2,214*		*N=817*		*N=1,397*		*N=1,298*		*N=916*		*N=301*		*N=615*		*N=516*		*N=782*		*N=677*		*N=720*								
	All		***MZ***		***DZ***		***Female***		***Male***		***MZm***		***DZm***		***MZf***		***DZf***		***Dzo***		***DZss***		***Zyg.***		***Sex***		***Zyg.*Sex***		***Tot.***
Measures	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*p*	η^2	*p*	η^2	*p*	η^2	R^2
g	.01	.99	-.03	.99	.04	.99	.03	.97	-.02	1.0	-.04	1.0	-.01	1.0	-.03	.99	.07	.96	.07	.99	.01	.98	.18	.00	.14	.00	.42	.00	.003
Mathematics	.03	1.0	.00	.99	.05	1.0	.02	1.0	.04	.96	-.02	.92	.08	.98	.01	1.0	.02	1.0	.05	.99	.04	1.0	.48	.00	.59	.00	.10	.00	.003
Number Line	.00	.98	.04	.93	-.02	.99	-.01	.97	.01	.98	.05	.92	-.01	1.0	.03	.94	-.03	.99	.00	1.0	-.04	1.0	.30	.00	.57	.00	.08	.00	.003
Weber Fraction	-.10	.85	-.07	.85	-.11	.84	-.10	.83	-.01	.87	-.08	.86	-.10	.88	-.07	.85	-.12	.81	-.10	.85	-.12	.83	.56	.00	.87	.00	.56	.00	.001

MZ = Monozygotic, DZ = Dizygotic, MZm = Monozygotic males, DZm = Dizygotic males, MZf = Monozygotic females, DZf = Dizygotic females, DZss = Dizygotic same sex (male-male pairs or female-females pairs), DZos = Dizygotic opposite sex (male-female pairs), SD = standard deviation, N = number of twins in each group, R^2 = variance explained in the model by sex and zygosity.

7.4.2 Sex differences

Twins' similarity on traits is indexed by the intraclass correlation coefficients (ICC, Shrout, & Fleiss, 1979). These are presented in Table 7.3 for all the twins, separate males and females; and by zygosity.

Table 7.3

Intraclass correlations with 95% Confidence Intervals

Measure	*MZ (all)*	*DZ (all)*	MZm	MZf	DZm	DZf	DZss	Dzos
g	.59 (.54-.64)	.36 (.31-.41)	.66 (.59-.72)	.55 (.49-.61)	.39 (.28-.49)	.42 (.33-.49)	.40 (.34-.47)	.30 (.23-.38)
Mathematics	.76 (.73-.79)	.46 (.41-.51)	.73 (.66-.78)	.78 (.74-.81)	.47 (.36-.56)	.49 (.41-.57)	.48 (.42-.54)	.43 (.35-.49)
Number Line	.26 (.20-.32)	.14 (.09-.19)	.32 (.22-.42)	.23 (.15-.31)	.10 (-.02-.21)	.19 (.10-.27)	.15 (.08-.22)	.13 (.06-.20)
Weber Fraction	.31 (.24-.38)	.15 (.09-.20)	.36 (.24-.46)	.29 (.20-.37)	.08 (-.05-.20)	.13 (.02-.23)	.11 (.03-.19)	.19 (.11-.27)

MZ = Monozygotic; DZ = Dizygotic; MZm=Monozygotic males MZf =Monozygotic females; DZm = Dizygotic males; DZf = Dizygotic females; DZss = Dizygotic same sex; DZos = Dizygotic opposite sex. Confidence Intervals are in brackets.

The correlations show higher values for MZ than DZ twins, suggesting genetic influences on all the measures. In addition, the DZ correlation for Mathematics and *g* is more than half of the MZ correlation, suggesting shared environmental influences for these two measures. The correlations of the dizygotic same sex twins are higher than the dizygotic opposite sex; this may indicate the presence of qualitative sex differences in the measures. However, the confidence intervals for the two groups almost overlap for all the measures, suggesting that the correlation of the DZ opposite and same sex may not be significantly different.

The univariate sex-limitation model fitting was performed to assess the presence of qualitative or quantitative sex differences in the four measures. The results are summarised in Table 7.4. The table shows that the fit of the Null Model is not significantly worse than the Full sex-limitation model, thus suggesting the absence of qualitative, quantitative or variance sex differences in the aetiology of individual differences in the four measures.

Table 7.4

Univariate Sex-Limitation Model and parameter estimates from the sex-limitation model fitting

Measure	Δ-2LL (Δdf)	p	AIC	BIC	Am	Af	Cm	Cf	Em	Ef	rg
g											
Full sex-limitation	--	--	2587.76	-19678.75	.38 (.22-.53)	.22 (.07-.37)	.27 (.15-.41)	.33 (.21-.46)	.34 (.29-.41)	.45 (.39-.51)	.00 (.00-.23)
Null model (no sex effects)	2.07 (4)	.72	2581.83	-19708.55	.41(.29-.54)		.19 (.08-.30)		.40 (.36-.44)		0.5
Mathematics											
Full sex-limitation	--	--	2148.82	-19502.99	.37 (.30-.51)	.46 (.40-.57)	.38 (.25-.50)	.33(.22 -.43)	.25 (.21-.31)	.21 (.19-.25)	.00 (.00-.11)
Null model (no sex effects)	-3.87 (4)	1.0	2136.95	-19538.73	.56 (.46-.67)		.21(.11-.31)		.22 (.20-.25)		0.5
Number Line											
Full sex-limitation	--	--	3381.57	-22155.39	.27 (.10-.38)	.15 (.00-.26)	.03 (.00-.15)	.10 (.00-.26)	.70 (.60-.80)	.77 (.69-.85)	.00 (.00-.50)
Null model (no sex effects)	1.53 (4)	.82	3375.10	-22185.73	.27 (.11-.33)		.00 (.00-.11)		.73 (.67-.79)		0.5
Weber Fraction											
Full sex-limitation	--	--	2011.30	-20434.25	.34 (.13-.42)	.29 (.07-.37)	.00(.00-.15)	.00(.00-.18)	.66 (.57-.76)	.70 (.63-.79)	.50 (.12-.50)
Null model (no sex effects)	.69 (4)	.95	2003.99	-20465.43	.31 (.16-.37)		.00 (.00-.11)		.69 (.63-.75)		0.5

Δ-2LL = Difference in log likelihood between the Full sex-limitation model and the Null Model; Δdf = difference in degrees of freedom between the Full and Null models; p = p-value; AIC = Akaike Information Criterion BIC = Bayesian Information Criterion; Am/ Af = additive genetic variance component, males/ females; Cm/ Cf = shared environmental variance component, males/females; Em/ Ef = non-shared environmental variance component, males/ females; rg = genetic correlation in opposite-sex pairs (fixed at .50 in the Null Model). AIC and BIC estimates suggest that the Null model (no sex differences) provides the best fit to the data. The p-value indicates that the Null Model is not significantly worse than the Full Model.

7.4.3 Multivariate genetic analysis

Table 7.5 shows the loading of the factors in the Cholesky decomposition with the variables ordered as follows: *g*, Mathematics, Number Line and Weber Fraction.

Table 7.5

Standardised Cholesky squared paths coefficients and 95% Confidence Intervals

Path	*g*	*Mathematics*	*Number Line*	*Weber Fraction*
A1	.45 (.15 - .57)	.34 (.24 - .47)	.11 (.05 - .20)	.11 (.04 - .21)
C1	.15 (.06 - .25)	.13 (.06 - .25)	.01 (.00 - .06)	.00 (.00 - .03)
E1	.40 (.35 - .43)	.01 (.00 - .02)	.01 (.00 - .02)	.01 (.00 - .02)
A2		.23 (.11 - .33)	.06 (.00 - .15)	.01 (.00 - .05)
C2		.08 (.00 - .16)	.01 (.00 - .05)	.01 (.00 - .01)
E2		.21 (.19 - .23)	.01 (.00 - .02)	.01 (.00 - .02)
A3			.07 (.01 - .13)	.16 (.01 - .26)
C3			.02 (.00 - .07)	.03 (.00 - .11)
E3			.70 (.66 - .75)	.00 (.00 - .00)
A4				.00 (.00 - .20)
C4				.00 (.00 - .01)
E4				.66 (.60 - .71)

The effects of the first latent variable are decomposed into genetic (A1), shared environmental (C1), and non-shared environmental factors (E1) - influencing *g*, Mathematics, Number Line and Weber Fraction. The variance unexplained by the first latent variable is further decomposed into A2, C2, E2 influences affecting Mathematics, Number Line and Weber Fraction independently from *g*. The next variables explain all the unaccounted variance.

From this table it is possible to observe that the first genetic latent variable A1 accounts for genetic influence on *g* (.45), and also influences Mathematics (.34), Number Line (.11) and Weber Fraction (.11). The influences of A1 on Mathematics, Number Line and Weber Fraction represent the common genetic influences between *g* and the three variables. For example, .34 of the genetic influences on Mathematics are in common with *g*. Similarly, .11 of genetic influences are shared between Number Line and *g*. Also .11 of the genetic influences are shared between Weber

Fraction and *g*. The second latent variable A2 explains .23 of the genetic variance in Mathematics, independent of genetic influences on *g*. This second variable contribute very little (and non-significantly) to the variance in Number Line (.06) and Weber Fraction (.01). A third genetic factor (A3) exerts a small (.07) but significant influence on Number Line, independently from *g* and Mathematics. A3 also contributes to Weber Fraction (.16). No specific genetic effects are present for the Weber Fraction (A4 factor has 0 effect on it).

To summarise, most of the genetic influences in the 4 variables are explained by two genetic latent variables (A1 and A2). All the genetic influences in Weber Fraction are explained by genetic factors shared with the other 3 variables with the absence of genetic specific influences on this trait. Only less than 10% of genetic influences in Number Line are independent from genetic factors in common with the other three traits (A3 influences in Number Line are .07). From the same table it is possible to observe that most of the shared environmental influences are very small in magnitude and specific to each trait. This suggests that between-families environments important to individual differences in mathematical skills are mostly different from the family environments important for numerosity estimation skills or *g*. Similarly, the non-shared environmental influences are specific to each trait.

Changing the order of the variables in the Cholesky decomposition did not affect the pattern of the results reported above. However, a Correlated Factor model was fitted in order to obtain information about genetic and environmental correlations and bivariate heritability. The results of this analysis are visually summarised in the path diagrams in Figure 7.3

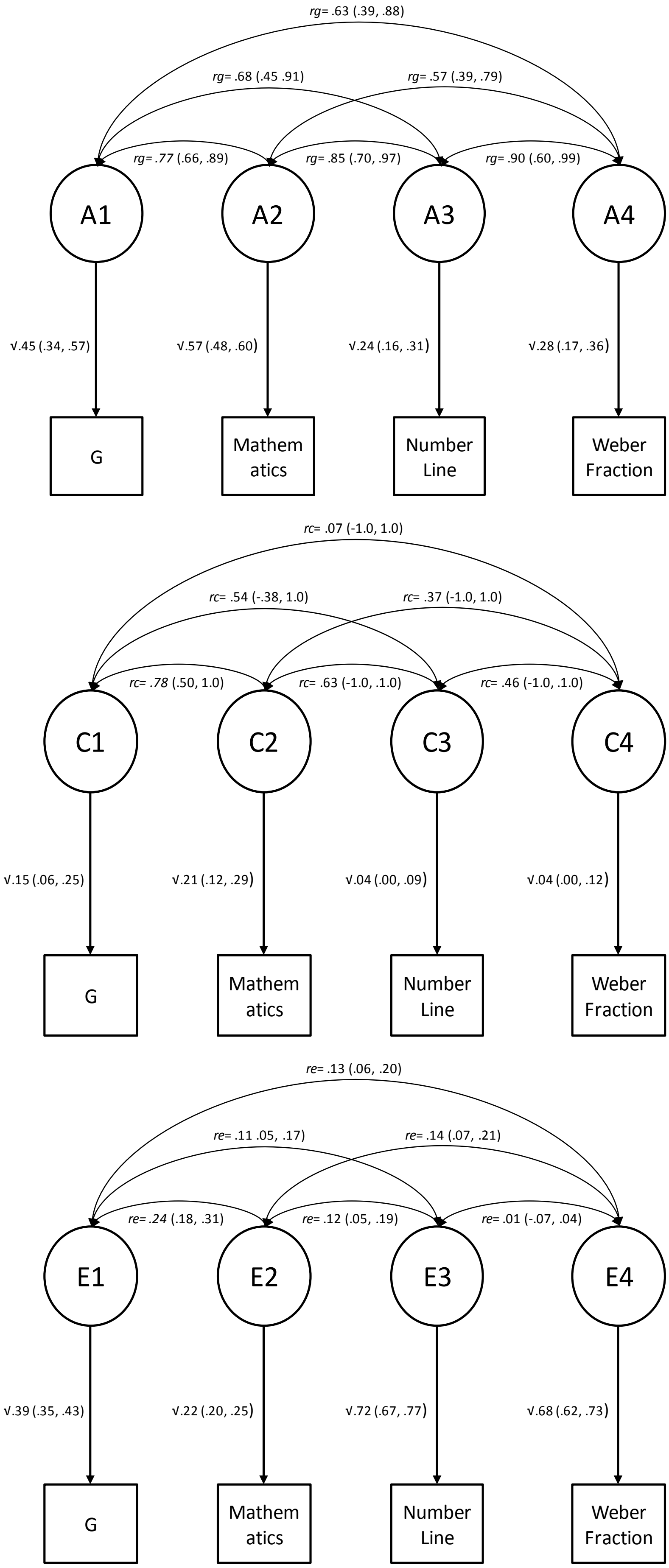


Figure 7.3: Cholesky Correlated Factor Solution
Each of the path reports the univariate estimate with the 95% Confidence Intervals. The double headed arrows between the latent variables represent the genetic correlation (*ra*), shared environmental (*rc*) and non-shared environmental correlation (*re*).

The first part of the diagram summarises the genetic influences, where each latent variable explains the entire genetic variance specific to the trait. Therefore, the path from the genetic latent variable to the trait represents the total genetic variance for that trait, the univariate estimate of the heritability. The heritability of *g* and Mathematics at 16 is .45 and .57, respectively. The two number sense measures, indexed by Number Line and Weber Fraction, both show modest heritability of .24 and .28, respectively. It has to be noted that the univariate estimates of the Weber Fraction reported here are slightly different compared to what is reported in the study in Chapter 6; the difference is within the tolerance of the different models fitted to the data. In the study described earlier, the best model describing the data did not include the shared environmental estimate as this was non-significant. By fitting an AE model, the small variance component of c^2 was incorporated in the heritability estimate h^2. Here we decided to fit a Cholesky ACE model to better illustrate the pattern of correlations among the measures, although some of these relationships are not statistically significant.

The genetic correlations among all the measures are quite substantial, ranging between .57 (between Mathematics and Weber Fraction); and .90 (between Number Line and Weber Fraction). For example, the genetic correlation between *g* and Weber Fraction is .63; meaning that if a gene is involved in *g* there is 63% probability that the same gene is also involved in estimation of numerosity skills (as indexed by Weber Fraction). Overall, there is a greater genetic overlap between Weber Fraction and *g,* compared to the genetic overlap between Weber Fraction and Mathematics. The genetic correlation between the two number sense measures is almost complete (90%).

The estimates of the genetic bivariate heritability derived from the model fitting are reported in Table 7.6 These estimates index the proportion of phenotypic correlation mediated by genetic factors.

Bivariate heritability can be calculated using the formula 7.1. For example, the genetic contribution to the phenotypic relationship between *g* and Mathematics is: $(\surd(.45) \; * \; .77 \; * \; \surd(.57))/(.60)$ = .65. This indicates that 65% of the .60 correlation between Mathematics and *g* is mediated by common genetic factors. From Table 7.6

it can be seen that the mediation of genetic factors is substantial for all of the phenotypic correlations. Although the correlation between Number Line and Weber Fraction is modest (.21), it is entirely mediated by common genetic factors, indicating that all the genes involved in individual differences in Number Line are also involved in individual differences in Weber Fraction.

Table 7.6

Correlated Factor Solution

Phenotypic correlations, genetic, shared and non-shared environmental correlation, bivariate heritability between *g*, Mathematics, Number Line and Weber Fraction with 95% Confidence Interval (in brackets)

	g and Mathematics	g and Number Line	g and Weber Fraction	Mathematics and Number Line	Mathematics and Weber Fraction	Number Line and Weber Fraction
Phenotypic correlation	.60 (.52, .63)	-.28 (-.32, -.24)	-.25 (-.29, -.21)	-.40 (-.43, -.36)	-.29 (-.33, -.25)	.21 (.15, .22)
Genetic, Shared & Non-shared Environ. Correlations						
ra	.77 (.66, .89)	.68 (.45 .91)	.63 (.39, .88)	.85 (.70, .97)	.57 (.39, .79)	.90 (.60, .99)
rc	.78 (.50, 1.0)	.54 (-.38, 1.0)	.07 (-1.0, 1.0)	.63 (-1.0, .1.0)	.37 (-1.0, 1.0)	.46 (-1.0, .1.0)
re	.24 (.18, .31)	.11 (.05, .17)	.13 (.06, .20)	.12 (.05, .19)	.14 (.07, .21)	.01 (-.07, .04)
Bivariate heritability estimates						
A	.65 (.52 - .78)	.69 (.44 - .93)	.75 (.46 - 1.0)	.75 (.56 - .92)	.72 (.46 - .97)	1.0 (.75 - 1.0)
C	.23 (.01 - .34)	.12 (-.05 - .32)	.01 (-.01 - .25)	.13 (.00 - .28)	.11 (.00 - .32)	.08 (.00 - .09)
E	.12 (.01 - .16)	.18 (.01 - .28)	.22 (.11 - .36)	.11 (.05 - .18)	.17 (.01 - .26)	.05 (.00 - .21)

ra = Genetic correlation, *rc* = Shared Environmental correlation, *re* = Non-shared Environmental correlation. The bivariate heritabilily estimates: A, C, E, indicate the proportion of phenotypic correlation mediated by genetic factors (A), shared environmental factors (C), non-shared environmental factors E).

The second part of Figure 7.3 illustrates the shared environmental correlations. The univariate estimates - represented by the arrows from the shared environmental latent factor to each variable - show a small contribution of shared environmental factors to individual differences in *g* and Mathematics (respectively .15 and .21) and almost nonexistent in the two number sense variables (.04). This diagram also shows that the only significant shared environmental correlation is between *g* and Mathematics. The confidence intervals of all the other correlation estimates cross zero, indicating that are non-significant. These shared environmental correlations suggest that the environment that make people more similar in Mathematics, are different from the environments that make people more similar in estimation skills, but some shared environments important for Mathematics are also important for *g*. That is to say, that the small influences of shared environments are largely specific to each measure. This is also reflected in the small bivariate shared environmental estimates, indicating that the contribution of shared environmental factors to the correlation among the measures is very small. These range between .01 (between *g* and Weber Fraction) and .23 (between *g* and Mathematics).

The third section of Figure 7.3 illustrates the Correlated Factor solution for the non-shared environmental influence. The univariate estimates indicate that individual differences in *g* and Mathematics are moderately affected by non-shared environmental factors (estimates respectively of .39 and .22). The correlation of these environments is also small (*re* = .24), indicating that the non-shared environmental influence are largely specific to each trait. The contribution of non-shared environmental influences to the phenotypic correlation between *g* and Mathematics (.60) is also small (.12). Conversely, the contribution of the non-shared environment to individual differences in number sense is very large (e^2 = .72 for Number Line and e^2 = .68 for Weber Fraction). The non-shared environmental correlation between the two number sense measures is almost nonexistent (.01, and statistically non significant as the confidence intervals cross zero), indicating that individual-specific experiences important for the Number Line skills are not important for the numerosity estimation skills (Weber Fraction). The results also suggest that the two tests may have uncorrelated measurement error. Overall, the non-shared environmental correlations among the four measures are low, ranging between .01 and .24, suggesting that the non-shared environmental influences are specific to each trait. Similarly, the contribution of the non-shared environment to the phenotypic correlations is small, as shown in Table 8.4, ranging from .05 (between Number Line and Weber Fraction) to .22 (between *g* and Weber Fraction).

7.5 Discussion

The aims of the present study were threefold. First, the study presented the first genetically sensitive investigation into the aetiology of number sense by examining the covariation of two number sense measures. The first univariate genetic and environmental estimates for symbolic estimation skills, as measured with a Number Line estimation task, were also obtained from the analyses. The second aim was to investigate the aetiology of the relationship between mathematics and number sense (as assessed by Number Line and Weber Fraction), examining the relationship with *g* at the same time The analyses also allowed for the first investigations of the genetic and environmental aetiology of the individual differences in mathematical abilities and *g* at 16. Lastly, sex-limitation models were used to investigate the aetiology of gender differences in number sense, mathematics and *g*.

7.5.1 Number sense measures

The results showed that individual differences in Number Line estimation at 16 years are largely driven by non shared environmental influences (.72), very small genetic influence (.24), and negligible shared environmental influences (.04). These heritability estimates are not surprising given the nature of this estimation ability. Many behavioural studies (e.g. Siegler & Booth, 2004; Siegler & Ramani, 2008, 2009) have suggested training and numerical activities as promoters of Number Line estimation skills. The findings of this study suggest that individual specific experiences and environments, rather than genes, play a major role in people's differences in estimation skills. This means that school, family and other objectively shared environments may not contribute to individuals' resemblance in estimation of numerical magnitudes beyond their genetic similarity. What also emerged from the analyses is the striking similarity in heritability estimates for Number Line and Weber Fraction. Being both regarded as number sense measures, the similarity in the aetiology of individual differences in these skills is not surprising. However, this is not a necessary condition for different aspects of a domain. In fact, other studies have found a degree of variability in heritability within measures of mathematics (Kovas et al., 2007c; Hart et al., 2009). Other twin studies have reported a trend of more shared environmental influences for untimed tests of mathematical skills and more genetic influences for timed tests (Hart et al., 2009). If the heritability estimates would have been influenced by the nature of the task, we would have expected Weber Fraction (scores derived from the timed Dots task) to have more genetic influences compared to Number Line (untimed task). This was not the case, as both measures showed the same heritability. Although there are many factors that contribute to differences in heritability estimates (see section 1.2.9.2), it is possible that the similarity in the magnitude of the heritability estimates is a reflection of a common aetiology. Evidence of a common aetiology in the two estimation abilities is provided by the multivariate results of this study. The genetic correlation between Number Line and Weber Fraction was .90; this suggests that there is 90% probability that the same genes are involved in individual differences in the two abilities. However, the low heritability indicates that these influences, although almost overlapping are very small, resulting in a low covariation between the two number sense measures. In the absence of shared environmental influences, and with the non-significant non-shared environmental correlation between the two measures, the phenotypic correlation between Number Line and Weber Fraction is entirely mediated by the small genetic influences.

In summary, heritability of number sense is modest, and lower than for mathematics and *g*. Individual differences in both measures of number sense are largely explained by non-shared environmental factors. The small relationship that does exist between the two measures is entirely explained by overlapping genetic factors. The two traits seem to be differentiated by effects of totally independent environmental influences.

7.5.2 Mathematics and number sense

Despite developmental changes and changes in the mathematical phenotype, heritability estimates for mathematics in TEDS have shown consistency throughout the school years. Overall, from early to middle childhood, individual differences in mathematics, are driven by moderate genetic influences with modest contributions from shared and non-shared environmental influences (e.g. Kovas et al., 2007a). At the age of 16, the heritability estimates are in agreement with previous results. With estimates $h^2 = .57$, $c^2 = .21$, $e^2 = .22$ mathematics still shows to be a moderate-high heritable trait.

Beyond the univariate estimates of genetic and environmental effects on mathematics, another goal of this study was to investigate the nature of the relationship between Mathematics and number sense. The results showed a higher genetic overlap between Mathematics and Number Line (.85), compared to that between Mathematics and Weber Fraction (.57). The environmental correlation was small, or non significant. This suggests that the extent to which number sense and Mathematics covary is due to genetic influences. In fact, genes explained ~ 74% of the phenotypic correlation between Mathematics and number sense. On the other hand, most of the environmental influences were largely specific to each trait and contributed almost non-significantly to the link between Mathematics and number sense.

7.5.3 Mathematics, number sense and g

The inclusion of *g* in the analyses gives a clearer picture of the relationship between number sense and Mathematics. From the results in Table 7.5 it is clear that there are almost no genetic influences specific to Weber Fraction beyond the small influences shared with *g* (.11) and Number Line (.16). Weber Fraction and Mathematics have no common genetic influence beyond what they share with *g*. Number Line seems to have a very small specific genetic influence (.07), but like

Weber Fraction, the remaining genetic influences are largely shared with *g*. These results partially agree with our predictions. We expected a genetic correlation between Number Line and *g* since the two mesures have previously been shown to be related (Geary et al., 2008). However, the results of this investigation do not confirm a unique relationship between Mathematics and Weber Fraction, as all the genetic variance in Weber Fraction is shared with *g* and Number Line, but not with Mathematics. Mathematics on the other hand, maintains a stable amount of genetic overlap with *g*, as shown at earlier ages (e.g. Kovas et al., 2005; Davis et al., 2008a). The shared environmental correlation between g and Mathematics was .78, indicating that the environments contributing to similarities in g in children in the same family, also contribute to similarity in mathematics. In practical terms this means that, although of small magnitudes (the bivariate heritability for the shared environments was .23) environmental effects contributing to people's similarity in *g* would also contribute to people's similarity in mathematics. This cannot be said for the number sense measures. There would be no common environments benefiting *g* and number sense as the shared environmental correlations with *g* were non-significant for both measures.

7.5.4 Sex-differences

The results of the sex-limiation models did not find qualitative or quantitative sex differences in any of the measures. This suggests that for number sense, Mathematics and g, the same genetic and environmental factors influence individual differences in boys and girls to the same extent. These results are consistent with previous findings in TEDS that did not find sex differences in the aetiology of individual differences in mathematics or *g* across school ages (e.g. Kovas et al., 2007a; Davis et al., 2008b). The absence of average or aetiological sex differences in Number Line and Weber Fraction suggest that genetic and environmetal influences important for estimation abilities are the same for boys and girls. More importantly, these results suggest that any observed mean sex differences in mathematics are not meadited by *g* or number sense.

7.5.5 Conclusion

The results of this study support the Generalist Genes Hypothesis (Plomin & Kovas, 2005), that proposes that largely the same set of pleiotropic genes influence congnitive abilities associated with learning. In this study we found that genetic contribution to the four traits is largely due to the same genetic effects.

It is therefore difficult to claim that number sense indexes a unique ability which is not part ofthe general factor of intelligence. While Mathematics shows some independent genetic influences, we may say that what there is in common between Mathematics and number sense is *g*. It also should be emphasised that although genetic influences on number sense greatly overlap with Mathematics, the overall genetic influence on number sense is very small. Specialist environments shape up estimation abilities, but the contribution of these environments to the correlation with Mathematics is minor. In fact, we could say that what it is "special" about number sense is the environment.

It needs to be noted that these estimates refer to the snapshot of time at age 16, with its specific environmental settings. In Chapter 5, the potential involvement of numerosity estimation at the early stage of mathematical learning was discussed. It is possible that at earlier ages different aetiological relationships exist between number sense and mathematics, such as more environmental and/or genetic influences in common between early mathematics and number sense. These hypotheses need to be investigated using samples of different ages and in different cultural and educational settings.

Chapter 8: General Discussion

This thesis set out to investigate various aspects of the number sense domain. The first aim was the creation and validation of an internet-based battery of tests designed to assess different aspects of number sense (estimation of non-symbolic numerosity and of magnitude of numerical symbols), mathematics, and general cognitive abilities associated with mathematical acquisition, in 16 year olds. This battery was then used in other studies presented in this thesis, to conduct the first large-scale genetically sensitive investigation into the number sense at 16 years of age and of its links with mathematics and general cognitive abilities. Taken together, the studies presented in this thesis provided further understanding of the nature of estimation abilities, as well as new insights into the origins of individual differences in number sense and of the relationship between number sense, mathematics and general cognitive ability.

8.1 General summary and findings

Chapter 1 provided an overview of the current understanding of number sense from a behavioural perspective. Previous research has shown that the number sense is involved in the development of early mathematical concepts beyond the contribution of other cognitive abilities such as reading or memory. The chapter also presented behavioural genetic findings showing the existence of genetic factors influencing mathematics, independently from reading and *g*. This thesis investigated whether these genetic factors were related to the variance in the number sense.

Most of the analyses reported in this thesis were conducted using the longitudinal TEDS sample. The sample and the longitudinal measures adopted in this thesis were described in Chapter 2. The studies reported in the thesis relied on data collected with a bespoke on-line battery. The development and validation of the online battery was based on a sample of 16 year-old singletons. This sample was also described in Chapter 2.

The study reported in Chapter 3 offered further evidence for the existence of mathematics-specific genetic influences. In previous studies, these specific genetic influences were identified by using multivariate genetic analyses. In the multivariate analyses, the overlapping variance among traits is decomposed into 3 components, respectively attributed to genetic, shared and non-shared environmental factors. The approach used in the study presented in Chapter 3 was to remove all the variance explained by reading and *g* from the mathematical scores. Consistent with the results of the multivariate analyses, this new mathematics variable showed to be heritable, thus suggesting that some genetic factors influence mathematics independently from reading and *g*.

The tests included in the online battery were selected to tap into abilities that have shown to be related to mathematical development in behavioural literature. Chapter 4 presented the methods used for the online implementation of these tasks, as well as evidence for the reliability and efficiency of the battery as an assessment tool.

Chapter 5 examined the relationship of number sense, measured at 16, with contemporaneous and past mathematical and cognitive abilities. In line with findings in the current literature, number sense showed a relationship with mathematics at 16 and previous mathematical school achievement. This relationship was stronger for estimation of numerical magnitudes than for estimation of numerosity. The study did not support the hypothesis that number sense (numerosity estimation) is uniquely related to mathematics. The results showed the relationship of numerosity estimation with other cognitive abilities. The results in Chapter 5 suggested that estimation abilities that make up the number sense construct are heterogeneous. It was also suggested that numerosity estimation may provide the basis for the automatic processes of mathematics by using the same mechanism of pattern recognition used in reading.

The contribution of genes and environment to individual differences in estimation of numerosity (indexed by Weber Fraction scores) were investigated in Chapter 6. Individual differences in numerosity estimation were found to be largely driven by non shared environmental factors (.68). The low heritability of the Weber

Fraction (.32) supported the hypothesis that this evolutionarily important trait developed through de-selection of negative variants, leading to reduced genetic variance relevant for this trait.

The aetiology of the relationship between number sense, mathematics and *g* was investigated in Chapter 7. The striking findings of this study were that the relationship between mathematics and number sense was entirely mediated by *g*.

8.1.1 Number sense

Number symbols and numerosity concepts are interlinked, as numbers are used to identify numerosities. However, the extent to which number processing and numerosity processing are similar to each other required further investigation. This research showed a low phenotypic correlation (.22) between the two estimation tasks (numerical magnitudes - Number Line, and numerosity estimation - Weber Fraction). This may mean that despite dealing with numerical material, estimation of numerosities and of numerical magnitudes reflect two different processes. Indeed, the multivariate genetic analyses suggested the same genetic aetiology for numerosity and magnitude estimation (the study presented in Chapter 7 showed the genetic correlation of .90 between Number Line and Weber Fraction). However, this common genetic aetiology was not sufficient to ensure a degree of correlation higher than .22, at least at 16 years of age. This low phenotypic correlation can be explained by low heritability of both measures (.32 for Weber Fraction and .24 for Number Line); as well as high (~70%) but uncorrelated non-shared environmental influences on the two traits. Overall, the results suggest that individual differences in estimation abilities, both numerosity and of numerical magnitudes are mostly driven by environments specific to the individuals, and to each trait. The genetic effects on individual differences in number sense are very modest.

The low heritability of numerosity estimation was confirmed by the GCTA analysis (reported in Chapter 6). Low heritability indicates that there is very little genetic variability relevant to estimation abilities in a particular population. If numerosity estimation has been evolutionarily selected, as suggested by evidence from non-human animals, low heritability should be expected. The capacity to discriminate items according to numerosity (simply, being able to tell apart more from less) is a useful skill for hunting, gathering and for social situations. For this reason, these skills must have provided important adaptive evolutionary advantages. From an evolutionary standpoint, the successful transmission of these abilities across generations is ensured by the natural selection of the protective alleles and de-selection of those genetic influences that negatively affect the trait. In other words, individuals with better estimation abilities may have been more adaptable to adverse conditions requiring estimations skills and perhaps this may have lead to their increased reproductive fitness. This process leads to a reduction in the genetic variability in this trait in the population. It is possible that one set of (largely invariant) genes is involved in the mechanisms enabling discrimination, such as the neural network involved in numerosity discrimination in intraparietal areas of the brain (e.g. Piazza et al., 2006). The same genes may also be involved in number magnitude estimation, as the same brain areas are involved in numerosity and number symbols processing (e.g. Cohen Kadosh et al., 2011). On the contrary, as the results of the present study suggests, individual differences in these traits may largely stem from environmental influences

8.1.2 Number sense, mathematics, and g

An unpredicted finding of this work was the lack of genetic influences that are uniquely shared between number sense and mathematics. As presented in Chapter 7, there was no genetic overlap between mathematics and either measures of number sense, aside common genetic factors shared with *g*. Although this finding is congruent with the Generalist Genes Hypothesis, this result suggests that number sense is not the source of the specific genetic influences on mathematics found by previous behavioural genetic studies (e.g. Kovas et al., 2005; Hart et al., 2009; Chapter 3); in fact genetic effects on number sense, as measured by estimation abilities, do not overlap with the unique (unshared with *g* and reading) genetic influences on mathematics.

In addition, the study did not find any overlap between the environmental influences on mathematics and number sense. This suggests that the observed moderate association between mathematics and number sense at age 16 (-.40 with Number Line; and -.29 with Weber Fraction) can be largely explained by genetic influences that mathematics also shares with *g*.

The analyses also showed that the two estimation measures shared a small genetic influence (.16) independently from mathematics and *g*. As seen in Chapter 5, Number Line and Weber Fraction, measured at 16, showed small associations with spatial abilities and reading across the ages. It is possible that the two estimation measures share some genetic influences with reading, spatial abilities or other abilities not included in this investigation. These genetic influences may be independent from mathematics and *g*, thus explaining the specific-number sense genetic variance. These results further suggest that number sense, as measured by estimation skills, is part of a general cognitive construct rather than specific to mathematical domain. The genetic overlap between Weber Fraction and *g* also suggests that numerosity estimation may account for some perceptual processes, together with numerical processes.

8.2 General implications and future directions

The low phenotypic correlation, as well as largely uncorrelated aetiology, between Number Line and Weber Fraction implies that, at 16 years of age, twins with low numerosity estimation abilities do not necessarily have low magnitude estimations skills. A normally developing child may start off with low estimation of numerosity skills. As individual differences in numerosity estimation and of numerical magnitudes are driven by the same genetic influences, there is a possibility that magnitude estimation skills may also be low in this child, at least early in development. However, the child may engage in activities boosting estimation of magnitudes, which will not affect numerosity estimation (the two traits are influenced by non-overlapping non-shared environments). This may explain the lack of a significant relationship between symbolic and non-symbolic estimation found in some studies (e.g. Rousselle, & Noël, 2007; Holloway & Ansari, 2009). As both abilities positively correlated with mathematical abilities, a deficit in either one of them is neither necessary nor sufficient to associate with low/high maths skills.

Further consideration can be made regarding the purpose of estimation skills. The results in Chapter 5 suggested that numerosity estimation (Weber Fraction) may be involved in the acquisition of the automatic mathematical processes that are important at the beginning of mathematical learning. Once these processes are in place, estimation of numerosity may still be relevant to mathematical learning, but may no longer be as important as estimation in the number symbols. In support of this hypothesis, a recent longitudinal study found that non-symbolic estimation measured in kindergarten children was associated with arithmetic achievement in the first but not in second grade. Symbolic estimation, measured in kindergarten, was found to predict arithmetic only in the second grade (Desoete et al., 2010). This result is congruent with the hypothesis that numerosity estimation is involved in the development of some processes relevant only for very early mathematics. In the same study, children diagnosed with mathematical difficulties, and displaying poor numerosity estimation skills when in kindergarten, still showed poor numerosity estimation in the second grade. However, they did not differ in symbolic estimation compared to normally achieving children. This last finding provides evidence of how numerosity estimation and numerical magnitudes estimation may follow different developmental trajectory, most likely due to environmental influences that are different for the two abilities.

Although this thesis is primarily concerned with normal abilities, inferences can be extended to disabilities. A further implication stemming from the findings of this thesis is that mathematical disabilities are more likely to be the product of general cognitive impairments rather than specific disability in number sense. Our results have shown a large genetic overlap between number sense and *g*. Therefore, any selective impairment in number sense (in the absence of a general cognitive impairment) stems from the influence of non-shared environments. Further research is necessary to identify specific environments involved.

As with all behavioural genetic findings, the results presented in this thesis are age and population specific. With regards to the sample, we need to consider that different settings may change the relative genetic and environmental contribution to estimation abilities. For example, studies have shown that the contribution of shared environmental factors to individual differences in mathematical abilities was greater in the WRRPM US sample compared to the UK TEDS sample (e.g. Hart et al., 2009; Kovas et al. 2007a). One of the possible causes for this discrepancy in heritability estimates is the differences in the type of mathematical instruction and teaching that children receive in the two countries. In the UK, the National Curriculum and the unified teachers' training ensure more homogenous educational standards compared to the US. It is possible that the unified educational environment in the UK reduces the inter-individual variations attributed to the shared environment.

Other samples may therefore show different heritability of estimation abilities. One reasonable question is whether Eastern cultures, whose population have higher mathematical performance compared to many Western countries (e.g. OECD, 2010b), benefit from some environments positively influencing number sense. For example, it has been suggested that languages influence the way we think and conceptualise numbers (Ito & Hatta, 2004; Dehaene et al., 1993). Based on the relationship between reading and Weber Fraction shown in Chapter 5, numerosity estimation may rely on pattern recognition processing to create the link between numerosity sets and the corresponding numerals. In the UK TEDS sample, the relationship between mathematics and number sense was largely mediated by genetic influences. However, it is possible that in samples with different languages (or other different cultural settings) the relationship between mathematics and number sense/estimation may be different. For instance, the correlation between number sense and mathematics may be mediated by environmental factors (e.g. influencing language and reading development), in addition to genetic factors. Such hypotheses can be investigated using genetically sensitive cross-cultural designs. These studies may help to advance our knowledge of mathematics and number sense processes.

They may also help to identify environments beneficial to the number sense-mathematics relationship.

The research reported in this thesis suggested that estimation processes are heterogeneous. Cross-cultural studies, using the same tools of assessment, can provide further insights into the nature of the different aspects of number sense. A large cross-cultural investigation in 3 countries, that uses twin and non-twin children of all ages, is currently underway. This investigation is adopting the battery of tests used in this thesis to address some of these issues. In addition to collecting new data, the results will be cross-analysed with existing TEDS' data in the UK. This investigation, to which the author of this thesis is contributing, aims to provide insights into those environments that are more likely to be beneficial for the development of number sense skills.

As mentioned previously, the two aspects of number sense follow different developmental trajectories. For this reason, studies addressing the relationship between number sense and mathematics both from a behavioural and behavioural genetic perspective need to use both, numerosity and magnitude estimation tasks.

This thesis addressed the origins of number sense skills within a normal range of abilities. However, low and high number sense abilities need further investigation. According to the Generalist Genes Hypothesis, the same generalist genes affect abilities and disabilities. As this investigation has supported other predictions of the Generalist Genes Hypothesis, it is possible that low and high number sense abilities are driven by the same genetic influences as normal variation. However, there is evidence that mathematical abilities in the high range are driven by some specific genetic influences (Haworth et al., 2009b). A new study, in which the author of the thesis is involved, is underway to investigate the aetiology of the relationship of mathematical, general, and number sense abilities at the low and high extremes of the distribution in the TEDS sample. Cross-cultural investigations in low and high number sense and its relationship with mathematics should also be designed. The investigation of different levels of number sense abilities seems appropriate if we consider this ability as the analogous of phonemic awareness (Gersten & Chard, 1999). As discussed in section 1.2.1, children with normal reading abilities seem to benefit more from phonemic awareness instructions compared to children with low reading abilities (Smith, Simmons, & Kameenui, 1998). Similarly, we may find different kinds of involvement of the environment with different levels of number sense skills.

Future genetically-sensitive studies in number sense will also need to consider the age of the sample. For example, twin studies have shown that the heritability of mathematics is stable across development (e.g. Kovas et al., 2007a; Haworth et al., 2007). Between the ages of 7 and 10, despite the changes in mathematical complexity, individual differences in mathematical abilities were influenced mostly by the same genetic factors. From behavioural studies we understand that estimation abilities change with development (e.g. Halberda et al., 2012; Booth & Siegler, 2006; Siegler & Opfer, 2003). It is possible that new genetic influences arise or different environments become relevant to estimation abilities potentially leading to different estimates of heritability at different ages. This study provides heritability estimates at 16 years of age, but future genetically sensitive studies should assess estimation abilities at earlier ages. This is particularly important considering that numerosity estimation may be more relevant to mathematics at the beginning of mathematical learning. Ideally, a longitudinal genetically sensitive study into the number sense and mathematics would help to understand the dynamic and nature of their relationship across development. At 16, this relationship is largely influenced by genetic factors, but at early ages the relationship between mathematics and number sense may be influenced more by environmental factors. This is a reasonable assumption since environmental factors have such strong influence on the development of number sense skills. Identifying the environmental sources driving this relationship could contribute to a more effective mathematical learning.

One of the implications of the large non-shared environmental influences on number sense is that number sense skills can be acquired and be manipulated through the environment. At the start, this investigation attempted to identify the genetic influences on estimation abilities as the unique genetic influences (unshared with reading and *g*) on individual differences in mathematics. We found no such influences. In fact, number sense showed genetic overlap with mathematics through *g*, suggesting that genetic influences on number sense may also act on such cognitive mechanisms as memory, speed of processing, and attention. Another possibility is that these genetic effects are the same as those acting on motivation and attitude towards numerical material. Motivation at early ages has been found to be associated with later mathematics; moreover this association was largely genetically mediated (Luo et al., 2011). It also has been found that some children, already at the age of 5/6 years show spontaneous focus of attention to numerosity and generally to numerical material (Hannula, Lepola, & Lehtinen, 2010). In these children, this attentional process was found to correlate positively with later arithmetic skills.

Lastly, this investigation did not find sex differences in the means or in the aetiology of individual differences in number sense, mathematics and *g* at 16 years of age. Earlier TEDS assessments showed a significant but very small male advantage in mathematical scores. Sex explained less than 2% of mathematical scores between the ages of 7 and 10 (Kovas et al., 2007a). The absence of mean sex difference at 16 could indicate that mathematical sex differences level out with development. However, the TEDS sample is not the only case where mean sex differences were not detected. The latest PISA assessment, carried out on 15 year old students in 65 countries, did not find measurable differences in mathematical performance between boys and girls in 23 countries (OECD, 2010b). It needs to be noted that the UK was not among these 23 nations. It is possible that mean sex differences are detected when mathematics is measured on a broader range of skills such as school achievement, rather than the tests used in this study. It is planned to extend this investigation in TEDS to the measures of school achievement, such as GCSE grades. Irrespective of whether small sex differences in mathematical achievement exist at this age, the aetiology individual differences was found to be the same for boys and girls at age 16 in this study, in line with results of TEDS at earlier ages. The absence of aetiological differences between the sexes in mathematics and in the abilities associated with it (number sense and *g*) suggests that any average differences in mathematical performance reported in the UK (the PISA assessment in 2009, OEDC, 2010b) may stem from completely different factors from those driving variation in these traits.

8.3 Limitations and strengths

General limitations of the twin method apply to the studies of this thesis. However, as discussed in Chapter 3, TEDS has minimised many of such concerns. For example, zygosity determination has been carried out by both questionnaire and molecular methods; and the sample has been shown as highly representative of the UK population, ensuring good generalisability of findings. Assortative mating and equal environment may be the two assumptions having significance for the studies of this thesis. Assortative mating implies non random pairing of spouses that may affect the genetic transmission from parents to offspring as well as environmental correlations. For example, partners tend to choose each other on the same level of education, which in turn correlates with *g* (Plomin et al., 2008). This phenomenon generally leads to an increase of shared environment and a decrease of genetic influences. If number sense abilities constitute a preferred trait in partners' selection, then we would observe a decrease in heritability and increase in shared environmental influence in the population. However, the analyses in this thesis showed very negligible shared environmental influence on number sense, suggesting the absence of assortative mating for this trait.

Equal environment assumption violation would be more difficult to detect, as different environments may decrease or increase the MZ and DZ twins' similarity. One environment concerning number sense abilities could be the class environment. Previous studies suggest that twins' similarity in mathematics does not increase as a function of being in the same class with the co-twin and being taught by the same teacher (Kovas et al., 2007a). This finding suggests that teachers and other aspects of the classroom environment may act as non-shared environments in influencing mathematical abilities. Although our research suggests that the non-shared environments between mathematics and number sense are unrelated, the influence of the teachers, especially at early age could be relevant for number sense. In this case, a larger (or smaller) number of MZ compared to DZ twins could lead to overestimation or underestimation of genetic influences. In the TEDS sample at the age of 7 and 9 there is similar proportion of MZ and DZ twins attending a class with their co-twin or separately. This ensures that any teacher effect does not influence the heritability estimates.

TEDS sample is homogeneous in age. This property ensured that the findings of this study were free from a developmental bias. However, as mentioned previously, the findings cannot be extended to other ages. This study has pioneered the investigation of number sense using behavioural genetic methodology; however results will need to be replicated across samples and ages to assess the change and continuity of this trait.

Furthermore the study in Chapter 5 indicated that a number of cognitive abilities had an association with mathematics across the school ages. This investigation was primarily concerned with the relationship between number sense and mathematics, also including *g* in the models. This investigation suggests that there is a small genetic variance shared between Number Line and Weber Fraction, independently from mathematics and *g*. This variance may be shared with other abilities not measured as part of this investigation. It is also possible that this variance uniquely pertains to estimation abilities. Future research should include other variables, such as reading-decoding as reading showed the most robust association with estimation of numerosity at the beginning of mathematical learning in this study.

The use of the large longitudinal representative TEDS sample in this investigation was its major strength. The study in Chapter 5 conducted the longitudinal-retrospective analyses on one twin in each pair and used the co-twins as replication sample – another major strength of the investigation. Another strength of the investigation was the extensive piloting of the measures used in the TEDS assessment. During the process of online implementation of the task, internal validity was continuously monitored. The validation study conducted on the subset of TEDS (illustrated in Chapter 4) confirmed the reliability of the battery, as all the measures showed a good test re-test reliability and internal validity.

8.4 Conclusion

This investigation has provided an unprecedented opportunity to use the large and representative UK TEDS sample to advance the knowledge of the origins of number sense abilities and the nature of their relationship with mathematics. These results chart the course for future research into the relationship of number sense and other cognitive abilities to better understand mathematical development. The number sense domain has proven to be more complicated than what previously thought. A more comprehensive understanding is likely to result from interdisciplinary, collaborative and cross-cultural investigation. Further understanding into mathematical abilities is likely to result from the cross-cultural investigations that use the same measurements and procedures. The test battery developed as part of this interdisciplinary PhD project, as well as new research directions charted by this thesis, are currently being used in new large scale cross-cultural investigations.

References

Agrawal, A., Heath, A. C., Grant, J. D., Pergadia, M. L., Statham, D. J., Bucholz, K. K. et al. (2006). Assortative mating for cigarette smoking and for alcohol consumption in female Australian twins and their spouses. *Behavioral Genetics,* 36*,* 553–566.

Agrillo, C., Dadda, M., Serena, G., & Bisazza, A. (2009). Use of number by fish. *PLoS One*, 4(3), e4786.

Agrillo, C., Piffer, L., Bisazza, A. (2011). Number versus continuous quantity in numerosity judgments by fish. *Cognition*, 119, 281-287.

Adams. M. J. (1990). *Beginning to read: Thinking and learning about print*. Cambridge, MA: MIT Press.

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika*, *52*, 317-332.

Al Aïn, S., Giret, N., Grand, M., Kreutzer, M., Bovet, D. (2009). The discrimination of discrete and continuous amounts in African grey parrots (*Psittacus erithacus*). *Animal Cognition,* 12, 145–154.

Alarcón, M., DeFries, J. C., Light, J. G., & Pennington, B. F. (1997). A twin study of mathematics disability. *Journal of Learning Disabilities,* 30, 617-623.

Alarcón, M., Knopik, V. S., & DeFries, J. C. (2000). Covariation of mathematics achievement and general cognitive ability. *Journal of School Psychology,* 38, 63-77.

Amato, P. R. (2004). The consequences of divorce for adults and children. *Journal of Marriage and Family*, 62(4), 1269-1287.

Bachot, J., Gevers, W., Fias, W., Roeyers, H. (2005). Number sense in children with visuospatial disabilities: orientation of the mental number line. *Psychology Science*, 47(1), 172 -183.

Baddeley, A. D. (1983). *Working Memory.* In Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, Vol. 302, No. 1110, Functional Aspects of Human Memory. pp. 311-324, The Royal Society Press.

Baddeley, A. D. (1992). Working Memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, 4 (3), 281-288.

Baddeley, A. D. (1996). Exploring the central executive. *Quarterly Journal of Experimental Psychology*, 49A, 5–28.

Badian, N. A. (1999). Persistent arithmetic, reading, or arithmetic and reading disability. *Annals of Dyslexia*, 49, 45–70.

Baroody, A. J. (1987). The development of counting strategies for single-digit addition. *Journal for Research in Mathematics Educatio*n, 18, 141 – 157.

Baroody, A. J., & Gatzke, M. R. (1991). The estimation of set size by potentially gifted kindergarten-age children. *Journal for Research in Mathematics Education*, 22, 59 – 68.

Barth, H.C. (2008). Judgments of discrete and continuous quantity: An illusory Stroop effect. *Cognition*, 109, 251-266.

Barth, H., La Mont, K., Lipton, J., Dehaene, S., Kanwisher, N., & Spelke, E. (2006). Non-symbolic arithmetic in adults and young children. *Cognition*, 98, 199–222.

Barth, H., La Mont, K., Lipton, J., & Spelke, E. S. (2005). Abstract number and arithmetic in preschool children. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 14116-14121.

Barth, H., Kanwisher, N., & Spelke, E. S. (2003). The construction of large number representations in adults. *Cognition*, 86, 201–221.

Beran, M. J. (2008). The evolutionary and developmental foundations of mathematics. *PLoS Biology*, 6, e19.

Berch, D. B. (2005). Making sense of number sense: implications for children with mathematical disabilities. *Journal of Learning Disabilities,* 38 (4), 333–339.

Besner, D., & Coltheart, M. (1979). Ideographic and alphabetic processing in skilled reading of English. *Neuropsychologia,* 17(5), 467–472.

Bijeljac-Babic, R., Bertoncini, J., Mehler, J. (1993). How Do 4-Day-Old Infants Categorize Multisyllabic Utterances? *Developmental psychology,* 29 (4), 711-721.

Boker, S., Neale M., Maes, H., Wilde, M., Spiegel, M., Brick, T., Spies, J., Estabrook, R., Kenny, S., Bates, T., Mehta, P, & Fox J. (2011). OpenMx: an open source extended structural equation modeling framework. *Psychometrika*, 76(2), 306-317.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology,* 42*,* 189–201.

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

Brannon, E. M., Abbott, S., Lutz, D. J. (2004). Number bias for the discrimination of large visual sets in infancy, *Cognition*, 93, B59–B68.

Breukelaar, J. W. C., & Dalrymple-Alford, J. C. (1998). Timing ability and numerical competence in rats. *Journal of Experimental Psychology: Animal Behavior Process*, 24, 84–97.

Bull, R., Johnston, R. S., & Roy, J. A. (1999). Exploring the roles of the visual-spatial sketch pad and central executive in children's arithmetical skills: Views from cognition and developmental neuropsychology. *Developmental Neuropsychology*, 15, 421–442.

Bull, R., & Johnston, R. S. (1997). Children's arithmetical difficulties: Contributions from processing speed, item identification, and short-term memory. *Journal of Experimental Child Psychology*, 65, 1–24.

Butterworth, B. (2005). The development of arithmetical abilities. *Journal of Child Psychology and Psychiatry*, 46(1), 3–18.

Butterworth, B., Zorzi, M., Girelli, L., & Jonckheere, A. R. (2001). Storage and retrieval of addition facts: The role of number comparison. *Quarterly Journal of Experimental Psychology, 54*A, 1005–1029.

Bruder, C.E. G, Piotrowski, A., Gijsbers, A. A. C. J., et al. (2008). Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. *The American Journal of Human Genetics*, 82, 763–771.

Byrne B, Coventry WL, Olson RK, Wadsworth SJ, Samuelsson S, Petrill SA et al. (2010). "Teacher Effects" in early literacy development: evidence from a study of twins. *Journal of Educational Psychology*; 102(1), 32-42.

Byrne, B., Samuelsson, S., Wadsworth, S. et al. (2006). Longitudinal twin study of early literacy development: Preschool through Grade 1. *Reading and Writing: An Interdisciplinary Journal*, 20, 77-102.

Byrne, B., Wadsworth, S., Corley, R. et al. (2005). Longitudinal twin study of early literacy development: Preschool and kindergarten phases. *Scientific Studies of Reading*, 9, 219-235.

Carnine, D. (1991). Reforming mathematics instruction: The role of curriculum materials. *Journal of Behavioral Education*, 1, 37–57.

Carpenter, T. P., & Moser, J. M. (1984). The acquisition of addition and subtraction concepts in grades one through three. *Journal for Research in Mathematics Education*, 15(3), 179–202.

Case, R., Kurland, D. M., & Goldberg, J. (1982). Operational efficiency of short term memory span. *Journal of Experimental Psychology,* 33, 386–404.

Cantlon, J., & Brannon, E. M. (2007). How much does number matter to a monkey (Macaca mulatta)? *Journal of Experimental Psychology: Animal Behavior Processes*, 33(1), 32–41.

Case, R., & Sowder, J. T. (1990). The development of computational estimation: A Neo-Piagetian analysis. *Cognition and Instruction, 7(2),* 79–104.

Chabris, C. F., Hebert, B. M., Benjamin, D. J., Beauchamp, J. P., Cesarini, D., van der Loos, . . . Laibson, D. (in press). Most reported genetic associations with general intelligence are probably false positives. *Psychological Science.*

Chen, J., Li, X., Chen, Z., Yang, X., Zhang, J., Duan, Q., et al. (2010). Optimization of zygosity determination by questionnaire and DNA genotyping in Chinese adolescent twins. *Twin Research and Human Genetics, 13*, 194-200.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A.M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: Follow-up study. *British Medical Journal,* 333*,* doi:10.1136/bmj.38959.650903.7C.

Clearfield, M. W., & Mix, K. S. (1999). Number versus contour length in infants' discrimination of small visual sets. *Psychological Science*, 10, 408–411.

Cohen Kadosh, R., Bahrami, B., Walsh, V., Butterworth, B., Popescu, T., & Price, C. J. (2011.)Specialization in the human brain: the case of numbers. *Frontiers in Human Neuroscience*, 5:62. doi: 10.3389/fnhum.2011.00062.

Cordell, H. J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Human Molecular Genetics*, 11(20), 2463-2468.

Corsi. P. (1972). Human memory and the medial temporal region of the brain. *Dissertation Abstracts International, 34* (02), 891B. (University Microfilms No. AAI05 -77717).

Cortina. J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology,* 78, 98–104.

Cunningham. A. (1990). Explicit vs. implicit instruction in phonemic awareness. *Journal of Experimental Child Psychology*, 50, 426–444.

D'Amico, T. A., Guarnera M. (2005). Exploring working memory in children with low arithmetical achievement. *Learning and Individual Differences*, 15, 189–202.

Davies, G., Tenesa, A., Payton, A., Yang, J., Harris, S. E., Liewald, D., . . . Deary, I. J. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry,*16*,* 996-1005.

Davis, H., & Memmott, J. (1982). Counting behavior in animals: A critical evaluation. *Psychological Bulletin*, 92, 547–571.

Davis, H., & Perusse, R. (1988). Numerical competence in animals: Definitional issues, current evidence and a new research agenda. *Behavioral and Brain Sciences*, 11, 561–579.

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009a). Dramatic increase in heritability of cognitive development from early to middle childhood an 8-year longitudinal study of 8,700 pairs of twins. *Psychological Science*, 20(10), 1301-1308.

Davis, O. S. P., Haworth, C. M. A., & Plomin, R. (2009b). Learning abilities and disabilities: Generalist genes in early adolescence. *Cognitive Neuropsychiatry*, 14, 312-331.

Davis, O. S. P., Arden, R., & Plomin, R. (2008b). g in middle childhood: Moderate genetic and shared environmental influence using diverse measures of general cognitive ability at 7, 9 and 10 years in a large population sample of twins. *Intelligence*, 36, 68-80.

Davis, O. S. P., Kovas, Y., Harlaar, N. et al. (2008a). Generalist genes and the Internet generation: etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior,* 7, 455-462.

Deary, I. J., Der, G., Ford, G. (2001). Reaction times and intelligence differences. A population-based cohort study. *Intelligence*, 29, 389–399.

Deary, I. J., Yang, J., Davies, G., Harris, S. E., Tenesa, A., Liewald, D., . . . Visscher, P. M. (2012). Genetic contributions to stability and change in intelligence from childhood to old age. *Nature, 482,* 212-215.

Dehaene, S. (1997). *The number sense: How the mind creates mathematics*. New York: Oxford University Press.

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

Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and number magnitude. *Journal of Experimental Psychology: General*, 122(3), 371–396.

Dehaene, S., Mehler, J. (1992). Cross-linguistic regularities in the frequency of number words. *Cognition*, 43, 1–29.

Dehaene, S., Piazza, M., Pinel, P., Cohen, L. (2003). Three parietal circuits for number processing. *Cognitive Neuropsychology*, 20, 487–506.

Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics*, 9, 403-411.

De Smedt, B., & Gilmore, C.K. (2011). Defective number module or impaired access? Numerical magnitude processing in first graders with mathematical difficulties. *Journal of Experimental Child Psychology*, 108(2), 278-292.

Desoete, A., Ceulemans, A., De Weerdt, F., Pieters, S. (2010). Can we predict mathematical learning disabilities from symbolic and non-symbolic comparison tasks in kindergarten? Findings from a longitudinal study.

British Journal of Educational Psychology; doi: 10.1348/2044-8279.002002.

Dirks, E., Spyer, G., van Lieshout ,E. C. D. M., & de Sonneville, L. (2008). Prevalence of Combined Reading and Arithmetic Disabilities *Journal of Learning Disabilities,* 41, 460-473.

Dowker, A. (2003). Young children's estimates for addition: The zone of partial knowledge and understanding. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: Constructing adaptive expertise* (pp. 243–266). Mahwah, NJ: Erlbaum.

Dowker A. (2005). Early identification and intervention for students with mathematics difficulties. *Journal of Learning Disabilities*, 38 (4), 324-332.

Duncan, G. J., Dowsett, C. J., Claessens, A., Magnuson, K., Huston, A. C., Klebanov, P., et al. (2007). School readiness and later achievement. *Developmental Psychology*, 43, 1428–1446.

Dunn, J. F., & Plomin, R. (1990). *Separate lives: Why siblings are so different.* New York: Basic Books.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor — The structure of adolescents achievement task values and expectancy-related beliefs. *Personality and Social Psychology Bulletin*, 21(3), 215–225.

Evans, D. M., & Martin, N. G. (2000). The validity of twin studies. *GeneScreen*, 1(2), 77-79.

Fagard, R.H., Loos, R.J., Beunen, G., Derom, C., & Vlietinck, R. (2003). Influence of chorionicity on the heritability estimates of blood pressure: A study in twins. *Journal of Hypertension*, 21, 1313–1318.

Falconer, D. S., & MacKay, T. F. C. (1996). *Introduction to Quantitative Genetics* (4th ed.). Longmans Green, Harlow, Essex, UK.

Feigenson, L. Carey, S., & Hauser, M. (2002). The representations underlying infants' choice of more: object-files versus analog magnitudes. *Psychological Science*, 13, 150–156.

Feigenson, L., Carey, S., & Spelke, E. S. (2002). Infants' discrimination of number vs. continuous extent. *Cognitive Psychology*, 44, 33–66.

Feigenson, L., Dehaene, S. & Spelke, E. S. (2004). Core systems of number. *Trends in Cognitive Science,* 58, 307–314.

Fias, W., Lammertyn, J., Reynvoet, B., Dupont, P., & Orban, G. A. (2003). Parietal representation of symbolic and nonsymbolic magnitude. *Journal of Cognitive Neuroscience*, 15(1), 47-56.

Fias, W., Lauwereyns, J., & Lammertyn, J. (2001). Irrelevant digits affect feature-based attention depending on the overlap of neural circuits. *Cognitive Brain Research*, 12(3), 415-23.

Fischer, M. H., & Rottmann, J. U. L. I. A. (2005). Do negative numbers have a place on the mental number line. *Psychology Science*, 47(1), 22-32.

Fuchs, L. S., Fuchs, D., & Prentice, K. (2004). Responsiveness to mathematical problem- solving instruction: Comparing students at risk of mathematics disability with and without risk of reading disability. *Journal of Learning Disabilities*, 37, 293– 306.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., & Bryant J. D. (2010a). The contributions of numerosity and domain general abilities to school readiness. *Child Development,* 81(5), 1520–1533.

Fuchs, L. S., Geary, D. C., Compton, D. L., Fuchs, D., Hamlett, C. L., Seethaler, P. M., Bryant, J. D., Schatschneider, C. (2010b). Do different types of school mathematics development depend on different constellations of numerical versus general cognitive abilities? *Developmental Psychology*, 46(6), 1731-1746.

Furst, A. J., & Hitch, G. J. (2000). Separate roles for executive and phonological components of working memory in mental arithmetic. *Memory an Cognition*, 28(5), 774–782.

Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, 44, 43–74.

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

Geary, D. C. (1993). Mathematical disabilities: Cognitive, neuropsychological, and genetic components. *Psychological Bulletin,* 114, 345–362.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities*, 37(1), 4-15.

Geary, D. C. (2011). Cognitive predictors of achievement growth in mathematics: a 5-year longitudinal study. *Developmental Psychology*, 47(6), 1539–1552.

Geary, D. C., Bow-Thomas, C. C., & Yao, Y. (1992). Counting knowledge and skill in cognitive addition: A comparison of normal and mathematically disabled children. *Journal of Experimental Child Psychology,* 54, 372–391.

Geary, D. C., Bailey, D. H., & Hoard, M. K. (2009). Predicting mathematical achievement and mathematical learning disability with a simple screening tool: the number sets test. *Journal of Psychoeducational Assessment*, 27(3), 265-279.

Geary, D. C., Brown, D. C., & Samaranayake, V. A. (1991). Cognitive addition: A short longitudinal study of strategy choice and speed-of-processing differences in normal and mathematically disabled children. *Developmental Psychology,* 27, 787–797.

Geary, D. C., Hoard, M. K, Byrd-Craven, J., Nugent, L., & Numtee, C. (2007). Cognitive mechanisms underlying achievement deficits in children with mathematical learning disability. *Child Development*, 78 (4), 1343 – 1359.

Geary, D. C., Hoard, M. K., & Hamson, C. O. (1999). Numerical and arithmetical cognition: Patterns of functions and deficits in children at risk for a mathematical disability. *Journal of Experimental Child Psychology,* 74*,* 213–239.

Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: a longitudinal study of process and concept deficits in children with learning disability, *Journal of Experimental Child Psychology*, 77, 236–263.

Geary, D. C., Hoard, M. K., Nugent, L., & Byrd-Craven, J. (2008). Development of number line representation in children with mathematical learning disability. *Developmental Neuropsychology*, 33(3), 277–299.

Geary, D. C., & Wiley, J.G. (1991). Cognitive addition: strategy choice and speed of processing differences in young and elderly adults. *Psychology and Aging,* 6 (3), 474-483.

Gebuis, T., Cohen Kadosh, R., de Haan, E., & Henik, A. (2009). Automatic quantity processing in 5-year olds and adults. *Cognitive Processing*, 10(2), 133–142.

Gersten, R., & Chard, D. (1999). Number sense: Rethinking arithmetic instruction for students with mathematical disabilities. *Journal of Special Education,* 33(1), 18 –28.

Gelman, R., & Gallistel, C. R. (1978). *The child's understanding of number* (1986 edn). Cambridge, MA: Harvard University Press.

Gersten, R., Jordan, N. C., & Flojo, J. R. (2005). Early identification and interventions for students with mathematics difficulties, *Journal of Learning Disabilities,* 38 (4), 293-304.

Gilmore, C. K., McCarthy, S. E., Spelke, E. S. (2010). Non-symbolic arithmetic abilities and mathematics achievement in the first year of formal schooling. *Cognition*, 115: 394–406.

Gibbon, J. (1977). Scalar expectancy theory and Weber's law in animal timing. *Psychological Review*, 84(3), 279-325.

Gibbon, J., & Church, R. M. (1981). Time left: linear versus logarithmic subjective time. *Journal of Experimental Psychology. Animal Behavior Processes,* 7, 87–107.

Ginsburg, H. P. (1997). Mathematics learning disabilities: A view from developmental psychology. *Journal of Learning Disabilities*,30, 20 – 33.

GOAL plc. (2002).GOAL Formative Assessment: Key Stage 3. London: Hodder & Stoughton.

Gosling, S. D., Vazire, S., Srivastava, S., John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59, 93–104.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence,* 24, 13–23.

Gottfried, A. E. (1985). Academic intrinsic motivation in elementary and junior-highschool students. *Journal of Educational Psychology*, 77(6), 631–645.

Griffin, S. (2002). *The development of math competence in the preschool and early school years: Cognitive foundations and instructional strategies*. In J. M. Roher (Ed.), Mathematical cognition (pp. 1–32). Greenwich, CT: Information Age Publishing.

Grilo, C. M., & Pogue-Geile, M. F. (1991). The nature of environmental influences on weight and obesity: A behavior genetic analysis. *Psychological Bulletin,* 10*,* 520-537.

Gross, J., Hudson, C., Price, D. (2009).*The Long Term Costs of Numeracy Difficulties*. London. Every Child a Chance Trust and KPMG.

Gross-Tsur, V., Manor, O., & Shalev, R. S. (1996). Developmental dyscalculia: Prevalence and demographic features. *Developmental Medicine and Child Neurology, 38,* 25–33.

Guay, F., Marsh, H. W., & Boivin, M. (2003). Academic self-concept and academic achievement: Developmental perspectives on their causal ordering. *Journal of Educational Psychology*, 95(1), 124–136.

Gullo, D., Burton, C. (1992). Age of entry, preschool experience and sex as antecedents of academic readiness in kindergarten. *Early Childhood Research Quarterly*, 7, 175–186.

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

Halberda, J., Ly, R., Wilmer, J. B. , Naiman, D. Q., & Germine, L. (2012). Number sense across the lifespan as revealed by a massive Internet-based sample. *PNAS* 109 (28) 11116-11120, doi:10.1073/pnas.1200196109.

Halberda, J., Mazzocco, M. M., & Feigenson, L. (2008). Individual differences in non-verbal number acuity correlate with maths achievement. *Nature*, 455, 665–668.

Hale, J. B., Fiorello, C. A., Kavanagh, J. A., Hoeppner, J. B., & Gaither, R. A. (2001). WISC-III predictors of academic achievement for children with learning disabilities: Are global and factor scores comparable? *School Psychology Quarterly*, 16, 31–55.

Halpern, D. F. (2000). *Sex Differences in Cognitive Abilities*. Erlbaum Associates, N.J.

Hammill, D. D., Brown, V. L., Larsen, S. C., & Wiederholt, J. L. (1994). Test of Adolescent and Adult Language (TOAL-3). Austin, TX, Pro-Ed.

Hannula, M. M., Lepola, J., & Lehtinen, E. (2010). Spontaneous focusing on numerosity as a domain-specific predictor of arithmetical skills. *Journal of Experimental Child Psychology*, 107, 394–406.

Harlaar, N., Hayiou-Thomas, M. E., & Plomin, R. (2005). Reading and General Cognitive Ability: A Multivariate Analysis of 7-Year-Old Twins. *Scientific Studies of Reading*, 9(3), 197–218.

Hart, S. A., Petrill, S. A., Kamp Dush, C. M. (2010a). Genetic influences on language, reading, and mathematics skills in a national sample: an analysis using the National Longitudinal Survey of Youth. *Language, Speech, and Hearing*

Services in Schools, 41, 118-128.

Hart, S. A., Petrill, S. A., Thompson, L. A. (2010b). A factorial analysis of timed and untimed measures of mathematics and reading abilities in school aged twins. *Learning and Individual Differences*, 20, 63-69.

Hart, S. A, Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of maths: a genetic analysis and its links with reading ability and general cognitive ability. *Journal of Educational Psychology,* 101 (2), 388-402.

Haworth, C. M. A., Dale, P. S., & Plomin, R. (2009b). Generalist genes and high cognitive abilities. *Behavior Genetics,* 39, 437-445.

Haworth, C. M. A., Harlaar, N., Kovas, Y., Davis, O. S. P., Oliver, B. R. Hayiou-Thomas, M. E., Frances, J., Busfield, P., McMillan, A., Dale, P. S., & Plomin, R. (2007a). Internet cognitive testing of large samples needed in genetic research. *Twin Research and Human Genetics*, 10(4), 554–563.

Haworth, C. M. A., Kovas, Y., Dale, P. S., & Plomin, R. (2008). Science in elementary school: generalist genes and school environments. *Intelligence*, 36(6), 694-701.

Haworth, C.M.A., Kovas, Y., Harlaar, N., Hayiou-Thomas, M.E., Petrill, S.A., Dale, P.S., & Plomin, R. (2009a). Generalist genes and learning disabilities: a multivariate genetic analysis of low performance in reading, mathematics, language and general cognitive ability in a sample of 8000 12-year-old twins. *Journal of Child Psychology and Psychiatry*, 50 (10), 1318-1325.

Haworth, C. M. A., Kovas, Y., Petrill, S. A. and Plomin, R. (2007). Developmental origins of low mathematics performance and normal variation in twins from 7 to 9 years. *Twin Research and Human Genetics,* 10,106-117.

Haworth, C. M. A., Wright, M. J., Luciano, M. et al. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry, 15*, 1112-1120. doi: 10.1038/mp.2009.55

Hayiou-Thomas, M. E., & Dale, P.S., in collaboration with Snowling, M. (available from theauthors)

Heathcote, D. (1994). The role of visuo-spatial working memory in the mental addition of multi-digit addends. *Current Psychology of Cognition*, 13, 207–245.

Hecht, S. A. (1998). Toward an information-processing account of individual differences in fraction skills. *Journal of Educational Psychology*, 90(3), 545-549.

Hetherington, E. M., & Clingempeel, W. G. (1992). *Coping with Marital Transitions: A Family Systems Perspective*. Monographs of the Society for Research in Child Development, 2-3, Serial No. 227.

Hitch, G. J., & McAuley, E. (1991). Working memory in children with specific arithmetical learning difficulties*. British Journal of Psychology*, 82, 375–386.

Hoard, M. K., Geary, D. C., & Hamson, C. O. (1999). Numerical and arithmetical cognition: performance of low- and average-IQ children. *Mathematical Cognition*, 5(1), 65-91.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgements of academic achievement: A review of literature. *Review of Educational Research*, 59(3), 297–313.

Howie, B. N., Donnelly, P., & Marchini, J. (2009). A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genetics, 5*: e1000529.

Hulshoff Pol, H. E., Posthuma, D., Baaré, W. F. C., De Geus, E. J. C., Schnack, H. G., van Haren, N. E. M., van Oel, C. J., Kahn, R. S., and Boomsma, D. I. (2002). Twin–singleton differences in brain structure using structural equation modelling. *Brain*, 125, 284-390.

Hurewitz, F., Gelman, R., & Schnitzer, B. (2006). Sometimes area counts more than number. *Proceedings of the National Academy of Sciences*, 103 (51), 19599–19604.

Hyde, J., Fennema, E., Lamon, S. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin,* 107, 139–155.

Ito, Y., & Hatta, T. (2004). Spatial structure of quantitative representation of numbers: Evidence from the SNARC effect. *Memory & Cognition*, *32*(4), 662–673.

Jacobs, N., Gestel, S.V., Derom, C., Thiery, E., Vernon, P.A., Derom, R. et al. (2001). Heritability estimates of intelligence in twins: Effect of chorion type. *Behavior Genetics, 31,* 209–217.

Jensen, A. R. (1978). Genetic and behavioural effects of nonrandom mating. In R.T. Osbourne, C.E. Noble, & N. Weyl (Eds.), *Human variation: The biopsychology of age, race, and sex* (pp. 51–105). New York: Academic Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Joram, E., Subrahmanyam, K., Gelman, R. (1998). Measurement estimation: learning to map the route from number to quantity and back. *Review of Educational Research, 68(4), 413–449.*

Jordan, N., Hanich, L., & Kaplan, D. (2003). A longitudinal study of mathematical competencies in children with specific mathematics difficulties versus children with comorbid mathematics and reading difficulties. *Child Development,* 74, 834– 850.

Jordan, N. C., Kaplan, D., Locuniak, M. N., & Ramieni, C. (2007). Predicting first-grade math achievement from developmental number sense trajectories. *Learning Disabilities Research & Practice,* 22(1), 36–46.

Jordan, N. C., Kaplan, D., Oláh, L. N., & Locuniak, M. N. (2006). Number Sense growth in kindergarten: a longitudinal investigation of children at risk for mathematics difficulties. *Child Development*, 77(1), 153–175.

Jordan, K. E., Suanda, S. H., Brannon. E. M. (2008). Intersensory redundancy accelerates preverbal numerical competence. *Cognition*, 108, 210–221.

Kail, R. (1991). Developmental change in speed of processing during childhood and adolescence. *Psychological Bulletin, 109*, 490–501.

Kail, R. (1992). Processing speed, speech rate, and memory. *Developmental Psychology,* 28, 899-904.

Kaminsky, Zachary A., Tang, T., Wang, S. , et al. (2009). DNA methylation profiles in monozygotic and dizygotic twins. *Nature Genetics*, 41(2), 240-245.

Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1999), WISC-III as a Process Instrument (WISC-III-PI), New York: The Psychological Corporation.

Kavale, K. A. (1984). Potential advantages of the meta-analysis technique for special education. *The Journal of Special Education*, 19, 443–458.

Kavale, K. A., & Forness. S. R. (2000). What definitions of learning disability say and don't say: a critical analysis. *Journal of Learning Disabilities*, 33, 239-256.

Kendler, K.S., Neale, M.C., Kessler, R.C., Heath, A.C., & Eaves, L.J. (1993a). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23,* 21–27.

Kirk, S. A., & Bateman, B. D. (1962). Diagnosis and remediation of learning disabilities. *Exceptional Children*, 29, 73-78.

Knopik, V. S., & DeFries, J. C. (1999). Etiology of covariation between reading and mathematics performance: A twin study. *Twin Research,* 2, 226–234.

Kovas, Y., Harlaar, N., Petrill, S. A., & Plomin, R. (2005). Generalist genes and mathematics in 7-year- old twins. *Intelligence,* 33, 473-489.

Kovas, Y., Haworth, C.M.A., Dale, P.S., & Plomin, R. (2007a). The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development,* 72, vii–160.

Kovas, Y., Haworth, C. M. A., Harlaar, N., Petrill, S. A., Dale, P. S., & Plomin, R. (2007d). Overlap and specificity of genetic and environmental influences on mathematics and reading disability in 10-year-old twins. *Journal of Child Psychology and Psychiatry*, 48, 914-922.

Kovas, Y., Haworth, C.M.A., Petrill, S. A., & Plomin, R. (2007b). Mathematical ability of 10-year-old boys and girls: genetic and environmental etiology of typical and low performance. *Journal of Learning Disabilities,* 40, 554-567.

Kovas, Y., Petrill, S. A., & Plomin, R. (2007c).The origins of diverse domains of mathematics: generalist genes but specialist environments. *Journal of Educational Psychology*, 99(1), 128-139.

Kovas, Y., & Plomin, R. (2006). Generalist genes: Implications for cognitive sciences. *Trends in Cognitive Science*, 10, 198-203.

Kranzler, J. H., Jensen, A.R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence,* 13, 329–347.

Lachance, J., Mazzocco, M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences*, 16, 195–216.

Landerl, K., Bevan, A., & Butterworth, B. (2004). Developmental dyscalculia and basic numerical capacities: A study of 8–9-yearold students. *Cognition*, 93, 99–125.

Landerl, K., Fussenegger, B., Moll, K., & Willburger, E. (2009). Dyslexia and dyscalculia: Two learning disorders with different cognitive profiles. *Journal of Experimental Child Psychology*, 103, 309–324.

Leahey, E., & Guo, G. (2001). Gender differences in mathematical trajectories. *Social Forces,* 80, 713–732.

Lee, K. -M., & Kang, S. -Y. (2002). Arithmetic operation and working memory: Differential suppression in dual tasks. *Cognition*, 83, B63–B68.

Lee, S. H., Wray, N. R., Goddard, M. E., & Visscher, P. M. (2011). Estimating missing heritability for disease from genome-wide association studies. *American Journal of Human Genetics, 88,* 294-305.

LeFevre, J. A., Fast, L., Skwarchuk, S. L., Smith-Chant, B. L., Bisanz, J., Kamawar, D., & Penner-Wilger, M. (2010). Pathways to mathematics: Longitudinal predictors of performance. *Child Development*, 81(6), 1753-1767.

LeFevre, J. A., Greenham, S. L., & Naheed, N. (1993). The development of procedural and conceptual knowledge in computational estimation. *Cognition and Instruction*, 11(2), 95–132.

Lemaire, P., Abdi, H., & Fayol, M. (1996). The role of working memory resources in simple cognitive arithmetic. *European Journal of Cognitive Psychology,* 8(1), 73–103.

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

Levine, S. C., Jordan, N. C., & Huttenlocher, J. (1992). Development of calculation abilities in young children. *Journal of Experimental Child Psychology*, 53, 72–103.

Lewis, C., Hitch, G. J., & Walker, P. (1994). The prevalence of specific arithmetic difficulties and specific reading difficulties in 9-year-old to 10-year-old boys and girls. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 35,* 283–292.

Libertus, M. E., Brannon, E. M. (2009). Behavioral and neural basis of number sense in infancy. *Current Directions in Psychological Science*, 18, 346–351.

Libertus, M. E., & Brannon, E. M. (2010). Stable individual differences in number discrimination in infancy. *Developmental Science*, 13(6), 900–906.

Light, J. G., DeFries, J. C., & Olson, R. K. (1998). Multivariate behavioral genetic analysis of achievement and cognitive measures in reading-disabled and control twin pairs. *Human Biology*, 70, 215-237.

Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense. Large number discrimination in human infants. *Psychological Science*, 14, 396–401.

Locuniak, M. N., & Jordan, N. C. (2008). Using kindergarten number sense to predict calculation fluency in second grade. *Journal of Learning Disabilities,* 41(5), 451–459.

Loehlin, J. C. (1996). The Cholesky approach: A cautionary note. *Behavior Genetics, 26*, 65-69.

Loehlin, J. C., & Nichols, R. (1976). *Heredity, environment and personality*. Austin & London: University of Texas Press.

Loehlin, J. C., Harden, K. P., & Turkheimer, E. (2009). The effect of assumptions about parental assortative mating and genotype-income correlation on estimates of genotype-environment interaction in the National Merit twin study. *Behavior Genetics*, 39, 165-169.

Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working memory in arithmetic problem solving. *Memory and Cognition*, 22(4), 395–410.

Lourenco, S. F., Bonny, J. W., Fernandez, E. P., &Rao, S. (2012). Nonsymbolic number and cumulative area representations contribute shared and unique variance to symbolic math competence. *Proceedings of the National Academy of Sciences, USA,* 109, 18737-18742.

Lourenco, S. F., Longo, M. R. (2009). Multiple spatial representations of number: evidence for co-existing compressive and linear scales. *Experimental Brain Research*, 193, 151 – 156.

Lubke, G. H., Hottenga, J. J., Walters, R., Laurin, C., de Geus, E. J., Willemsen, G., . . . Boomsma, D. I. (2012). Estimating the genetic variance of major depressive disorder due to all single nucleotide polymorphisms. *Biological Psychiatry.* Advanced online publication. doi: 10.1016/j.biopsych.2012.03.011.

Luo, Y. L. L., Kovas, Y., Haworth, C. M., Dale, P. S., & Plomin, R. (2011). The etiology of mathematical self-evaluation and mathematics achievement: Understanding the relationship using a cross-lagged twin study from ages 9 to 12. *Learning and Individual Differences,* 21, 710-718.

MacGillivray, I., Campbell, D. M., & Thompson, B. (1988). *Twinning and twins.* Chichister, UK: John Wiley & Sons.

Markowitz, E. M., Willemsen, G., Trumbetta, S. L., van Beijsterveldt, T. C. E. M., & Boomsma, D. I. (2005). The etiology of mathematical and reading (dis)ability covariation in a sample of Dutch twins. *Twin Research and Human Genetics,* 8, 585-593.

Markwardt, F. C. (1997), *Peabody Individual Achievement Test - Revised, Circle* Pines, MN: American Guidance Service.

Marsh, H. W., & Yeung, A. S. (1997). Causal effects of academic self-concept on academic achievement: Structural equation models of longitudinal data. *Journal of Educational Psychology*, 89(1), 41-54.

Martin, N. G., & Eaves, L. J. (1977). The genetical analysis of covariance structure. *Heredity*, 38, 79–95.

Martin, N. G., Jardine, R., & Eaves, L. J. (1984). Is there only one set of genes for different abilities? A reanalysis of the National Merit Scholarship Qualifying Tests (NMSQT) data. *Behavior Genetics, 14*, 355-370.

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

McCarthy, D. (1972). *McCarthy Scales of Children's Abilities*; New York: The Psychological Corporation.

McGraw, K. O., Tew, M. D., Williams, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, 11, 502–506.

McGue, M., & Bouchard, T. J. (1984). Adjustment of twin data for the effects of age and sex. *Behavior Genetics, 14,* 325–343.

McKrink, K., & Wynn, K. (2004).Large number addition and subtraction by 9-month-old infants. *Psychological Science*. 15 (11), 776 – 781.

McLean, J. F., & Hitch, G. J. (1999). Working memory impairments in children with specific arithmetic learning difficulties. *Journal of Experimental Child Psychology*, 74(3), 240- 260.

Meck, W. H., Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, 9(3), 320–334.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex frontal lobe tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100.

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

Muller, C. (1998). Gender differences in parental involvement and adolescents' mathematics achievement. *Sociology of Education*, 71,336–356.

Mullis, I.V.S., Martin, M.O., Gonzalez, E.J., O'Connor, K.M., Chrostowski, S.J., Gregory, K.D., Garden, R.A., & Smith,T.A. (2001). *Mathematics benchmarking report: The Third International Math and Science Study—Eighth Grade*.Boston, MA: Boston College International Study Center.

Murphy, M. M., & Mazzocco, M. M. M (2008). Mathematics learning disabilities in girls with Fragile X or Turner Syndrome during late elementary school. *Journal of Learning Disabilities*, 41 (1), 29-46.

Murphy, M. M., Mazzocco, M. M. M. Hanich, L. B., & Early, M. C. (2007). Cognitive characteristics of children with mathematics learning disability (MLD) vary as a function of the cutoff criterion used to define MLD. *Journal of Learning Disabilities*, 40(5), 458–478.

Neale, M. C. & Maes, H. H. M. (2003). *Methodology for genetic studies of twins and families.* Dordrecht, The Netherlands: Kluwer Academic Publishers B.V.

Neisser, U., Boodoo, G., Bouchard Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., ... & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American psychologist*, 51(2), 77

Neubauer, A. C. (1997). The mental speed approach to the assessment of intelligence. In J. Kingma, & W. Tomic, *Advances in cognition and educational practice: Reflections on the concept of intelligence,* (pp. 149- 174). Greenwich, Connecticut: JAI Press.

Nieder, A., & Miller, E. K. (2003). Coding of cognitive magnitude: compressed scaling of numerical information in the primate prefrontal cortex. *Neuron*, 37, 149–157.

Neil, P. A., Chee-Ruiter, C., Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, 9, 454–464.

Nuerk, H. C., Iversen, W., & Willmes, K. (2004). Notational modulation of the SNARC and the MARC (linguistic markedness of response codes) effect. *Quarterly Journal of Experimental Psychology Section A*, 57(5), 835-863.

Nuerk, H. C., Weger, U., & Willmes, K. (2001). Decade breaks in the mental number line? Putting the tens and units back in different bins. *Cognition*, 82(1), B25-B33.

Nuerk, H. C., Wood, G., Willmes, K. (2005). The universal SNARC effect: the association between number magnitude and space is amodal. *Experimental Psychology*, 52(3), 187 – 194.

nferNelson. (1994). Maths 5–14 series. London: nferNelson Publishing Company Ltd.

nferNelson. (1999). Maths 5–14 series. London: nferNelson Publishing Company Ltd.

nferNelson. (2001). Maths 5–14 series. London: nferNelson Publishing Company Ltd.

Nys, J., Content, A. (2012). Judgement of discrete and continuous quantity in adults: Number counts! *The Quarterly Journal of Experimental Psychology*, 65(4), 675-690.

OECD. (2010a). *The High Cost of Low Educational Performance: The Long-Run Economic Impact of Improving Educational Outcomes*. Paris. OECD.

OECD. (2010b). *PISA 2009 Results: What Students Know and Can Do – Student Performance in Reading, Mathematics and Science (*Volume I). http://dx.doi.org/10.1787/9789264091450-en.

Oliver, B., Harlaar, N., Hayiou-Thomas, M. E. et al. (2004). A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology, 96,* 504-517.

Oliver, B., & Plomin, R. (2007). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence. *Twin Research and Human Genetics*, 10(1), 96−105, doi:10.1375/twin.10.1.96.

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

Pagulayan, K. F., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental Normative Data for the Corsi Block-Tapping Task. *Journal of Clinical and Experimental Neuropsychology*. 28 (6),1043 - 1052.

Passolunghi, M. C., & Siegel, L. S. (2001). Short-term memory, working memory and inhibitor control in children with difficulties in arithmetic problem solving. *Journal of Experimental Child Psychology*, 80, 44–57.

Paterson, S. J., Girelli, L., Butterworth, B., & Karmiloff-Smith, A. (2006). Are numerical impairments syndrome specific? Evidence from Williams syndrome and Down's syndrome. *Journal of Child Psychology and Psychiatry*, 47, 190–204.

Penner, A. M., Paret, M. (2008). Gender differences in mathematics achievement: Exploring the early grades and the extremes. *Social Science Research*, 37, 239–253.

Petrill, S. A., Kovas, Y., Hart, S. A., Thompson, L. A., Plomin, R. (2009). The genetic and environmental etiology of high math performance in 10-year-old twins. *Behavior Genetics,* 39 (4), 371-379.

Phillips, D.I.W. (1993). Twin studies in medical research: Can they tell us whether diseases are genetically determined? *Lancet*, 341, 1008–1009.

Piazza, M., Izard, V., Pinel, P., Le Bihan, D., Dehaene, S. (2004). Tuning curves for approximate numerosity in the human intraparietal sulcus. *Neuron*, 44(3), 547–55.

Piazza, M., Mechelli, A., Price, C. J., Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, 177-188.

Pica, P., Lemer, C., Izard, V., Dehaene, S. (2004). Exact and approximate arithmetic in an Amazonian indigene group. *Science*, 306, 499-503.

Pisa, P. E., & Agrillo, C. (2009). Quantity discrimination in felines: A preliminary investigation of the domestic cat (Felis silvestris catus). *Journal of Ethology*, 27(2), 289–293.

Plomin, R., & Daniels, D. (1987). Why are children in the same family so different from each other? *Behavioral and Brain Sciences, 10,* 1-16.

Plomin, R., DeFries, J. C., McClearn, G. E. and McGuffin, P. (2008). *Behavioral Genetics*. 5th ed. New York: Worth Publishers.

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2013). Behavioral genetics (6th edition). New York: Worth.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin,* 131, 592–617.

R Development Core Team. (2011). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. R Foundation for Statistical Computing.

Price, T. S., Freeman, B., Craig, I., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3*, 129-133.

The National Curriculum. (2004). *The National Curriculum. Handbook for secondary teachers in England*. Qualifications and Curriculum Authority.

Race, J. P., Townsend, G. C., & Hughes, T. E. (2006). Chorion type, birthweight discordance and tooth-size variability in Australian monozygotic twins. *Twin Research and Human Genetics*, 9(2), 285–291.

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales*, Oxford: Oxford University Press.

Raven, J., Raven, J.C., & Court, J.H. (1998). *Mill Hill vocabulary scale*. Oxford: Oxford University Press.

Rathbun, A., West, J., Germino-Hausken, E. (2004). *From Kindergarten Through Third Grade: Children's Beginning SchoolExperiences*. National Center of Education Statistics, Washington, DC.

Reznikova, Z., & Ryabko, B. (2011). Numerical competence in animals, with an insight from ants. *Behaviour,* 148, 405–434.

Rijsdijk, F. V., & Sham, P. C. (2002). Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*, 3(2), 119-133.

Ronalds, G. A., De Stavola, B. L., & Leon, D. A. (2005). The cognitive cost of being a twin: Evidence from comparisons within families in the Aberdeen children of the 1950s cohort study. *British Medical Journal,* 331(7528): 1306.

Rodic, M. Zhou, X., Tikhomirova, T. , Wei, W., Malykh, S., Ismatulina, V. , Sabirova, E., Davidova, Y., Tosto, M. , Lemelin, J-P. , & Kovas, Y. (submitted). Cross-cultural perspective on cognitive underpinnings of individual differences in early mathematics. *Developmental Science*.

Roitman, J. D., Brannon, E. M., & Platt, M. L. (2007). Monotonic coding of numerosity in macaque lateral intraparietal area. *PLoS biology*, 5(8), e208.

Rousselle, L., & Noël, M. P. (2007). Basic numerical skills in children with mathematics learning disabilities: A comparison of symbolic vs. non-symbolic number magnitude processing. *Cognition*, 102, 361–395.

Rousselle, L., & Noël, M. P. (2008). The development of automatic numerosity processing in preschoolers: Evidence for numerosity-perceptual interference. *Developmental Psychology*, 44(2), 544-560.

Rubinsten, O. (2009). Co-occurrence of developmental disorders: The case of developmental dyscalculia. *Cognitive Development*, 24, 362−370.

Rubinsten, O., Henik, A., Berger, A., & Shahar-Shalev, S. (2002). The development of internal representations of magnitude and their association with Arabic numerals. *Journal of Experimental Child Psychology*, 81(1), 74–92.

Rugani, R., Fontanari, L., Simoni, E., Regolin, L., & Vallortigara, G. (2009). Arithmetic in newborn chicks. *Proceedings of the Royal Society B- Biological Sciences*, 276, 2451–2460.

Rugani, R., E., Regolin, L., & Vallortigara, G. (2011). Summation of Large Numerousness by Newborn Chicks. *Frontiers in Psychology*, 2, 179. doi: 10.3389/fpsyg.2011.00179.

Rutter, M., & Redshaw, J. (1991). Annotation: Growing up as a twin: twin-singleton differences in psychological development. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 32*, 885-895.

Sadker, M., Sadker, D. (1994). *Failing at Fairness: How America's Schools Cheat Girls*. Charles Scribner's Sons, New York.

Saxe, G.B. Gearhart, M., & Nasir, N. S.(2001). Enhancing students' understanding of mathematics: a study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education* , 4 (1) 55-79.

Saxe, G. B., Guberman, S. R., Gearhart, M., Gelman, R., Massey, C. M., & Rogoff, B. (1987). *Social processes in early number development.* Monographs of the Society for Research in Child Development, 52(2), 162.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics, 9,* 527–542.

Shalev, R. S., Manor, O., & Gross-Tsur, V. (2005). Developmental dyscalculia: A prospective six-year follow-up. *Developmental Medicine & Child Neurology*, 47, 121 – 125.

Shrout, P. E., & Fleiss, J. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86, 420−428.

Siegler, R. S. (1987). The perils of averaging data over strategies: An example from children's addition. *Journal of Experimental Psychology: General,* 116, 250–264.

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

Siegler, R. S., & Mu, Y. (2008). Chinese children excel on novel mathematics problems even before elementary school. *Psychological Science,* 19, 759 – 763.

Siegler, R. S., & Opfer, J. E. (2003). The development of numeral estimation: evidence for multiple representations of numerical quantity. *Psychological Science*, 14, 237–243.

Siegler, R. S., & Ramani, G. B. (2008). Playing linear numerical board games promotes low-income children's numerical development. *Developmental Science*, 11(5), 655– 661.

Siegler, R. S., & Ramani, G. B. (2009). Playing linear numerical board games - but not circular one - improves low-income preschoolers' numerical understanding. *Journal of Educational Psychology*, 101(3), 454– 560.

Siegel, L. S., & Ryan, E. B. (1989). The development of working memory in normally achieving and subtypes of learning disabled children. *Child Development*, 60, 973–980.

Smith, P., Fernandes, C., & Strand, S. (2001). *Cognitive Abilities Test 3 (CAT3).* Windsor: nferNELSON.

Smith, S. B.,Simmons. D. C., & Kameenui, E. J. (1998). Phonological awareness: Research bases. In D. C. Simmons &: E. J. Kameenui (Eds,). *What reading research, tells about children with diverse learning needs: Bases and basics* (pp. 61-127). Mahwah NJ; Erlbaum.

Sokol, D. K., Moore, C. A., Rose, R. J., Williams, C. J., Reed, T., Christian, J. C. (1995). Intrapair differences in personality and cognitive ability among young monozygotic twins distinguished by chorion type. *Behavior Genetics,* 25(5), 457-66.

Soltész, F., Szűcs, D., Szűcs. L. (2010). Relationships between magnitude representation, counting and memory in 4- to 7-year-old children: A developmental study. *Behavioral and Brain Functions*, 6(13). doi:10.1186/1744-9081-6-13.

Sowder, J. T., & Wheeler M. M. (1989). The development of concepts and strategies used in computational estimation. *Journal for Research in Mathematics Education,* 20(2), 130 –146.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science. *A critical review. American Psychologist*, *60*(9), 950-958.

Spelke, E. S., & Grace, A. D. (2006). Abilities, motives, and personal styles. *American Psychologist*, 61(7), 725.

Spuhler, J. N. (1968). Assortative mating with respect to physical characteristics. *Eugenics Quarterly, 15,* 128–140.

Starkey, P., Klein, A., & Wakeley, P. (2004). Enhancing young children's mathematical knowledge through a pre-kindergarten mathematics intervention. *Early Childhood Research Quarterly*, 19, 99 – 120.

Stevens, J., Wood, J., Hauser, M. (2007). When quantity trumps number: discrimination experiments in cotton-top tamarins (*Saguinus oedipus*) and common marmosets (*Callithrix jacchus*). *Animal Cognition,* 10 , 429–437.

Stevenson, H. W., & Newman, R. S. (1986). Long-term prediction of achievement and attitudes in mathematics and reading. *Child Development, 57,* 646–659.

Steffens, M. C., Jelenec, P., & Noack, P. (2010). On the leaky math pipeline: comparing implicit math-gender stereotypes and math withdrawal in female and male children and adolescents. *Journal of Educational Psychology*, 102(4), 947–963.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption and linkage studies. *Language*, 77, 647-723.

Swanson, H. L., & Sachse-Lee, C. (2001). Mathematical problem solving and working memory in children with learning disabilities: Both executive and phonological processes are important. *Journal of Experimental Child Psychology,* 79, 294–321.

Tanner, J. (1978). *Foetus into Man: Physical Growth from Conception to Maturity*. Harvard Univ. Press, Cambridge, MA.

Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science*, 2, 158–165.

Thompson, C. A., & Opfer, J. E. (2008). Costs and benefits of representational change: Effects of context on age and sex differences in symbolic magnitude estimation. *Journal of Experimental Child Psychology*, 101(1), 20-51.

Tiu, R. D., Jr., Wadsworth, S. J., Olson, R. K., & DeFries, J. C. (2004). Causal models of reading disability. *Twin Research*, 7, 275–283.

Torgesen, J. K. Morgan, S. T., & Davis, C. (1992). Effects of two types of phonological awareness on word learning in kindergarten children. *Journal of Educational Psychology*, 84, 364–370.

Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999): *TOWRE (Test of Word Reading Efficiency)*. Austin, Texas: Pro-Ed.

Tzelgov, J., Meyer, J., & Henik, A. (1992). Automatic and intentional processing of numerical information. *Journal of Experimental Psychology: Learning, Memory & Cognition*, 18(1), 166–179.

Uller, C., Jaeger, R., Guidry, G., & Martin, C. (2003). Salamanders (Plethodon cinereus) go for more: Rudiments of number in a species of basal vertebrate. *Animal Cognition*, 6, 105–112.

Van Selst, M., Jolicoeur, P. (1994). A solution to the effect of sample size on outlier elimination. *Quarterly Journal of Experimental Psychology Section A*, 47(3), 631–650.

Vinkhuyzen, A. A. E., van der Sluis, S., Maes, H. H. M., & Posthuma, D. (2012b). Reconsidering the heritability of intelligence in adulthood: Taking assortative mating and cultural transmission into account. *Behavior Genetics, 42,* 187-98.

Visscher, P. M., Medland, S. E., Ferreira, M. A. R., Morley, K.I., Zhu, G., Cornes, B. K., Montgomery, G. W., Martin, N. G. (2006). Assumption-free estimation of heritability from genome-wide identity-by-descent sharing between full siblings. *PLoS Genet*ics, 2(3), e41.

Vukovic,R. K., Lesaux, N.K., & Siegel, L. (2010). The mathematics skills of children with reading difficulties. *Learning and Individual Differences*, 20, 639-643.

Vukovic, R., K., Siegel, L. S. (2010). Academic and cognitive characteristics of persistent mathematics difficulty from first through fourth Grade*. Learning Disabilities Research & Practice*, *25*(1), 25–38.

Walker, A., Maher, J., Coulthard, M., Goddard, E., & Thomas, M. (2001). *Living in Britain: Results from the 2000/2001 General Household Survey.* London: TSO.

Weber, E. H. (1834). *De Pulsu, Resorptione, Auditu et Tactu: Annotationes Anatomicae et Physiologicae.* Koehler, Leipzig, Germany.

Wechsler, D. (1992): *Wechsler intelligence scale for children* (3rd Ed. UK); The Psychological Corporation.

Whalen, J., Gallistel, C. R., & Gelman, I. I. (1999). Nonverbal counting in humans: the psychophysics of number representation. *Psychological Science,* 2, 130–137.

Wiig, E. H., Secord, W., & Sabers, D. (1989). *Figurative language subtest. From the test of language competence – Expanded Edition*. San Antonio, TX, The Psychological Corporation.

Wood, J.N., Hauser, M.D., Glynn, D.D., Barner, D. (2008). Free-ranging rhesus monkeys spontaneously individuate and enumerate small numbers of non-solid portions. *Cognition,* 106, 207–221.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III.* Itasca, IL: Riverside Publishing.

Xu, F. (2003). Numerosity discrimination in infants: Evidence for two systems of representations. *Cognition*, 89(1), B15–B25.

Xu, F., & Arriaga, R. (2007). Number discrimination in 10-month-old infants. *British Journal of Developmental Psychology,* 25*,* 103–108.

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

Xu, F., Spelke, E. S., & Goddard, S. (2005). Number sense in human infants. *Developmental Science*, 8 (1), 88–101.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011a). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics,* 88*,* 76-82.

Yang, J., Manolio, T. A., Pasquale, L. R., Boerwinkle, E., Caporaso, N., Cunningham, J. M., . . . Visscher, P. M. (2011b). Genome partitioning of genetic variation f or complex traits using common SNPs. *Nature Genetics*, 43, 519-525.

Zorzi, M., Priftis, K., Umiltá, C. (2002). Brain damage: Neglect disrupts the mental number line. *Nature*, 417, 138–139.

Appendix 1: Information sheet validation study

Goldsmiths
UNIVERSITY OF LONDON

Numerical Ability in 16-year olds: Tests validation study

We would like to invite you to take part in a special project conducted by Dr Yulia Kovas and Maria Grazia Tosto at Goldsmiths College. The information below should help you understand why this study is being done and what it will involve. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

Thank you for reading this.

<u>The purpose of the study</u>

Little is known about why people differ in their capacity to recognise and evaluate numerosity. We are planning a large study that aims to improve our understanding in this area, which will involve thousands of 16-year olds. This study will also help us to understand whether this ability of recognising numerosity and the ability to approximate are related to the development of mathematical ability. Ultimately, the results of this large study might help us to understand why some people struggle with mathematics and others do well in it. We have developed a battery of tests that assess numerical ability, but before we can start the large study, we need a 100 volunteers to help us to evaluate how well this battery of tests works and to find the best way of administering these tests.

<u>Do I have to take part?</u>

It is up to you to decide whether or not you would like to take part in this study. If you do decide to take part you will be asked to sign a consent form. We will also ask your parent or guardian to provide consent for your participation by signing the consent form. You are still free to withdraw from the study at any time and without giving a reason. We would like to emphasise that the participation in this research is entirely voluntary.

What does taking part involve?

We will make two appointments with you. At the first appointment, we will ask you to complete several tests on a computer in the presence of one researcher. The researcher will come to your school; alternatively, you can come to Goldsmiths College if you prefer. The procedure should take between 30 and 60 minutes. Approximately one months later, we will send you a web-link and ask you to complete a battery of tests on the internet using log-in information that we will provide. This procedure should take approximately 30 minutes.

Reward

We will give you two £10 Love2Shop vouchers (one for each assessment) as appreciation of the time and effort that you devoted to our research. These vouchers can be redeemed in many high street stores, including HMV, Waterstone's, and New Look, as well as some leisure activities. You can look up Love2Shop vouchers on the internet or ask us for more information. In addition, if you choose to come to Goldsmiths College for the first assessment, we will reimburse all your travel expenses.

Will my taking part in this study be kept confidential?

All information collected during the course of the research will be kept strictly confidential. The data that you will provide will be entered into a dataset together with other participants' data, and you will be identified in this dataset only by participant number. Your name and contact information will be stored separately and will be deleted from our records as soon as you have completed both assessments and have received the reward, or if you choose to withdraw from the study earlier.

What will happen to the results of the research study?

It will take us a few weeks to collect the data from all 100 participants and to analyze the results. Once we have done that, we will choose the tests and methods that have worked the best in order to conduct the study with thousands of participants. This bigger study will take a couple of years to complete, and once the results are available, they will be published in reputable journals.

Goldsmiths
UNIVERSITY OF LONDON

Who is organising and funding the research?

The validation study in which we are asking you to participate is organised by Dr Yulia Kovas. Goldsmiths College is funding this project.

Who has reviewed the study?

The Ethics Committee of Goldsmiths College has reviewed the ethical aspects of this study.

Contacts for Further Information

Please contact Dr Yulia Kovas and Maria Grazia Tosto for further information (see contact details below).

If you would like to participate in this study, please complete and sign the consent form. Your parent or guardian should also sign this form if agree to your participation.

Dr Yulia Kovas
Department of Psychology
Goldsmiths, University of London
SE 14 6NW
Tel. (0)207 078 5025
Fax (0)207 919 7873

Maria Grazia Tosto
Department of Psychology
Goldsmiths, University of London
SE 14 6NW
Tel. (0)207 7919 7171 4286
e-mail : ps_n_16@gold.ac.uk

Appendix 2: Consent form pilot study

CONSENT FORM: Numerical Ability in 16-year olds: Tests validation study

Please tick appropriate box:

Participant

☐ **Yes, I would like to participate in this study.**

☐ **No, I do not want to participate in this study.**

Parent/Guardian

☐ **Yes, I would like my child to participate in this study.**

☐ **No, I do not want my child to participate in this study.**

Participant

If Yes, please complete the following:

☐ I have read the Information Sheet about the study.

☐ I understand that I do not have to take part in this study if I do not want to.

☐ I understand that I can withdraw from the study at any time without giving a reason.

☐ I have had the opportunity to ask any questions I wish to ask.

☐ I have access to the names and telephone numbers of the research team in case I have any questions.

Participant's Name: ____

Participant's Signature: _________________________ Date: _____________

Goldsmiths
UNIVERSITY OF LONDON

Parent's/Guardian's Name: ____________________________________

Parent's/Guardian's Signature: _____________________ Date: _____________

Contacts for further information:

Dr Yulia Kovas ,Department of Psychology, Goldsmiths, University of London SE 14 6NW

Tel. (0)207 078 5025

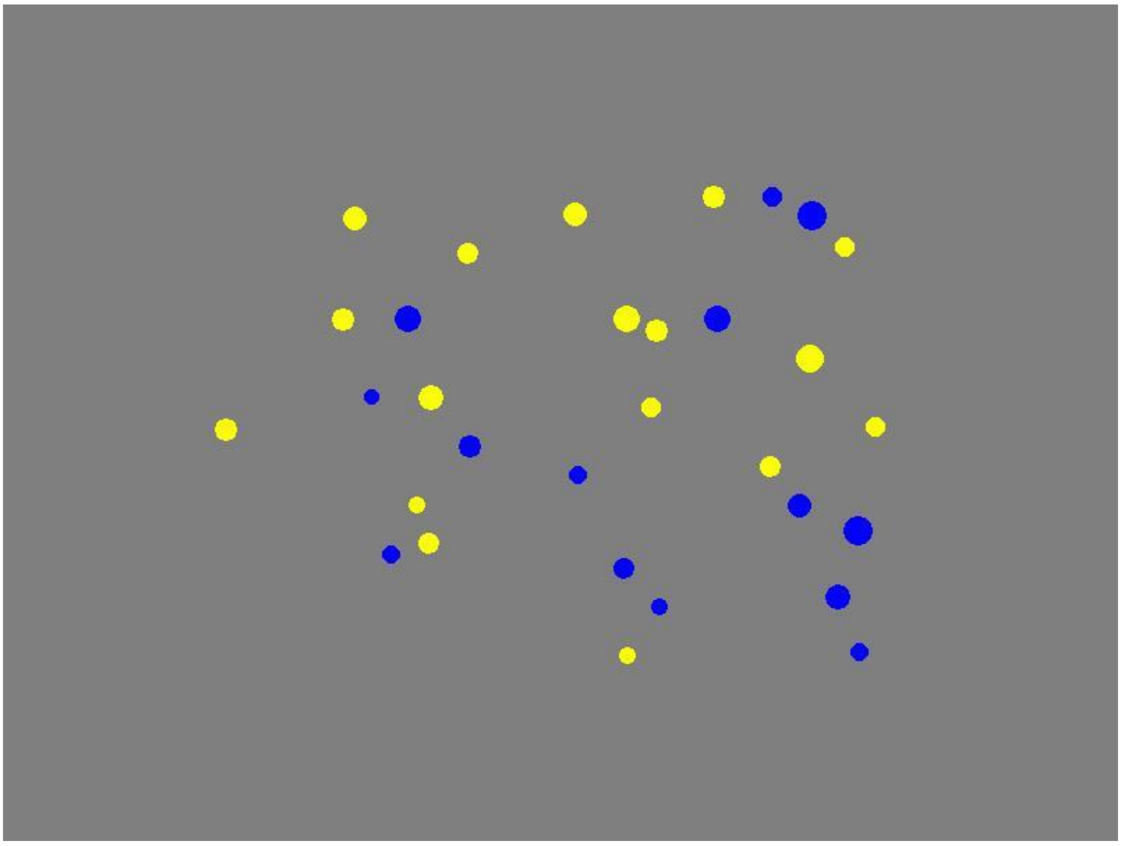
Fax (0)207 919 7873

Maria Grazia Tosto Department of Psychology Goldsmiths, University of London SE 14 6NW

Tel. (0)207 078 4286

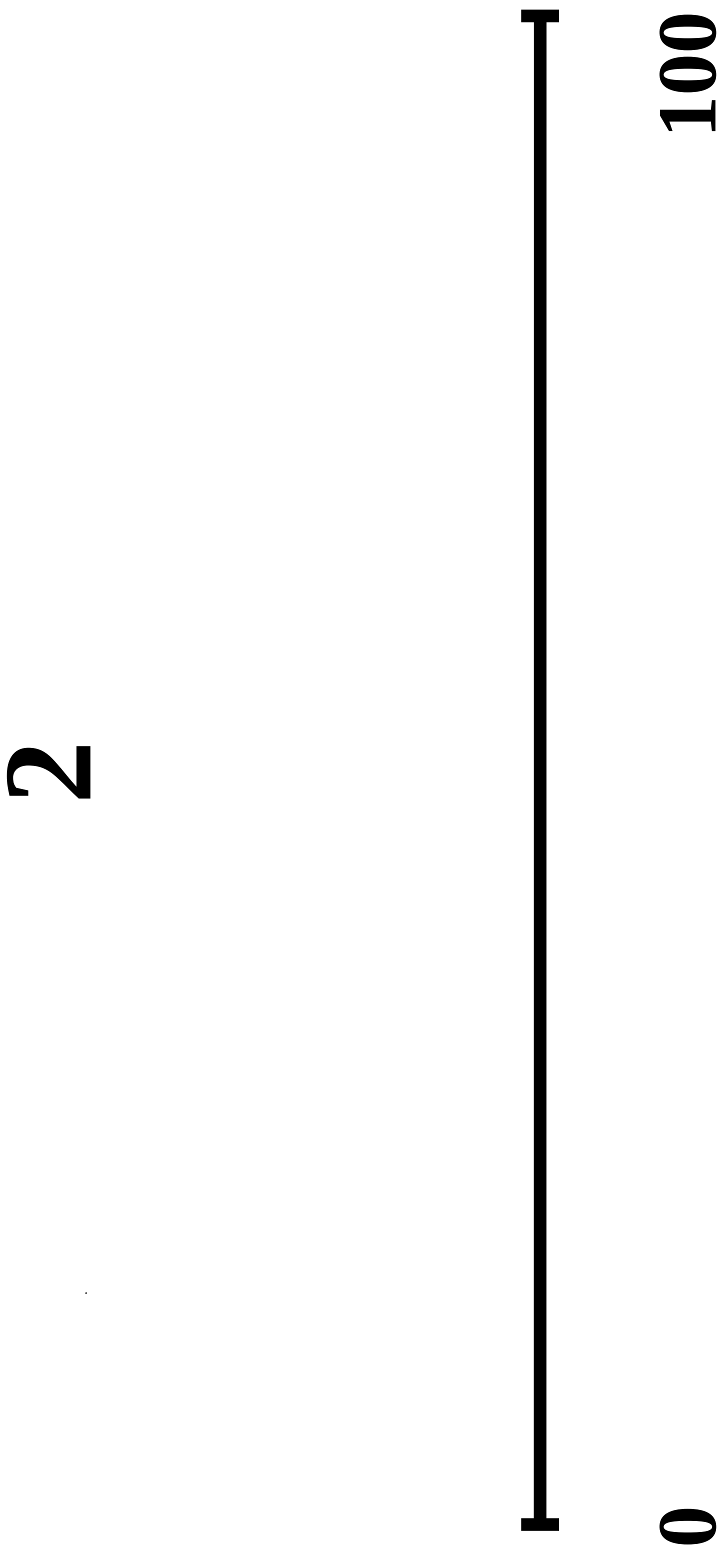
e-mail : ps_n_16@gold.ac.uk

Appendix 3: Trial sample Dot Task

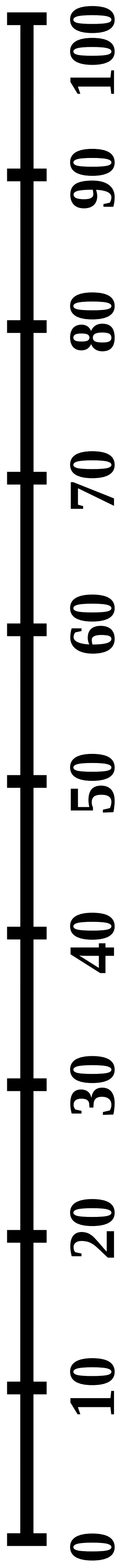


17 yellow and 14 blue dots

Appendix 4: Number Line test trial

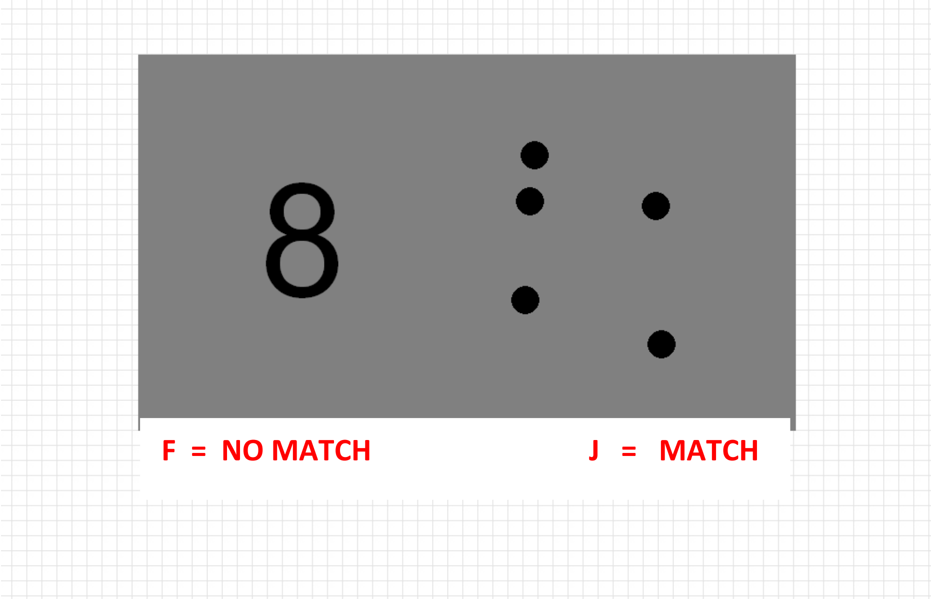


Appendix 5: Number Line practice trial

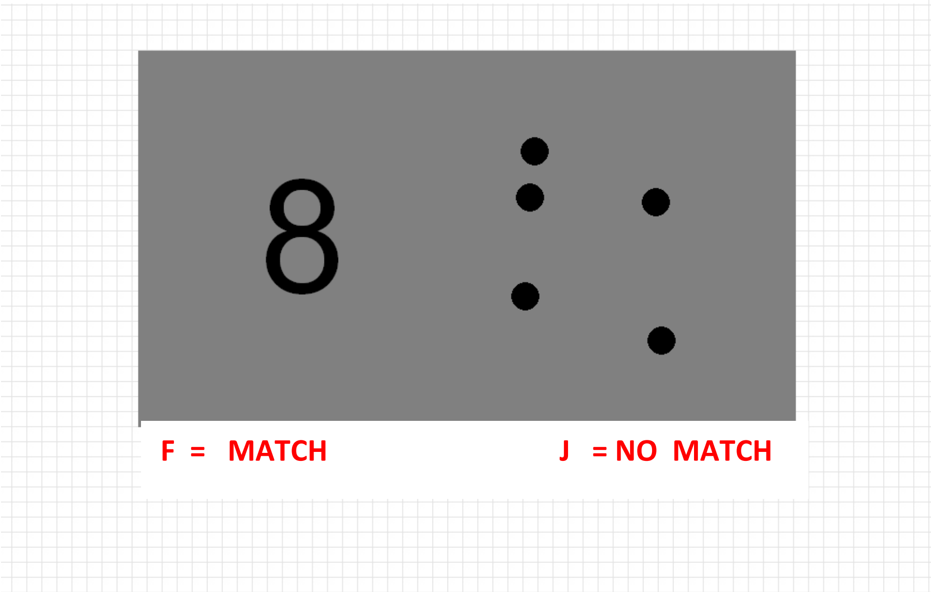


Appendix 6: Dot Matching test trial sample

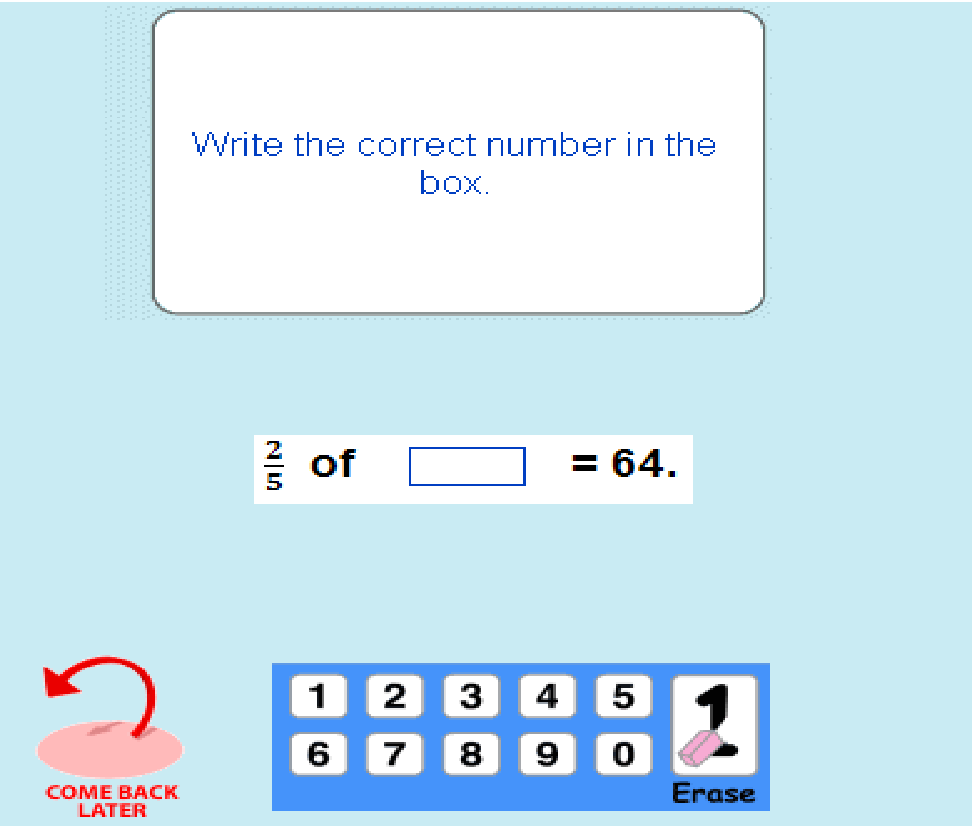
Presentation trial right handed participants



Presentation trial left handed participants



Appendix 7: Understanding Number test trial sample



Appendix 8: Problem Verification Task trial sample