The Formal Disciplinary Value of Advanced Mathematical Study: A Focus on Spatial Skills

Sara Humphries

A thesis submitted for the degree of
Doctor of Philosophy

Loughborough University

Contents

List of Figures

List of Tables

Acknowledgements

I would like to thank both of my supervisors, Matthew Inglis and Ian Jones, for all of their support throughout my PhD. In additional, everyone at the Mathematics Education Centre at Loughborough University has always been happy to give me guidance when needed. I would also like to thank all of the schools and sixth form colleges for kindly inviting me to collect data with their students.

Declaration

I, the author, declare that the work presented in this thesis is my own and has not been submitted for a degree at any other institution. None of the work has previously been published in this form.

Overview

With the current school leaving age of children having been raised to 18 years old for those students born after 1st September 1997 (Department for Education, 2014b), research into the benefits that this extra emphasis on mathematics education has for students is essential. One justification used by the Department for Education is that employers in the UK place the status of advanced mathematical skills very highly when selecting candidates for jobs and promotions. Research conducted for the Conservative Government into the state of mathematics education in the UK in 2011 reported that 24% of economically active adults were "functionally innumerate" and that employers felt that the mathematics skills of most school leavers were not adequate (Vorderman et al., 2011). In order to address these shortcomings, and in an attempt to ensure that all students leave compulsory education with the skills required to have a successful career, students will have to continue in some type of education or training for longer. As part of these plans, students that fail to achieve a mathematics GCSE of at least a grade C must continue some level of mathematics education until the age of 18. Matthew Hancock, the Conservative Party Under-Secretary for further education, skills, and lifelong learning, is quoted as saying that

> "For those who fail to get a grade C at GCSE, it's a huge impairment to their future life, their ability to participate not just in work but also as a citizen" (BBC, 2013).

The notion that finishing key stage 4 education with a grade D or below in mathematics could be a 'huge impairment' to the future life of an individual is a true statement, but not necessarily due to the amount of mathematics content that has been learnt and retained. The grade C cut off point that is imposed by colleges and employers as an entry requirement, and the emphasise that the UK Government places on 'grade C or above' as a universal measure of success makes it inevitable that students falling below this cut-off will suffer greatly.

What is an interesting additional to Hancock's reasoning is that this lack of mathematical knowledge will affect a person's ability to participate as a citizen; a notion that is used as a justification for the high-level of complex mathematical content that makes up the GCSE mathematics examinations. Although very few students are likely to ever use Pythagoras' Theorem, or nth term sequences outside of their mathematics classroom, the idea that learning and practising these topics will improve some more general reasoning and problem solving skills is a belief held by educators, policy makers, teachers, and students. It is, perhaps, for this reason that employers are so keen for their potential employees to possess a 'grade C or above' in mathematics: because it suggests that they have successfully acquired skills that can be transferred to resolving work-based issues effectively.

To function as a citizen on a day-to-day basis, an average person is unlikely to use mathematical skills much more advanced than simple arithmetic. Few children leave primary school lacking the skills to compute simple additions and subtractions, multiplications and divisions and certainly, without the majority of the general pubic being fluent in these skills, civilised society would struggle. It is true that possessing some knowledge of fractions and percentages would be of an advantage if a person is faced with mortgages, or loan payments, but this person will, most likely, have access to some tool that will help with calculations, and to experts that are employed to offer advice in these situations. Very few careers require mathematical knowledge up to the level that is taught at compulsory GCSE. Even aircraft designers and roofing contractors do not derive trigonometrical laws, or solve complex equations every day; the formulas that they use require little more than simple arithmetic to employ (Dudley, 2010).

The idea that a mathematics education has the capacity to teach children not only mathematics, but also how to function as a citizen is not a modern one. Throughout the history of mathematics education, theorists and philosophers have referred to an assumed "higher-level cognitive advantage" of learning mathematics and used this assumption to validate and defend the teaching of mathematics. The current Government's enthusiasm for students to continue to study mathematics for more hours a week, and for more years than other equivalent school subjects, is partially rooted in the idea that this will prepare them for life after school in some way additional to the mathematical content taught. In a report to the UK Government, Professor Adrain Smith endorses mathematical study for its own sake because

> "mathematical training disciplines the mind, develops logical and

> critical reasoning, and develops analytical and problem-solving skills to a high degree." (Smith (2004))

Considering the strength of belief in this claim and the influence that it has over the way in which UK schools are run, there is surprisingly little evidence to support it. The intuitive idea that learning to think in a 'mathematical way' will improve the way in which real-life problems are approached is difficult to argue with, as it seems a rational claim. This thesis documents research that aims to enrich the evidence base for these claims in order to further inform decisions that are made concerning mathematics education.

This additional worth that is potentially learnt through mathematical study is a concept that educational psychologists have coined as a 'formal discipline' value. The theory of formal discipline is based on the premise that the inclusion of particular subjects in a school's syllabus is justified by the mental capacities that it trains, rather than what is being learnt. Seymour Papert, a psychologist who worked with Jean Piaget, in a speech about 21st Century education, stated that:

> "We need to produce people who know how to act when they're faced with situations for which they were not specifically prepared" (Papert, 1998).

It is thought by supporters of the formal discipline theory that the learning of subjects such as mathematics has the potential to train the brain in reasoning skills, or problem solving (Stanic, 1986) and that these skills can be transferred for use in everyday situations outside of school.

To study the ideas of formal discipline in a scientific way, it is, of course, necessary to have an outcome measure; a construct that is expected to enrich the students' life above and beyond the content of their school subjects studied. A number of different measures will be discussed throughout this thesis, from the crude measurements of the early 1900s to the highly refined IQ testing available today, but the main focus will be that of spatial reasoning, based on its strong associations with mathematics in the literature, and its links to success in science and engineering careers. This thesis will begin by covering the literature, past and current, in relation to the theory of formal discipline, links between spatial reasoning and mathematics, and explore the possible associations between them. The literature discussed will show that students that choose to study mathematics do excel in certain cognitive constructs, particularly spatial, and the data collected for Studies One and Two of this thesis will shed light on whether this advantage is due to an effect of the study of advanced mathematics, in line with the theory of formal discipline, or whether individuals with higher levels of these constructs are more likely to choose to study mathematics at an advanced level: a filtering effect that results in individuals with higher levels of spatial skills being more likely to enrol themselves into advanced mathematical education. Study Three further explores the nature of the relationship between mathematics and spatial skills.

Chapter 1

Formal discipline: A literature review

The theory of formal discipline is based on the premise that the purpose of educating children is not to just teach the content of the subject matter, but to train useful life skills and cognitive abilities through the act of studying itself.

In the late 1800s, it was believed by educators that the human brain was comprised of distinct faculties that could be strengthened and improved through 'exercise and correct use'. As a consequence of this understanding of human cognition, the role of education was to exercise the intellectual capacity of the brain to the point that it would be strong enough to control students' wills and emotions. This preparation would leave students equipped with the skills needed to deal with every aspect of life outside of education (Brooks, 1883). The study of Latin was thought to strengthen skills in memory and, by learning geometry, a student could improve their reasoning ability (Henderson, 1911). School subjects that were thought to be the most useful and effective in exercising the faculties of the brain made up what was recognised as a Liberal Arts Education; the only type of education available in the Western world until roughly 1870 (Schmidt, 1958).

At the turn of the 20th Century, the concept of faculty psychology and the formal discipline value of education were being called into question by educators and psychologists. In particular, a series of studies conducted by behavioural psychologist Edward Thorndike cast doubt on the feasibility of transferring skills from one domain to another (Thorndike and Woodworth, 1901; Woodworth and Thorndike, 1901). Thorndike and Woodworth's studies found that, when participants were trained in one mental function (for example, estimating the areas of triangles), transfer to even very similar tasks (for example, estimating the areas of rectangles) was minimal. Thorndike and Woodworth concluded from this set of studies that transfer of learning was only possible when the contexts shared 'identical elements' and, if the elements were anything but completely identical, transfer was not a possibility.

As the beliefs and concepts of behaviourist psychology became the principle school of thought in educational research during the first half of the 20th Century, a number of studies sought to determine the extent to which the transfer of skills was, in fact, a feasible concept and, in turn, whether this transfer of skills could be applied to the learning that takes place through the study of school subjects.

Whether or not the transfer of training in one area to performance in another is possible continues to be a question that is asked by more contemporary psychologists. The viability of transfer of skills is in no way agreed upon in the literature and the following section reviews relevant research in this area.

1.1 Viability of transfer of skills

Transfer of skills differs from ordinary learning in that it requires a change of context between the learning and performance stages. Within this classification, most examples of learning can be said to involve an element of transfer, for example, skills learned in a classroom can be transferred to performance in an examination situation. This type of transfer is commonly referred to as 'near transfer' and, being clearly possible in many settings, is of little interest to researchers (Perkins and Salomon, 1992). What is of more interest, particularly in an educational setting, is the idea of 'far transfer', described by Perkins and Salomon as the "transfer between contexts that, on appearance, seem remote and alien to one another" (p.4). For far transfer to occur, skills from one context are required to be abstracted and generalised in order for them then to be applied successfully in a different context. This is the type of transfer that is required for the concept of formal discipline to have any strength as a theory and is one of the main aims of the education system. For there to be a benefit of students studying advanced geometry, for example, the skills obtained through the learning of the geometry must transfer to other useful situations, or the knowledge is nearly useless.

Although far transfer cannot be assumed, and apparently very often fails to happen, situations in which it does occur can be thought to share some important elements which should be strived towards for maximum success. Perkins and Salomon (1992) classified the following as being features associated with successful far transfer, based on a systematic review of the transfer literature:

- *Thorough and diverse practice* — If reasoning skills are going to be learnt through school mathematics, and transferred to unlike contexts, students must have experience of reasoning in many mathematical situations, e.g. spatially, algebraically, and therefore begin to recognise the skills that are common to all.
- *Abstraction* — The suggestion of Thorndike and Woodworth (1901) that transfer is only possible when the contexts share 'identical elements' can be applied to far transfer when the elements under consideration are assumed to be highly abstract. If a student is able to abstract the skills that are common to both solving a classroom-based mathematical problem, and solving a work-place unrelated problem, then these abstracted elements are identical enough to be viably transferred.
- *Explicit abstraction* — When students are more fully aware of the specific principles that will be useful in other contexts, far transfer becomes more likely.
- *Active self-monitoring* — Metacognitive reflection about the skills that are being learnt increases the chances of far transfer. When students are able to recognise the thinking processes that were successful in one situation, they are more likely to recognise when it might be useful to apply them again.
- *Arousing mindfulness* — Alert, rather than passive, learning will foster the environment required for far transfer, promoting deeper and more reflective learning.
- *Transfer by affordances* — If the learning situation brings about opportunities for particular interactions between the learner and the environment, the student is able to build new action schemas that can be applied in other contexts. For example, if a students learns during his mathematics class the usefulness of using diagrams to illustrate and solve problems, the student may be more likely to employ this technique in other problem-solving situations.
- *High road transfer* — A term coined by Perkins and Salomon, 'high road transfer' refers to the linking of remote contexts through the investment of mental effort, and mindful abstraction. This is in contrast to 'low road

transfer' in which well-developed, semi-automatic responses are triggered by similar contexts (Perkins and Salomon (1992), p.8).

Research has suggested that far transfer is a rare phenomenon that must be carefully fostered for it to occur, for example Thorndike and Woodworth (1901). The extent to which education systems in the UK and around the world are doing this successfully is undetermined, and has been debated by many educational psychologists without any firm conclusions being reached. The following section discusses the transfer of skills and formal discipline literature related specifically to the study of school subjects.

1.2 Transfer of skills from school subjects to more general cognitive skills

In 1924, Edward Thorndike conducted a study into the formal discipline value of high school education, finding no significant transfer to general thinking skills for any particular school subject (Thorndike, 1924a,b). The study involved 8,564 students aged 9 to 13 years, recruited from 26 schools in 11 different cities across the United States of America (USA). The students were tested on a number of reasoning skills using the Institute of Educational Research (IER) Tests of Generalisation, Organisation, and Selective and Relational Thinking (Thorndike, 1922), considered at the time to be the best measure of general reasoning in high school students. Some examples of the questions that made up this test can be seen in Figure 1.2. In order to answer many of these questions successfully, students would have to depend on many skills, including literacy and general knowledge. It therefore could be argued that this was not an effective measure of general reasoning in the way that more modern tests are. This is discussed further in Section 1.5.

The students were tested at the beginning and the end of a one year period (September 1922 to June 1923) and the scores were analysed in relation to the school subjects that they were studying over this time period. The results showed that the students who scored highly on the tests at the beginning of the year tended to gain more points by the end of the year than those who scored lower, but that no particular subject pattern, as studied by a particular student, had any large or significant effect on gains in reasoning. In terms of support for the formal discipline theory and the Liberal Arts Education system, none of the focal subjects, such as Latin and geometry, showed any superiority over any other subject taught in American schools, such as physical sciences, or

Test 4. Thirty-five Wylie opposites

"Look at each of the words in the list below. Then write a word after each one which means just the opposite and which also begins with the letter *b*. If you come to any word which you cannot do, then go on to the next one. These three samples are given as they should be:"

girl — boy
covered — bare
upset — balance

Numbers 1, 6, 11, 16, 21, 26, and 31 are:[3]

1. good 6. white 11. straight 16. loose
21. cultured 26. exhume 31. spiritual

Test 1. Twenty-one Arithmetical Problems

"Find the answers to these problems. Write the answers on the dotted lines. Use the blank sheets to figure on." Nos. 1, 4, 7, 10, 13, 16, and 19 are:

1. What is the cost of four tickets at 50 cents each? Answer........
4. How much will 24 lemons cost at 30 cents a dozen? Answer........
7. At 6 for 25c, what is the cost of 3 dozen? Answer........
10. What number minus 7 equals 23? Answer........
13. 4 percent of $600 equals 6 percent of what amount? Answer........
16. A family spends $600 on rent, $3,000 on other expenses and saves $200. If they increase their total expenses to $4,200 and their savings in the same ratio, how much will they save? Answer........
19. Jofas are 4 for 25c. Kelas are 2½c each.
A Jofa costs............as much as a Kela.

Figure 1.1: Example questions from the IER Tests of Generalisation, Organisation, and Selective and Relational Thinking

sewing. Despite finding no evidence to support the theory of formal discipline or transfer of skills, Thorndike advised that more research was necessary before any genuine conclusions could be drawn regarding the formal discipline advantages of any particular school subjects. In an attempt to clarify Thorndike's results, Broyler et al. (1927) repeated the study using the same tests and procedures and obtained the same, inconclusive, results. For the following decades, subjects such as Latin, Greek and geometry continued to be taught in the majority of schools, despite the usefulness of these subjects being undetermined and the theory of formal discipline enjoying fewer and fewer followers in the field of psychology (Stanic, 1986).

The following decades saw a small number of studies attempt to find effects of transfer of skill (for examples see Dorsey and Hopkins (1930); Gadske (1940)). These studies claimed significant but small gains in general cognitive skills due to the study of particular school subjects, or specific training, but in many cases the methods were criticised for being weak. The samples were often small and the participants not representative and therefore the studies had little influence on general opinions about the theory of formal discipline (Hartung, 1942). A more convincing example of these transfer studies was conducted by Ulmer (1939).

These studies involved three groups of students (total $N = 330$), matched for age and initial scores on an intelligence test and a reasoning test. Two of the groups took a course in plane geometry (one focused on the application of logical reasoning and reflective thinking, and one did not), and another group, used as a control, took no geometry classes at all. The results showed that the experimental group that were exposed to training in logical reasoning displayed significant improvements on the reasoning test after the teaching period. This finding provided evidence of some transfer of skill, but only between quite similar constructs (the students were taught reasoning, and then tested on reasoning), and only when the training was adapted to facilitate the testing. It should be noted, however, that the reasoning test used by Ulmer was relatively far removed from the regular geometry subject content being taught in the geometry classes. The test consisted of a discussion of some controversial matter followed by a choice of various statements, of which the students were asked to choose the one that best supported the conclusion of the discussion. The fact that the teachers of the experimental groups were able to teach in a way that did promote general reasoning skills, without jeopardising the mathematical content of the lessons is a noteworthy finding. It is possible that the skills being transferred in the Ulmer study, from the logical reasoning in geometry, to the construction of a convincing argument in the testing phase, were genuine abstracted reasoning skills, although it is notable that similar examples of transfer have been hard to replicate, e.g. Brooks (1924); Thorndike (1924a,b); Wesman (1945).

A second repetition of the Thorndike study of mental discipline in high school studies tested 643 high school students in one New York City school using the same IER tests administered at the beginning and the end of an 8-month period (Wesman, 1945). At the same time, the students were tested on proficiency in various school subjects in order to assess the learning of the subjects, and not purely exposure to them. As well as comparing course patterns with gains in general reasoning, as performed by Thorndike, initial and final correlations between reasoning and subject proficiency, and correlations between gains in test scores were analysed. Again, Wesman found that the students that scored highly initially also displayed the higher gains over time, but that, in terms of formal discipline value, the results were inconsistent between groups and no evidence for transfer of skill from school subjects to general thinking skills was found. Considering the Thorndike (1924a,b), Broyler et al. (1927) and Wesman (1945) studies, by the 1950s there existed a sizeable amount of research to suggest that the study of particular school subjects did not have any transfer value to general intelligence or reasoning. It was, however, unclear what this meant for liberal arts education, and little notice was taken of the findings.

For the decades following these studies, the state of the Western education system changed in many ways. In America, there were rapid growths in urbanisation, industrialisation, and immigration occurring all over the country and, by 1940, the number of children in education was twenty times that recorded in 1890 (Stanic, 1986). Not only did the sheer number of extra students impact the way in which the education system could run, but the increasingly diverse school population meant that many educators started to doubt the appropriateness of the content of a liberal arts education for the masses, and felt that a more practical-based education would be more beneficial. Unsatisfied with the school content of Latin, Greek and advanced mathematics, many parents and educators felt that these more modern and diverse students would find more benefit in spending their school hours being taught functional and vocational skills that would enable them to more successfully seek remunerative work when leaving school. Furthermore, advocates of the formal discipline value of subjects such as Latin and Greek, at this point, did not have a strong research-based argument for their inclusion in main-stream education. In addition, many of the young men drafted into service for World War II in the USA did not pass army entry exams due to illiteracy, and those who did enrol created vacancies in service jobs that the 'left-behind' population were not skilled enough to fill (Schmidt, 1958). A combination of these factors exposed the Western education system at the time as inadequate in training young people for the jobs that needed filling, and highlighted a need for an organisation and standardisation of the system.

This overhaul was implemented over the following half century with the introduction of nationwide examinations and the teaching of more practically useful subjects such as business studies and computing, as well as an increase in vocational and technical education which makes up the education system that we would recognise today (Institute of Education, 2010).

Although this review has focused on the American education system, the same can be applied, on the whole, to all Western countries. Currently in the UK, there are government initiatives that reflect both sides of the argument about the purpose of education. From 2017, measurement of a school's performance will be more weighted towards students' achievement in the English Baccalaureate (EBacc) subjects which are similar to the liberal arts subjects (Department for Education, 2016b). At the same time, the government are implementing plans to improve vocational courses and apprenticeships and promoting these as routes into employment (Wolf, 2011).

1.3 Mathematics and the theory of formal discipline

The research that has been discussed so far has, as a whole, applied the theory of formal discipline to education as a whole, with no particular focus on mathematics. It can, however, be quite easily argued that the subject of mathematics makes up the core of secondary education, and is perhaps the only subject that has held this esteem throughout the history of education and educational research. In the current education system, very few schools offer Greek or Latin, two of the major components of a Liberal Arts Education, but mathematics has held its place as probably the most important subject taught in schools in terms of teaching time and the emphasis put on results in school league tables. In fact, in England from 2017, mathematics results are given double weighting when judging how well a school's students are performing at the end of compulsory education (Department for Education, 2016b).

The idea that mathematics education in particular holds a formal discipline value originates from the work of Plato, who philosophised that:

> "... those who have a natural talent for calculation are generally quick at every other kind of knowledge; and even the dull, if they have had an arithmetical training, although they may derive no other advantage from it, always become much quicker than they would otherwise have been... arithmetic is a kind of knowledge in which the best natures should be trained, and which must not be given up" (Plato, 375BC/2008)

In some part as a legacy of this way of thinking, in the current Western education system, more money, research, and training is focused on mathematics teaching and learning than any other school subject, and children are expected to learn mathematics for more hours per week, from a younger age than ever before (Vorderman et al., 2011) and now until the age of 18 (Education Funding Agency, 2014).

As a comparison to the way in which mathematics has held its important role in education so stably, it is interesting to look at the decline in Latin as a school-taught subject; treated in a very different way, despite the research evidence for formal discipline value being similar. Weisert (1939) wrote about the challenges that Latin as a school subject faced in light of the formal discipline research activity of the early 1900s. He described the situation as being that educators were placed into two camps: those that believed that the learning of Latin helped to “lay the foundation for that higher intellectual and spiritual life which constitutes humanity’s full stature” (Weisert (1939), p.62), and those that believed in the specificity of learning. Rather than abandon the learning of languages completely, Latin was replaced by modern languages, thought of as inferior as a means of discipline, but useful in developing direct skills. At the same time as the study of languages was undergoing these changes, there were also educational psychologists that were fighting for similar reorganisation of mathematics, advocating topics that were more focused on the students’ specific, practical needs, and making close links to other, more vocational courses (Hutson, 1935). However, over time, the disappearance of Latin from the curriculum was accepted as a sensible step towards a more effective and relevant education system, but the inclusion of advanced mathematical subjects, above teaching knowledge needed to function successfully in society, was rarely questioned.

Despite research evidence throughout the 20th Century giving no solid conclusions regarding formal discipline and mathematics education, this was still used as a reason for its inclusion in the curriculum (Stanic, 1986). Today, with mathematics education being enforced on every child to the age of 18, and teachers being pressured into getting improved results year on year, it is essential for education research to further the understanding of the benefits of studying mathematics and its impact on skills outside of the classroom. The following section discusses some of the more contemporary research into the theory of formal discipline and transfer of learning from education, particularly in regard to mathematics.

1.4 Recent research in the area of formal discipline

More recently, formal discipline has again become a focus of educational research, with new evidence emerging for the possibility of far transfer from school education to general skills.

In the 1970s and 80s, the increasing use of computers in schools, and the integration of this into the education system became a new focus for much education research. It was important to answer the question of whether or not the advantages of using information technology (IT) spread to wider cognitive benefits: did using IT and computer programming increase the formal discipline value of education? As previously with mathematics and Latin, this theory held some presumed validity and began to influence decisions about the inclusion of

IT in mathematics lessons ahead of solid evidence being found to support them.

Pea and Kurland (1984), in a review of the potential cognitive benefits of computer programming in education, quote an educational report of the time as arguing that "the teaching of the set of concepts related to programming can be used to provide a natural foundation for the teaching of mathematics and indeed for the notions and art of logic and rigorous thinking in general" (Feurzeig et al., 1981). These claims of the formal discipline value of computer programming sparked a number of research studies with mixed results. For example, Clements and Gullo (1984) found that, after a 12 week intervention, children that had had computer programming training scored higher on measures of divergent thinking[1]. However, other studies continued to find no transfer effects of computer programming education on higher thinking and reasoning skills, although these findings often found little increase in actual programming knowledge during the intervention period either, suggesting that no thorough learning had occurred, an essential condition for far transfer as described by Perkins and Salomon (1992) in Section 1.1 of this thesis.

This new interest in education and formal discipline led to further research in the field of reasoning and cognition to be conducted in regard to A level (post-16) [2] and university study. Lehman and Nisbett (1990), for example, studied the effect of undergraduate training on inductive reasoning and reasoning in conditional logic. Inductive reasoning was measured using a test of statistical and methodological principles, taken from Fong et al. (1986). These questions required the students to reason about given situations, for example, they were asked "Why is it that promising new major-league baseball players tend not to do as well in their second year?". In reply to this problem, and in order to show a high level of inductive reasoning, the students were expected to display knowledge and application of the regression principle by indicating that, because the player's performance is so unusually exceptional in that first year, having two years in a row that are so exceptional is extremely unlikely. Reasoning in conditional logic was tested using a mixture of the Wason selection task (See Figure 1.4 for a description of this task) and written problems that could be solved using a conditional or bi-conditional interpretation. For an extensive account of the theories surrounding logical reasoning and conditional inferences see Evans et al. (1993).

[1] Divergent thinking refers to a problem-solving method that involves finding multiple successful solutions to a task as opposed to only one, which would be referred to as convergent thinking.

[2] see Appendix 9.4 for a full explanation of the UK education system

The most famous of Wason's experimental paradigms is the selection task. The experimenter puts four cards down on the table in front of you, bearing respectively: A, D, 4, 7. You know that every card has a letter on one side and a number on the other side. Your task is to choose just those cards that need to be turned over to determine whether the following conditional rule is true or false about the four cards: *If a card has a vowel on one side then it has an even number on the other side*. The cards are to be turned over simultaneously, so that what is on the other side of a card has no bearing on the selection of another card. Nearly everyone selects the A card, and some in addition select the 4 card. What is surprising is how few people select the 7 card. Yet if it has an A on its other side, the rule is false.

Figure 1.2: An example of the Wason selection task, taken from Newstead (2003)

Lehman and Nisbett had, in a previous study (Lehman et al., 1988), found an improvement in inductive reasoning associated with the study of medicine and psychology, but not law and chemistry, and improvements in conditional reasoning associated with law, psychology, and medicine, but not with chemistry. It was hypothesised by Lehman and Nisbett that the reason for the observed improvements in inductive reasoning was that the psychology and medical students were exposed to research articles and had experience in dealing with variability and uncertainty in causal relations. Law students were thought to have improved in logical reasoning because of exposure to contractual relations that have the form of the conditional, such as permissions and obligations. It was suggested that psychology and medical students improved in the same reasoning tasks because of the checking procedures necessary in probabilistic science; similar to those of a conditional statement. Lehman et al. concluded that the only subject not linked to any improvements in more general thinking skills was chemistry and that this was due to that fact that this subject was vastly content based, required few higher thinking skills, and held no formal discipline value. The students studying subjects in which they were encouraged to analyse situations, and follow thorough conclusions, improved their reasoning skills outside of their undergraduate subject matter, supporting the idea of formal discipline and transfer of training. Two of the key graphs from this 1988 study can be seen in Figure 1.3 and Figure 1.4.

For the follow-up study, Lehman and Nisbett (1990) focused on a comparison of deterministic, or natural sciences (such as chemistry), probabilistic, or social sciences (such as psychology) and non-sciences, or humanities (such as law) as studied by undergraduate students over the period of four years. The study

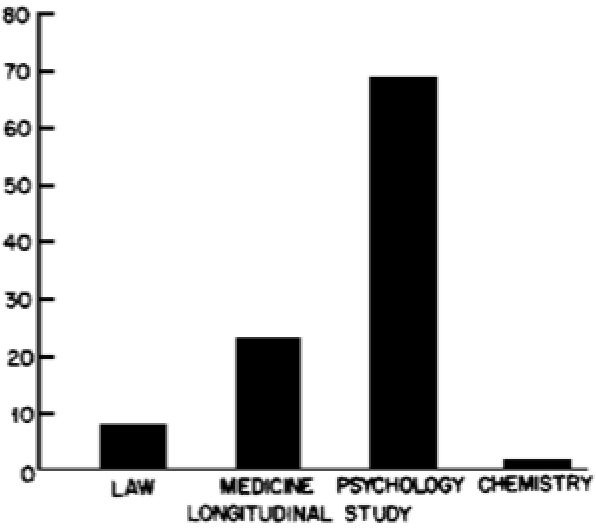


Figure 1.3: Percentage gains on scores of statistical and methodological reasoning found by Lehman et al. (1988)

replicated the previous results, showing that the social science students displayed large significant gains in inductive reasoning. In addition, significant gains in logical reasoning from the natural sciences students and the humanities students were found. Lehman and Nisbett suggested that the observed improvement in logical reasoning seen for the natural science students may be related to the large number of mathematics courses that the students were required to take over the four years. Further analysis of the course taken by the students found that the number of mathematics courses taken by the natural science undergraduates did, in fact, correlate with their gains on the conditional reasoning task, suggesting that this explanation could be valid. The gains seen for the humanities students were not explained by Lehman and Nisbett.

The study of mathematics at an undergraduate level involves understanding and practising mathematical proofs which have a very similar set of rules to conditional logic. Jackson and Griggs (1988) found that individuals with expertise in mathematics scored higher on an abstract Wason selection task in comparison to experts in other fields, with no effect of level of education. They suggested that the

> "mathematics students' greater likelihood of using a disconfirmation strategy and greater familiarity with the relevant propositional logic ... account(ed) for their superior performance" (Jackson and Griggs,

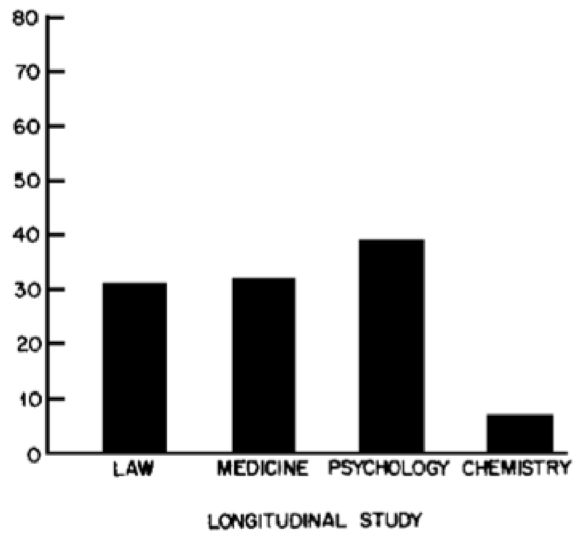


Figure 1.4: Percentage gains on scores of conditional reasoning found by Lehman et al. (1988)

1988, p. 327).

However, mathematicians' performance on selection tasks overall has been found to be surprisingly poor, with over a third of mathematics university students, and a half of mathematics staff, making logical errors on the task (Inglis and Simpson, 2004).

In reaction to this emerging evidence for formal discipline, Inglis and Simpson (2007) investigated the effects of advanced mathematical study on students' behaviour when making inferences about abstract conditional statements. Inglis and Simpson compared a group of undergraduate students studying mathematics with those studying arts or social science. Both groups were presented with 32 problems in the form:

> **Rule**: If the letter is X then the number is 1
> **Premise**: The letter is not X
> **Conclusion**: The number is not 1

and asked to state whether they thought that the conclusion necessarily followed, Yes or No. Inglis and Simpson hypothesised from theory and from previous evidence that the mathematics students' answers would differ from that of the arts and social sciences students in terms of their reasoning behaviour and the types of inferences that would be endorsed.

It was found that mathematics students answered significantly more of the questions in the correct way compared to the students who were not studying mathematics, where 'correct' is defined as being in line with the material conditional[3]. Both groups displayed 'negative conclusion bias' by answering 'Yes' to more negative conclusions (e.g. The number is not 6) than affirmative conclusions (The number is 2). This is a fallacy often found in logical inference behaviour along with another, coined 'affirmative premise bias': the tendency to endorse more inferences with affirmative premises (e.g. The letter is A) as opposed to negative premises (The letter is not T). Interestingly, this affirmative premise bias was only observed in the non-mathematics students, suggesting that the mathematics students were less likely to be misled by the more confusing wording of the premises.

Inglis and Simpson concluded that these results were consistent with the idea that studying higher-level mathematics did affect the development of logical reasoning skills which could potentially be transferable to other situations. This study, although providing evidence that mathematics students reasoned in a different way to non-mathematics students, did not stretch to establishing whether it was actually the study of advanced mathematics that had resulted in this, or whether mathematics students as a group were less prone to logical fallacies outside of any influence of the study of mathematics. This theory of the filtering of individuals by reasoning abilities through the decision to study advanced mathematics is explored further throughout this thesis.

Alternative explanations for the findings were investigated by the same authors in a two part analysis of further evidence (Inglis and Simpson, 2009). Firstly, groups of mathematics and non-mathematics undergraduate students were compared following a method similar to the first study. One possible explanation for the findings of Inglis and Simpson (2007) was that the differences between the groups was due to a difference in general intelligence. In order to address this, Inglis and Simpson (2009) used a measure of intelligence to balance the two groups. The sixteen highest scores from the mathematics students group were removed, as were the seven lowest from the comparison group, leaving two groups with very similar mean intelligence scores. The findings replicated the results of the original study, rebutting the claim that the results might be due purely to the mathematics students having higher general intelligence (an argument also suggested by Thorndike (1924a,b)). Secondly, a longitudinal study

[3]The material conditional interpretation of such logical problems would consider, given that 'If the letter is X then the number is 1', when the letter is X, the number must be 1, and when the letter is not X, the number may or may not be 1.

of the mathematics students was conducted in order to explore the developmental aspects of the differences in reasoning skills between the two groups and the possibility of formal discipline effects. After a year of undergraduate mathematics study, the reasoning behaviour of the mathematics students did not change significantly, suggesting that the differences found between the two groups occurred because of an influence pre-university. This influence could be the study of A level mathematics (of which all mathematics students had been exposed to, but so had 36% of the comparison group) or from individual differences in 'thinking dispositions': the tendency toward the use of different cognitive behaviours. For example, an individual might possess the ability to reason and think critically, but might not be disposed to do so (Facione, 2000). This theory would match the idea of a filtering effect of mathematical study; those that have a disposition to reason in a certain logical way choose to study mathematics at university, those that do not, do not.

In consideration of the possible influence of A level mathematics study and thinking dispositions on more general reasoning behaviour, Attridge and Inglis (2013) compared two groups of A level students. The first group studied mathematics, among other subjects, and the other, used as a comparison, studied English literature without mathematics. These groups were tested on a conditional inference task (Evans et al., 1995) and a Cognitive Reflection Test (CRT) (Frederick, 2005) before and after one year of A level study. The four possible conditional inferences are illustrated in Table 1.1.

Table 1.1: The four possible conditional inferences. For an extensive account of conditional logical inferences, see Evans et al. (1993).

Inference	First premise	Second premise	Conclusion
Modus Ponens (MP)	If p then q	p	q
Denial of the Antecedent (DA)	If p then q	$\neg p$	$\neg q$
Affirmation of the Consequent (AC)	If p then q	q	p
Modus Tollens (MT)	If p then q	$\neg q$	$\neg p$

Attridge and Inglis found that the mathematics students endorsed significantly more MP inferences at Time 1 than they did at Time 2, and rejected more DA and MT inferences. The comparison group did not change their reasoning behaviour from Time 1 to Time 2. This change in behaviour observed in the mathematics students represents a move to a more defective conditional interpretation of the conditional (In which MP inferences would be endorsed, AC and DA inferences would be seen as invalid, and MT inferences would be seen as irrelevant, and therefore be endorsed less of the time (Manktelow, 2012)). The authors hypothesised that this change could be interpreted as an effect of pre-university mathematics involving logical arguments such as solving equations, which require only forward-direction logical manipulations, similar to MP inferences. However, advanced mathematical study involves to additional learning of proof by contradiction, for which deductions are required to be of a MT nature. This could explain the reduced MP inferences and the increased MT inferences seen amongst the mathematics students. Attridge and Inglis found that none of the measures taken, including the CRT score, intelligence, and prior achievement predicted the change in reasoning behaviour and concluded that this change was very probably related to the experience of advanced mathematical study between Time 1 and Time 2; evidence of far transfer, and of the formal discipline value of A level mathematics.

Other A level subjects chosen to be studied by the students were not taken into consideration for this study. It is likely that the students that chose to study mathematics at A level would also choose more additional science subjects than those students that chose to study English, and not mathematics. The effects that these additional subjects had on the students' reasoning behaviour might have partially contributed to the results and therefore would be worthy of investigation. As with much of the research into the effects of transfer and formal discipline, it is very difficult to be certain of the causal direction of any significant results obtained. Potentially, the students that chose to study advanced mathematics could have had a fundamentally different reasoning development pattern than the students that did not choose to study advanced mathematics. These students may have displayed changes in their reasoning behaviour even in the absence of any advanced mathematical study. Although this alternative interpretation is possible, it seems unlikely as the mathematics and the non-mathematics students did not differ significantly in their reasoning behaviour at Time 1, so there was no filtering effect in place at that time. This was also the case for the Lehman and Nisbett (1990) study in which the groups of students did not differ significantly on any of the reasoning measures at Time 1.

The more recent research into mathematics and formal discipline highlights the need for further study into this area. It is clear that the idea of formal discipline and the transfer of skills from the learning of school subjects to more general reasoning skills may have been rejected too hastily in the 20th Century. More recent advances in the understanding of cognitive processes and human reasoning behaviour have enabled researchers to study the effects of formal discipline in a way that was not possible for the psychologists of the early 1900s. It would be of interest to approach the research questions posed by early education psychologists using more modern methodologies and instruments of measure.

The research that has been discussed so far has focused on the effect of education on a number of reasoning constructs and thinking skills. Divergent thinking (Clements and Gullo, 1984; Pea and Kurland, 1984; Perkins and Salomon, 1992), inductive reasoning (Lehman et al., 1988; Lehman and Nisbett, 1990), and conditional reasoning (Attridge and Inglis, 2013; Inglis and Simpson, 2007, 2009) have all proved to have links with mathematics learning in the literature, and provide evidence for a 'quickening of the mind', as described by Plato, (375BC/2008). However, it has not been made completely clear from the research whether the differences observed between mathematicians' and non-mathematicians' reasoning behaviour can be attributed to a formal discipline value of studying the subject, or to a filtering effect of individual thinking dispositions. This filtering effect would, in fact, be represented by the first part of the previous quote from Plato:

> "... those who have a natural talent for calculation are generally quick at every other kind of knowledge..." (Plato, 375BC/2008)

This forms an alternative plausible hypothesis for the differences seen in reasoning behaviours between mathematicians and non-mathematicians.

In order to assess any formal discipline effects of studying mathematics on an individual's cognitive abilities, the measurement of these abilities must be valid. The following section summarises the development of tasks designed to measure these.

1.5 Measuring cognitive constructs

An advantage of more modern research is the development of instruments to measure cognitive constructs in a more valid manner than in the past.

The use of tests designed to measure the cognitive, or mental, abilities of individuals can be traced back to the mid-1800s when the speed and accuracy rates of cognitively impaired children were measured using form boards. In its most simple form, a form board consists of a number of differently shaped wooden blocks and a large board with recessed corresponding shapes. The speed and accuracy with which children could complete versions of this task produced relatively valid measurments of intelligence. An example is illustrated in Figure 1.5.

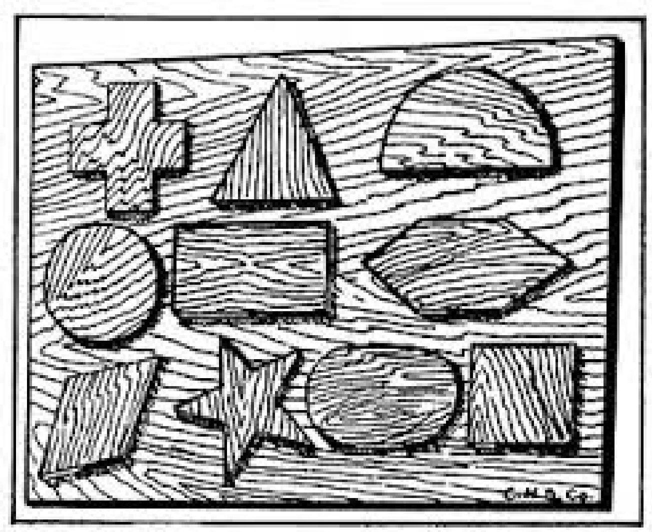


Figure 1.5: An illustration of the Seguin Form Board (see Venkatesan (2014) for an account of the development of form boards)

Originally developed as a training tool for these children, the form board's usefulness as a measurement for research soon became apparent (Boake, 2002). Other crude measurements, such as line-bisecting, and digit-span were being used to measure a construct close to intelligence throughout the 17th Century.

In the late 1800s, the development of psychometric testing escalated rapidly. Alfred Binet was asked by the US government to construct a test that could be used to identify children that were in need of specialist help at school. The result of this was the Binet-Simon scale (Binet and Simon, 1916), the first set of standardised tests of 'general intelligence', which brought with them much attention and interest in various fields of psychology. The aim of the tests was to measure the "faculty of adapting one's self to circumstances ... nothing to do either with (one's) past history or with (one's) future" (Binet and Simon, 1916, p.42-43). The ability to label an individual with an ordinal score of cognitive ability meant that people could be ranked, excluded, diagnosed, and classified in a way that was not possible previously. The Binet-Simon scale, mostly consisting of questions unrelated to the subjects that were being taught in school, gave a mental age for the children. The ratio between the child's mental and chronological ages resulted in an intelligence quotient, or IQ. Although the tests were initially designed to identify children with cognitive impairments, they were soon being used to test the intelligence of all typically developing children as well as adults (Boake, 2002). Being able to measure the level of 'intelligence' that an individual possesses made it possible for psychologists to predict behaviour when faced with certain academic and social tasks such as school achievement and job performance; a useful tool for educators and employers (Schmidt and Hunter, 1998). Developing reliable and accurate ways of measuring intelligence and its cognitive components became an essential goal for intelligence researchers, and many psychologists have contributed toward the most common methods used in present research.

Although the development of cognitive testing has progressed much since the 17th Century, researchers are still continually faced with the fundamental difficulties that come with the attempted measurement of cognitive constructs as they have no physical expression that is directly measurable. Walter Lippmann, a political commentator and amateur philosopher, remarked that

> "...psychologists have never agreed on a definition (of intelligence)...The intelligence tester cannot confront each child with the thousand and one situations arising in a home, a workshop, a farm, an office, or in politics, that call for the exercise of these capacities which in a summary fashion we call intelligence. He proceeds, therefore, to guess at the more abstract mental abilities which come into play again and again. By this rough process the intelligence tester gradually makes up his mind that situations in real life call for memory, definition, ingenuity, and so on. He then invents puzzles, which can be employed quickly and with little apparatus, that will according to his best guess test memory, ingenuity, definition and the rest." (Spearman, 1927, p. 11)

Although this quote is intended to express Lippmann's thoughts on the practice of intelligence testing back in the 1920s, theorists today are still attempting to achieve the same objective: to design a simple, inexpensive, easily administered test which is capable of predicting an individual's performance in a wide range of situations.

Looking at the measure of general ability that was employed in the Thorndike (1924a,b) formal discipline studies, there are many aspects that would raise concerns today regarding its validity. Thorndike measured general ability using the IER tests. These tests, although considered the best for this purpose at the time, did not measure a pure form of intelligence in the way that more modern cognitive tests aim to do. Figure 1.2 is an example of the questions used by Thorndike and shows that the IER tests were measuring a combination of reading skills, vocabulary, culture and mathematical knowledge among other constructs and not 'quickness of thought' as described by the early formal discipline theorists.

The notion of a single measure of intelligence was first suggested by Spearman (1904) when he noticed positive correlations between children's performances on a range of different tasks, some academic and some cognitive. These abilities, broadly classed into two groups: crystallised intelligence (the ability to use knowledge and skills) and fluid intelligence (the ability to solve novel problems), have been shown through factor analysis to load onto a single factor, often denoted g. Because so many different measurements of intelligence correlated with one another, Spearman proposed a 'Unity of the Intellectual Function', meaning that performance on a wide range of different tasks, from school performance to a simple task such as pitch discrimination, depended on a single measure of general ability.

The existence of g is widely accepted by the scientific community and has influenced the majority of intelligence theories. The single factor theory of intelligence, however, is not without its critics. Physically speaking, g means very little; its presence is purely statistical. This has led scientists to question the status that g has as a measurable entity. Gould (2006) argued that factor analysis could not be used to draw meaningful conclusions about direction of causality, or about the reasons underlying positive correlations. Gould stated that:

> "We cannot reify g as a 'thing' unless we have convincing, independent information beyond the fact of correlation itself." (Gould, 2006, p. 281)

Spearman himself was eager to define g as something physical, anticipating that in the future psychologists would discover some type of 'mental energy' that would exemplify g; that the mathematical abstraction of factor analysis should correspond with some material reality. Carroll (1997), in his explanation of the 'Three-Stratum Theory', also insisted that the factors identified were representative of physiological elements, such as nerve-firing speed, rather than purely mathematical processes. There is no material manifestation of g that can be measured directly but, although often debunked as a statistical myth, g proves itself to be the measurement of some useful construct in its correlations with other wide-ranging human experiences.

The reliance on g as a measure of general ability is thought by some, however, to be simplified to the point of uselessness. The measure disregards other strengths such as creativity, character, and practical knowledge and propagates the idea that individuals are born with an unchangeable potential intellectual capacity that will determine many of their life outcomes (Benson, 2003). This view is seen by many as pessimistic and unhelpful and as possessing the potential to be detrimental to individuals that are assessed or classified in this way. The use of cognitive testing to classify groups of individuals, particularly when related to social and racial issues, has contentious political and ethical implications. The general factor g has been used questionably to support eugenics (Galton, 1883) and come under accusations of inciting scientific racism (Jensen, 1969), resulting in the theory being unpopular in many circles. Past psychometric tests, designed to measure g, have also been criticised as suffering from a design and structure that might have give unfair advantage to certain races, genders, social classes and cultures, although more modern tests, such as RPM[4], are promoted as being free from these biases (Raven et al., 2000). The large amount of controversy surrounding research into these socially sensitive topics has meant that scientific exploration of the practical uses of g in psychology and social theory can come under excessive criticism (Gottfredson, 1986).

With the concept of g as a single factor of intelligence are two independent domains, first suggested by Cattell (1963), crystallised intelligence and fluid intelligence, often denoted as Gc and Gf respectively. The following section discusses the predictive strength that Gf has been shown to have in regard to positive life-outcomes, in particular academic achievement.

Gf as a predictor of life-outcomes

Correlations between a measure of Gf and various valuable life skills, such as school achievement and job performance, have been shown to exist through extensive research. Correlations between school grades and Gf are about $r = .50$, with Gf scores predicting scores on school achievement tests (Neisser et al., 1996; Nisbett et al., 2012). There are, of course, many other factors that will affect a child's performance at school, such as the teaching styles, attitude of family and peers, and willingness to work hard, but it remains established that children with higher scores of Gf tend to perform better academically than those with lower scores. Colom and Flores-Mendoza (2007), for example, found that scores on an intelligence test significantly predicted scholastic achievement (as measured by standard school tests), irrespective of socioeconomic status (SES), in an analysis of 641 Brazilian children. There is also a correlation of approximately $r = .55$ between Gf and the number of years that a child stays in school (Neisser et al., 1996), although an alternative viewpoint on the direction of the causal relationship of this correlation is discussed in the following section.

It is likely that a child with higher intelligence will have a different school experience to a child with lower intelligence. The majority of taught classes

[4]Raven's Progressive Matrices: A task designed to measure non-verbal intelligence

in UK schools are ability grouped. A number of educational researchers have suggested that the higher ability groups are allocated the 'better' teachers, and therefore have more opportunity to learn (Boaler et al., 2000). Children in the higher sets (likely to be those with higher Gf scores), will be surrounded by peers who are enthusiastic about studying and that will be influential in any decision made to take further studies. As a consequence of this, students with a higher Gf will be more likely to choose to continue into non-compulsory education. A similar explanation makes up the alternative hypothesis to the results of some of the formal discipline studies discussed in Section 1.4; that students with better (or different) reasoning abilities chose to study mathematics at an advanced level. This is one of the possible reasons for the predictive power of intelligence scores when looking at correlations with adult occupation and salary (Neisser et al., 1996). Schmidt and Hunter (1998) found, through an analysis of 85 years of research findings, that the strongest pre-employment predictor of job performance was a combination of a test of general mental ability and either an integrity test (r = .65), or structured interview (r = .63). Measures of previous education and personal interests were found to have no predictive power. For a discussion of many other studies that have found evidence of strong links between performance in a variety of jobs, and general mental ability or intelligence, see Schmidt (2000).

Longitudinal studies into the predictive nature of early intelligence scores suffer from a plenitude of confounding environmental factors that are impossible to control for, due to their close relationships with intelligence itself. Strenze (2007) conducted a meta-analysis of studies into intelligence as a predictor of academic achievement, occupation and income. He found intelligence to be only a little stronger as a predictor of these measures than parental SES, or school grades, although intelligence itself *was* a very powerful predictor.

Intelligence has been shown to predict many health behaviours, both good and bad. Physical fitness, better diets, and longevity of life increase as intelligence does; alcoholism, smoking, infant mortality, and obesity increase with lower intelligence (Gottfredson, 2004). Gottfredson and Deary (2004) argued that self-care, in regards to an individual's health, requires the same skills that define intelligence, or Gf; effective and efficient learning, problem solving, reasoning and abstract thinking. Gottfredson and Deary claim that this is why measures of Gf predict health behaviours even when SES is taken into account. Considering that intelligence is so influential on many measurements of success in life, the possibility of being able to increase this construct at an early age is of major interest. Therefore, the possibility of increasing any of the cognitive constructs that contributed to Gf through the formal discipline value of the study of mathematics has wide and significant implications. If there does exist a formal discipline effect that transfers to another cognitive ability that increases an individual's chances of experiencing positive life-outcomes, this should be honed in on and exploited as far as possible.

The question of whether cognitive constructs can be influenced by external factors at all comes down to a debate between the balance of genetic and environmental factors. This topic has enjoyed much debate within psychology, with the simple answer that both have a significant effect. Many studies have sought to quantify the impact of both separately, but observing the effects of genes and the environment independently is not a straightforward task as biological parents contribute 100% of an individual's genes, but also influence a large percentage of their children's environment. Few research opportunities are found in which this is not the case. Twins separated at birth is one such opportunity which is always utilised keenly by researchers. Bouchard (1998) studied twins and triplets that had been separated during childhood and had spent the majority of their lives apart. He found that identical twins, reared apart, had intelligence scores (measured using the Weschler Adult Intelligence Scale) that correlated at $r = .69$, not much lower than those that had been reared together which correlated at $r = .88$ (Deary, 2001)). It is certainly true that factors other than the shared genes might account for these correlations. For example, it is likely that twins born in the same hospital would be adopted by similar families and therefore the twins would be likely to grow up in separate, but very similar, environments. However, if these similar family environments do play a role in affecting the children's cognitive development, the same effects would be found between close friends, or cousins, which they were not. Bouchard's study suggests that genetics can explain about 70% of the differences in people's intelligence scores. Across all available studies on this area, this percentage averages at approximately 50% (Deary, 2001), meaning that half of the differences between people's intelligence scores can be attributed to genetics, leaving the other half to be accounted for by the environment.

The theory of formal discipline assumes that a certain amount of the cognitive constructs that make up Gf can be influenced by the environment, and that these are not set in stone from birth. From the evidence available, it can be quite confidently concluded that approximately half of an individual's cognitive abilities are a result of external factors and influences. The external influencing factor that is of particular interest to the theory of formal discipline and transfer of training, and therefore to this thesis, is that of schooling. The following section summarises some of the literature surrounding the influence that schooling can have on a measure of general ability.

Effects of schooling on Gf

The environment

The relationship between schooling and intelligence (Gf levels) was extensively discussed by Ceci (1991) in a review of research studies. Ceci offered an alternative directional explanation for the high correlations between the amount of schooling and Gf, suggesting that more years of schooling, and higher achievement, fosters higher Gf, as opposed to the reverse causal pattern. The evidence for this claim was based on the high correlation between Gf and both school grades achieved, and number of years in schooling, even when SES was a covariate. Adding evidence to this argument is the fact that the increase in Gf that occurs during schooling reverses during the summer holidays, especially when the children's activities over this holiday period are very different from their school environment (Downey et al., 2004). This implies that the act of being at school and studying school subjects has a positive developmental effect on Gf. Ceci also reported that children who did not attend school regularly had lower Gf, as did children that were delayed in starting schooling. When comparing children of very similar ages, some having attended an extra year of schooling due to birthday-related entry times, the children that had undergone more schooling had higher Gf (Ceci, 1991; Neisser et al., 1996). Clouston et al. (2012) also studied British and USA populations from the age of 15 or 16 years old, to their mid-fifties and found that non-compulsory education had a significant impact on adult Gf score, even after adolescent cognition was accounted for.

Considering the evidence supporting the idea that a child's schooling plays a large part in their intelligence development, it is important that the education being received by students is of a high quality and is as effective as possible in fostering cognitive abilities.

> "... It is clear that sheer amount of schooling, even in backward countries and of low quality, helps to promote both school achievement and the kind of reasoning measured by non-verbal tests ... if such schooling is unduly delayed, the possibilities for mental growth deteriorate" (Vernon, 1969, p. 350)

A number of cognitive skills might mediate the relationship between schooling and scores of Gf such as memory, concept formation, reasoning abilities, and perceptual abilities and it can be very difficult to be certain of the direction of the relationship between any of these constructs. It could be true that schooling helps to develop these skills, and these skills are needed to score highly on intelligence tests, or that highly intelligent children have higher levels of these skills, and therefore find schooling more rewarding and beneficial as a similar set of skills are needed. More recent evidence relating to working memory training (discussed in Chapter 2) has provided the former explanation with more validity through the study of increasing intelligence through training in working memory capacity.

If education is assumed to have some formal discipline value, there are two questions to be answered. One is related to which subjects within the education system possess this value, of which mathematics has enjoyed much research attention and is the focal school subject of this thesis. The other is related to which cognitive constructs are benefiting from the transfer of skills associated with learning the subject. Looking back at the 18th Century psychologists and educators who were advocates of the formal discipline theory, this construct was alluded to as being a 'quickness of thought, sharpness, or general intelligence' (Stanic, 1986) and certainly the first studies of formal discipline value (Thorndike, 1924a,b) focused on this general ability. More recently, research has shown links between the learning of mathematics and a number of different reasoning abilities, as well as links between schooling and Gf in general.

The most recent research into the causal effects of education on cognitive abilities comes from Ritchie et al. (2015). They investigated the magnitude of the influence of 'years of education' on general ability (or Gf), with the additional theoretical tier of de-constructing this into measuring the influence on sub-categories of Gf. Ritchie et al. used structural equation modelling, with a sample of over 1,000 participants, to assess whether education was associated with improvements in general cognitive abilities, or whether the benefits were better classified in terms of the effect on specific skills.

The data were obtained from the The Lothian Birth Cohort, 1936 (LBC1936); a collection of cognitive measures taken from every child born in Scotland in 1936 who was attending school in June 1942. At age seventy, 1,091 of the cohort were tested again[5], allowing a unique opportunity to examine cognitive changes over a lifetime. For a full description of this data set, see Deary et al. (2012).

A previous study that had utilised this dataset had suggested that, although

[5]The measures employed at both time points were logical memory; digit symbol substitution; matrix reasoning; block design; verbal paired associates; symbol search; letter-number sequencing; digit span backwards and spatial span. For a full description of these tasks see Ritchie et al. (2015).

years of education was associated with Gf in later life, significant relationships between some cognitive constructs could not be found, such as choice reaction time and visual information processing (Ritchie et al., 2013). Ritchie et al. (2015) therefore aimed to, with this study, address the question of whether education had domain-general or domain-specific effects on the development of Gf.

Ritchie et al. statistically tested three theoretical models of the way that IQ (age 11), years of education, fluid intelligence (Gf), and a subtest of specific cognitive abilities influenced and interacted with one another, see Figure 1.6.

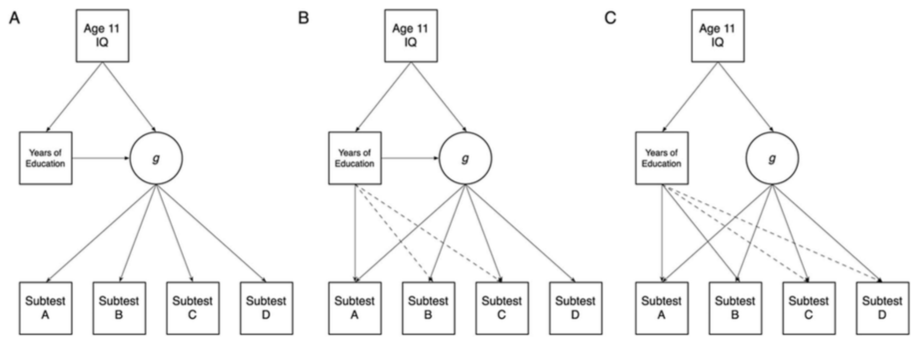


Figure 1.6: Three possible theoretical models of the links between education, g and specific cognitive skills (Ritchie et al., 2015)

Model A assumes an effect of education on Gf which, in turn, affects the specific cognitive abilities. Model B introduces the addition of direct effects of education on specific abilities and Model C proposes that there is no direct effect of education on Gf but instead there is an effect on all, or some, of the subsets directly.

Ritchie et al. found that all the cognitive subtest measures at age ~ 70 correlated positively and significantly with years of education and with IQ at age 11. It was found that Model C explained the largest amount of variance in the data (55%), suggesting that the effect that education had on specific skills is not mediated by a general factor Gf.

The findings of this study introduce the importance of studying the formal discipline effect of school subjects on specific cognitive skills. The research that has been discussed previously (Section 1.4) included studies on divergent thinking (Clements and Gullo, 1984; Pea and Kurland, 1984; Perkins and Salomon, 1992), inductive reasoning (Lehman et al., 1988; Lehman and Nisbett, 1990) and conditional reasoning (Attridge and Inglis, 2013; Inglis and Simpson, 2007, 2009). The next chapter of this thesis explores the links between mathematics education and another cognitive construct, spatial reasoning, and the relevance of investigating the formal discipline effects that advanced mathematical study could have on this construct. The next chapter will discuss why spatial skill in particular is worthy of in depth investigation regarding its relationship to mathematics.

1.6 Mechanisms of formal discipline

The section will discuss some possible mechanisms though which the transfer of skills from learning advanced mathematics to more general spatial skills may happen in practice. The identical element theory of transfer (Thorndike and Woodworth, 1901) states that transfer is dependent on how similar the training (learning mathematics) and the performance (spatial tasks) environments are: transfer will only happen if the activities share some identical element. In the case of advanced mathematics and spatial skills, these identical elements might be in the form of quite tangible skills, established in training and therefore more automatic and more easily executed in the performance task, akin to a type of 'low road' transfer, described in Section 1.1 (Perkins and Salomon, 1992). Alternatively, the identical elements might be more abstracted and represent 'general principles' that are associated with better performance in both activities, taking the form of some more 'high road' transfer occurrence. What form these low or high road transfer elements take, and how to contextualise and describe them, are discussed in the following section.

1.6.1 Transferred elements from learning advanced mathematics to performance on spatial tasks

Low road transfer

There are elements of learning mathematics that involve the acquisition of particular tangible skills that are common to those needed to complete a spatial task. An inspection of the mathematics syllabus identifies a number of topics that have an easily identifiable spatial quality:

- Trigonometry
- Graph sketching
- Coordinate geometry (two- and three-dimensional)
- Vectors

- Polar coordinates
- Revolutions around the x-axis
- Kinematics in one- and two-dimensions
- Newton's laws of motion
- Projectiles
- Critical path analysis

Some of the topics mentioned above are compulsory, and will be covered by all students taken A level mathematics, and some make up optional modules. One topic which is compulsory is 'graph sketching' and Figure 1.7 shows an example of the question that students might face in an examination of the topic. Figure 1.8 shows a question from the non-compulsory topic of 'projectiles'.

On separate diagrams, sketch the graphs of

(a) $y = (x + 3)^2$,

(3)

(b) $y = (x + 3)^2 + k$, **where** k **is a positive constant.**

(2)

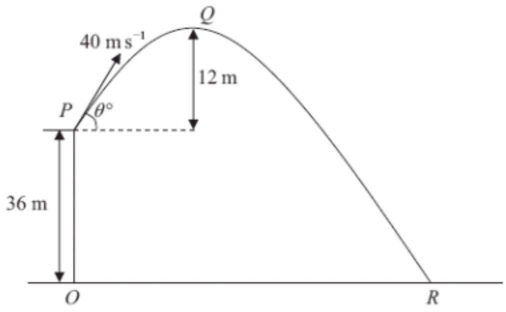
Show on each sketch the coordinates of each point at which the graph meets the axes.

(Total 5 marks)

Figure 1.7: An example of an A level mathematics question about sketching graphs

From these examples, it can be seen that students would be required to represent mathematical situations spatially, and to reason about particular manipulations of these representations in order to find the correct answers. These domain-specific skills could be directly transferable to a context of performing a spatial task. For example, both examples require a student to be aware of the spatial relationship between two points in two-dimensions.

A further inspection of the schools mathematics syllabus prior to A level reveals more topics that appear to involve domain-specific skills that could be directly transferable to performance on a spatial task. For example, the key stage 3 mathematics specification requires students to be able to rotate objects by a given amount of degrees on a cartesian coordinate grid. This skill is in some way identical to common rotation tasks used to measure spatial skills (see Figure 4.6 for an example). It is possible that there are more examples of these domain-specific skills present in the early mathematics syllabus than there are



A ball is projected with speed 40 m s^{-1} from a point P on a cliff above horizontal ground. The point O on the ground is vertically below P and OP is 36 m. The ball is projected at an angle $\theta°$ to the horizontal. The point Q is the highest point of the path of the ball and is 12 m above the level of P. The ball moves freely under gravity and hits the ground at the point R, as shown in the diagram above. Find

(a) the value of θ,

(3)

(b) the distance OR,

(6)

(c) the speed of the ball as it hits the ground at R.

(3)

(Total 12 marks)

Figure 1.8: An example of an A level mathematics question about projectiles

in the A level syllabus. This being the case, transfer of skills in this low road way might be expected to happen at an earlier stage than this thesis focuses on. It would therefore be predicted that this transfer would happen at a similar rate and magnitude for all UK students up to the end of compulsory education. However, there are likely to exist mediators and obstacles to the process, some of which are discussed in the section below.

High road transfer

Another way in which transfer could happen between learning mathematics and performing spatial tasks is in a more abstracted way, referred to as 'high road transfer' by Perkins and Salomon (1992). The common elements involved in this type of transfer are much more difficult to identify and to define. These elements may not have anything spatial about their nature and could be described as abstracted strategies, or domain-general skills, as opposed to the tangible and domain-specific skills discussed in reference to low road transfer. These elements might come in the form of strategies such as 'looking for and recognising patterns', or 'reasoning logically about problem', and would be associated more with advanced level mathematics than earlier school mathematics. For example, looking for patterns in number sequences in order to describe them through mathematical formula might require similar skills to those needed to recognise patterns in a task such as Raven's Progressive Matrices and choose the correct image to complete the pattern.

Possible mediators of transfer

If the transfer of skills from learning mathematics to performance in a spatial reasoning task is achievable, possibly through the mechanisms described above, then there exist some possible mediators to the magnitude and rate of the effect. Perkins and Salomon (1992), in their list of features that lead to successful far transfer, talk about 'thorough and diverse practice'. How through and diverse this is will depend a certain amount on how motivated a student is, whether this be through their teacher, parents, or self-motivation. Meta-cognition about their own mathematical abilities and skills could also be a mediator of transfer for students, akin to Perkins and Salomon's mention of 'active self-monitoring' as an important feature. The amount of attention that an individual student is able to pay during mathematics lessons, and their working memory capacity for learning and performing tasks both also have the potential to be mediators to transfer. Two final, and more external factors that have the potential to affect the process of transfer from learning advanced mathematics to spatial skills are the quality of the design of the syllabus being taught, and the delivery of the content by teachers. It could be possible that some advanced mathematics modules, or courses, do not require the learning of the sort of skills that can be transferred, or that some teaching lacks the methods to encourage students to develop the necessary skills.

Chapter 2

Spatial skills: A literature review

This chapter will discuss the development, measurement and training of spatial skills, and their links to mathematics. In recent years, the study of spatial skills has become of more interest to educationalists, with research showing that a measure of spatial ability might be as, if not more, predictive of later success in science and engineering careers than the more often measured constructs of mathematical and verbal abilities (Wai et al., 2009). Tests of mathematical and verbal reasoning are used in schools to identify intellectually talented children early on in their school careers, partly to enable them to be directed towards careers in which they can make the greatest contribution. Evidence linking spatial skills in particular to success in STEM (Science, technology, engineering and mathematics) careers suggest that spatial thinking should not be overlooked in the education system (Ministry of Education, 2014).

2.1 Spatial skills and education

In the USA, Project TALENT, and the Study of Mathematically Precocious Youth (SMPY) are longitudinal studies set up to investigate the best ways of identifying and developing abilities for children destined for STEM careers (Wai et al., 2009). Initiated as a consequence of America's efforts to stay ahead of the rest of the world in the 'Space Race', Project TALENT tracked approximately 400,000 American high school students on a variety of measures. The students completed tests of cognitive abilities, academic abilities, personality traits and questionnaires about interests and opinions. The students completed this set of tasks in 1960 and then again in the early 1970s.

The SMPY was started in 1970 by Julian Stanley, an advocate for education for gifted children. It follows a cohort of over 5,000 children and teenagers, identified as intellectually talented, on a similar variety of tasks to Project TALENT. The SMPY is on-going and data from both studies has provided evidence for the importance of spatial abilities in successful STEM careers (Lubinski, 2010; Shea et al., 2001; Wai et al., 2009, 2010). Figure 2.1 shows data from the 20-year report of the SMPY study (Shea et al., 2001). The graph shows the trivariate means for mathematical, verbal, and spatial abilities at adolescence by occupation at age 33. Mathematical ability is shown on the x-axis, verbal ability on the y-axis, and spatial ability represented by the length of the arrowed lines. An arrow pointing towards the right indicates a positive mean spatial ability in relation to the other participants. It can be seen that spatial abilities in those who continued into STEM careers were, on average, a lot higher than those that pursued other careers.

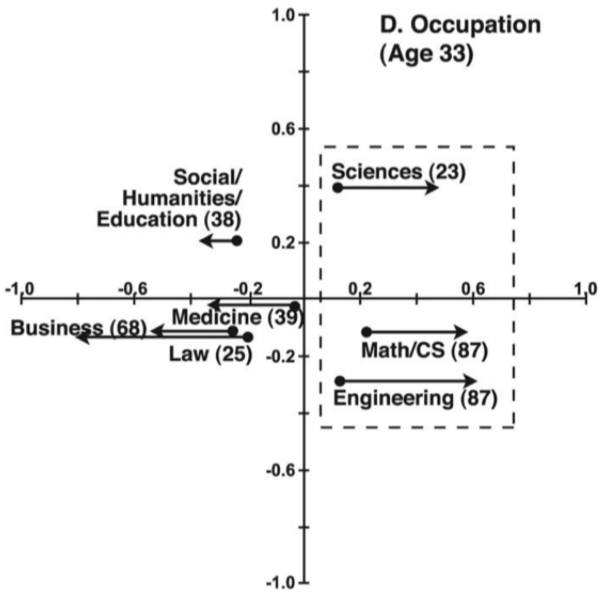


Figure 2.1: Trivariate means (X/Y/Z = Mathematical/Verbal/Spatial abilities in adolescence) for occupation at age 33 (Shea et al., 2001)

Wai et al. (2009) summarised data from both Project TALENT and the SMPY which comprised over 50 years worth of 'cumulative psychological knowledge'. The authors replicated the findings of Shea et al. (2001), finding that spatial abilities, as measured during adolescence, were an important attribute of those that went on to successful STEM qualifications and occupations, in addition to verbal and mathematical skills. These analyses of extensive amounts of data, accumulated over decades, point towards the possibility of an untapped source of future STEM candidates that possess high spatial skills, but not exceptional levels of verbal and mathematic achievement. Wai et al. stressed that educational programmes needed to be better structured to nurture those students with high spatial skills.

However it is very possible that there exist some elements of the structure of the current education system, in America and elsewhere, that do already foster these skills. Mathematics is a potential candidate for this role. Many mathematical topics that are taught in schools have an inherently spatial nature, such as area, volume, and geometry. In addition, the spatial representations of more abstract concepts such as fractions and algebra are often used as teaching tools when first being introduced to children. Being able to visualise, manipulate and reason about spatial representations are all useful in dealing with mathematical concepts, and the specific links that these have with mathematic achievement will be discussed in Section 2.2. Strong links between infants' and children's early understanding of spatial and numerical elements are thought to exist, with an argument for a generalised magnitude system (Newcombe et al., 2015). An example of the way in which these magnitudes are intertwined is the spatial-numerical association of response codes: an internal representation of the basic number line. Referred to as SNARC, the effect can be observed behaviourally when participants are asked to use their left or right hand to indicate whether a number is smaller or larger than a target number (Dehaene et al., 1993). Dehaene et al. found that participants that were asked to react to a smaller number with their left hand were quicker than those asked to use their right hand, indicting that the mental positioning of the representation of smaller numbers was towards the left, and larger numbers towards the right. Much research has confirmed the existence of a left-to-right SNARC effect in cultures that write and read from left to right (See Wood et al. (2008) for a meta-analysis of 46 studies measuring the SNARC effect).

In further support of the links between spatial representation and early mathematical development, there exists neurological evidence that, when performing spatial tasks, the same areas of the brain are activated as those that are associated with number processing. Göbel et al. (2001), for example, applied rTMS[1] to the left and right angular gyrus, a region of the brain in the parietal lobe, and

[1]rTMS (repetitive transcranial magnetic stimulation) involves using a magnetic coil to send small electrical currents to specific brain areas.

found that this disrupted performance on both a number comparison task, and a visual search task. It was suggested by Göbel et al. that a possible explanation for the poorer performance on the numerical processing task was because of the interference with spatial processing. In terms of investigating the formal discipline value that the study of advanced mathematics might have, this evidence from both a behavioural and neurological perspective places spatial skills as a likely cognitive construct to be affected. The studies that make up this thesis will investigate this possibility. The next section introduces each of the spatial skill tasks that are used in Study One of this thesis, their links with mathematics, development, measurement, and viability of training and transfer.

2.2 Measuring spatial skills

2.2.1 Rotation tasks

One type of spatial skill is the ability to perform mental rotations. Rotation tasks require participants to move 2D or 3D objects around an axis, holding the manipulations in their minds. Often these tasks are of the form of a target object, or image, and a choice of rotated shapes, only one of which is identical to the target. Participants must identify this by mentally rotating the target shape and using this to match with the other shapes. Figure 2.2 shows a 2D and a 3D example of these type of tasks.

It is obvious from the examples in Figure 2.2 that mentally rotating objects in three dimensions is more challenging than in two dimensions and this will be discussed in more detail later in this chapter.

Links with mathematics

It is unsurprising that the construct of rotation, or spatial manipulation of any kind, has links with mathematics. In the UK, as with other education systems across the world, the early mathematics syllabus is made up of, in part, teaching and learning in the area of 'shape and space', or 'geometry and measures'. In the UK, at key stage 3[2] (KS3), this topic comprises a quarter of the lesson and assessment content of school mathematics. As well as covering subjects such as measuring lengths and calculating areas, which certainly require an amount of spatial awareness, included in the curriculum is a particular focus on rotation. The programme for study at KS3 includes the following subject content point:

[2]See Appendix 9.4 for an explanation of key stages of the UK examination system.

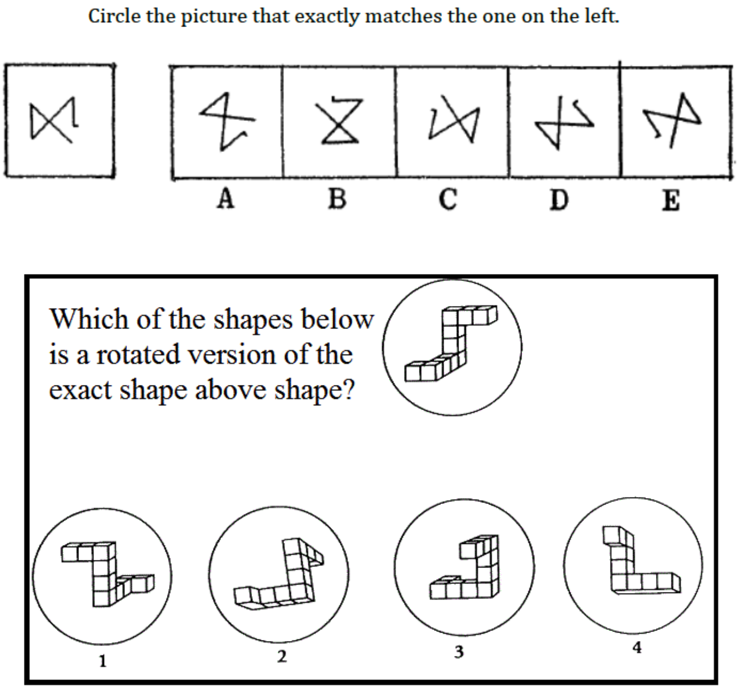


Figure 2.2: An example of a 2D and a 3D mental rotation task. The correct response to the top example is C and for the bottom example, also C.

> ‘identify properties of, and describe the results of, translations, rotations and reflections applied to given figures’ (Department for Education, 2013, p.8).

Figure 2.3 shows an example of the kind of 2D rotation problems that students are expected to tackle early on in the KS3 syllabus.

By the time that students progress to key stage 4 (KS4) they are expected to tackle 2D-rotation problems that involve a higher level of mathematics, but that still depend on the basic manipulations. The programme of study at KS4 states that students should be able to:

> ‘describe the changes and invariance achieved by combinations of rotations, reflections and translations’ (Department for Education, 2013, p.9)

Figure 2.4 shows an example of a more advanced KS4 problem.

Considering the fact that rotation skills make up part of the mathematics syllabus, it is unsurprising that there exists well-founded correlations with

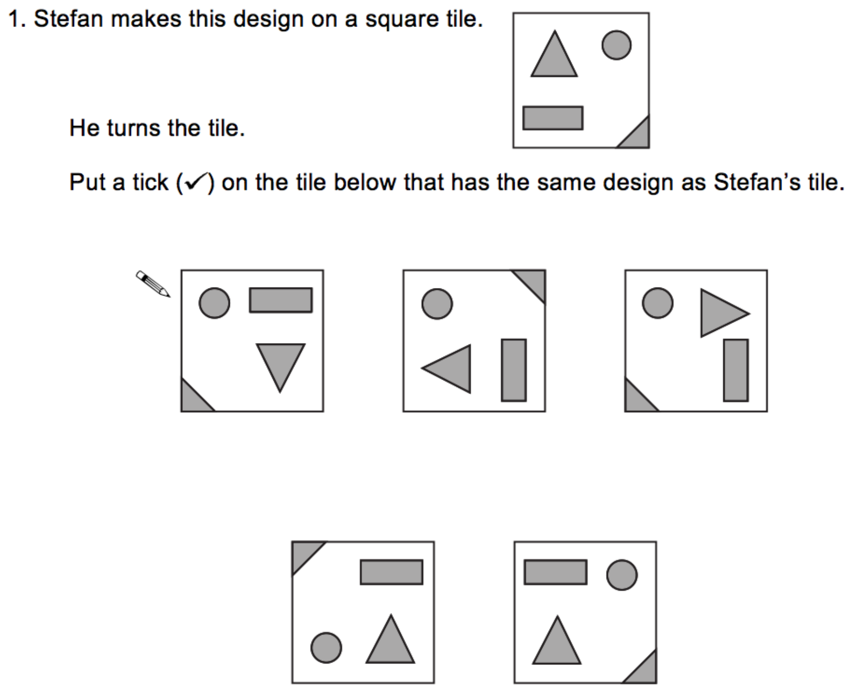


Figure 2.3: An example of a KS3 mathematics 2D-rotation problem. The correct response would be the middle of the top row.

mathematical achievement. For example, Delgado and Prieto (2004) found correlations of $r = .19$ ($p < .01$) between 3D rotation and school mathematics achievement in 13-14 year olds. A number of other studies discussed in this thesis used a spatial ability measure that included a rotational task, and found significant relationships between this and mathematical ability. Mathematical links with rotation skill are found not only for simple calculation, but also more abstract areas of mathematics. For example, Hegarty and Kozhevnikov (1999) found a correlation of $r = .52$ ($p < .01$) between performance on a 2D-rotation task, and mathematical problem solving skills in 11-13 year olds. 3D rotation skills have also been found to have links with mathematics ability, for example Tolar et al. (2009) found correlations between 3D rotation and algebra achievement of $r = .30$ ($p < .05$) in undergraduate students.

The literature discussed in the following sections includes that of 2D and 3D rotation. Similar links are found between mathematics and both dimensions of rotation. However, younger children often find 3D-rotation tasks too challenging, and adults display ceiling effects on 2D-rotation tasks (Jansen et al., 2013).

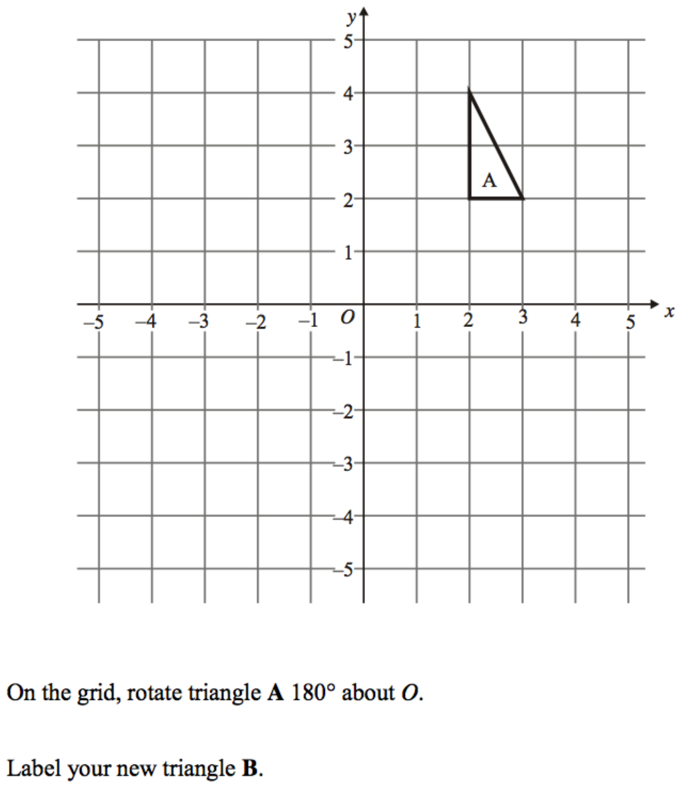


Figure 2.4: An example of a KS4 mathematics 2D-rotation problem. The correct answer would be a triangle drawn between the points (-2,2), (-2,4) and (-3,2).

Jansen et al. suggested that the ages differences in 2D rotation skills could be due to increases in general processing speed.

Viability of training and transfer

To question whether the study of mathematics has the potential to influence the development of spatial skills, it has to be assumed that spatial skills can be trained at all. In terms of rotation skills, many studies have reported improvements as a function of training and some recent research is discussed below.

Studies of the effects of training on mental rotation have reliably found an effect. For example, in children aged 10-11 years old, Wiedenbauer and Jansen-Osmann (2008) found that an intervention group, trained on rotation, improved their reaction times and error rates significantly on a 2D rotation task compared with a control group, although the stability of this improvement was not tested over time. Terlecki and Newcombe (2008) also studied the effects of spatial training on mental rotation performance in 1,300 undergraduate students and found that those students who underwent repetitive practice on a rotation task, and those who had more general spatial training, improved. Terlecki and Newcombe tested 79 of the students again after 2-4 months and found that performance on the rotation task remained significantly higher than pre-test measures, suggesting that the effects were durable.

It seems quite established in the literature that rotation skills can be trained successfully. However, the studies discussed above reported increases in rotation skills after specific training on rotation tasks; what would be labelled as near transfer. For the study of mathematics to hold any formal discipline value in terms of spatial skills, an element of far transfer needs to be possible. Next are discussed some studies that claim to find more far transfer effects of training rotation skills.

Bruce and Hawes (2015) studied the effects of a Lesson Study intervention on the 2D and 3D rotation skills of young children aged 4-8 years. Bruce and Hawes, along with teachers, designed a short syllabus intervention programme that emphasised a spatial approach to mathematics. Around 40 children took part in these activities as a replacement of their regular mathematics lessons for four months. After this period, it was found that children of low, middle and high ability, as well as at each age group, improved their performance on a 2D and 3D rotation task. Whether this example can be considered as far transfer is not clear. Some of the activities that made up the four month intervention were very similar to the pre- and post-tests, and no control group was used as a comparison. However, the teachers that were involved in the study expressed very positive feedback about the study, and the amount of spatial learning that they witnessed in the children. Certainly, a positive feature of this study was that the research was entirely submersed in the practical application of the theory, as opposed to being lab-based as many transfer studies are. This gives any findings the advantage of having the potential to be practically implemented in schools.

Feng et al. (2007) compared performance on a 3D-rotation task before and after 10 hours of action video game training. The task involved the adult participants choosing which of four 3D shapes was a rotation of a target shape, similar to the example in Figure 2.2. A control group spent the same amount of time playing a non-action video game. Feng et al. found an increase in 3D rotation skills, even at a five month re-test compared to the control group. Interestingly, Feng et al. found bigger improvements for females than males in the intervention group, concluding that the findings could be used to attract more females into careers that required a higher level of spatial skills. However, the female group of undergraduates did display much lower pre-test scores on the rotation task, suggesting that the data might be displaying a case of regression to the mean[3]. However, similar findings in terms of the potential of training mental rotation skills and using this to close the gender gap between males and females have been replicated, for example by Okagaki and Frensch (1994).

The literature surrounding interventions to train mental rotation skills suggests that there does exist some potential for far transfer, and therefore helps to confirm one of the basic assumptions necessary for the theory of formal discipline of the study of school mathematics.

Development

Much of the research on the development of mental rotation skills is based on the investigation of sex differences which have been found in infants (Quinn and Liben, 2008), as well as pre-schoolers (Levine et al., 1999) and older participants (Feng et al., 2007). Higher levels of testosterone are thought to be responsible for these sex differences, with higher levels of the hormone in females associated with higher levels of mental rotation skill (Voyer et al., 2016).

Quinn and Liben (2008) tested the mental rotation skills of 24 infants aged 3-4 months. The task involved the infants being familiarised with various rotations of the symbol 1, and then visual preference tested[4] with images of either a mirror image of the symbol (compared to a stimulus figure) or a novel rotation. Quinn and Liben found that male infants displayed a stronger preference than female infants for the mirror image, indicating a very early emerging sex difference in mental rotation. The robustness of findings in the field of visual preference testing in very young infants has been criticised recently for sometimes being too heavily influenced by the theoretical opinions of the researchers, and for employing scientifically unrigorous procedures to achieve significant results (Peterson, 2016). In the case of the Quinn and Liben study, the p-value that the conclusions are based on was only $p = 0.03$. However, despite this not necessarily holding up as evidence of sex differences at this age, it does suggest that infants as young as 3-4 months have started to develop some sense of mental rotation.

[3]Regression to the mean, and other issues concerning the interpretation of data in quasi-experimental studies are discussed in Chapter 3

[4]The study of infants often relies on preference testing to measure cognitive constructs. From birth, infants are known to prefer certain stimuli over others Kirkham et al. (2002). Visual preference testing involves displaying a choice of images to the infant, and measuring the amount of time that the infant fixates on the images. Conclusions about the cognitive processes of the infant are then inferred from which image(s) the infant had preference for.

In slightly older children, Frick et al. (2013) found that 5-year olds were considerable better at rotating a figure to fit into a hole than 3-year-old children (95% compared to 10%). This suggests very large leaps in development of mental rotation skills between these ages. In addition, Kail (1986) found that 8-year-old children were twice as quick at solving mental rotation problems than 5-year-olds. The pattern of development of mental rotation skills seems to follow an almost exponential track from infancy to later childhood, and then to plateau to some extent, as with many other measurable cognitive constructs. Frick et al. (2009) found an effect of age on the performance in a mental rotation task with 5-, 8-, 11-year-olds and adults. See Figure 2.5 for a plot of the error rates and response times of the participants.

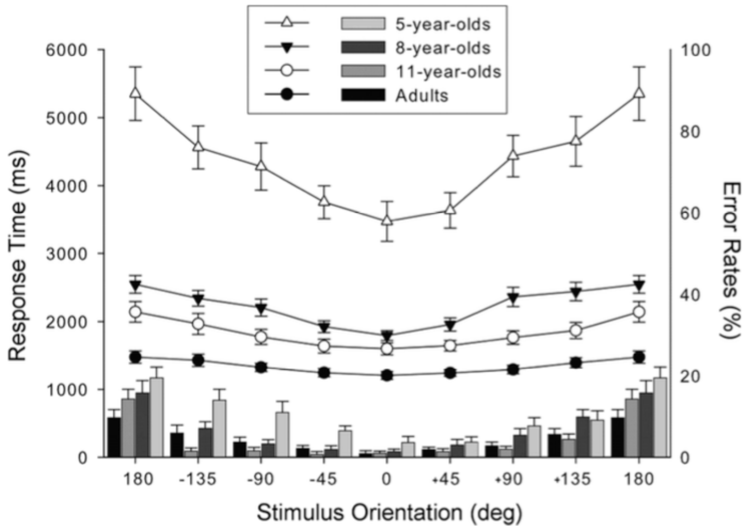


Figure 2.5: Response times and error rates for four ages groups on a mental rotation task (Frick et al., 2009). The response times are represented by the lines and the left-hand axis.

Although sex differences in spatial abilities are not as established through research in children as they are in adults (Spelke, 2005), and have even been found to be non-existent in some cases (Lachance and Mazzocco, 2006), of the studies that claim evidence of sex differences, mental rotation in particular has been shown to display the strongest effect (Levine et al., 1999).

Kaufman (2007) tested 50 males and 50 females on measures of 3D rotation, spatial visualisation, and spatial and verbal working memory in order to establish the extent to which the reported sex differences in spatial abilities were due to working memory capacity differences. Kaufman found that differences in performance between males and females on spatial visualisation was mediated by working memory measures, but not performance on the 3D rotation tasks, on which there was a direct effect of sex. Kaufman concluded that working memory capacity could explain much of spatial skills in general, including 2D rotation, but that 3D rotation involved an additional construct that should be researched separately.

Measurement

The tasks used in the measurement of mental rotation depend largely on the participants that are being studied. For younger populations, 2 dimensional, rather than 3 dimensional tasks are often used. Jansen et al. (2013) found that children aged 6-12 years old were able to distinguish whether 2 dimensional rotated pictures were the same or different, but could not transfer this skill to solve problems in 3 dimensions. See Figure 2.6 for an illustration of the task that was used. Jansen et al. found that the children performed below chance on the last of the examples in Figure 2.6 and therefore was unable to analyse the data in relation to this task.

For older participants, 3 dimensional tasks are more commonly used to measure mental rotation, avoiding ceiling effects. The Mental Rotation Test (MRT), first created by Vandenberg and Kuse (1978) is the most commonly used measure of mental rotation skills in adult populations, and consists of 3 dimensional items, as illustrated in Figure 2.7.

Vandenberg and Kuse (1978) reported that the MRT had an internal reliability of 0.88, and a test re-test correlation of 0.83 after a year. Hirschfeld et al. (2013) tested the reliability of a number of 2D and 3D rotation tasks with different target objects and found that all but one of the ten tasks had split-half reliabilities of between 0.5 and 0.8, and test/re-test reliabilities of between 0.4 and 0.7 after six weeks.

The rotation task chosen to use in Study One of this thesis (see Chapter 4) was adapted from the one used throughout the Project TALENT research discussed earlier in this chapter. Wai et al. (2009) found substantial links between scores on this rotation task (in combination with a number of other tasks) for school-age participants and later success in STEM careers. Project TALENT involved testing the participant at age 13, and then again in college, and later life and did not find ceiling effects at any time point. The participants involved in Study One of this thesis were aged between 16 and 17 years old. As this was younger than the adults involved in Project TALENT, it was possible to be quite confident that ceiling effects would not be found in Study One using the

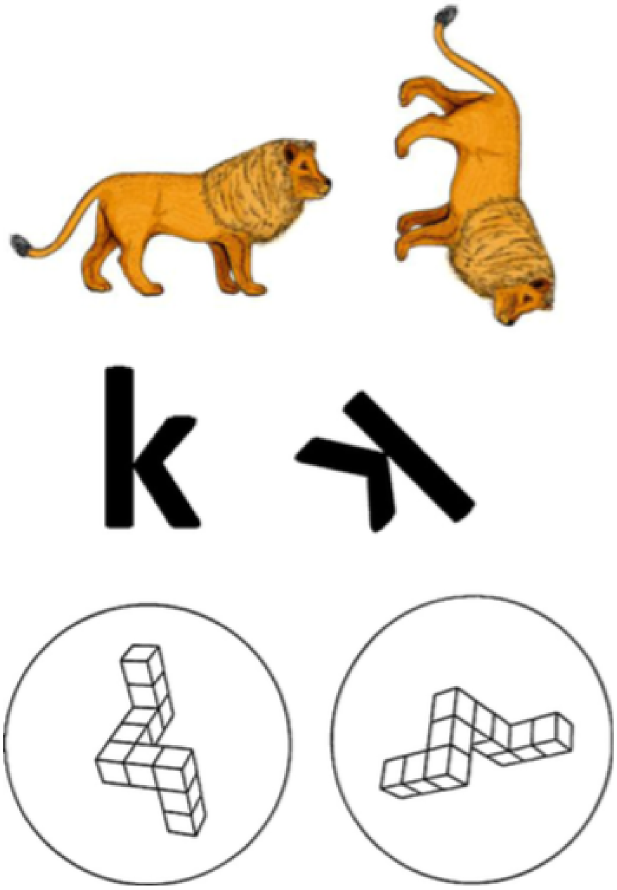


Figure 2.6: Examples of 2D and 3D mental rotation tasks (Jansen et al., 2013). In all example, the rotated image is the same as the original.

same rotation task. An illustration of this task can been seen in Figure 2.8.

2.2.2 Spatial reasoning

As well as spatial measures that involve the manipulation of spatial material, such as mental rotation, often measured is an individual's ability to reason about spatial objects and draw conclusions based on information given. This is referred to as spatial reasoning. A person who is adept at spatial reasoning will very likely also perform well in tasks of mental rotation and visualisation, as these tasks will also involve an element of spatial reasoning.

An example of a commonly used measure of reasoning spatially is the non-verbal task Raven's Progressive Matrices (RPM) (Raven et al., 2000). This task involves a participant being shown a grid of patterns. The last place in the grid

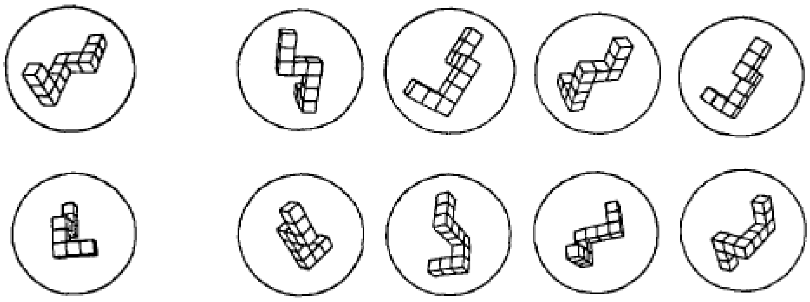


Figure 2.7: An example item from The Mental Rotation Test (MRT) (Vandenberg and Kuse, 1978). In the top example, the correct response would be the third image. In the bottom example, the correct response would be the first image.

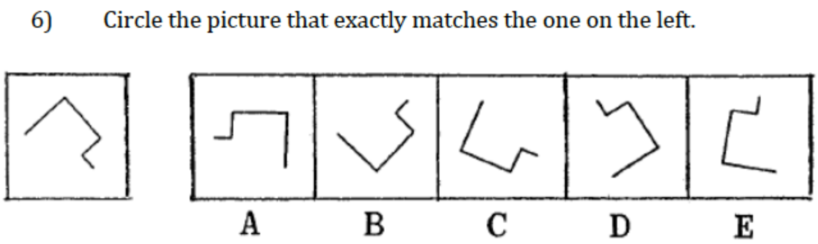


Figure 2.8: An example of the rotation task used in Study One. The correct answer is E

is blank and participants choose which of a selection of images should replace the blank to fit the rules of the patterns. See Figure 2.9 for an example.

RPM are progressive, meaning that the difficulty of the items increases from the beginning to the end of the test. The items nearer to the end of the task require more complex levels of reasoning skills in order to arrive at the correct solution. A number of researchers have attempted to pinpoint the specific cognitive capacities that are captured by an individual's score on RPM. Kunda et al. (2009), a group of Artificial Intelligence (AI) researchers, investigated the information processing demands of RPM and produced a number of algorithms which could be used to solve them. Kunda et al. used evidence of individual differences on RPM performance, due to differing representations of the problems, to program algorithms based on three types of representation; fractal, visual spatial-symbolic, and propositional. These algorithms reproduced a similar pattern of results to that found in human performance and were thought by Kunda et al. to be representative of inbuilt processes of spatial reasoning cognition. Ravens 'Standard' Progressive Matrices, the version of RPM that was

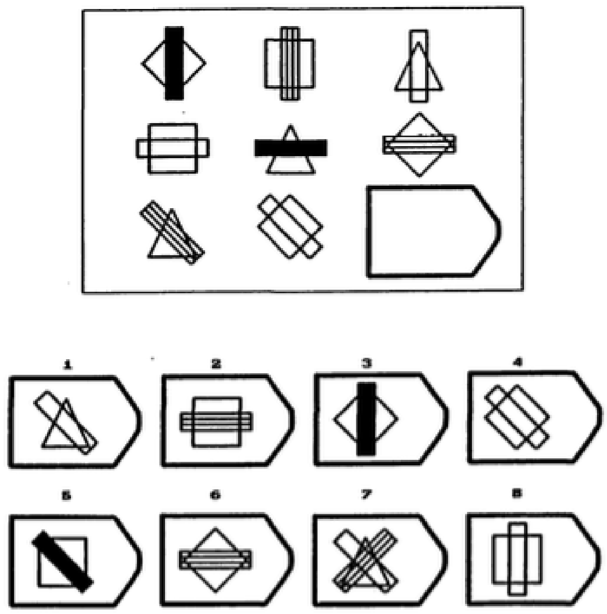


Figure 2.9: An example of Raven's Standard Progressive Matrices. The correct answer is 5

used as a measure of spatial reasoning throughout this thesis, consists of 5 sets of 12 matrices, labeled A to E. The 60-item test has a cyclical format in which each set begins with a very easy introductory item. The subsequent 11 items of the set build on this theme and become progressively more difficult. The five sets individually focus on subtly different reasoning strategies with sets D, E and the later items in set C requiring reasoning by analogy (Raven et al., 2000). It is thought that the AI analogies found by Kunda et al. (2009) represent these different strategies.

Ravens Progressive Matrices as a spatial reasoning measure

RPM are most established as a measure of fluid intelligence, or Gf. RPM is thought of by many as a pure measure of non-verbal intelligence, and that any factorial loadings onto other constructs are negligible (Jensen, 1980). However, more recently, much research has shown that the items that make up RPM require spatial strategies to be solved (Carpenter et al., 1990; Colom et al., 2004; Mackintosh and Bennett, 2005; van der Ven and Ellis, 2000). From the example in Figure 2.9, it can be seen that RPM is a visual task. It is therefore not surprising that success on the task is reliant on a certain amount of spatial reasoning capacity. Carpenter et al. describes the example above as following three rules:

1. Each row contains three geometric figures (a diamond, a triangle and a square) distributed across its entries
2. Each row contains three textured lines (dark, striped and clear).
3. The orientation of the lines is constant within a row, but varies between rows (vertical, horizontal then oblique) (Carpenter et al., 1990, page 4)

Each of these rules require visual and spatial analogies to be made to solve successfully. Lim (1994) also found evidence for the spatial nature of RPM in that sex differences on RPM could partly be explained by the variance of participants' performance on a number of spatial tasks. Lim suggested that the poorer performance of females on RPM could be due to their tendency to be over reliant on verbal reasoning skills. Colom et al. (2004) also found that the male advantage in performance on RPM was non-significant when spatial skills were controlled for.

Links with mathematics

The fact that the construct of spatial reasoning will involve elements of the other construct discussed in this chapter meant that much of the research described in the links between them and mathematics is also relevant to spatial reasoning. Specifically focusing on the relationship between mathematics and performance on RPM as a spatial reasoning measure, reliably significant correlations are found.

Attridge and Inglis (2013) researched whether studying one academic year of mathematics at AS-level was associated with a change in a number of reasoning skills. As reported in Section 1.4, it was found that the 44 mathematics students behaved differently when faced with conditional inference problems in comparison to a group of 38 students who were studying English literature, and not mathematics. As a measure of non-verbal intelligence for the groups, an 18 item subset of Raven's advanced progressive matrices (RAPM) was used. Table 2.1 shows the mean RAPM scores for both groups. Attridge and Inglis found that the mathematics group scored significantly higher than the comparison group at Time 1 ($t(80) = 3.43, p = .001$) and at Time 2 ($t(80) = 4.94, p < .001$), despite the prior achievement (a summed score of their GCSE results) of the two groups not being significantly different ($t(112) = 3.89, p = .089$).

Interestingly, a re-analysis of the raw data obtained by Attridge and Inglis (2013) found that the mathematics students significantly improved on RPM

Table 2.1: Mean(SD) scores on RAPM from Attridge and Inglis (2013)

Time	Mathematics	English Lit.
Time 1	9.57(3.26)	7.03(3.45)
Time 2	10.64(2.93)	7.34(3.11)

from Time 1 to Time 2 ($t(43) = 2.59, p = .013$), whereas the comparison group did not ($t(37) = .71, p = .484$). Attridge and Inglis did not report this data as an important aspect of their findings as there was no significant Group × Time interaction effect found ($ps > .20$). Figure 2.10 illustrates this interaction.

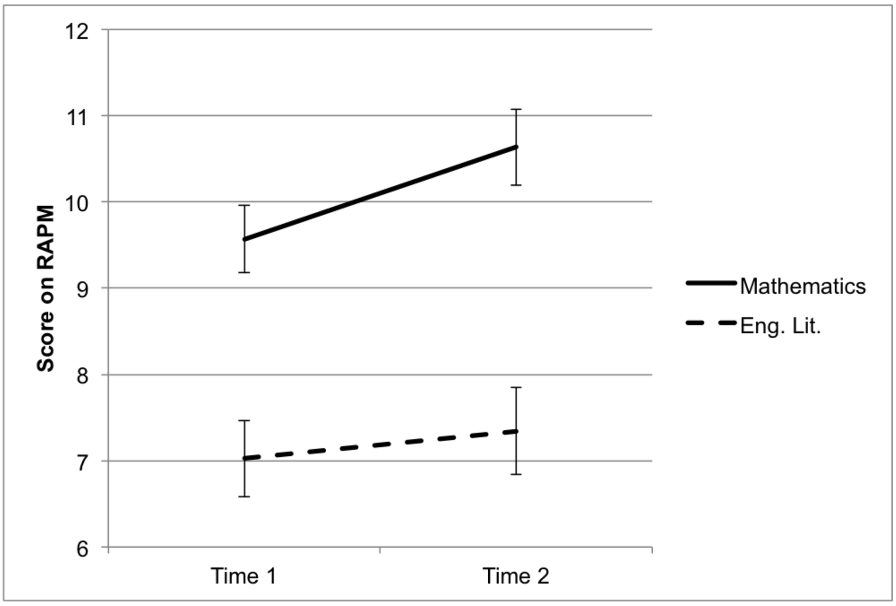


Figure 2.10: A plot of the interaction between Group and Time, obtained from data in Attridge and Inglis (2013)

This finding suggests that the study of mathematics has the potential to improve performance on RAPM, evidence of far transfer, and strong links between mathematics and spatial reasoning.

Viability of training and transfer

The possibility of being able to train performance on RPM is of great interest to researchers because of the strong links that RPM performance has with many positive life outcomes (Gottfredson and Deary, 2004). Much of the research evidence for the viability of training participants on spatial reasoning, as measured by RPM, comes from the context of 'brain fitness' programs which are designed to train the brain with repetitive working memory style tasks. This has become a large, money making business in the last few years (SharpBrains, 2013) and it is predicted by SharpBrains, a company that tracks the brain-fitness industry, that by 2020 the market will be worth $6 billion. These 'brain fitness' programs are targeted at a wide audience, but have been shown to be most effective for elderly people (Angelakis et al., 2007), or as an alternative to drug treatment for children (and adults) suffering from disorders such as attention deficit/hyperactivity disorder (ADHD) (Klingberg et al., 2002b).

Many of the companies that stand to gain from the sale of these 'brain-training' programs claim to have strong evidence for their effectiveness in healthy adults and the general population (see Rabipour and Raz (2012) for a summary of these companies and their research), but much of the data that these claims are based on has been obtained using questionable research designs, small numbers of participants, and insubstantial control measures (Buschkuehl and Jaeggi, 2010; Chooi and Thompson, 2012; Rabipour and Raz, 2012; Redick et al., 2012)

Despite the claims made from studies based on dubious theoretical rationales and poor research designs, there are a number of findings that should be noted. Klingberg et al. (2002b) trained children diagnosed with ADHD on a variety of working memory (WM) tasks such as backwards digit span[5], visual span[6], and go-no go tasks[7]. It was found that, after 25 weeks of training, the children improved significantly on a test of RPM. Within the same paper, Klingberg et al. documented another working memory (WM) training study in which adult men were tested on RPM before and after 26 days of training. These men also showed a significant positive increase in spatial reasoning. Although this study provides evidence of training spatial reasoning, it must be noted that the groups of participants were small (7 children in the treatment group for the first experiment, and only 4 males in the adult study) and that, although the first study included an active control group of 7 children, there was no control used in the adult experiment. Flawed methodologies such as these are prevalent in brain-training research, and any claims made by large companies such as Cogmed, the leading producer of these 'brain fitness' programs, should be investigated fully before drawing conclusions.

[5]The participant heard a series of single digit numbers read out aloud, and then was required to key them in in reverse order.

[6]Circles were presented one at a time in a four-by-four grid. After a delay the subjects indicated the positions of the circles. The number of circles in the sequence was successively increased until the subject missed two trials in a row.

[7]Two grey circles were presented on a screen. Participants were then required to press a spatially congruent key when one of the circles became green, and to withhold responding when one of the circles became red.

Jaeggi et al. (2008) showed that training on a WM task improved performance on RPM. Thirty four healthy adults performed n-back tasks[8] daily for 25 minutes and were tested on RPM before and after intervals of 8, 12, 17 and 19 days. Jaeggi et al. found an improvement in the RPM scores from Time 1 to Time 2, as well as an effect of training period (the more days the participants trained for, the more their scores on RPM improved), irrespective of initial intelligence scores. Jaeggi et al. believed that the gains in spatial reasoning occurred due to the fact that a number of executive processes, such as memory, attention, and inhibition were being continuously engaged whilst the adults were occupied by the tasks. The experimental groups were compared with an inactive control group of 35 healthy adults that did not engage in any tasks but were tested on RPM at the same times as the experimental groups. This control group also showed significant gains in spatial reasoning over the time period, which the authors put down to being a test-retest effect. This study, although cited in many articles as evidence for transfer of training (e.g. Sternberg (2008)), has also been criticised (Chooi and Thompson, 2012; Moody, 2009; Rabipour and Raz, 2012). The most worrying criticism of the study is the mis-use of the RPM task which should have an administration time of much longer than 10 minutes, the shortened test-time which has been employed a number of times by the same research group (Jaeggi et al., 2008, 2010, 2011). By cutting down the time allowed to answer the questions, the participants would not have a chance to reach the more challenging questions of the progressive test which are designed to fully test spatial reasoning. Instead, the test that Jaeggi et al. used could be better described as a speed test (Moody, 2009). This methodological issue, as well as the lack of an active control group, should be remembered when considering the claims that this paper is making.

Development

Developmental differences on performance on RPM have been often linked with the development of working memory capacity in children, and to the decline of both in elderly populations (Fry and Hale, 1996). The two constructs have been shown to share as much as 50% of their variance, and therefore the development of them follows very similar trajectories (Kane et al., 2005). Correlations between age and performance on RPM are found to be approximately $r = .6$ (e.g. Fry and Hale (1996)). A meta-analysis of 57 studies of general population

[8]A n-back task requires participants to decided whether a stimulus matches another from n steps earlier in the sequence of stimuli. The larger the number n is, the more difficult the task is, requiring the storage and manipulation of information in WM.

samples found that there were no sex differences found on performance on RPM from ages 6-14, but that from 15 to adulthood, there existed a male advantage equivalent to 5 IQ points (Lynn and Irwing, 2004)

Measurement

In all three studies of this thesis, spatial reasoning is measured using RPM. This task is very widely used in psychological research and is available in versions that are suitable for testing children as young as 5 years old to elderly populations. First developed in the mid 1930s, RPM has been revised and standardised many times among young people across USA and Europe. Studies into the reliability and validity of RPM in clinical and normally developing populations have covered a wide range of ages and cultural groups. In general, internal consistency, retest reliability, and concurrent and predictive validities are good (Pearson, 2007; Raven et al., 2000).

Employing RPM as a comparison measure in a longitudinal study of the effects of formal discipline requires the test to be taken at least twice by the same group of participants and, consequently, practice effects must always be taken into account. Bors and Vigneau (2003) tested 67 adults using RPM at three separate time points in order to evaluate the effects of practice. They found that the total scores on RPM did increase significantly across the three time points, but that this increase was due to a reflected learning improvement as opposed to remembering specific items or improved time strategy. The participants' spatial reasoning scores increased by an average of approximately 3% each time. Bors and Vigneau used an identical version of RPM at each time point, a procedure that is avoided in most longitudinal research studies. However, if the increased scores were due to increased reflective learning, practice effects will be present even when RPM are employed as two sub-sets (e.g. odd and even numbered items) and should be considered in any analysis of results.

2.2.3 Visuo-spatial working memory

Working memory (WM) refers to a person's ability to hold and manipulate information mentally over short periods of time. WM capacity is limited, varies with age, and is sensitive to individual differences. The most common model of WM was developed by Alan Baddeley and Graham Hitch which stemmed from the concept of short-term memory (STM), the capacity for holding, but not manipulating, information in an accessible state, with added functions for processing as well as storage. (Baddeley and Hitch, 1974). The Baddeley and

Hitch model consists of three sub-components[9]. See Figure 2.11 for a simplified representation of the model.

- **the central executive** - the attention-controlling system. The central executive allocates tasks to the following two slave sub-components as well as being thought to play a crucial part in problem solving and reasoning.
- **the visuospatial sketch pad** - controls the manipulation and storage of visual and spatial information.
- **the phonological loop** - stores and rehearses sound-based information. The phonological loop is thought to consist of two parts: The phonological store which holds information for only a few seconds, and the articulatory control process which is used to rehearse information. (Baddeley, 1992)

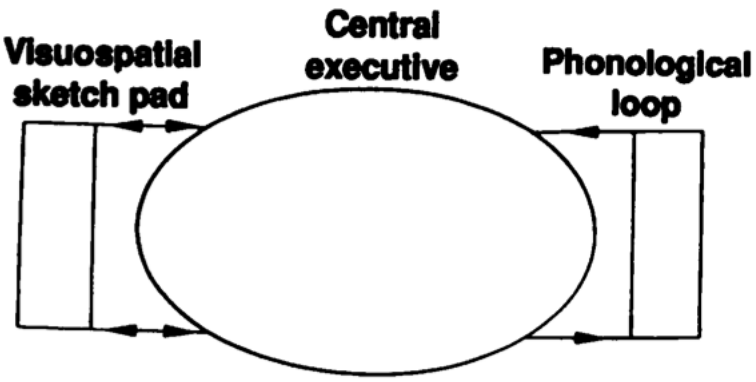


Figure 2.11: A simplified representation of the Baddeley and Hitch working memory model (Baddeley and Hitch, 1974)

WM that is controlled by the visuospatial sketch pad is known as spatial WM and that which is controlled by the phonological loop is classified as verbal WM. Evidence for the presence of two, distinct subsystems within the WM model came from Baddeley and Hitch's research that showed that, although performance on a variety of tasks such as comprehension and reasoning suffered when the participant was asked to concurrently remember strings of digits, it was by much less than was predicted. It was still possible for individuals to learn effectively even when their 'digit memory' was at full capacity (Baddeley,

[9]Baddeley (2000) later updated the model to include an 'episodic buffer' which is thought to help to make links between the other three components and to long-term memory (LTM)

1992). These findings suggested that STM consisted of more than just one element, and that these elements could take on different roles in tasks.

The suggestion that there are two distinct systems of working memory, spatial and verbal, is well established in the literature. For example, Myerson et al. (1999) provided evidence for separate verbal and spatial elements to WM in adults by asking participants to perform either a digit (verbal) or location (spatial) memory task. The participants that were asked to name a selection of colours as a dual task performed worse on the verbal task, but not on the spatial one, and those that were asked to point to the matching colour performed worse on the spatial task, but not on the verbal one. Evidence for the separate nature of verbal and spatial processing in WM also comes from neuropsychology. Smith et al. (1996) used position emission tomography (PET) to show significant differences in cerebral blood flow in the brain when undertaking a spatial task compared to a verbal task. When participants were asked to remember the names of four letters, the left-hemisphere of the brain was activated. When they were asked to remember the position of four dots on a grid, the right-hemisphere was activated. Although the tasks employed by Smith et al. may strictly be considered STM tasks, rather than WM tasks, the study provided strong evidence for the presence of separate neural structures for verbal and spatial WM.

A typical WM task would require the participant to remember a string of words, numbers, letters, or spatial positions at the same time as completing a mental processing task. Often, participants are asked to repeat the list backwards, requiring processing over and above just remembering the list, or are asked to answer questions about the content (e.g. which number was the biggest?) before repeating the list back. A high WM capacity has been strongly linked to better performance in many tasks, as well as real-life skills such as mathematics learning (Passolunghi and Siegel, 2004) and reasoning (Kyllonen and Christal, 1990). The following sections will elaborate on the development of verbal and spatial WM and their links to mathematics.

Links with mathematics

To perform almost any task, including that of mathematical processing, a level of WM capacity is required. It is therefore not surprising that links can be found between WM and mathematical achievement (Cragg and Gilmore, 2014; Raghubar et al., 2010). In order to successfully complete mathematical procedures both at a simple and more advanced level, information needs to be held and processed in order to arrive at a solution.

In particular, spatial working memory (SWM) is found to correlate with simple arithmetic skills, and more complex mathematical word problem solving. Andersson (2007) found correlations of $r = .26$ and $r = .35$ respectively in 9-10 year old children, finding also that a measure of verbal working memory (VWM) correlated by similar amount with arithmetic ($r = .29$), but by much lower with word problem solving ($r = .13$). This particular SWM link to mathematics has been studied in children and adults, often through dual task experiments in which the participants perform a mathematical processing task at the same time as a WM task. If the WM task interferes with performance on the mathematical task, this is assumed to be because the mathematical task is requiring some WM capacity. In terms of simple arithmetic in adults, there is little evidence of a role of SWM, and the specific SWM element in children's mathematical processing is not fully understood. Hubber et al. (2014) investigated the role that SWM played in mental arithmetic in adults and found that dual task performance on the arithmetic task was affected by SWM load, particularly when the participants were employing counting strategies, but similar effects were also present for retrieval.

Research into the role of SWM, as opposed to VWM, and mathematics is lacking in the literature, particularly for mathematical processes more complex than arithmetic, and particularly for adults. One example is Wei et al. (2011) who found correlations between SWM and advanced mathematical processing in undergraduate students in China. The measure of advanced mathematics included a number of high-level topics such as algebra and geometry. A regression model from the Wei et al. study showed that SWM predicted mathematical performance over and above VWM. However, the SWM task employed by Wei et al. simply required the participants to remember the location of a series of dots on a grid, and not to process any information, meaning that the task could be considered as more of a short-term memory task than a working-memory task. Another recent study that linked more complex mathematics with SWM measures was an investigation of whether mathematicians have superior WM capacities in comparison to non-mathematicians (Hubber, 2016). Forty-four adult participants (27 mathematicians) were tested on VWM and SWM. Hubber found that the mathematicians scored significantly higher than the non-mathematicians for SWM but there were no differences in VWM. This result contradicted a number of studies which have found a general working memory advantage for mathematics, for example Dark and Benbow (1990), but provided evidence of an important link between mathematics and SWM.

Viability of training and transfer

Due to measures of working memory capacity's high correlations with performance in a range of tasks, there has been much research interest about whether it is possible to increase WM skills, in the hope of this leading to an increase in the performance in more general tasks. Therefore, much of the literature surrounding the viability of training WM is concerned about the ways in which the training can be transferred to other non-WM tasks, as discussed in the previous section relating to transfer to performance on RPM. In a review of the plasticity of working memory capacity, Klingberg (2010) found strong evidence of training effects. Klingberg referenced studies in which a computerised WM training task transferred to a number of other WM tasks, and to other cognitive constructs such as inhibition and reasoning. A number of studies discussed earlier aimed to train performance on RPM also found that performance on working memory tasks improved, but this transfer can only be considered as very near (e.g. Jaeggi et al. (2008)). Melby-Lervåg and Hulme (2013), in a meta-analysis of 23 working memory training studies, concluded that both VWM and SWM could be trained, but that there was only evidence for a sustained effect for SWM. The meta analysis found no evidence of transfer of skills from the working memory tasks to other skills, putting into doubt the usefulness of working memory training programmes for enhancing any cognitive functioning in typically developing individuals. The research literature provides strong evidence that both SWM and VWM can be trained, as well as some evidence that this training is durable over time. However, there is less evidence of far transfer. If the study of advanced mathematics is to have any formal discipline value in training SWM, it is likely that this will be from the more spatial aspects of the mathematics syllabus.

Development

Less is known about the development of working memory in relation to the visuo-spatial sketchpad than the phonological loop, as described in the working memory model (Baddeley, 1992). Unsurprisingly, performance on SWM tasks is found to increase with age, alongside increases in the other spatial tasks discussed in this chapter. Pickering (2001) reviewed a number of studies of children's SWM development in which marked increases were found between the ages of 5 and 15 on a variety of SWM tasks. It is suggested by a number of studies reviewed by Pickering (2001) that these increases are not only due to an increased WM capacity, but also due to attentional capacity and the use of more advanced memory strategies, with older children being more likely to employ some verbal as well as visual methods when memorising and classifying visual patterns.

As with many other cognitive constructs, SWM seems to develop at a faster rate in childhood, and then to plateau, before declining in elderly populations (Rowe et al., 2008). There does exist evidence that developments in SWM are still happening during early adulthood. Zald and Iacono (1998) found that male participants aged 20 years performed significantly better on a SWM task compared to 14 year olds.

The large amount of literature that focuses on sex differences in the development of all spatial skills often suggests that the observed differences are due to differences in SWM. Kaufman (2007) found that sex differences between students aged 16-18 on a variety of spatial measures were completely mediated by SWM. Voyer et al. (2016) performed a meta-analysis of 98 samples of non-clinical male and female populations aged 3 to 86 years old. It was found that a significant male advantage was present, but that the effect size was small. Age was also found to be a significant moderator of the effect, with sex differences in SWM not appearing until 13 years old.

Measurement

A pure measurement of SWM is reportedly hard to achieve due to participants tending to employ a combination of verbal and spatial strategies when memorising and processing information (Pickering, 2001). Hitch et al. (1988) studied SWM in young children and found that only the youngest children (five to six years old) displayed impaired performance when the task was manipulated to be more confusing with the use of visually similar stimuli. Hitch et al. suggested that this was because the youngest children were relying heavily on their visuo-spatial sketchpad capacity, whereas the older children were also using resources from their phonological loop.

The most common form of SWM task involves a participant being asked to memorise a sequence of locations of an object in order, which requires short-term memory, at the same time as performing some kind of processing task. This processing task can be spatial or non-spatial but its additional presence is what makes the task a measure of working memory, rather than just short-term memory. A participant's working span is commonly calculated as the longest chain of items that they can reliably remember. Traditionally, working memory tasks consist of presenting participants with increasingly longer lists of items, with the task cutting off when a predetermined number of mistakes are made. However, this method may not encapsulate an individual's capacity completely accurately. Another method is to present all of the lengths of lists of items randomly to the participant, and then calculate the number of correctly remembered lists. Chapter 4 details this method that was employed for the research of this thesis.

2.2.4 Spatial visualisation

The definition of spatial visualisation would very likely include the skills of mental rotation, spatial reasoning, and perhaps even spatial working memory. In the research literature, these things are not always referred to as separate constructs, and the definition of spatial visualisation is not clear-cut. For example, one definition is provided by Salthouse et al. (1990):

> "... the mental manipulation of spatial information to determine how a given spatial configuration would appear if portions of that configuration were to be rotated, folded, repositioned, or other- wise transformed" (Salthouse et al., 1990, p. 128).

This definition would certainly include elements of all of the spatial skills previously discussed in this chapter. Spatial visualisation could be thought of as the ability to use all of these successfully in conjunction with one another. Therefore, much of the relevant research has already been discussed in previous sections.

Links with mathematics

The ability to visualise a situation spatially is something that many mathematicians claim to be the key to understanding complex and abstract concepts. In the teaching of mathematical concepts to children, such as fractions, successful learning activities often involve an element of representing the ideas in pictures and diagrams. The way in which children of different abilities used spatial visualisation in solving mathematical problems was investigated by van Garderen (2006). The author found that the most mathematically gifted students performed best on a measure of spatial visualisation, and that the use of visual images was positively correlated with higher mathematical word problem solving performance. Tolar et al. (2009) developed a structural model of students' algebra achievement which considered a number of cognitive abilities and arithmetic skills. The authors found that performance on a 3D spatial visualisation task explained some of the variance of both algebra achievement, and SAT-M scores[10]. The effects of working memory on these higher-order mathematical measures was mediated by 3D visualisation and computational fluency. These research studies confirm that the ability to visualise situations spatially is of benefit when attempting to solve mathematical problems, and this is linked more strongly with higher-order mathematics as opposed to arithmetic skills.

Viability of training and transfer

A small amount of research has been conducted into the potential training of spatial visualisation skills through the study of academic subjects (Blade and Watson, 1955; Burnett and Lane, 1980). The literature is dated, but is particularly relevant to this thesis, and to the theory of formal discipline. Blade and Watson (1955) tested 89 engineering students' spatial visualisation skills before and after one year's study, and 46 of these students again at the end of their four-year course. The same test was also given to 77 non-engineering students who acted as a control group. Blade and Watson found that the engineering students' spatial visualisation scores improved approximately three times more than the control group after one year of study. This finding was replicated in another group of 593 engineering students at another university, confirming that the effect was not due just to a particular teaching style. The 46 students who took the test a third time, after four years, maintained the accuracy rates that they had displayed after one year. This suggests that the training was durable, but does bring into question why no further gains were seen after an additional three years of study.

Burnett and Lane (1980) tested 142 college students' spatial visualisation skills before and after two years of study. The task used was the Guildford-Zimmerman (G-Z) Spatial Visualisation Test, designed to measure "the process of imagining movements, transformations, or other changes in visual objects" (Guildford and Zimmerman, 1948). As well as completing the spatial visualisation task, the students also reported the college courses that they were studying. Burnett and Lane found that the students who studied mathematics and physical sciences improved significantly in terms of their scores on the pre- and post-test, whereas those students who were studying humanities and social sciences did not. Burnett and Lane also performed a multiple regression to predict the gains in spatial visualisation scores, with number of courses taken in each academic area as predictor variables. The only variables found to be significant in the model were the number of mathematics courses, and the number of engineering laboratories.

[10]Scholastic Assessment in Mathematics

Both the Blade and Watson (1955) and the Burnett and Lane (1980) studies provide evidence for the viability of training spatial visualisation skills, and for far transfer from academic subjects. In the discussion of both studies, the authors made an attempt to hypothesise about what elements of the courses that showed an effect were responsible for the training. Blade and Watson suggested that one aspect in common to all the engineering students was thier experience in mechanical drawing. Burnett and Lane suggested that the observation of the predicting power of the number of mathematics courses taken was the common element responsible, and that this could also explain the findings of Blade and Watson as these students that shared the mechanical drawing experience would also have had a large amount of exposure to mathematics. More recently, Sorby (1999) reported on a pre- and post-testing of engineering students on spatial visualisation skills after studying a variety of graphic design modules, similar to the mechanical drawing described by Blade and Watson. Sorby found that the students who had spent more time sketching designs by hand, rather than using computer simulations, improved most in their spatial visualisation performance. This provides further evidence for the potential far transfer of skills from education to more general spatial skills.

Development

The development of spatial visualisation skills is understandably very closely linked to the development of the more fine-grained spatial skills such as rotation skills, and spatial working memory (Sorby, 1999). These constructs develop with age, and sex differences at different ages have been reported (Rowe et al., 2008; Levine et al., 1999). It is suggested by research that children are able to cope with 2D spatial visualisations sooner than they are 3D ones (Sorby, 1999), and that the skill declines in older adults, with 60 year olds performing 1-2 standard deviations below the performance of 20 year olds (Salthouse et al., 1990). Again, this is consistent with other constructs such as working memory.

Measurement

The most commonly used measures of spatial visualisation involve a process of mentally rotating, folding or cutting an object, and imagining what the resulting object would look like. Figure 2.12 illustrates two examples of a folding task.

For the first task, the participant is required to state which of the 5 images on the right would be the result of printing the dot by folding the paper, the correct answer being C. The second task would ask which of the edges 1-5 of the 2D net correspond with which edges A-G of the 3D shape. Another commonly

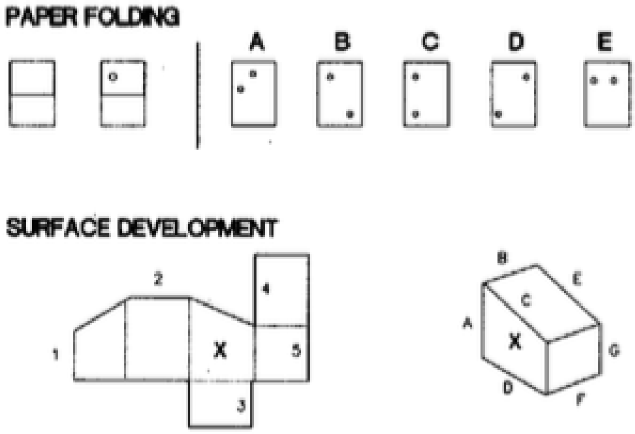


Figure 2.12: Examples of folding spatial visualisation tasks. For the paper folding task, the correct answer would be C, and for the surface development task, an example would be that edge 5 corresponds to edge F.

used task of spatial visualisation is the DAT:SR[11] which is illustrated in Figure 2.13.

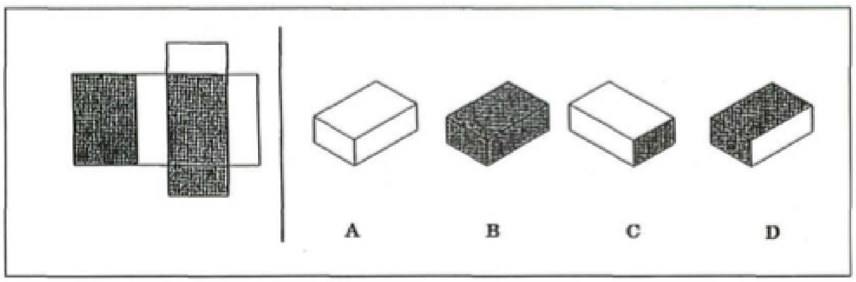


Figure 2.13: An example of an item from the DAT:SR. The correct response would be D.

Here, the participant is required to fold the 2D net into a 3D shape mentally, and then select the correct image from the choices on the right. The correct answer is D. The DAT:SR has, from a variety of spatial tasks, been found to be the best predictor of success in an engineering university course (Medina and Sorby, 1998). The spatial visualisation task used in Study One of this thesis (see Chapter 4) uses a task very similar to the DAT:SR, adapted from the same Project TALENT booklet of tasks as the 2D rotation task (Wai et al., 2009).

[11]Differential Aptitude Test: Spatial Relations.

2.3 Classification of spatial skills

Establishing a clear typology of spatial skills is not a straightforward task. Past research has strived to identify distinct sub-skills within the construct through exploratory factor analysis, some identifying as many as ten sub-skills (although intercorrelations between these are very high), and some identifying just one general spatial factor (Guttman et al., 1990). Uttal et al. (2013a) argue that the reason that this exploratory factor analysis approach has failed to be successful in establishing a clear and agreed-upon typology is that "tests of spatial ability did not grow out of a clear theoretical account or even a definition of spatial ability" (p. 353). Uttal et al. therefore proposed a typology built on linguistic, cognitive, and neuro-scientific evidence, placing spatial sub-skills, and their related tasks, along two dimensions: intrinsic vs. extrinsic, and static vs. dynamic. Figure 2.14 illustrates this.

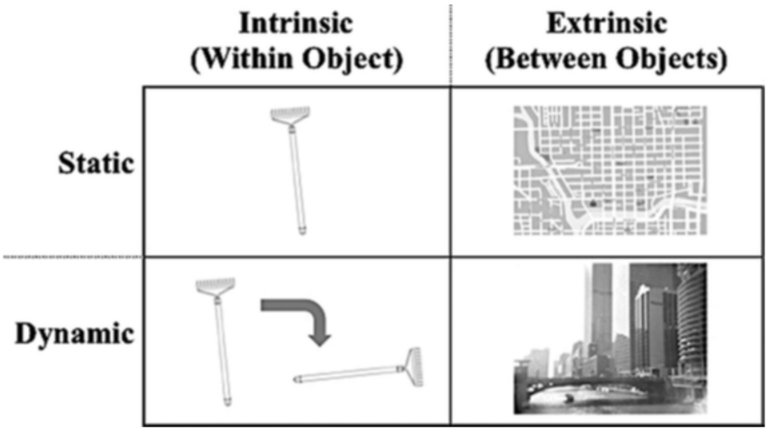


Figure 2.14: A suggested typology of spatial skills (Uttal et al., 2013a)

Intrinsic spatial tasks focus on the the spatial characteristics within an object, and extrinsic tasks on the spatial relationship between objects. Static tasks involve no movement, whereas dynamic tasks do. The following sections will discuss how the four spatial sub-tasks used in this thesis might be classified in this typology.

2D rotation task

This task is clearly dynamic, as the movement of rotation is required, and as only one object needs to be considered, it can be classified as intrinsic, placing it in the bottom left corner.

Spatial reasoning task

The matrix reasoning task used involves no movement of objects, and therefore is static in nature. As with the rotation task, there is no requirement to consider the spatial relationship between objects, as there would be in an example such as a map reading task, and therefore would also be classified as intrinsic. However, the matrix task does involve an element of comparison between objects in relation to their spatial features, and so is not as clearly defined by the 2×2 typology.

Visuo-spatial working memory task

This task is also a little difficult to define with the typology proposed by Uttal et al. (2013a), most likely owing to the fact that it is very much a memory task as well as a spatial task. Participants are not required to perform any movement of objects, making it a static task. However, whether or not the task should be classified as intrinsic or extrinsic is, in part, dependent on how it is assumed that the participant approaches the task. It could be the case that an individual would memorise the position of each dot alone, and therefore intrinsically (although even this would be expected to be in the context of the spatial relationship between the dot and the grid). What is more likely is that the participant would consider the spatial relationship between the dots in order to aid the memorisation of them, making the task extrinsic.

Spatial visualisation task

The spatial visualisation task required participants to mentally manipulate a 2D net into a 3D shape, making it dynamic in nature. As only one object, and its internal spatial features, was under consideration, the task can be classified as intrinsic.

Figure 2.15 illustrates where the tasks used in this thesis might be placed on the typology suggested by Uttal et al. (2013a).

It can be seen from Figure 2.15 that the tasks of this thesis cover the span of the typology very well, with the exception of the inclusion of an dynamic extrinsic task. In terms of the focus on the theory of formal discipline, the transfer of skills from school mathematics, and the mechanisms through which this might be possible (see Section 1.6), this exclusion is justifiable. Uttal et al. (2013a) describe tasks within the 'dynamic extrinsic' category as "thinking about how ones perception of the relations among objects would change as one moves through

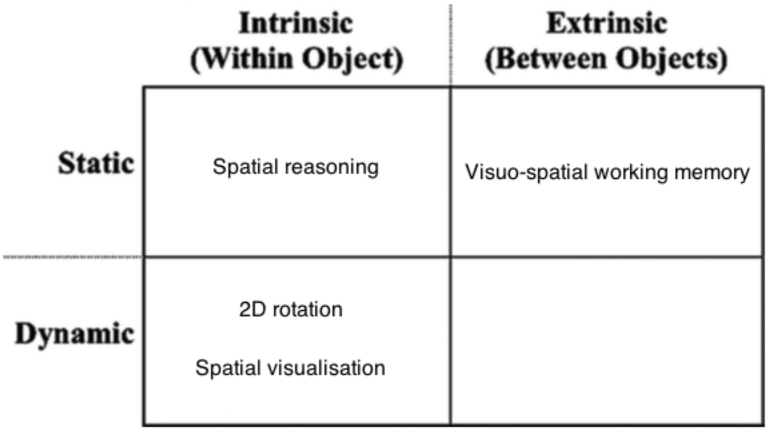


Figure 2.15: Placement of the tasks of this thesis on the typology proposed by Uttal et al. (2013a)

the same environment" (p. 354), or "visualising an environment in its entirety from a different position" (p. 355). These types of skills, and their related tasks, would be expected to have very little measurable relationship with the learning of school mathematics, and no explainable mechanism through which one could influence the other. In conclusion, the tasks chosen for use in this thesis span a theoretically complete space as defined by this 2×2 typology.

2.4 Overview – Spatial skills, training, mathematics and research questions

The literature reviewed in this chapter regarding the measurement of different spatial skills, their development, and training and transfer potential, reveals a number of common findings. The training potential of all of the skills discussed is agreed to be relatively high, but the transfer of training in one construct to performance in another is less evidenced. Uttal et al. (2013a) conducted a meta-analysis of 217 spatial training studies in order to determine the magnitude, moderators, durability, and generalisability of the training. Uttal et al. found that spatial skills were moderately malleable, with training resulting, on average, in improving task performance by half a standard deviation. Combined with data from the Wai et al. (2009) study of spatial skills as a predictor of future success in STEM careers, in which it was found that individuals with degrees in engineering had spatial skills 1.58 standard deviations above the general population. Uttal et al. (2013b) calculated that this improvement of half a standard deviation after training would, if implemented in all American schools, lead to a doubling of the number of engineers in America. This calculation does assume a causal relationship between spatial training and STEM attainment which may not be the case.

In terms of durability of training, Uttal et al. (2013a) found evidence of similar effect sizes at immediate post-test and at delays of one week and one month. Of the studies included in the meta-analysis that made some attempt at transfer, there was an average improvement of around half a standard deviation on the transfer task(s). A distinction was made between those studies that attempted near transfer and those that attempted far transfer. The effect sizes for both types of transfer differed significantly from zero, suggesting that both near and far transfer were possible. The findings of the meta-analysis challenge the general view that spatial training can only lead to very limited transfer. A number of the studies that provided evidence for far transfer were from an educational setting, essentially testing the formal discipline potential of the study of particular academic subjects. Some features of the studies that displayed far transfer were identified by Uttal et al. as more intensive training sessions and longer training sessions. In the studies that compared performance across males and females, a male advantage was found in all of the results. However, the effect sizes for the improvements after training were similar for both sexes. Uttal et al. also found that the malleability of spatial skills was not significantly different for children, adolescents, or adults. However, as this comparison could only be made across studies, as very few involved children and adults, factors such as differences in study design and outcome measures could have contributed to this non-significant finding.

From this extensive meta-analysis, it can be concluded that spatial skills can be trained, transferred, and that the effects are durable (Uttal et al., 2013a), and longitudinal data has shown that higher spatial skills in early life lead to successful careers in STEM areas (Wai et al., 2009). A key consideration is therefore how to connect these two features, and incorporate spatial training into existing educational settings. A number of approaches have started to be developed for this purpose, for example CogSketch (Forbus et al., 2011), an education tool that uses sketching to encourage spatial development, a concept based initially on a successful training programme with engineering students (Sorby, 2009). Studies One and Two of this thesis investigate the potential that an advanced mathematics course has in terms of training spatial skills. The literature reviewed in this chapter has established that a certain amount of training is possible, and that mathematics achievement and spatial skills are closely linked. An A level[12] mathematics qualification in England is required to include a number of aspects that may have the potential to train spatial skills, for example:

- reason logically and recognise incorrect reasoning,
- draw diagrams and sketch graphs to help explore mathematical situations and interpret solutions, and
- make deductions and inferences and draw conclusions by using mathematical reasoning (Department for Education, 2014a).

Students that take an A level mathematics qualification are required to sit examinations in a combination of modules in four main strands:

- Core: the fundamental building blocks of the subject, e.g. algebra, geometry and calculus
- Mechanics: forces, energy and motion
- Statistics: probability, data handling and hypothesis testing
- Decision: networks, algorithms and sorting

A circle with centre $C(-3, 2)$ has equation

$$x^2 + y^2 + 6x - 4y = 12$$

Find the y-coordinates of the points where the circle crosses the y-axis. ***(3 marks)***

Find the radius of the circle. ***(3 marks)***

The point $P(2, 5)$ lies outside the circle.

Find the length of CP, giving your answer in the form $\sqrt{n}$, where n is an integer. ***(2 marks)***

The point Q lies on the circle so that PQ is a tangent to the circle. Find the length of PQ. ***(2 marks)***

Figure 2.16: An example of an assessment item from a Core module

All of these modules have some potential to train spatial skills, whether it be in a more general reasoning capacity, or development of visualisation skills and

[12]In the case of Study One of this thesis, some students took A level mathematics, and some the International Baccalaureate, which consists of much of the same content.

representing situations in a spatial manner. Figure 2.16 shows an example of an item from the assessment of a Core module that is compulsory for all students. Although this item does not specifically require the candidate to sketch the situation, representing and visualising the problem spatially would be of great benefit when attempting to solve it.

The remainder of this thesis details the design and execution of three studies of advanced mathematics' potential to train spatial skills. Throughout this thesis, the label 'mathematician' will be used to describe a student that has chosen to study advanced mathematics, as opposed to a student who has not chosen to study the subject, referred to as a 'non-mathematician'. The research questions that the studies aim to provide evidence for are as follows:

1. Do mathematicians perform better on spatial tasks?
2. Is there evidence of developmental differences [13] between mathematicians and non-mathematicians?

A relatively large amount of literature that has been discussed so far has shown that mathematics achievement has positive correlations with performance on spatial tasks, and therefore it would be expected that the answer to the first question will be positive. The answer to the second question is less apparent. There is evidence that spatial skills are malleable, and that interventions, some educational, can have an effect. However, the results of these intervention studies vary, and it is not clear what aspects of them lead to successful and durable transfer. If the answer to the second question is positive, this would provide evidence for the formal discipline value of advanced mathematical study. Otherwise, if not, then an alternative hypothesis must be considered: that there is a filtering effect that results in individuals with higher levels of spatial skills being more likely to enrol themselves into advanced mathematics education.

[13] Developmental differences, in this context, refer to differences in the way that spatial skills develop during advanced study

Chapter 3

Study design and methodology

The remainder of this thesis is comprised of discussions of three studies which explore the relationship between advanced mathematical study and spatial skills in regard to the issues that have been raised in the previous literature review chapters.

The validity and generalisability of a study is directly affected by the way in which it is designed and carried out. It therefore is essential to clearly outline the methodology employed throughout this thesis prior to describing the studies in full. This chapter will discuss experimental and quasi-experimental designs, cross-sectional and longitudinal designs, and the implications of using them. Statistical methods for analysing data will also be discussed in terms of interpreting the results in a meaningful way.

3.1 Experimental and quasi-experimental design

The main two studies of this thesis follow a quasi-experimental design, and the last is experimental, with both designs having implications for the conclusions that can be legitimately drawn from the data. The following sections discuss the details of these implications.

The ultimate aim in designing an experiment is to be as confident as possible about the conclusions that are drawn from it. This means being sure that what was measured was truly the variables that were intended to be measured, and that no other variables had an effect on the results. This strength of validity[1]

[1] Validity refers to the extent to which the effects observed in a study are due to the

in a research study is only possible through a properly controlled, and strictly experimental design. However, for much of education research, a strict experimental design is not possible. For example, in order to measure the effect that one year of schooling had on a group of 12 year olds, it would be necessary to compare with a statistically identical group of 12 year olds that did not attend school for a year. This is clearly an unethical design, and such a group would be impossible to source. Due to this common setback, much quantitative education research follows a quasi-experimental design which, when implemented appropriately, can still be relied upon to draw certain conclusions.

The intention of a properly designed experiment is to isolate the independent variable (IV) of interest in order to observe the effects of any changes in this variable on another variable, or variables, which are dependent (DV). In the context of the main analyses of this thesis, the independent variable of interest was 'learning mathematics' and the dependent variables were the measures of spatial skills, as described in Chapter 2.

The key difference between experimental and quasi-experimental design is the random allocation of participants to the experimental and control groups. In the case of a pure experiment, the act of randomisation is essential for establishing a cause-effect relationship between the variables, and for determining the true magnitude of the effectiveness of the IV on the DV. When participants are randomised, it is reasonable to assume that any changes in the DV, from one time point to another, observed in the experimental group but not the control, are due only to the variable that has been experimentally manipulated. Figure 3.1 illustrates the classic experimental design.

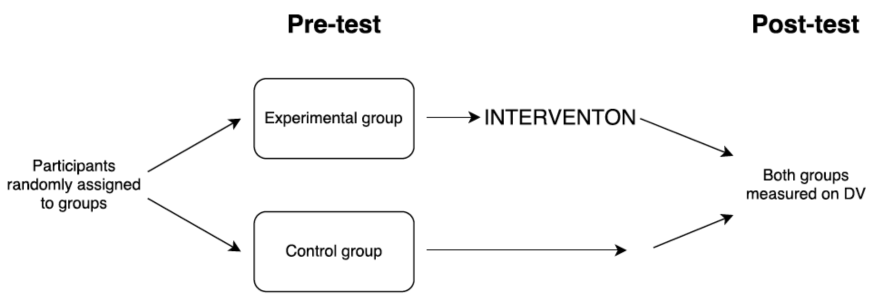


Figure 3.1: The classic experimental design

When participants are randomly assigned to the experimental and control groups, any differences between the groups will be random. However, when

manipulation of the independent variable and not some other factor, and how generalisable the findings are to the population at large participants are not randomly assigned to either group, there is a risk that the two groups will differ in some systematic way that is associated with the independent variable. In this quasi-experimental case, it is possible to identify associations between the intervention and the post-test data, but not to completely rule out the possibility that the observed differences are due to some confounding variable. The following section describes some of these threats to validity that could occur in a quasi-experimental study. All of these have been taken into consideration in the design and interpretation of results in this thesis, and the more relevant of these are discussed in more detail in Section 3.1.2.

3.1.1 Threats to internal validity

History effects

History effects refers to the influence of events that have not been factored into the experimental design during the intervention period. For example, the development of spatial skills over time may be affected by some of the other school subjects being studied at that time, not only mathematics. The mathematics students are more likely to have been studying complementary subjects such as physics and chemistry and therefore any effect that these subjects had on spatial skills would be observed only in the mathematicians, but would not be directly due to the study of mathematics.

Maturation effects

Maturation effects are due to natural developmental changes that happen over time, with or without the introduction of an intervention. It is important to consider whether there is any reason that the spatial skills development of mathematicians would be expected to be different to that of non-mathematicians.

Statistical regression

If participants display extreme results at one measurement, over time, they will tend to regress towards the mean for the second measurement. This effect is discussed further in Section 3.1.2.

Selection effects

Selection effects are due to pre-existing differences between the intervention group and the control group. In a cross-sectional quasi-experimental study, without random allocation of participants, this cannot be avoided entirely. Commonly, however, groups are matched on appropriate variables in order to try and minimise selection effects.

Attrition

In a longitudinal study, attrition is the effect of unequal drop-out rates in the intervention and control group. For example, if the participants that were not available to be tested at Time 2 had lower levels of motivation, and the drop-out rate for non-mathematicians was higher, this could disguise or create a between-group difference.

Hawthorne effects

Hawthorne effects occur when the participants are aware that they are involved in an evaluation, and that they are members of either the intervention or control groups. To avoid Hawthorne effects, it is very important to make sure that the participants are as blind as ethically appropriate to the nature of the research question and that experimenter bias is avoided.

3.1.2 Possible outcomes of a quasi-experimental intervention study

For a quasi-experimental design, it is assumed that the groups might not be non-equivalent prior to the intervention, and therefore the results must be interpreted in a different way to that of pure experimental data. The following figures illustrate some of the possible outcomes from a quasi-experimental intervention study involving two groups at two time points.

Outcome One

In the case illustrated in Figure 3.2, there are pre-test differences present between the two groups. In terms of the focus of this thesis, this would represent the situation in which the students that had chosen to study advanced mathematics displayed higher levels of spatial skills prior to any intervention. The spatial skills of the intervention group increase, whereas the control group show no change. At a first glance, it would be intuitive to interpret this result as an effect of the intervention but it is important to consider other possible explanations that the data are particularly susceptible to in quasi-experimental design. A possible alternative explanation for this pattern of results is that of maturation effect. This would be a case in which the intervention group were

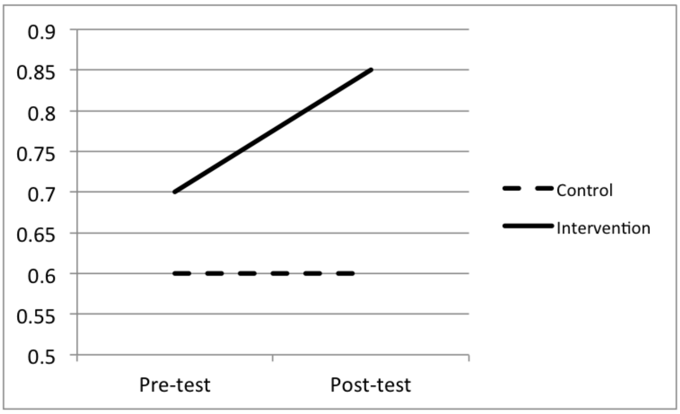


Figure 3.2: Outcome 1: the intervention group shows a small advantage at pre-test and an increase after the intervention period. The control group show no changes.

maturing at an increased rate in comparison to the control group and therefore the increase would have been observed only in one group, with or without the intervention. In the case of this thesis, this would mean that the spatial skills of the students that had chosen to study advanced mathematics were developing at a faster rate than those that had chosen to study other subjects, due to some extraneous variable, confounded with choosing to study mathematics. One possibility in this particular case could be general intelligence. The effects of other non-confounding extraneous variables can be ruled out in the case of outcome 1, in which the control group shows no change, as these effects would be observable in both groups. It is also possible that the observed change in the intervention group is due to some history effect: events that the intervention group, but not the control group, were exposed to during the intervention period that have had an effect on their spatial skills. There are a number of possible history effects which could affect the studies of this thesis that are due to the quasi-experimental nature of the research. For example, the students that chose to study mathematics might have been more likely to have also chosen other scientific subjects such as physics and chemistry, or the students might have experienced differences in the way in which teachers and/or family supported their academic career, affecting their motivation regarding their school work. The effect seen in outcome 1 is unlikely to be due to statistical regression as the control group show no movement towards a shared mean.

Even with consideration of the possible confounds that might result in an outcome such as in Figure 3.2, if the data showed this pattern, the possibility of an effect of the intervention could also be a valid conclusion.

Outcome Two

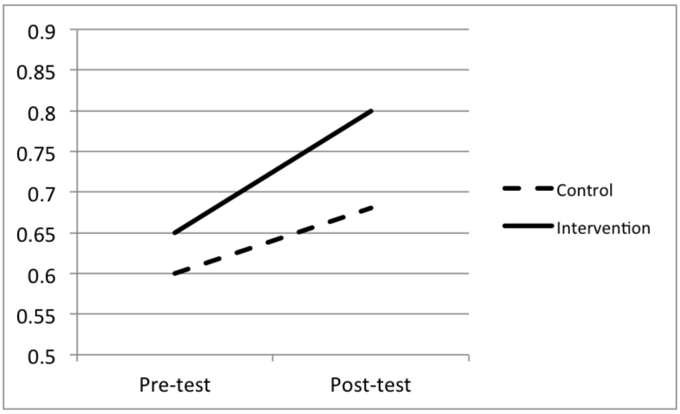


Figure 3.3: Outcome 2: the intervention group shows a small advantage pre-test which is still present post-test. Both groups increase from pre- to post-test.

In the situation illustrated in Figure 3.3, both groups increase from pre-test to post-test. The differences between the groups that are observed at pre-test, and at post-test, shows that the development of the DV previous to this testing was not equal, suggesting maturation effects already. At post-test, both groups have increased on the DV, the intervention group slightly more than the control group. This sort of pattern could very easily be explained by maturation effects, and even history effects, as explained above.

The pattern seen in outcome 2 could also be explained by attrition, or possibly Hawthorne effects. Attrition would be present if there was a tendency for students that were lower scoring on the spatial skills measure to drop out of the study at a higher rate. An analysis of only the students that participated at both pre-test and post-test, as a longitudinal design, would eliminate the risk of this. Possible Hawthorne effects may be an explanation of the observed data if the mathematics students were made aware of the links between spatial skills and mathematics learning and the justification for the study and, for this reason, practiced and developed their skills more than the control group through the intervention period, outside of their mathematics lessons. As with outcome 1, it is unlikely that this result could be explained by statistical regression.

Outcomes Three and Four

In both outcomes 3 and 4 (Figures 3.4 and 3.5), one group scored higher than the other at pre-test but, by post-test, the groups are more similar. These could both be examples of statistical regression and are situations which must

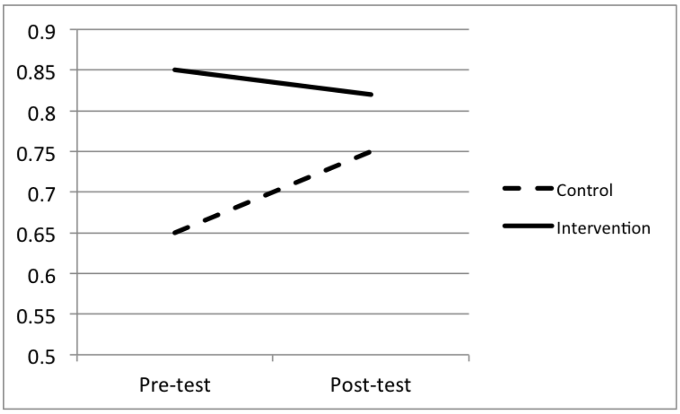


Figure 3.4: Outcome 3: the intervention group score higher than the control group pre-test, but decrease to a more similar level by the post-test

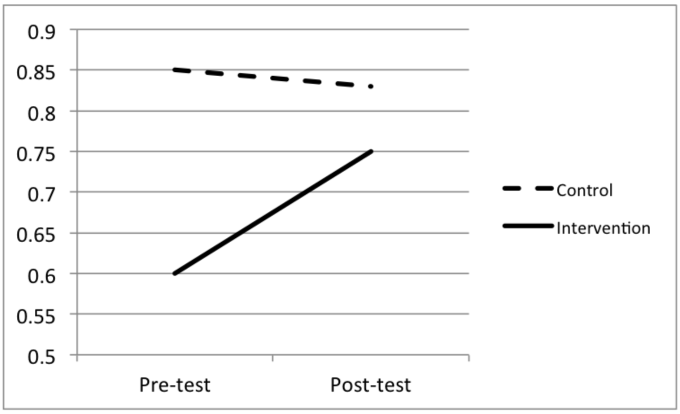


Figure 3.5: Outcome 4: the control group score higher than the intervention group at pre-test, but decrease to a more similar level by the post-test

be treated with much caution when looking at the relationship between the DV and IV in a quasi-experimental design. In terms of this thesis, outcome 3 is much more of a likely situation considering the literature, with the mathematics group scoring higher on spatial skills than the non-mathematics group at the pre-test stage. The apparent observed increase in skills for the control group after the intervention can be attributed simply to a movement toward the population mean.

Outcome Five

Figure 3.6 represents the strongest evidence for an effect of the intervention in a quasi-experimental study. Here, the intervention group start at a lower level

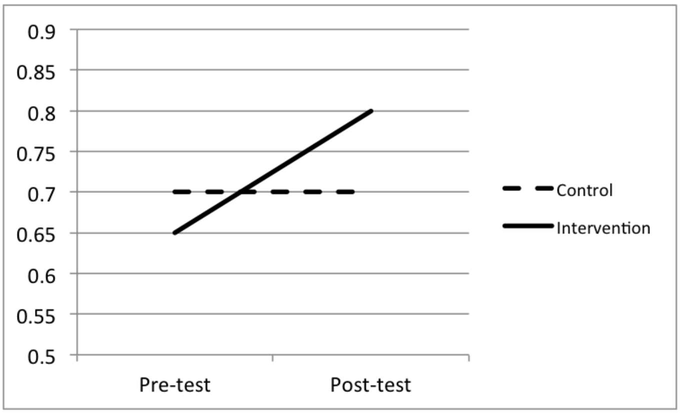


Figure 3.6: Outcome 5: the control group does not change. The intervention group score lower than the control, but score higher at the post-test point

of the DV pre-intervention, and end higher, with the control group showing no change. This can be referred to as a 'cross over effect'. It is possible to argue that this pattern could be a result of a maturation effect, where the intervention group started with lower levels of the DV but are more quickly affected by the intervention than the control group. This argument would be, however, quite implausible, and this pattern of results seen here cannot easily be attributed to any threats to internal validity. However, for any intervention being studied, it is unlikely that the expectation would be to see such a clear effect. In the case of measuring spatial skills, it would not be expected for the mathematicians to display lower levels of ability at a pre-test point, making this outcome improbable.

3.2 Cross-sectional and longitudinal designs

The two main studies that make up this thesis differ in their design and each have advantages and disadvantages in terms of methodology and the potential of drawing conclusions from the data. Both Study One and Study Two investigate developmental changes over time. Study One follows a cross-sectional design, where two groups are compared between subjects, whereas Study Two makes a similar comparison in terms of the variables measured, but within the same group of students at two time points. Both the cross-sectional and longitudinal studies of this thesis are of a quasi-experimental design and, therefore, interpretation of results must consider the points in the previous section. This section will describe some of the features, strengths, and weaknesses of cross-sectional and longitudinal design.

A cross-sectional design gives an opportunity to collect a large amount of data within a relatively small time-frame, allowing a number of variables to be tested within samples of populations. A snapshot of two population samples, at two different developmental points, allows a convenient way for hypotheses to be tested, but the conclusions that can be drawn are limited. Although every effort should be made to ensure that the two samples are comparable, inevitably there will be differences that will affect the results, and that are difficult to control for. The most prominent of these effects is what is known as the 'age-cohort effect'. Because a cross-sectional study looks at different groups of individuals, or different cohorts, it cannot be assumed that the two (or more) cohorts have had the same experiences. A cross-sectional study could measure, for example, the level of intelligence in two cohorts — one of age 25, and one of age 50, to investigate how intelligence changes over time. If the 50 year olds scored lower than the 25 year olds, it might be concluded that intelligence decreases over this developmental period. However, these differences could be attributed to many other factors that relate to the different experiences that the cohorts will have had. For example, the 50 year olds will have been schooled through a very different education system, and will have had much less access to technology than the 25 year olds. The fact that these confounding age-cohort effects exist in cross-sectional studies means that limited conclusions can be made about the effects of developmental age on a dependent variable, although many of these effects can be assumed to be very small if the two cohorts are similar in age. Other ways in which the cohorts differ may be due to factors such as sampling technique, motivation, Socioeconomic status (SES) background, and are all confound threats to the validity of the study.

A longitudinal study design measures a variable in the same cohort at different time points and is therefore immune to age-cohort effects. Cause and effect relationships between variables can be more easily identified than with a cross-sectional design and therefore present a more valid assessment of developmental changes. A quasi-experimental longitudinal study will, of course, still be vulnerable to many of the confounding factors previously discussed. In addition, there are some particular aspects of using a longitudinal design that can be disadvantageous. As the same group of participants are required to perform the same task twice or more, possible testing effects must be taken into account. The results of the second time point might be directly affected purely by the fact that the participants took part in the Time 1 testing. Another aspect to consider is the vulnerability to drop-out. A number of the participants that were involved in the study at Time 1 will not be available to be tested again at

Time 2, for a variety of reasons. A smaller pool of participants at Time 2 is not only a problem because of the reduced numbers, affecting the statistical power, but also because of the specific reasons that the participants dropped out. For this reason, it is vital to compare the drop-outs with those that remain in the study to ensure that the drop-outs have not produced a bias in the data.

3.3 Bayesian statistics

Another plausible result of an intervention study is illustrated in Figure 3.7. In this case, both groups improve on the DV at the same rate.

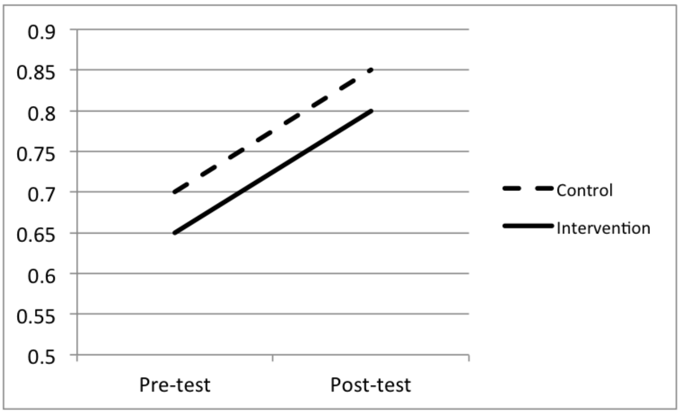


Figure 3.7: The control group and the intervention group improve at the same rate

Classical statistical methods do not have the capacity to satisfactorily test for this non-effect with p-value significance testing, but Bayesian statistics offers an alternative to this more common frequentist approach. It is based on the idea of adjusting the existing probability of a certain belief, or 'state of nature' being true, based on additional evidence. In essence, the concept is based on the calculation of a Bayes Factor (see Figure 3.8). Here, H_0 represents the null hypothesis and H_1 the alternative.

$$\underbrace{\frac{p(\mathcal{H}_1 \mid \text{data})}{p(\mathcal{H}_0 \mid \text{data})}}_{\substack{\text{Posterior beliefs} \\ \text{about hypotheses}}} = \underbrace{\frac{p(\mathcal{H}_1)}{p(\mathcal{H}_0)}}_{\substack{\text{Prior beliefs} \\ \text{about hypotheses}}} \times \underbrace{\frac{p(\text{data} \mid \mathcal{H}_1)}{p(\text{data} \mid \mathcal{H}_0)}}_{\substack{\text{Predictive} \\ \text{updating factor}}}$$

Figure 3.8: Calculation of a Bayes factor

Although Bayesian statistics can be traced back to Thomas Bayes' work in the late 1700s, its use as an alternative to the frequentist approach has only been prominent since the 1950s and onwards.

3.3.1 How the Bayes factor works

Prior beliefs about hypotheses are updated to posterior beliefs about the hypotheses through an 'updating factor' which takes into account the probabilities of the new data occurring, given the assumption of the alternative (H_1) and the null (H_0) hypotheses. This 'updating factor' is known as the 'Bayes factor' and provides information about the extent to which one should adjust one's belief about either hypothesis in light of the new evidence. From Figure 3.8, the posterior beliefs are calculated using both the 'predictive updating factor', or Bayes factor, and some prior beliefs about the hypothesis.

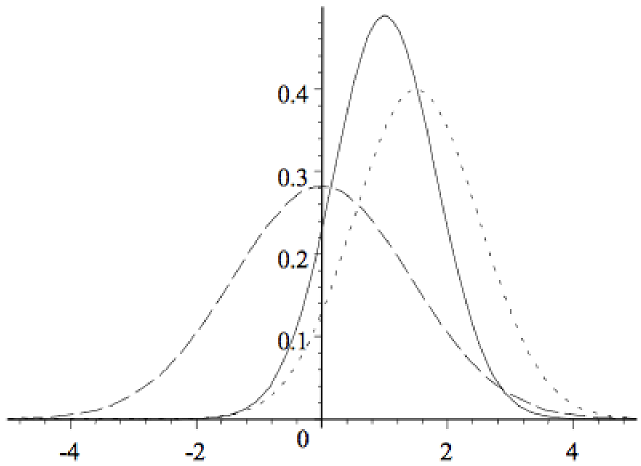


Figure 3.9: An illustration of prior density (dashed), additional data (dotted), and posterior density (solid)

Figure 3.9 illustrates the way in which Bayesian statistics works. The dashed line represents the prior density, which is based on previous data of beliefs. In the example, the prior belief is that the true value is between -4 and +4, and that zero is the most likely value. The additional data (dotted) gives further information: that the true value is between -2 and +5, and that the most likely value is around 2. Bayesian statistics works by using a combination of all of this information to update to a new, better informed belief that is represented by the solid line. The resulting posterior density covers a narrower range of possible values, meaning that one can be more confident that the true value is close to the most probable value (approximately 1 in this example). In addition, Bayesian statistics recognises that the strength (represented by the narrowness of the curve) of the prior density is less than that of the additional data, and therefore gives this less weight when calculating a posterior density.

3.3.2 Bayesian statistics and null hypotheses

Because the Bayes factor is concerned with the evidence that exists for both H_0 and H_1, it possesses the great advantage over traditional p-value testing of providing evidence for a null hypothesis as well as against it. If, for example, a frequentist statistician was to hypothesise that there was no link between two variables, a p-value of more than 0.05 would be a suggestion that the null hypothesis should not be rejected, but would not provide any evidence for accepting it, or any information about how likely it is that there is, in fact, no link. However, if a Bayesian statistician was to have a prior belief that there was no link between these variables (that the prior density was centred at zero), a Bayes factor calculated from the additional data, and the resulting posterior density, would tell the researcher more about the probability of the effect being zero. The size of the Bayes factor, and the narrowness of the posterior density, would determine how strong the evidence was. For instance, if $\frac{P(data|H_1)}{P(data|H_0)}$ is substantially less than 1, then this provides evidence in favour of H_0.

3.3.3 Criticism of Bayesian statistics

The main difference between Bayesian approach and a frequentist approach is the use of prior belief, as well as current data to reach conclusions. This aspect of the approach leads some critics to reject it as subjective. Two researchers, starting from different places in terms of their prior beliefs, would reach different posterior densities from the same data set. The counter-argument to this is twofold. Firstly, prior density beliefs should be based on the most well-informed and robust data possible and therefore should not differ drastically between researchers. Secondly, as Bayesian posterior densities are calculated again and again with additional data, the original prior belief loses the weight of its influence, eventually moving towards a consensus being reached, in spite of the researchers' prior beliefs. In fact, it could be argued that, in the case of differing opinions between researchers, Bayesian statistics would be the most rational way in which to interpret any new data. Each researcher would be provided with information about to what extent, and in which direction, they should update their beliefs. Ultimately, a repeated calculation of posterior densities, based on additional data, will move everyones' beliefs towards the 'truth'.

3.3.4 Interpreting the outputs of Bayesian statistics

More recent advances in statistic software have meant that Bayesian statistics have become the preferred statistical method in many research areas. The software employed to perform Bayesian statistics with the data in this thesis, JASP (Wagenmakers and Jove, 2016), offers a number of output plots that can be used to interpret the data. Figure 3.10 shows an example of a plot of prior and posterior densities.

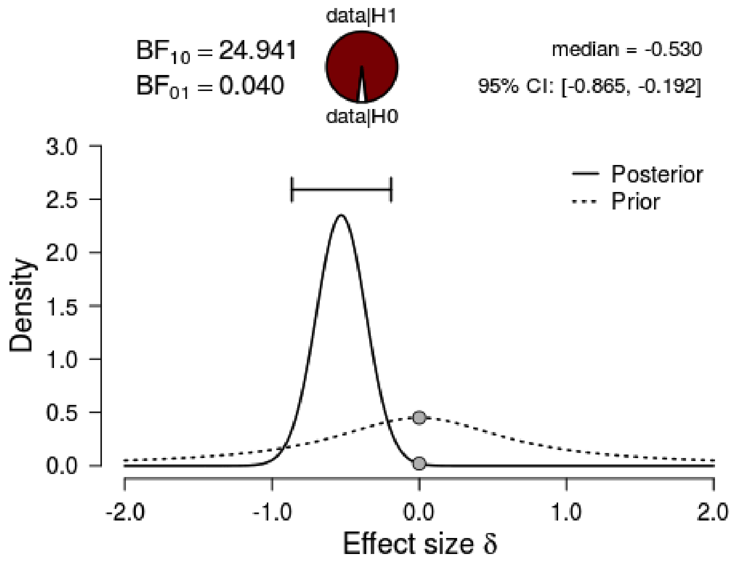


Figure 3.10: An example prior and posterior density plot

Here, two Bayes factors are calculated. BF_{10}[2] informs as to what extent belief should be shifted towards the alternative hypothesis, and BF_{01}[3] towards the null hypothesis. The additional data that the Bayes factor has been calculated from in this case provides strong evidence for the alternative hypothesis. A Bayes factor of $1 < BF < 3$ is considered anecdotal evidence, $3 < BF < 10$ moderate evidence, and $BF > 10$ strong evidence (Jarosz and Wiley, 2014).

[2] $BF_{10} = \frac{P(data|H_1)}{P(data|H_0)}$
[3] $BF_{01} = \frac{1}{BF_{10}}$

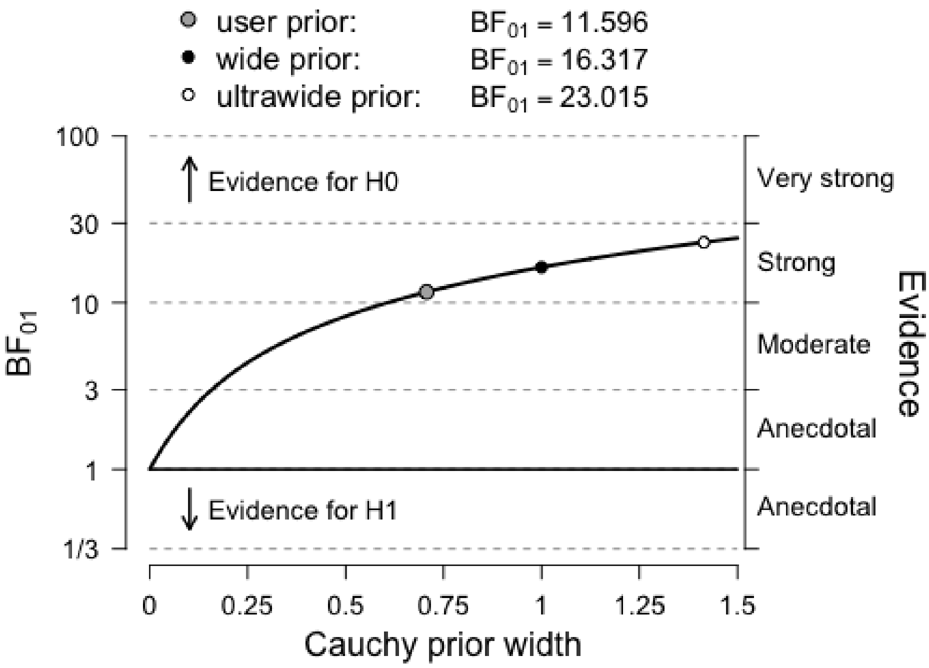


Figure 3.11: An example of a Bayes factor robustness check

Figure 3.11 illustrates the way in which the strength for H_0 and H_1 changes dependent on the prior width that is specified. The prior distribution of effect sizes is typically modelled as a Cauchy width, with a median of zero and a given width parameter (often set as a default of 0.707). One way of interpreting the Cauchy width is as a definition of the alternative hypothesis: the larger the prior width, the more likely it is that there will be an effect size further away from zero.

A combination of p-value testing and Bayesian statistics are used throughout this thesis to analyse most appropriately the data collected in relation to the research questions described in Chapter 2. The next chapter introduces and discusses Study One which approaches these questions using a cross-sectional quasi-experiment.

Chapter 4

Study One — A cross-sectional study of spatial skills pre- and post-A-level mathematical study

4.1 Introduction and theory

UK government policymakers support the idea that more students should study mathematics to an advanced level, partly due to the wider cognitive benefits that are associated with it, but there is surprisingly little scientific evidence to support this rationale. This study aims to assess whether the study of advanced mathematics has any effect on the development of cognitive constructs, specifically spatial skills, which are closely linked to mathematical abilities.

The acquisition of spatial skills follows a developmental pattern similar to that of many other cognitive constructs and can be reliably measured at a very young age and followed through into adulthood. An individual with a high level of spatial ability will have enhanced skills in remembering, manipulating, and reasoning about spatial information. Spatial skills are considered separate to other skills such as IQ, or memory abilities, and have only relevantly recently been of interest to education policy makers (Lubinski, 2010; Ministry of Education, 2014; Uttal and Cohen, 2012). Being in possession of a high level of these skills has the potential to affect mathematical achievement, from internal spatial representations of number lines (SNARC, see Dehaene et al. (1993)), to the ability to visualise and understand situations 3-dimensionally, to the skill of holding and manipulating spatial information successfully (known as spatial working memory (SWM)). These constructs can be measured using a variety of computerised or pen-and paper tasks. These tasks require the maintenance and manipulation of spatial information (the role of SWM) as well as more general spatial reasoning skills. SWM is limited and varies individually, but is thought to peak at 18-25 years and is of limited capacity (Baddeley, 1992). As well as the observable fact that many elements of mathematics have a spatial component, for example comparing areas, or geometry, a number of studies have claimed that training spatial skills can, in fact, improve mathematics ability. Cheng and Mix (2014), for example, found that training 6-8 year olds on 3D rotation activities improved their arithmetic skills, particularly for missing number problems (e.g. $4 + \square = 7$), and Holmes et al. (2008) found that performance on a spatial reasoning task was a strong predictor of mathematical achievement for 7-10 year olds. Early spatial abilities have also been found to predict future success in STEM (science, technology, engineering and mathematics) careers, over and above the verbal and numerical skills that are more commonly monitored in schools (Wai et al., 2009). This finding is of importance when considering the next generation of potential scientists and engineers. Early identification increases the potential for fostering and strengthening the development of relevant skills.

In the UK, children are expected to study mathematics up to year 11 (15-16 years old) and to sit a GCSE (General Certificate of Secondary Education) in order to receive a grade from A*-G, where below G is a fail[1]. After this point, students may choose to continue studying a number of subjects to a higher level. This non-compulsory stage of education is the focus of this study. The main syllabus studied in the UK, AS and A levels, award students with a qualification after one or two years of study. The International Baccalaureate (IB) system is becoming more common recently and of the participants involved in this study, all of the years 12s were planning on studying a combination of AS and A level subjects, whereas the undergraduates had studied a mixture of AS/A levels and IB.

[1]At the time of writing this thesis, the GCSE system was being reformed in a number of ways, including a change to the grading scale to include an additional grade point to distinguish students' achievement better.

The importance of spatial skills throughout education, particularly in relation to mathematics and in identifying STEM leaders is clear from the literature. What is now an important question to ask is what aspects of the curriculum are supporting the development of spatial skills, and whether they can be improved through education. The potential of training spatial skills has, in the past, had little research attention but, more recently, since psychologists have come to believe that certain cognitive constructs are less fixed than previously thought, the idea of being able to train transferable skills has become more established. Uttal et al. (2013a) performed a meta-analysis of 217 intervention studies to ascertain the malleability of spatial skills. The authors concluded that both near and far transfer of spatial skills was possible in adults and children, and that effect sizes were moderate and persistent over time.

The current study was cross-sectional, measuring spatial skills before and after advanced (A level or IB) study. The research questions posed for this study were:

1. Do mathematicians perform better on spatial tasks?
2. Is there evidence of developmental differences between mathematicians and non-mathematicians?

A comparison between those students that study mathematics and those who do not will provide evidence for the first question. The second question was investigated by considering education level in the analysis to establish whether the mathematicians' spatial skills develop in a different way.

4.2 Methods

4.2.1 Participants

4.2.1.1 Pre-A level study participants

Fifty-nine students (31 male) aged 16.10 - 17.8 years ($M = 16.8$ years; $SD = 0.3$ years) were recruited from the first year (year 12) of three sixth forms that are attached to schools, two in Nottinghamshire and one in Lincolnshire, UK. The latest Office for Standards in Education, Children's Services and Skills (Ofsted) ratings for the sixth forms of the schools A (235 students) and B (101 students) were 'good' and for school C (35 students) was 'requires improvement'[2]. All

[2]Ofsted is an independent government office that inspects and regulates services that deal with children and young people in the UK. Schools are rated on number of requirements, such as quality of teaching and safety of students and giving a grading from 1: outstanding, 2:

schools were made up of a below average number of students that are eligible for free school meals. School A was classified as having specialist mathematics and computing status.

All pre-A level study students had, previous to the study, obtained a General Certificate of Secondary Education (GCSE) in mathematics, amongst other subjects. The curricula that were followed by the three schools were very similar in terms of their content and grades and were awarded through 100% examination which covered topics relating to number, data, shape and space, and algebra. All of the students sat the 'higher' assessment tier examination paper which awards grades A*-D (A* being the highest).

4.2.1.2 Post-A level study participants

Seventy students (33 male) aged 17.50 - 26.90 ($M = 19.2$ years; $SD = 1.1$ years) were recruited from their first year of study at Loughborough University in Leicestershire, UK.

For the current study, it was of interest whether or not the students had studied advanced mathematics previous to taking part. Of the participants that had, the majority reported that they were awarded a UK-issued A level in mathematics, some an A level outside of the UK, some an International Baccalaureate in mathematics (IB), and a very few another advanced mathematics qualification equivalent. As A level and IB mathematics were the most reported qualifications, a description of those will follow:

- **A level mathematics** — Students are required to study six modules from the topics of 'core' (previously 'pure'), 'mechanics', 'statistics', and 'decision'. Although the four 'core' modules are compulsory, students may choose their other modules to suit their strengths and needs. Students are awarded a grade from A* to U (fail).
- **IB mathematics** — Students are required to participate in 190 hours of 'core' mathematics learning, 40 hours of chosen options, and to submit two pieces of 'mathematical investigation' coursework. Assessment is 80% examination (three papers) and 20% coursework. Students are awarded a mark out of 7.

For the purposes of this study, IB scores were coded as equivalent to A level grades with 7 and 6=A*, 5=A, 4=B and 3=C. These equivalences were chosen to match the findings of a study of a number of schools' IB and A level results,

good, 3: requires improvement, to 4: inadequate (Ofsted, 2016) of which the main focus was to compare the difficulty of the two qualifications (Handscombe, 2013). For ease, only A level mathematics will be referred to in the remainder of this thesis.

4.2.1.3 Mathematicians and non-mathematicians groups

Both education level groups were made up of some students that had (undergraduates), or were planning on (year 12), studying A level mathematics, and some that had not. The research study was undertaken at the beginning of an academic year, meaning that the pre-A level study group had been exposed to minimal advanced study at the time of testing, and that the post-A level study group had recently completed their advanced study, but had had minimal influence of their university study. Table 4.1 shows the gender mix of the mathematicians and non-mathematicians groups across both education levels.

$N = 129$	pre-A level study	post-A level study
Mathematics group	33 (19 male)	36 (24 male)
Non-mathematics group	26 (12 male)	34 (9 male)

Table 4.1: Gender mixes for Pre- and post-A level study students

4.2.2 Design

A between-subjects design cross-sectional study was conducted, with education level (year 12 or undergraduate) and group (either mathematics group (planning on studying, or had studied mathematics, dependent on education level) and non-mathematics group (not planning on, or had, studied mathematics)). The dependent variable was spatial skill, defined as a measure on the tasks described in Chapter 2, and recapped below.

4.2.3 Measures

All groups were tested on tasks designed to assess the following constructs:

- Working memory (verbal and spatial)
- 2D rotation
- 3D visualisation
- Spatial reasoning

In addition, a measure of mathematical fluency was taken to enable manipulation checks. Participants were also asked to self-report their GCSE (or equivalent) grades, and the advanced subjects that they planned on, or had studied, post-16.

Working memory tasks

Built using Psychopy, a software package used to write psychology experiments (Peirce, 2008), and adapted from Hubber et al. (2014), the tasks consisted of a processing element and a storage element. The processing element required the participants to decide whether two faces, presented side by side on a computer screen, were the same person or not. These faces were sourced from the Glasgow Unfamiliar Face Database (Burton et al., 2010), and were presented side by side on a computer screen. A face matching task was chosen for the processing element of the task due to it being as neutral as possible in terms of interference with either of the spatial or verbal storage elements. Figure 4.5 shows some examples of these faces. The faces were displayed for a maximum of 3000 milliseconds although the images disappeared if the participant responded on the keyboard before this time.

After participants responded to the pairs of faces with either a Y (yes, the faces are the same person), or N (no, the faces are not the same person) on the computer keyboard, the storage section of the task started.

- **Spatial working memory**: A 3×3 grid appeared on the computer screen for 5000 milliseconds, with a red dot located in one of the possible 9 positions. The participants were asked to remember the red dot's position on the grid. They were then shown a pair of faces and asked to respond with whether they thought they were matching or not. This was again followed by a red dot on the grid. This happened between 3 and 8 times, after which the participants were instructed to remember the positions of the red dots in the order that they were presented and relay this by clicking a mouse cursor on a similar 3×3 grid (see Figure 4.2(B) for an illustration of the task). Three of each repetition lengths were presented, comprising 18 trials in total.

- **Verbal working memory**: A number between 1 and 9 was displayed for 500 milliseconds. The participants were asked to remember the number. They were then shown a pair of faces and asked to respond with whether they thought they were matching or not, as with the spatial working memory task. This happened between 3 and 8 times within each trial. The

Figure 4.1: Examples of the faces used in the working memory tasks, all taken from the Glasgow Unfamiliar Faces Database (Burton et al., 2010) A — matching case; B — un-matching case

participant was then prompted to type in the numbers that they saw, in the order that they were presented. This resulted in the participants being required to remember lists of numbers between 3 and 8 items long. Three of each list length were presented, comprising 18 trials in total. Figure 4.2(A) illustrates this task.

2D rotation

A 10-item task that required the participants to state which of a selection of images is the exact rotation of the target image. Figure 4.3[A] shows an example of this task.

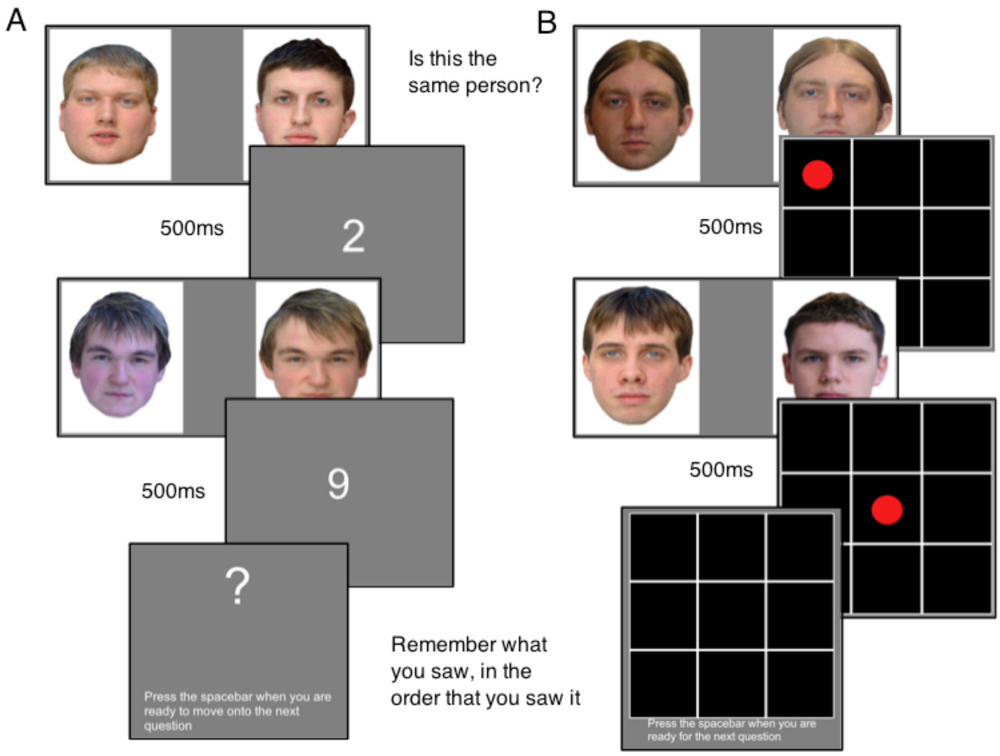


Figure 4.2: [A]: Verbal working memory task; [B]: Spatial working memory task

3D visualisation task

A 10-item task that asked the participants to select the 3D shape that could be constructed from the given 2D net. Both the 2D and 3D tasks were adapted from the tasks used for Project TALENT (Wai et al., 2009). Figure 4.3[B] shows an example of the 3D task.

Spatial (matrix) reasoning

An 11-item task in which participants were asked to select the correct picture, from a choice of six, that correctly completed the pattern (Raven et al., 2000). The 11 items were taken from Raven's Advanced Progressive Matrices (RAPM), designed to measure ability in the general population. The full set of RAPM items consists of five sets of six items which increase in complexity from the first to the last item, with set A being the first, and therefore easiest, and set E being the most challenging. The 11 items used in this study consisted of the odd numbered items taken from sets C & D, and so were of medium difficulty. Due to copyright reasons, it is not possible to replicate items from the matrix reasoning task, but Figure 4.4 shows a very similar item.

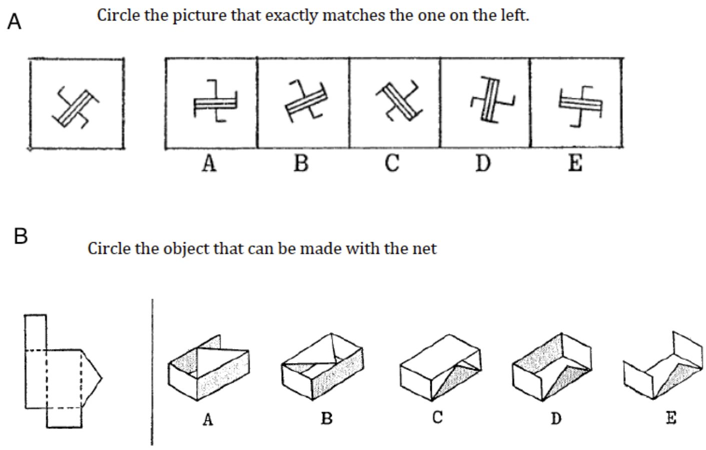


Figure 4.3: [A]: Two-dimensional rotation task; [B]: Three-dimensional visualisation task

Mathematical fluency

The Woodcock Johnson mathematics fluency test was administered using the standard procedure (Woodcock et al., 2001) to measure group differences at both education levels. The participants were given three minutes to answer as many items as they could of a possible 160 simple addition, subtraction, and multiplication questions, e.g. "3 + 2 = ?". All of the digits used were between 0 and 10 and the questions increased in difficulty very slightly during the task. Accuracy on the task was calculated as the number of correct answers minus the number of incorrect answers. Items that were not attempted were not counted.

The verbal and spatial working memory tasks were presented on a laptop, whereas the other four tasks were presented in a paper booklet for the participants to complete[3]. The order in which the participants were asked to complete the tasks was counter-balanced as far as possible. Participants did either the computer tasks, or the paper tasks first. The order of the paper tasks within the booklet was identical for each participant, whereas the working memory tasks were counterbalanced, with half of the participants attempting the spatial task first, and half the verbal.

[3] This booklet can be found in Appendix 9.1.

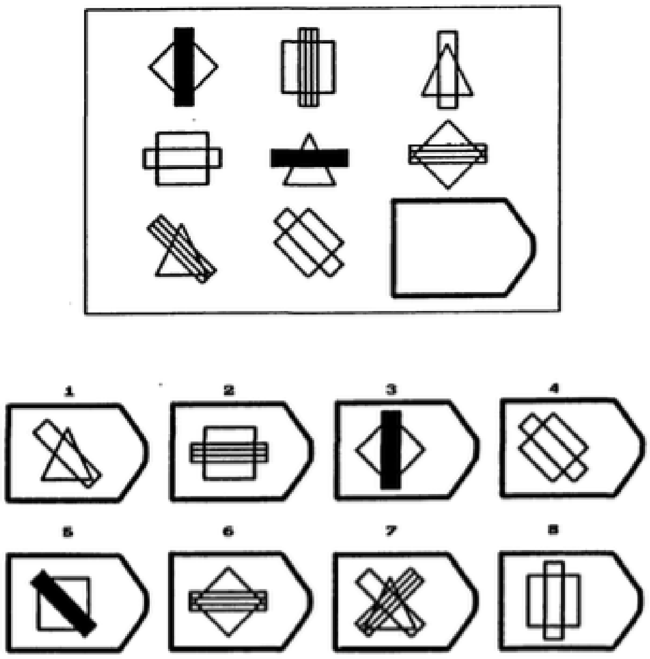


Figure 4.4: Examples of a similar matrix reasoning task

4.2.4 Procedure

For the pre-A level students, the tasks were completed in a small room of the school/college that they attended, during a free period of their timetable. The tasks were completed individually, but with between two and four students in the room at the same time. For the post-A level group, the tasks were completed in a small room of Loughborough University at a time convenient to the participants, with between one and four students in the room at the same time. The whole procedure was completed in one session per participants and took approximately 45 minutes from start to end.

4.2.5 Scoring the working memory tasks

Working memory span tasks are widely used in the field of cognitive psychology, and the way in which they are coded differs somewhat across research groups. Traditionally, working memory tasks were scored as the longest list of items that a participant can remember correctly without making mistakes. The participant would see the shortest lists first, for example, lists of 3 numbers, then 4 numbers, then 5, etc. At the point at which the participant failed to recall the list correctly, the task would stop, and the participant was awarded a score equal to the longest list remembered. Within the working memory literature, there is concern about the way in which different dimensions of the task might affect this 'absolute span scoring' method. For example, it is hard to compare the absolute scan score across two tasks where the display time might differ (Conway et al., 2005).

An alternative way to score working memory tasks is on a binary scale, so that a complete correct recall of the list of numbers, or position of dots, is scored as 1, and anything else is scored as 0. For this method, there is no need for the participant to see the lists of numbers in increasing order. Conway et al. (2005) go a step further and suggest that a method they term 'partial-credit scoring' is the most appropriate. For partial-credit scoring, incorrect answers are not given 0. Instead, they are given scores for partially correct answers. For example, if the list to be remembered was '$3, 5, 1, 2, 7$' and the participant recalled the incorrect answer of '$3, 5, 6, 2, 7'$, they would be allocated a score of $4/5$, as four out of the five numbers were remembered correctly. This method is a more sensitive measure of individual differences in working memory. Scoring recall as a proportion of correct to total items within a list has the added feature of giving a higher weighting to the items with a higher load: one incorrect element in a list of five items would result in a score of $4/5 = 0.8$, whereas one incorrect item in a list of two items would give a score of $1/2 = 0.5$.

For the current study, the working memory tasks were scored using this partial-credit model. In addition, the order of the elements of the list was considered when scoring. For example, looking at the list $3, 5, 1, 2, 7$ again; If a participant incorrectly recalled the list as $5, 3, 1, 2, 7$, they were not scored as having two numbers wrong, but as having one 'ordering error' and so were allocated a score of $4/5 = 0.8$. This was calculated using a 'Levenshtein distance' calculation, often used in spell checkers, which calculates the number of changes (letter substitutions, deletions, or additions) that need to be made to turn one list into another. For example, the words 'PAINT' and 'PINT' have a Levenshtein distance of 1, as do the words 'PAINT' and 'FAINT', whereas the words 'PAINT' and FAITN' have a Levenshtein distance of 2. A final score was obtained by subtracting the Levenshtein distance from the total list length, and then dividing by the list length to find the proportion correct.

For the example list '$3, 5, 1, 2, 7$', if participant A recalled the list as '$3, 1, 7, 2$', the Levenshtein distance would be 2, calculated as one missing number, and one ordering error. The final score for this list would be

$$(ListLength - LevenshteinDistance)/ListLength = (5-2)/5 = 3/5 = 0.6 \quad (4.1)$$

The total working memory span score for participant A would then be the average of these scores for all of the lists in the task (18 in the case of the tasks of this study), giving each participant a score between 0 and 1. This method, unique to this research, was used to calculate both the spatial working memory score (SWM) and the verbal working memory score VWM).

4.3 Results

For performance on the working memory tasks, average processing (face matching) accuracy rates below 50% were considered below chance, and taken as an indication that the participant was not engaging in this processing element of the task. In this case, if the participant was focusing their attention only on the memory element of the task, this would be deemed a short-term memory task, rather than a working memory task, and would not be accurate in revealing variation in the intended cognitive construct. Participants that displayed average reaction times longer than 3 seconds for the face processing elements of the working memory tasks were also excluded from the analysis as it was assumed that these individuals were not engaging in the task fully. This resulted in seven (4 pre-A level, and 3 post-A level) participants being excluded from the main analyses.

4.3.1 Reliability of measures

Cronbach's alpha was calculated for each of the measures employed in the study. Table 4.2 shows these.

Measure	Cronbach's alpha
2D rotation	0.83
3D visualisation	0.65
Matrix reasoning	0.35

Table 4.2: Reliability statistics for spatial skills measures: Study One

The Cronbach's alpha for the matrix reasoning task here is unusually low. The Raven's Advanced Progressive Matrices task that this measure was adapted from is a well established task that often reports much higher reliabilities. It is possible that the reason for the reliability being so low in this instance is due to the fact that the task was made up of a relatively small number of items, although a number of studies have reported higher alphas for short versions of the task. Study Two of this thesis, reported in full in Chapter 6, finds a much higher reliability ($\alpha = 0.84$) on a longer version of the matrix reasoning task.

4.3.2 Gender differences

Due to the large amount of literature on gender differences in regard to spatial skills (see Coluccia and Louse (2004) for a review), it was regarded appropriate to begin the analysis of these findings with a comparison across gender.

Pre-A level

Table 4.3 shows a break down of all measures in terms of gender for the pre-A level group.

Measure	Male (N=30) Mean (SD)	Female (N=27) Mean (SD)
GCSE grade mathematics	4.61 (.92)	4.54 (1.07)
GCSE grade English	3.77 (.96)	4.96 (.82)
Mathematics fluency	101.06 (22.03)	104.46 (25.69)
Faces processing verbal	.89 (.06)	.92 (.06)
Faces processing spatial	.73 (.07)	.84 (.06)
Verbal working memory	.84 (.10)	.85 (.06)
Spatial working memory	.73 (.18)	.85 (.06)
2D rotation	.80 (.18)	.62 (.30)
3D visualisation	.75 (.19)	.64 (.19)
Matrix reasoning	.56 (.15)	.53 (.15)

Table 4.3: Gender differences for the pre-A level group. For the GCSE grades, 3=C 4=B, 5=A. The mathematics fluency score is out of a maximum of 160. All other scores are represented as proportion correct.

Gender differences were found for GCSE English, with female participants reporting a higher grade ($t(55) = 4.97, p < .001$). Of the cognitive tasks, males displayed an advantage for 2D rotation ($t(55) = 2.76, p = .008$) and 3D visualisation ($t(55) = 2.31, p = .025$). The difference was approaching significance for accuracy in the face processing task for the spatial working memory task ($t(55) = 1.68, p = .098$). All other $ps > .165$. Although only the GCSE English difference remains significant after Bonferroni correction, this suggestion of gender differences for performance on the spatial tasks means that all further analysis will consider gender as far as possible.

Post-A level

A similar analysis was not possible for the post-A level group as the mathematics and non-mathematics groups were considerably uneven in terms of gender make-up. The non-mathematics group consisted of 9 males and 25 females, whereas the mathematics group contained 23 males and 12 females. Due to this imbalance, any comparison considering gender would be hard to interpret.

4.3.3 Preliminary analyses

The following two sections are preliminary analyses of the data, which do not directly address the main research questions.

Descriptive statistics

Table 4.4 shows a summary of the mean proportion scores that participants displayed in each of the experimental measures, confirming no ceiling effects on any of the measures for either group, and mean GCSE grades, where 3=C, 4=B, 5=A, etc.

Measure	Maths group mean (SD)	Non-maths group mean (SD)
GCSE maths	5.4 (0.65)	4.1 (0.80)
GCSE English	4.7 (0.98)	4.7 (1.04)
Maths fluency	0.78 (0.15)	0.63 (0.13)
Verbal working memory	0.90 (0.07)	0.69 (0.17)
Spatial working memory	0.79 (0.14)	0.69 (0.17)
2D rotation	0.64 (0.30)	0.81 (0.22)
3D visualisation	0.80 (0.20)	0.67 (0.20)
Matrix reasoning	0.65 (0.16)	0.54 (0.15)

Table 4.4: The mean proportion scores for each of the experimental measures for the maths and non-maths groups

The mathematics fluency measure confirmed a significant difference in mathematical capacity between groups for the pre-A level study cohort ($t(55) = 3.90$, $p < .001$) and the post-A level study cohort ($t(67) = 5.65$, $p < .001$), with the group that chose to study mathematics scoring highest in both instances.

A comparison of GCSE English grades between the mathematics and non-mathematics groups using a Mann-Whitney U test showed that there was no difference for the pre-A level group ($U(57) = 370, Z = .501, p = .616$) or the post-A level group ($U(69) = 593, Z = .026, p = .980$). This analysis confirms that the participants in the mathematics group, for both education levels, were not of a general higher ability, but that they possessed particular higher mathematical abilities.

A comparison of GCSE grades of the pre- and post-A level groups, again using a Mann-Whitney U test, found that the post-A level group had received higher grades in mathematics ($U(129) = 1607, Z = 2.26, p = .024$), English ($U(129) = 1230, Z = 4.12, p < .001$), and science ($U(129) = 1563, Z = 3332, p = .012$). This indicated that the post-A level group of students were generally higher achieving academically. This difference is considered in the interpretation and discussion of the results.

Results from the face processing element of the working memory tasks

For both working memory (WM) tasks, every participant was asked to perform a face processing task at the same time as remembering either a list of single digit numbers (verbal), or the position of a number of red dots on a grid (spatial). For this element of the tasks, each participant's accuracy (correct or incorrect) and reaction time (seconds) were recorded. Although this data is not part of the working memory measure, it is important to analyse this component of the task to ensure that participants were fully engaging in the processing element, and therefore using their working memory, rather than short-term memory capacity. In addition, looking at the relationship between them provided information about any possible trade-offs between accuracy and speed.

Accuracy

Overall, participants were more accurate at deciding whether the two faces displayed were the same person during the verbal WM task than the spatial WM task ($t(125) = 9.69$, $p < .001$). This difference is most likely due to the processing skills needed to distinguish similarities and differences between two faces having some spatial reasoning element to them, and therefore having more potential to require some spatial working memory load. This hypothesis is supported by the field of research that has found that spatial frequency of faces affects the way in which people process human faces (see, for example, Goffaux and Rossion (2006))

There were no significant differences between the mathematics and non-mathematics groups overall on either processing task (all $ps > .244$). The post-A level study group scored significantly higher than the pre-A level study group on the face processing element for both the verbal WM task ($t(100.0) = 2.82$, $p = .006$) and the spatial WM task ($t(98.1) = 2.92$, $p = .004$).

Correlations between the working memory tasks and the processing element of those tasks showed a moderate positive correlation for both the verbal task ($r(126) = .317, p < .001$) and the spatial task ($r(126) = .208, p = .019$). Positive correlations suggest individual differences in ability to perform the overall tasks without distraction, rather than a trade-off in effort allocated to the different elements of the task. Figure 4.5 shows the mean proportion correct scores for the processing task for all of the participants.

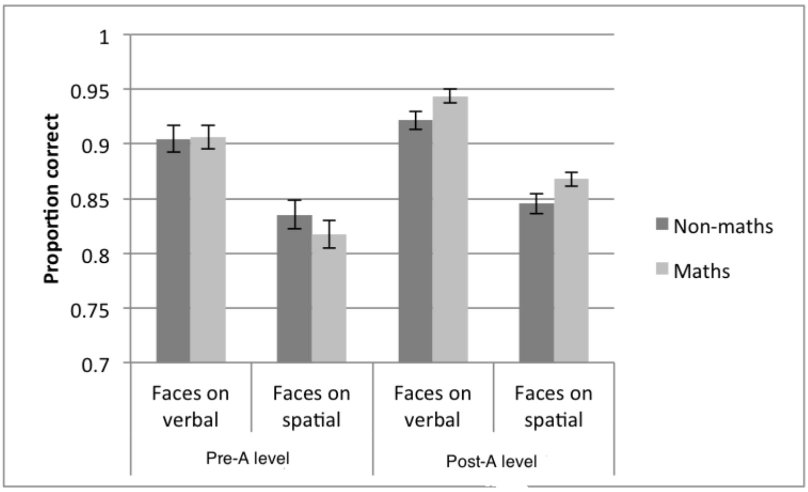


Figure 4.5: Mean accuracy for the face processing task. The error bars represent standard error

For the participants that did/were not study(ing) advanced mathematics, there were no differences between education levels in terms of the accuracy scores on the processing task ($ps > .237$). For the mathematics group however, the post-A level study group scored significantly higher on the face processing tasks during both the verbal WM task ($t(50.1) = 2.94$, $p = .004$) and the spatial WM task ($t(46.3) = 3.49$, $p = .001$).

Reaction times

In order to eliminate instances in which the participants may have been guessing, reaction times were analysed for only the trials in which the participants correctly decided whether or not the pictures were of the same person. Overall, participants were quicker at deciding correctly whether or not two faces were the same person during the verbal task than during the spatial task ($t(125) = 4.25, p < .001$). This result is in line with that found for overall accuracy on the task and is thought to be for the same reason. There were, again, overall, no significant differences between the mathematics and non-mathematics groups' reaction times on either of the processing tasks ($ps > .670$). As opposed to the comparison of accuracy rates, analysis between education levels did not find that the post-A level group were any quicker than the pre-A level group ($ps > .197$).

Correlations between reaction times and proportion correct scores on each of the working memory tasks revealed no relationship between the processing element and the working memory element of the task (all $ps > .744$). This finding confirms that there was no trade-off tactic between accuracy and speed being employed by the participants.

4.3.4 Analysis of data at each education level: do mathematicians outperform non-mathematicians on spatial tasks?

Preliminary analysis so far has revealed that the post-A level group tended to score better on most of the measures than the pre-A level group. This is to be expected as they were an average of 2.57 years older and a certain amount of cognitive development can be expected to happen in that time. It should also be taken into consideration that the post-A level group were a university cohort, whereas the pre-A level group, although choosing to study some post-compulsory education (A levels), may not be as academic. There is no way of knowing whether the participants that made up the pre-A level group will go on to study at university, but it can be assumed that this is likely to be less than 100% of them.

The research questions that are addressed in this results section are as follows:

1. Do mathematicians perform better on spatial tasks?
2. Is there evidence of developmental differences between mathematicians and non-mathematicians?

The reporting of the results will start with an overall analysis of all five dependent variables (2D rotation, 3D visualisation, matrix reasoning, and verbal and spatial working memory) between the mathematics and non-mathematics groups and across education levels. Following that analysis, group differences for each dependent variable across education levels will be looked at in more detail in order to understand more fully the way in which each of these constructs interacted. An analysis considering gender will be performed for the pre-A level group only, due to the confound between gender and group at post-A level. Finally, an analysis of working memory type, educational level and mathematics group will determine whether the mathematicians had a particular spatial advantage within this construct.

Overall MANOVA analysis

A 2(group: mathematicians/non-mathematicians) $\times$ 2(education level) MANOVA of the five dependent variables revealed an effect of group ($F(5, 119) = 6.708, p < .001$), and of education level ($F(5, 119) = 6.166, p < .001$), but no group $\times$ education level interaction ($F(5, 119) = 1.656, p = .151$). The significant effect of group provides support for the first research question, confirming that the mathematicians did perform better. However, the absence of a group $\times$ education level interaction suggests that there may not be support for the existence of developmental differences between the groups.

Variable-level ANOVA analysis

Next, each dependent variable was explored with 2(group) $\times$ 2(education level) ANOVAs. The main effects and interactions are reported in turn.

Figure 4.6 shows the interaction plot for 2D rotation. There exists a significant main effect of group ($F(1, 125) = 12.459, p < .001$), but no effect of education level ($F(1, 125) = 0.775, p = .380$) and no interaction ($F(1, 125) = 2.045, p = .155$). Table 4.5 shows the 2D rotation task accuracy scores for each group at both education levels.

Education level	Non-mathematicians	Mathematicians
Pre-A level	.66 (.27)	.75 (.24)
Post-A level	.63 (.33)	.86 (.19)

Table 4.5: Mean (SD) accuracy scores for the 2D rotation task

For the pre-A level group (year 12), scores on the 2D rotation task did not differ significantly between the mathematics and non-mathematics groups ($t(57) = 1.46$, $p = .150$), whereas they did for the post-A level (undergraduate) group ($t(68) = 3.55$, $p = .001$). From Figure 4.6, it can be seen that the data

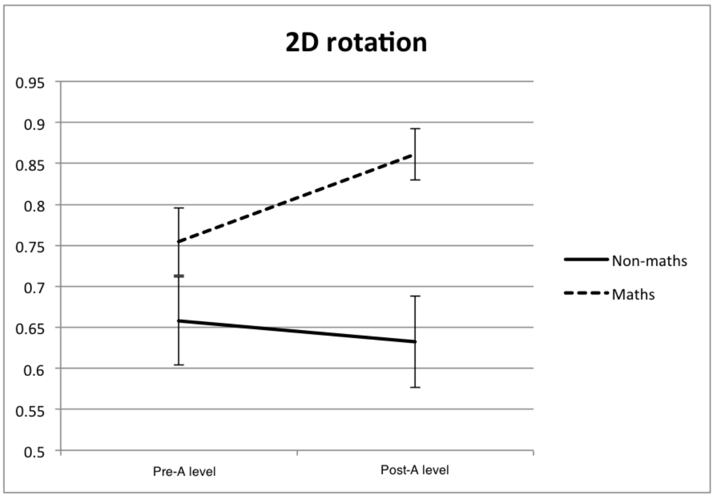


Figure 4.6: Effects of education level and group on 2D rotation

does appear to approach an interaction between education level and group. This is the case for a number of the dependent variables, and will be discussed further in Section 4.4.

Figure 4.7 shows the interaction plot for 3D visualisation. There exists a significant main effect of group ($F(1, 125) = 14.455, p < .001$) and a signifiant effect of education level ($F(1, 125) = 5.304, p = .023$), but, again, no interaction ($F(1, 125) = 2.402, p = .124$). Table 4.6 shows the 3D visualisation task accuracy scores for each group at both education levels. Again, there is no significant difference between groups at pre-A level ($t(57) = 1.51$, $p = .137$), but there is at post-A level ($t(68) = 4.01$, $p < .001$).

Education level	Non-mathematicians	Mathematicians
Pre-A level	.65 (.18)	.73 (.20)
Post-A level	.68 (.21)	.86 (.17)

Table 4.6: Mean (SD) accuracy scores for the 3D visualiisation task

Figure 4.8 shows the interaction plot for matrix reasoning. There exists a significant main effect of group ($F(1, 124) = 16.442, p < .001$) and of education level ($F(1, 124) = 13.089, p < .001$), but no interaction ($F(1, 124) = 3.787, p = .054$). Table 4.7 shows the matrix reasoning task accuracy scores for each group at both education levels. As with the other spatial measures, there is no difference between the groups at pre-A level ($t(57) = 1.47$, $p = .146$), but there is at post-A level ($t(67) = 4.34$, $p < .001$).

Figure 4.9 shows the interaction plot for verbal working memory. There

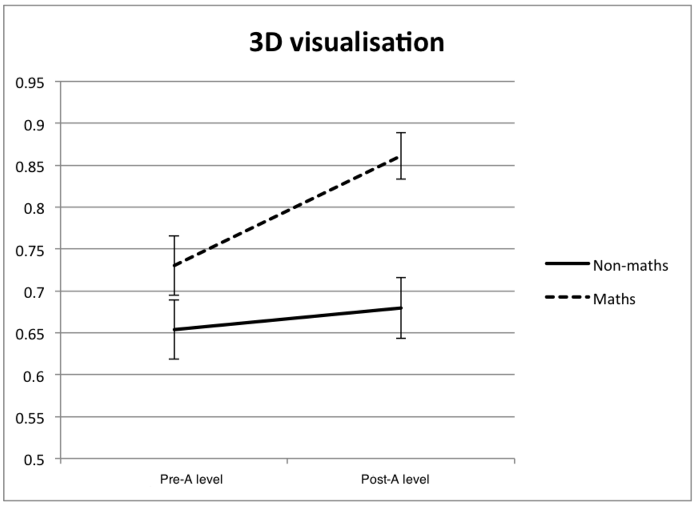


Figure 4.7: Effects of education level and group on 3D visualisation

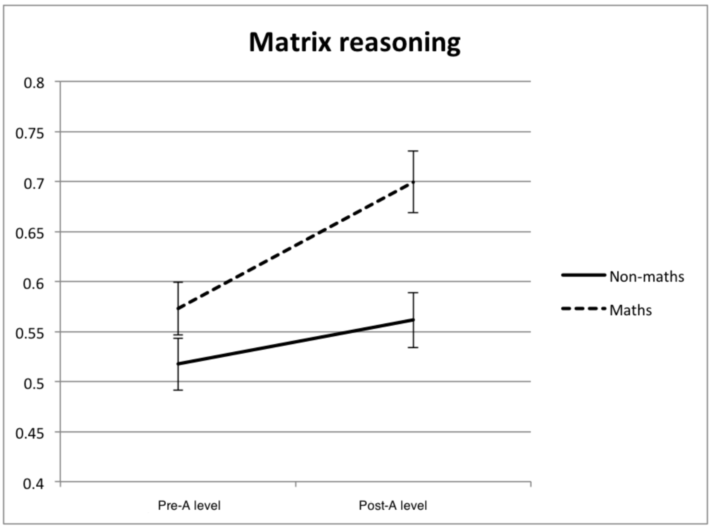


Figure 4.8: Effects of education level and group on matrix reasoning

Education level	Non-mathematicians	Mathematicians
Pre-A level	.52 (.13)	.57 (.15)
Post-A level	.56 (.16)	.70 (.18)

Table 4.7: Mean (SD) accuracy scores for the matrix reasoning task

exists a significant main effect of group ($F(1, 124) = 7.269, p = 008$) and of education level ($F(1, 124) = 20.934, p < .001$), but no interaction ($F(1, 124) = 0.036, p = .850$). Table 4.8 shows the verbal working memory task accuracy scores for each group at both education levels. At pre-A level, there was no difference between groups ($t(56) = 1.76$, $p = .084$), and at post-A level there was ($t(68) = 2.07$, $p = .042$), although this was marginal.

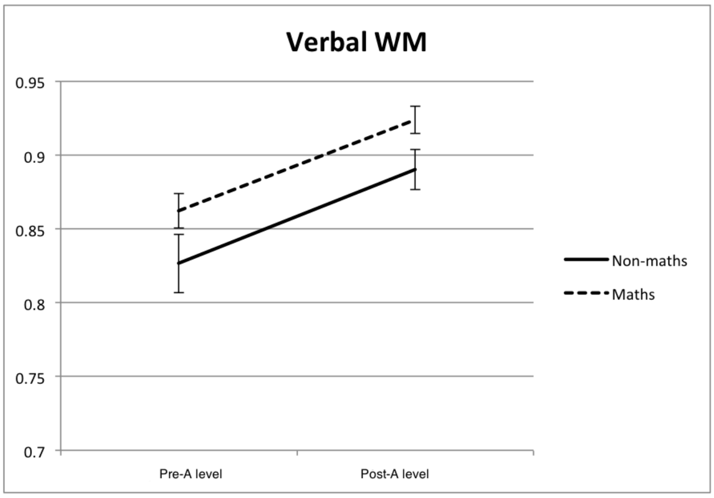


Figure 4.9: Effects of education level and group on verbal working memory

Education level	Non-mathematicians	Mathematicians
Pre-A level	.83 (.10)	.87 (.07)
Post-A level	.89 (.08)	.92 (.05)

Table 4.8: Mean (SD) accuracy scores for the verbal working memory task

Figure 4.10 shows the interaction plot for spatial working memory. There exists a significant main effect of group ($F(1, 124) = 13.155, p < .001$) and of education level ($F(1, 124) = 10.224, p = .002$), but no interaction ($F(1, 124) = 1.178, p = .280$). Table 4.9 shows the spatial working memory task accuracy scores for each group at both education levels. There were significant differences between the groups at both pre-A level ($t(56) = 2.83$, $p = .006$) and post-A level

($t(68) = 2.13$, $p = .037$), although, again, this was marginal.

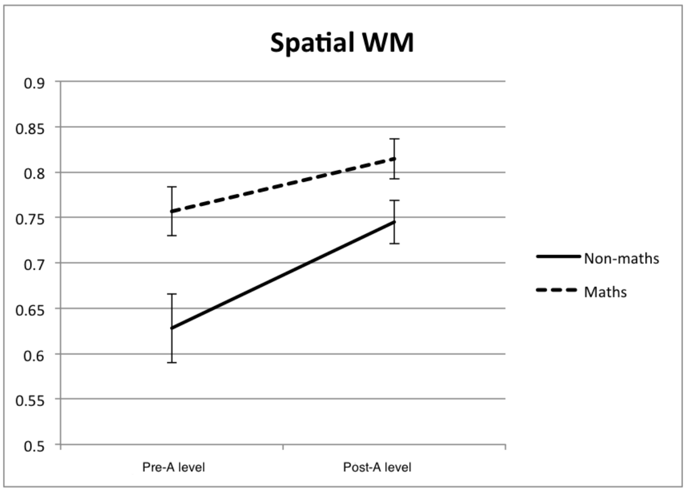


Figure 4.10: Effects of education level and group on spatial working memory

Education level	Non-mathematicians	Mathematicians
Pre-A level	.63 (.19)	.76 (.15)
Post-A level	.74 (.14)	.81 (.13)

Table 4.9: Mean (SD) accuracy scores for the spatial working memory task

Group × gender interactions pre-A level

An analysis to include gender was only possible at pre-A level as at post-A level, gender and group were, unsurprisingly, confounded. A gender × group MANOVA, including all five dependent variables revealed no main effects of either group ($F(5, 49) = 1.524, p = .200$), or gender ($F(5, 49) = 1.726, p = .146$) and no interaction ($F(5, 49) = .470, p = .796$). An absence of interaction between gender and group pre-A level gives confidence that this confound at post-A level should not affect any interpretation of results. As there were no effects or interactions present, ANOVAs for each dependent variable were not necessary.

Working memory × group × education level analysis

Hubber (2016), in a study of the working memory capacities of undergraduate mathematicians versus non-mathematicians, found that the students that were studying mathematics displayed no advantage for verbal working memory, but did for spatial working memory. In order to establish whether the data of the current study replicated this finding, this section reports the results of a 2(working memory type: spatial/verbal) × 2(group) × 2(education level) ANOVA.

This ANOVA revealed a main effect of working memory type ($F(1, 123) = 97.877, p < .001$), group ($F(1, 123) = 15.187, p < .001$) and education level ($F(1, 123) = 19.353, p < .011$). There was a significant interaction between working memory type and group ($F(1, 123) = 4.326, p = .040$), but no interaction between working memory type and education level ($F(1, 123) = 0.404, p = .471$) and no three-way interaction between working memory type, group and education level ($F(1, 123) = 0.599, p = .440$). The interaction between working memory type and group mirrors the findings of Hubber (2016), suggesting a particular spatial advantage for mathematicians in terms of working memory capacity. Figure 4.11 shows this interaction.

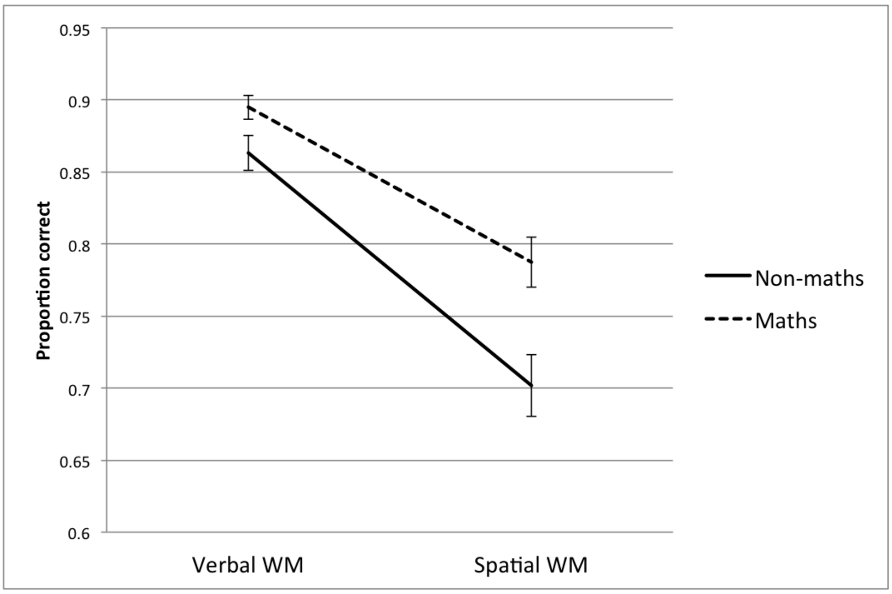


Figure 4.11: Interaction between working memory type and mathematics group

Bayesian analysis

In terms of the two research questions posed at the beginning of this chapter, the data had provided evidence for the first: that mathematicians perform better on spatial tasks. For the second question, that of whether or not there are developmental differences between mathematicians and non-mathematicians, and an effect of studying advanced mathematics, the evidence is ambiguous. To further investigate this second question, a 2(group) × 2(education level) Bayesian

ANOVA was performed for each of the dependent variables. This Bayesian ANOVA calculates the likelihood of five possible models of effects, given the data. Model 4 would represent the strongest evidence for a developmental difference between the mathematicians and non-mathematicians.

- **Model 0** — The null model as a comparison, in which there are no effects present.
- **Model 1** — An effect of group.
- **Model 2** — An effect of education level.
- **Model 3** — An effect of group *and* education level.
- **Model 4** — An effect of group and education level, *and* an interaction between group and education level.

The previous p-value testing ANOVAs established an effect of group for 2D rotation, and an effect of both group and education level for 3D visualisation, matrix reasoning, verbal working memory and spatial working memory. No significant interactions were found, although some approached significance.

Table 4.10 shows the calculated Bayes factors for each of the above models for performance on each of the dependent variable tasks, indicating how likely each were given the current data.

Dependent variable	Model 0 BF_{00}	Model 1 BF_{10}	Model 2 BF_{20}	Model 3 BF_{30}	Model 4 BF_{40}
2D rotation	1.00	60.30	0.25	16.83	11.36
3D visualisation	1.00	94.13	1.44	235.81	176.33
Matrix reasoning	1.00	109.39	23.41	8540.64	10991.66
Verbal working memory	1.00	671.41	1.95	3122.49	767.38
Spatial working memory	1.00	23.81	7.40	341.31	130.32

Table 4.10: Bayes factors for all possible ANOVA models

In the case of 2D rotation, the most likely model is one that only includes an effect of group. In the case of 3D visualisation and verbal and spatial working memory, the most likely model is one that includes an effect of both group and education. For matrix reasoning, the most likely model also includes an interaction between group and education level.

The Bayes factors for the likelihood of the null hypotheses can be calculated as the reciprocal of these. For example, for 2D rotation, the relative evidence for Model 1 over the null model is $BF_{01} = \frac{1}{BF_{10}} = \frac{1}{60.30} = 0.017$. Because these Bayes factors are transitive, it is possible to combine them as follows: $BF_{xy} = BF_{xa} \times BF_{ay}$. A combination of these two features allows a further investigation of the Bayes factors for each of the dependent variables. For example, it is possible to calculate how much more likely Model 4 is over Model 3 for matrix reasoning, which would be represented by $BF_{43} = BF_{40} \times BF_{03} = \frac{10991.66}{8540.64} = 1.29$. Table 4.11 shows the Bayes factors for two situations for each of the dependent variables:

- BF_{14} — How much more likely Model 1 (an effect of group only) is than Model 4 (a full model with an interaction), and
- BF_{43} — How much more likely Model 4 is than Model 3 (effects of both group and education level, but no interaction).

Dependent variable	BF_{14}	BF_{43}
2D rotation	5.31	0.67
3D visualisation	0.53	0.75
Matrix reasoning	0.01	1.29
Verbal working memory	0.87	0.25
Spatial working memory	0.18	0.38

Table 4.11: Bayes factors to compare the likelihood of Models 1 and 4, and Models 4 and 3

For all but 2D rotation, the values of BF_{14} are very small, indicating that Model 1 is no more likely than Model 4. In the case of matrix reasoning, in which Model 4 was the mostly likely given the data, the BF_{43} indicates that this was only 1.29[4] times more likely than Model 3, which is not strong enough evidence to be certain of any conclusions.

4.4 Discussion

The aim of this study was to collect data to help to answer two research questions:

1. Do mathematicians perform better on spatial tasks?
2. Is there evidence of developmental differences between mathematicians and non-mathematicians?

[4]This is considered a very small Bayes factor.

Each of the main analyses will be discussed in relation to the strength of evidence that they provide for each of these questions.

Education level $\times$ group ANOVAs for each DV

The first question was confirmed by the results, with mathematicians performing better on all of the spatial measures. This finding supports much of the research literature discussed in Chapter 2 which highlighted many links between the development of spatial skills and mathematical achievement. There are a number of explanations that would be worth investigating of why this advantage is apparent. Possibly, although the students had all studied the same UK syllabus for GCSE mathematics, some may have been exposed to more mathematics (e.g. extra work at home) or may have interacted with the mathematics being taught and learned in a more meaningful way, allowing more general spatial skills to develop. In order to gain a GCSE grade, students are required to sit an examination that is made up of a set of quite predictable questions. The design of mathematics examinations, and their similarity year-on-year, have influenced the way in which the subject is taught in many schools. The tendency for teachers to 'teach to the test' in order to help students obtain the best grades is a worry to educators, many of whom are concerned that this technique fails to foster creativity and deep understanding in students. On the other hand, students who are encouraged to make meaningful links between areas of mathematics, and to enjoy aspects of the subject outside of the classroom, are more likely to obtain further benefit from any formal discipline value. If this explanation does hold any truth, then it may be possible that the answer to the second research question would be yes, but this development happens before students decide whether or not to study advanced mathematics. It is also possible that the students who chose to study advanced mathematics were born with an innate level of spatial skills above that of the non-mathematicians. This idea of a genetic influence on the development of many cognitive skills is widely researched and discussed, and it can be quite confidently assumed that part of the variance in spatial skills will be due to this. A study of 4,174 twin pairs found that 60% of the relationship between spatial skills and mathematics could be explained by genetic factors (Tosto et al., 2014). Even when bearing in mind this seemingly large percentage, close to half of the associated variance between a persons mathematical skill and spatial skill is responsive to external factors, of which educational experience is certain to be a major one.

For the majority of the DVs, there was also an effect of education level, confirmed by both the p-value and Bayesian analyses. For 2D rotation, however, there was no effect of group. The literature discussed in Chapter 2 suggested that 2D rotation tasks could be completed too easily by adults, in which case, no difference would be expected to be seen between education levels. However, these sort of ceiling effects seem an unlikely explanation for the current data, as even the top-performing group (post-A level mathematicians) only scored an average of approximately 0.85 accuracy. The non-effect of education level observed for the 2D rotation task could be explained by its similarity to the type of lower-level mathematics tasks that might make up a GCSE question, e.g. Figure 2.3 in Chapter 2. Students studying A level mathematics are unlikely to be exposed to these types of problems through their A level course, as the content is focused on higher-order mathematical skills. Therefore, it is perhaps understandable that no significant improvement is seen between education levels.

For the remainder of the DVs, an effect of education level was present, with both the mathematicians and non-mathematicians performing better at post-A level. This suggested that these skills develop between pre- and post-A level, whether or not a student chose to study advanced mathematics or not. There are two possible explanations for this. Firstly, the cross-sectional nature of the study means that education levels cannot be compared without considering difference between the groups. The cohort of pre-A level students, although all choosing to continue into non-compulsory education, are likely to not be as academically gifted as the post-A level cohort of students who all had completed their advanced study, and chosen to continue to a degree level of education. The students' self-reported GCSE results showed that the post-A level students achieved higher grades than the pre-A level group, confirming this first explanation. A second explanation for the effect of education level is that students do genuinely continue to develop spatial skills between pre- and post-A level, whichever subjects that they chose to study. This would suggest that a certain amount of development is happening at this fairly late stage of maturity. The development of cognitive constructs such as spatial abilities at a young age is of particular interest to researchers due to the fact that development is happening quickly, and is sensitive to other internal and external factors. Certainly, a large amount of mathematics cognition research is focused on pre-secondary school aged children. However, there is a body of evidence that suggests a certain amount of plasticity of cognitive abilities in older children and adults (Dahl, 2004). Assuming that a certain amount of development is possible in older children, investigating what interventions might support and further this development are of interest, particularly when the importance of spatial skills STEM careers is considered. This explanation is explored more fully with a longitudinal study in Chapter 6, which eliminates the issue of comparing academically different cohorts of students.

Although the interaction plots for the education level × mathematics group ANOVAs for each DV suggested that the mathematical advantage for the post-A level students was more pronounced than for the pre-A level students, there were no significant interactions between education level and mathematics group. However, some were close to significance, with performance on the matrix reasoning task being the closest. Bayesian ANOVAs for each of the DVs also revealed some discrepancies about which model best suited the data. For the matrix reasoning task, a model that included both main effects, and an interaction was best suited, but only by a small margin. It could be argued that conclusive evidence was not found to support the second research question of developmental differences between the groups because of the small number of participants involved in the study. This additional concern is also addressed through the longitudinal study in Chapter 6, which employs a larger number of students.

Gender differences in spatial skills

The current data did not provide evidence for the existence of gender differences, although the analysis was unable to be performed at post-A level because of the gender imbalance between groups. There exists a reasonable amount of literature associating spatial skills with a male advantage from a young age, and suggesting that this might be a reason for the observed higher achievement of males in mathematics (e.g. see Baron-Cohen (2003)). This male advantage has also been reported in adulthood, for example, Geary et al. (2000) found that the male advantage seen when undergraduates were asked to solve worded arithmetic problems was, in part, mediated by spatial skills. However, Spelke (2005) published a critical review of the evidence surrounding sex differences in mathematics and science, asserting that the male dominance that is seen in high-level mathematics and science careers is not due to differences in cognitive development. Spelke claimed that much of the evidence of sex differences in infants' behaviour lacked replication validity, and concluded that men and women develop and possess equal cognitive capacities for mathematics. At pre-A level, the current data supports these conclusions. The fact that it was not possible to fully explore the effects of gender post-A level because of the high proportion of males in the mathematics group is representative of university mathematics course cohorts nationally, and was therefore unavoidable. Study Two of this thesis, which measures the spatial skills of students longitudinally, avoids this possible confound.

Working memory type × education level × group analysis

This analysis found main effects of group and of education level, which can be explained through the feature discussed previously. In addition, there was an effect of working memory type, with participants performing better on the verbal working memory task. As the tasks were designed to be as similar as possible, with identical processing elements, and the same number of items to remember, this result indicates that the spatial task was genuinely more challenging for the participants. The presence of a significant interaction between working memory type and group echoed the findings of Hubber (2016), showing that the mathematicians had a specifically spatial working memory advantage. This provides further evidence for mathematicians possessing higher levels of skills in spatial tasks, and rules out the possibility that the mathematicians are simply more skilled in all tasks of cognitive abilities. There were, however, no interactions found with education level which, again, suggests that there may not be a developmental difference between the mathematicians and non-mathematicians.

4.5 Conclusions

The purpose of Study One was to answer the question of whether the spatial advantage linked to mathematicians is due to an effect of studying advanced mathematics, or whether this filtering effect is in place pre-A level. Mathematicians from both education levels displayed advantages in all of the spatial measures in comparison to the non-mathematicians, but there were no interactions between education level and mathematics group. The data provided evidence, mostly, for a filtering effect, in place before the students chose to study advanced mathematics.

The cross-sectional nature of this study allowed a large amount of data, on a variety of measures, to be collected and analysed, and some conclusions to be drawn, in a relatively short time period. This design suited the explorative purpose of the study but had inescapable drawbacks regarding any inferences to be made about the causal direction of any findings. Chapter 6 presents a longitudinal study which aimed to research the same questions, allowing more definite conclusions to be drawn about whether or not there exists a developmental difference between the groups. The next chapter describes the identification of a single spatial skills construct to be used in Study Two.

Chapter 5

Identifying a 'general spatial reasoning skills' construct

Study One measured the spatial reasoning skills of year 12 and undergraduate students using four tasks: 2D rotation, 3D visualisation, matrix reasoning, and spatial working memory. The conclusions of Study One suggested that mathematicians possessed a particular spatial advantage over non-mathematicians (compared to a measure of verbal working memory), and revealed a general trend that suggested that these group differences may be more pronounced at undergraduate (post-A level) than in year 12 (pre-A level). Although no significant interactions were found to suggest that the study of A level mathematics in particular was having an effect on the development of these spatial skills, there were aspects of Study One that may have led to such effects not being identified. The relatively low power of the study is a possible reason for the interactions not being significant at the $p = .05$ level.

An ideal way in which to investigate this further was to run a similar study with a larger number of participants. To practically achieve this, it was necessary to reduce the time that participants were required to spend completing the spatial reasoning tasks. For Study One, each participant took up an hour of the experimenter's time, a resource impossible to reproduce on a larger scale. Ideally, the larger-scale study would employ just one measure of spatial reasoning that was simple, and quick, to administer.

Before deciding on a single measure of spatial reasoning, it was vital to identify more detail about the specific constructs that were being measured by the four tasks in Study One. Although each was classified as a measure of spatial reasoning, it is possible that they were honing in on subtly different aspects of this construct. For example, scores on the 3D visualisation task could be highly dependent on spatial awareness, whereas the spatial WM task could be dependent on access to short term memory. In order to establish whether there existed a 'general spatial reasoning skill' construct that all four spatial tasks were measuring, a factor analysis was performed to establish the shared variance on the tasks.

5.1 Principal components analysis of the four spatial skills measures

Table 5.1 shows the correlations between the four measures of spatial reasoning used in Study One.

Variables	1	2	3	4
1. Spatial working memory	-			
2. 2D rotation	346**	-		
3. 3D visualisation	.288**	.278**	-	
4. Matrix reasoning	.369**	.275**	.318**	-

Table 5.1: Correlations between the four spatial reasoning measures.
*** Correlation is significant at the 0.01 level*

A principal components analysis was performed to establish the amount of shared variance between the measures of spatial skills used in Study One.

The four spatial skills measures: spatial working memory (SWM), 2D rotation, 3D visualisation, and matrix reasoning, were subjected to a principal components analysis. Prior to performing this, the suitability of the data was established. The correlation matrix reveal that all of the measures were significantly correlated, with all $rs > 0.26$. The Kaiser-Meyer-Oklin value was .704, and Barletts Test of Sphericity reached statistical significance.

The principal components analysis revealed just one component with an Eigenvalue exceeding 1, explaining 47.2% of the variance, and an inspection of the scree-plot showed a clear elbow-break after the first component. These two pieces of information were used to conclude that all four spatial skills measures were loading onto one single factor, that could be considered to represent a general spatial factor.

Table 5.2 shows the correlation matrix output of this analysis and Figure 5.1 shows the scree plot for the extraction.

Measure	Component 1
SWM	.722
2D rotation	.665
3D visualisation	.654
Matrix reasoning	.705

Table 5.2: Loading on component 1

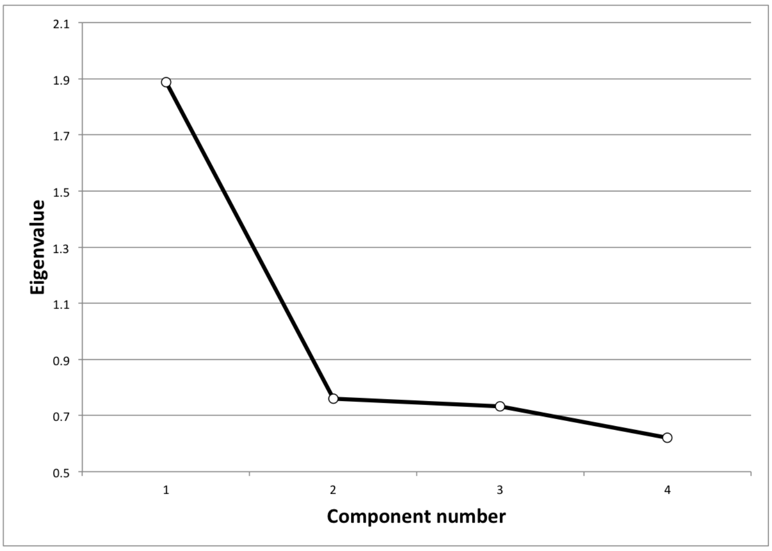


Figure 5.1: Screeplot of the principal components analysis of the four spatial skills measures used in Study One

Interactions using a single measure

A principal components analysis suggested that the four spatial skills measures used in Study One were measuring the same construct, and allowed a single general spatial measure to be calculated for each participant as an average of their scores on each of the measures. Table 5.3 shows the results of this new composite measure.

The interactions between education level and mathematics group and their effects on this single spatial measure can be seen in Figure 5.2. Here there is a significant effect of education level ($F(1, 124) = 10.2, p = .002, \eta_p^2 = .077$), and of group ($F(1, 124) = 29.7, p < .001, \eta_p^2 = .197$), but no interaction ($p = .088$).

	Pre-A level	Post-A level
Mathematicians	.704 (.125)	.811 (.091)
Non-mathematicians	.623 (.137)	.655 (.129)

Table 5.3: Mean 'general spatial reasoning' ability (Standard deviations in brackets) — Study One

This analysis using a single general spatial factor reveals a similar picture to that of the separate analyses in the previous chapter, and the interaction between group and education level is borderline significant, suggesting a possibility of a developmental difference between the spatial skills of the mathematicians and non-mathematicians.

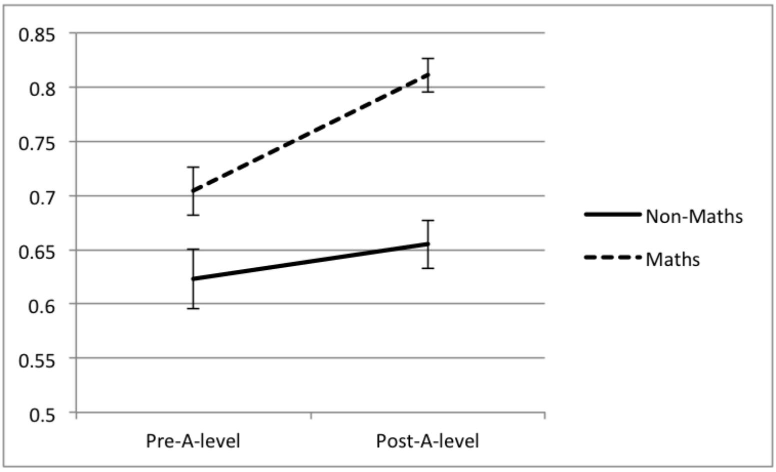


Figure 5.2: Interactions between education and group on combined spatial skills score — Study One

A Bayesian 2(group) × education level ANOVA with this single measure as the DV produced the Bayes factors displayed in Table 5.4. It can be seen that the most likely model was Model 3, but that $BF_{34} = 1.34$ indicated that this was only 1.34 times more likely than Model 4.

A further study, designed to explore the possibility of a developmental difference between group, using a longitudinal design, is reported in Chapter 6. A large number of year 12 students were tested on one measure of spatial skill (a method justified by the principal components analysis) at the very start of the academic year. After one year of advanced study, a second measure was taken and compared. The large sample size, as well as the longitudinal design, gave the study more power and helped to draw more solid conclusions. The problems

Model	BF_{x0}
Model 0	1
Model 1	52,946
Model 2	8.65
Model 3	2,560,000
Model 4	1,906,000

Table 5.4: Bayes factors for the five possible interaction models, see Section 4.3.4 for a description of these models

relating to group differences in the year 12 and undergraduate cohorts in Study One were also eliminated. Results from both studies provided an insight into the possible transfer value of advanced mathematics in terms of spatial skills, and are fully discussed in Chapter 8.

Chapter 6

Study Two — A longitudinal study of spatial skills during A level mathematical study

6.1 Introduction and theory

Study Two confronted the same research questions as Study One, but with a longitudinal design. As the literature surrounding this has been introduced and discussed at length previously, this chapter will summarise the justification for the study, and follow with a description of the methodology. The results will be reported and conclusions will be discussed based on data from both studies.

In Study One, the spatial skills of two education levels of students were measured using a variety of tasks; one group of year 12s (pre-A level), and one group of undergraduates (post-A level). These two groups were then both split into two: a mathematics group, and a non-mathematics group. For each of the dependent variables (DVs), the data provided evidence of a main effect of mathematics group which provided evidence for the first of the research questions:

1. Do mathematicians perform better on spatial tasks?

with the mathematicians performing better on all of the spatial measures. This mathematical advantage appeared to be more prominent post-A level than pre-A level, but the group $\times$ education level interactions did not reach significance. In addition, there was a specific spatial working memory advantage found compared to the verbal working memory measure for mathematicians. The cross-sectional nature of Study One, and the relatively small number of participants, meant that definite conclusions about the nature of the relationship between advanced mathematical study and spatial skills could not be drawn, and that further research was needed. Study Two therefore aimed to further investigate this relationship in regard to the second of the research questions:

2. Is there evidence of developmental differences between mathematicians and non-mathematicians?

The results helped to establish whether the study of advanced mathematics improves spatial skills through the process of formal discipline, via a transfer of skills, or whether the spatial advantage which is witnessed in mathematicians is the result of a filtering effect. Of these two opposing hypotheses the first, relating to the theory of formal discipline, makes the assumption that an intervention (in this case the study of advanced mathematics), can increase an individual's cognitive ability (in this case, spatial skills): a phenomenon that has been suggested as possible by a number of scientific studies as discussed in Chapter 1. Chapter 2 also discussed a number of working memory training studies that claimed to increase general reasoning skills (see Buschkuehl and Jaeggi (2010) for a full review of recent research). In addition, effects of studying certain school subjects on reasoning behaviour have been found (Attridge and Inglis, 2013; Inglis and Simpson, 2007; Lehman and Nisbett, 1990), and on spatial skills specifically (Blade and Watson, 1955; Sorby, 1999). The second of these two hypotheses is that of a filtering effect: that those students with more highly developed spatial skills, or a disposition towards utilising these skills, are more likely to choose to study advanced mathematics, effectively filtering them into two groups with observably different levels of spatial skill.

In addition to the advantages that the design of Study Two has over Study One in terms of its longitudinal nature, and the additional power of using a large sample, it was decided to incorporate the collection of data relating to all of the advanced-level school subjects that the participants were studying, not only mathematics. This allowed the relationship between spatial skills at Time 1, gains in spatial skills, and different school subjects to be explored.

A principal components analysis of the measures used in Study One revealed that they loaded onto one factor, representing spatial skill. In order to ensure that Study Two was properly powered, it was necessary to collect a large amount of data. Therefore, it was not possible to individually test participants on all of the measures previously employed in Study One as this would have been unrealistically time consuming. Therefore, one of the Study One tasks was chosen to act as a spatial reasoning measure for Study Two. The matrix reasoning task was chosen for three main reasons:

1. The potential for the task to be administrated by classroom teachers within the school/colleges, allowing a large amount of data to be collected at one time. The instructions for the matrix reasoning task were straightforward, the task was relatively self-explanatory, and the completed tasks were simple and quick to score.

2. The task was adapted from a very well established task, Raven's Progressive Matrices (RPM) (Raven et al., 2000). Although the Cronbach's alpha for the matrix reasoning task in Study One was low, RPM does have well documented validity and reliability values and established links in the literature with many positive 'real-life' outcomes such as creativity and job success (Ritchie, 2015).

3. In Study One, 2(group(mathematics; non-mathematics)) $\times$ 2(education level(pre-A level; post-A level)) ANOVAs for matrix reasoning revealed an interaction that approached significance (p = .054). A Bayesian analysis of the data suggested that a model that included both mains effects and an interaction was 1.26 times more likely to fit the data than a model which included only the main effects. Performance on this task therefore warrants further investigation.

6.2 Methods

6.2.1 Participants

The data was collected in two cohorts, from two consecutive academic years.

Cohort 2013/14

Seven hundred and fifty eight (758) year 12 students (aged from 15 years, 1 month to 20 years, 7 months, M = 16 years, 4 months at Time 1) from two Leicestershire colleges were recruited in September 2013. Of the 758 participants, 317 were male, 419 were female and 22 declined to report their gender.

Cohort 2014/15

One thousand, two hundred and eighty (1,280) year 12 students (aged from 15 years, 1 month to 21 years, 11 months, M = 16 years, 11 months at Time 1) from ten schools and colleges across the UK were recruited in September 2014. Of the 1280 students, 760 were male, 511 were female, and 9 declined to report their gender.

Neither the participants, nor the colleges were told of the specific aims of the study so as to avoid any influence of pre-held conceptions. All participants provided written informed consent and the study was approved by Loughborough University's Ethical Advisory Committee.

6.2.2 Design

The study followed a longitudinal, quasi-experimental design, over the period of one academic year. Participants were tested on a measure of spatial skills (matrix reasoning) as close to the beginning of the academic year of AS level study as possible, and again as close to the end of the academic year as possible. The quasi-experimental nature of the study was unavoidable because randomly assigning participants to the two groups was impossible, as the school subjects chosen to study at AS/A level by the participants was beyond the experimenter's control. These issues relating to quasi-experimental design were discussed in Chapter 3.

All students completed the spatial skills 20 minute paper and pen task at Time 1 and were invited to complete the task again at Time 2. Inevitably, there was a dropout at Time 2 that is discussed in the results section. This task was administered by teachers of the schools and colleges, following the instructions provided (see Figure 6.1).

6.2.3 Measures

The spatial reasoning skills of the participants were measured at Time 1 and Time 2. These were assessed using a matrix reasoning task very similar to that employed in Study One, adapted from Raven's Standard Progressive Matrices (RSPM) (Raven et al., 2000). All of the items from RSPM were used, split into two subsets using the even items at Time 1 and the odd items at Time 2. In addition, as RSPM are designed to be used to test reasoning skills in age groups of 6-years to 17-years, it was decided that 6 items from Raven's Advanced Progressive Matrices (RAPM) be added to the end of the tasks to ensure that the participants engaged with the task for the full time allocated.

THE TASK

The first page is information for the students to read and sign their consent to take part in the study.
The second page asks for some information from the students such as age and GSCE grades and the AS-levels that they are studying. Could you make sure that they carefully fill all of this in (some others have missed out some important information!)

The rest of the booklet is the task. The task is designed to last for 20 minutes (excluding filling in the second page). Given this time, it is unlikely that the students will finish all of the questions – this is fine, it's not designed to be finished.
Please ask the students to give an answer for each of the problems that they attempt, even if it is just a 'best guess' and to attempt the questions in the order that they are set out in the booklet.
You might want to check that everyone understands the example before they start the test (although the concept is pretty straight forward).

All information will be kept anonymously and not linked to any personal information.
If any of the students, or the school/college as a whole, wishes to withdraw from the study at any time, they are free to do so by contacting me.

Thank you again, so much, for participating in this research.
If you let me know when all of the completed tasks have been collected together and I will arrange for them to be picked up.

If you have any questions at all about the task or how to administer it, please let me know

Figure 6.1: Instructions given to teachers for the administration of the task

Again, the even items were used at Time 1, and the odd items at Time 2. This produced a 36 item task. The participants were given 20 minutes to complete the task. During this time, it was not expected, or desired, that the students finished all of the items. Any incorrect or uncompleted items were scored as a zero. A spatial reasoning score was taken as the number of items that were completed correctly (a theoretical maximum 36). An example of an item similar to the RPM test items is illustrated in Figure 2.9.

Participants were also asked to self-report the subjects that they had chosen to study at AS/A level. The standard practice in the UK was for students to study four subjects at AS level, dropping one as they progress to A level the next academic year. Therefore, the majority of the students self-reported four subjects, all of which had equal weighting in the student's timetable. Across the colleges, 42 distinct course titles were reported. Using the Higher Education Statistics Agency (HESA) classification, these subjects were coded into 14 subject groups which can be seen in Table 6.1.

If a student was studying one or more of the sub-courses within a subject, they were coded as studying that subject. In some cases, an individual student had chosen to study up to 3 sub-courses within a subject group. Instead of coding this situation as a '3', a binary coding was chosen as the most appropriate system. This was because an equal study time allocation across the subjects could not be assumed. Therefore each student was coded with either a 1 or a 0 for each subject.

Table 6.1: Categorisation of AS level courses

Subject group	**Sub-courses**
Biological Sciences	*Biology; Psychology*
Business & Administrative Studies	*Accounting; Business Studies*
Computing Science	*Computing; ICT*
Creative Arts and Design	*Art; Art and Design; Design and Technology; Drama and Theatre Studies; Fashion and Textiles; Film Studies; Graphic Communication; Music; Photography; Sculpture and Ceramics; Performing Arts; Food Technology*
Electronics & Technology	*Electronics*
English	*English Language; English Literature; Combined English*
Humanities	*History; Ancient History; Religious Studies; American History*
Languages	*French; German; Spanish; Italian*
Law	*Government and Politics; Law*
Librarianship & Information Science	*Media Studies*
Mathematical Science	*Mathematics; Further Mathematics; Statistics*
Physical Science	*Chemistry; Geology; Physics*
Social, Economic & Political Studies	*Geography; Health and Social Care; Sociology; Travel and Tourism; Citizenship*
Sport Science	*Physical Education*

A measure of prior attainment was also obtained from the participants as their self-reported GCSE grades in mathematics, English and science. The grades were coded for analysis (e.g. A* = 8, A = 7, B = 6, ...). A total score for prior attainment was calculated as a sum of these three grades (theoretical maximum of 24).

6.3 Predictions

Before discussion of the results of Study Two, it is worth recapping the two hypotheses that are being investigated, and what these would look like in terms of the data produced. Based on the theory discussed in Chapters 1 and 2, there are different outcomes that would be expected, dependent on whether the data supported the theory of formal discipline value of advanced mathematics: that studying mathematics improves spatial skills, or the alternative filtering account: that individuals with better spatial skills are more likely to choose to study mathematics.

A 2(mathematics or non-mathematics group) × 2(Time 1 and Time 2) mixed ANOVA will reveal any significant effects of a year of advanced mathematical study on spatial reasoning skills. Figure 6.2 illustrates the possible outcome of this ANOVA if the study of advanced mathematics had some formal discipline value in terms of the development of the students' spatial skills. Figure 6.3 shows the result if the second hypothesis were supported.

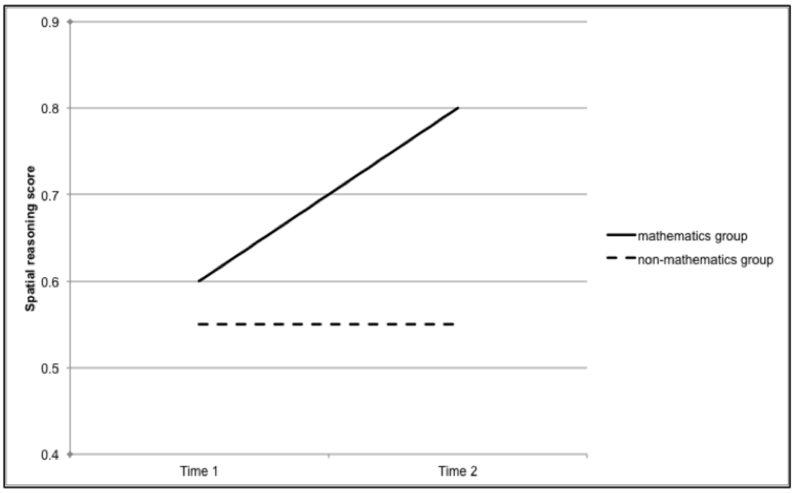


Figure 6.2: Predicted interaction plot 1

6.4 Results

6.4.1 Exclusions

Of the 2,038 participants, 14 were excluded as outliers as they scored more than 3 standard deviations from the mean on the spatial reasoning task at Time 1, leaving 2,024 participants. Of these, 1,140 participants completed both Time 1

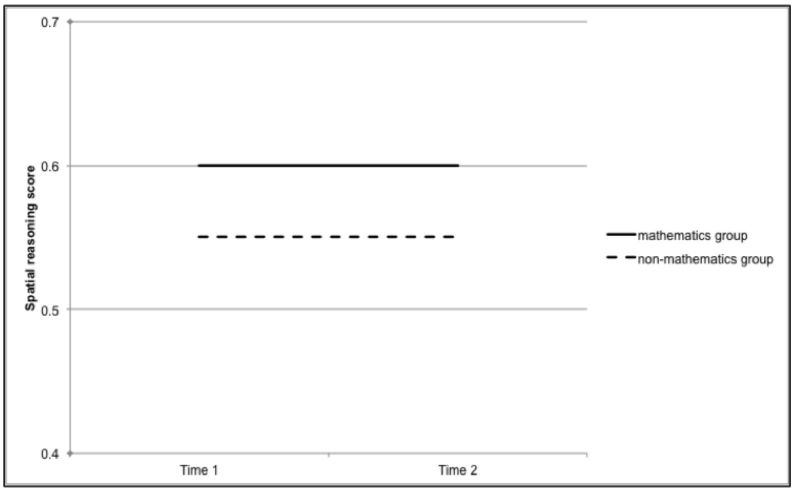


Figure 6.3: Predicted interaction plot 2

and Time 2. Table 6.2 shows the spatial reasoning scores at Time 1 for these two groups, as well as other descriptive statistics and the Cohen's d effect sizes for the independent t-tests comparing the Time 1 & Time 2 group, and the Time 1 only group.

The group of students that completed the data collection at both Time 1 and Time 2 scored significantly differently on all of the measures above. Although it is surprising, and difficult to explain why the 'dropout' group were significantly older ($t(2004) = 4.00, p < .001$) and more male ($\chi^2(1, N = 2008) = 10.16, p < .001$), the actual differences between the average age and gender between the two groups was very small (0.2 of a year, and 7.2% respectively) and effect sizes were small[1] (see Table 6.2). The dropout group also reported significantly lower GCSE grades in mathematics ($t(1982) = 3.68, p < .001$), science ($t(1956) = 4.33, p < .001$) and English ($t(1983) = 3.78, p < .001$), and scored lower on the Time 1 spatial reasoning measure ($t(2023) = 4.22, p < .001$), but again with small effect sizes.

The dropout group will have included students that did not complete their first year of AS level subject courses and therefore were not present to complete the data collection at Time 2. It is understandable that this group will have scored significantly lower in their GCSEs, as well as on a measure of spatial reasoning at Time 1 as students with lower general reasoning skills are less likely to cope with the academic challenge of A level study.

[1] A value of Cohen's d of $d = 0.2$ is considered small, $d = 0.5$ medium and $d = 0.8$ large (Cohen, 1977).
A value of ϕ of $\phi = 0.1$ is considered small, $\phi = 0.3$ medium, and $\phi = 0.5$ large (USGS, 2016).

Table 6.2: Ages, genders, GCSE grades and Time 1 mean proportion correct spatial reasoning scores (SD) for those that completed both time points, and those that only completed Time 1. For GCSE grades, 8=A, 7=A, 6=B, etc...*

	Time 1 & Time 2 (N = 1,140)	Time 1 only (N = 897)	Effect size
Age (years)	16.7	16.9	$d = 0.18$
Gender	50.5% male	57.7% male	$\phi = 0.071$
GCSE Mathematics	6.42	6.22	$d = 0.17$
GCSE Science	6.43	6.19	$d = 0.17$
GCSE English	6.49	6.31	$d = 0.20$
Spatial score Time 1 (prop. corr)	0.630	0.605	$d = 0.19$

This explanation does not, however, fully account for the amount of missing data at Time 2. As the data was most often collected in form groups by the form tutors at the schools and colleges, it was the case that a number of whole form groups were missing at Time 2. A possible explanation for this is that those form groups that did not complete the task at Time 2 were made up of students that, for various reasons, may have been disengaged with the task. These students would therefore score lower at Time 1 and also be less willing to participate at Time 2. The form tutors were informed that they were free to opt out of the data collection at any point; an opportunity that a number of tutors took at Time 2, presumably because their tutor group did not want to complete the task at Time 2, or because they were unable to find time.

It should be noted that the raw difference in the scores between the 'drop-out' group and the 'completed' group at Time 1 was very small (0.878 on a 36-point scale) and the effect size small (d = 0.189). Although statistically significant due to the large number of participants involved, the 'complete' group correctly answered, on average, less than one item more than the 'drop-out' group. Also, the differences in GCSE grades is most prominent for science, for which the 'complete' group reported only the equivalent to a quarter of a grade higher.

As the research focus of this study was the longitudinal effects of specific subject groups, it was of even more importance to determine whether or not the two groups (dropout and complete) differed in terms of the subjects taken. Chi-squared tests of association on each of the defined 14 subject groups showed that there was no significant difference between the two groups for the majority of subjects. However, the test did reveal a larger drop-out rate for students taking the following subjects: biology, computer science, physical sciences, and mathematics. This drop-out rate reflects the fact that mathematics and all of the sciences feature in the list of subjects with the highest drop-out rates for post-16 education in the UK (LGA, 2015). The fact that the drop-out group in this study tended to have taken more of these subjects is therefore expected.

Overall, any differences between the 'drop-out group' and the students that completed the task at both time points can be attributed to the large N, and should not be considered a major cause for concern in regard to the rest of the analysis.

6.4.2 Descriptive Statistics

The following sections report data on the 1,140 participants that completed both Time 1 and Time 2 data collection. Table 6.3 shows the means and standard deviations for scores on the spatial reasoning measure at Time 1 and Time 2.

Table 6.3: Mean proportion correct spatial reasoning scores (SD) at Time 1 and Time 2 for participants that completed the task at both time points - Study Two

	Mean	SD
Time 1	0.630	0.127
Time 2	0.725	0.143

There was a significant gain in spatial reasoning scores from Time 1 to Time 2 ($t(1, 139) = 26.34, p < .001, d = 0.963$). In raw scores, this increase was by 3.42. This means that the students answered 3.42 more items correctly at Time 2 than Time 1, but does not relate to any standardised reasoning scores.

Table 6.4 shows the number of students taking each of the 14 subject groups. Students reported that they were studying an average of 3.26 (SD = 0.802) AS level subjects (minimum = 1, maximum = 6).

Table 6.4: The number of students taking each subject group - Study Two

Subject Group	Number of students
Biological Sciences	584
Business & Administrative Studies	195
Computing Science	92
Creative Arts and Design	260
Electronics & Technology	26
English	353
Humanities	291
Languages	72
Law	140
Librarianship & Information Science	53
Mathematical Science	576
Physical Science	466
Social, Economic & Political Science	515
Sport Science	92

The remainder of the results will be split into two parts:

- **Analysis One** — A comparison of two groups: those that took mathematics, and those that did not.

- **Analysis Two** — An investigation into the subject groups as predictors of spatial reasoning scores.

Each analysis will include a short discussion of the most noteworthy results, followed by a more general discussion in Section 6.5.

6.4.3 Analysis One: group comparisons

The first analysis considers two groups: those who took mathematics AS level, and those who did not. Five hundred and seventy six (50.5%) of the students took AS level mathematics. Tables 6.5 and 6.6 show summaries of the measures for these two groups.

Table 6.5: Prior attainment levels for mathematics and non-mathematics groups (A = 8, A = 7, etc.)*

	Maths group (N=576)	Non-maths group (N=564)
GCSE mathematics	7.09 (.977)	5.75 (.871)
GCSE English	6.70 (1.03)	6.28 (.948)
GCSE science	6.91 (1.06)	5.97 (1.09)

The mathematics group had achieved significantly higher GCSE grades in mathematics ($t(1119) = 24.28, p < .001, d = 0.964$), English ($t(1119) = 7.03, p < .001, d = 1.11$), and science ($t(1111) = 14.50, p < .001, d = 0.988$) than the non-mathematics group.

It would be expected that the group of students that had chosen to study mathematics would have achieved higher grades in mathematics at GCSE, and arguably science GCSE as well. However, the fact that the mathematicians also achieved higher GCSE grades in English suggests that they possessed a higher general academic ability. Therefore a general ability, calculated as an average of GCSE mathematics and GCSE English grades were controlled for in the following analyses to try to ensure that any statistical findings were not due to this confound. English and mathematics grades, and not GCSE science grades were used in this measure because of the close relationship between abilities in science and mathematics, with the science GCSE content consisting of an amount of mathematics itself. Therefore the most representative measure of general ability is an average of English and maths.

Spatial reasoning score at Time 1 was significantly correlated with mathematics GCSE grade ($r = .489, p < .001$), English GCSE grade ($r = .320, p <$

Table 6.6: *Spatial reasoning scores for mathematics and non-mathematics groups (SD)*

	Maths group (N=576)	Non-maths group (N=564)
Spatial Time 1	0.669 (0.126)	0.589 (0.115)
Spatial Time 2	0.766 (0.135)	0.682 (0.138)
Spatial Gain	0.0968 (0.118)	0.0931 (0.126)

.001) and science GCSE grade ($r = .406, p =< .001$).

A 2(mathematics or non-mathematics group) × 2(Time 1 and Time 2) mixed ANCOVA with GCSE mathematics and English as covariates was performed with spatial skills as the dependent variable. Figure 6.4 shows the interaction plot of this ANCOVA.

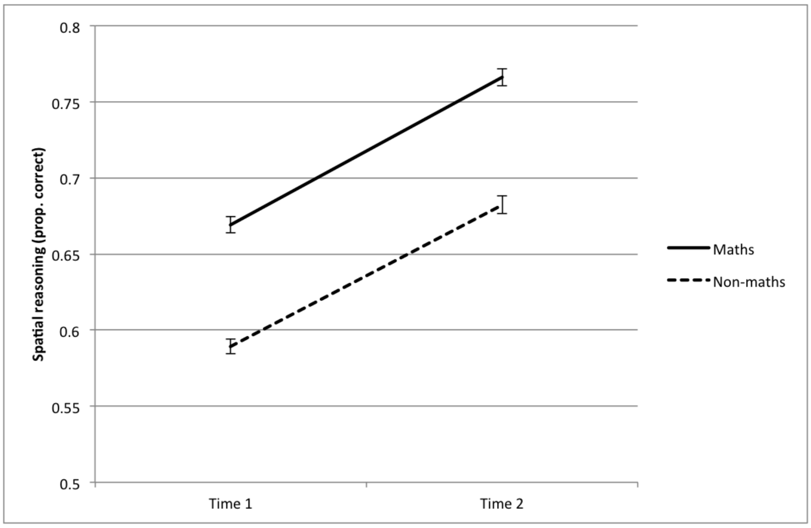


Figure 6.4: A graph to show the interaction of mathematics and non-mathematics groups on RPM score at Time 1 and Time 2

There was a significant main effect of time ($F(1, 1121) = 22.43, p < .001, \eta_p^2 = 0.02$) and of group ($F(1, 1121) = 5.29, p < .001, \eta_p^2 = 0.005$) but no interaction ($p = .603$). This result reflects the findings of Study One.

Although it is not customary to report effect sizes to this degree of accuracy, it is worth noting that the actual value for η_p^2 for the interaction was 0.000238. This effect size is considered extremely small and means that only 0.2% of the difference in spatial skills seen is due to the interaction between group and time.

Table 6.6 shows the spatial skill scores of the mathematics and non-mathematics groups at Time 1 and Time 2. Independent t-tests at each time point revealed that the mathematics group scored significantly higher than the non-mathematics group on spatial reasoning at Time 1 ($t(1132.32) = 11.16, p < .001, d = 0.988$) and at Time 2 ($t(1138) = 10.34, p < .001, d = 0.989$).

The next section of the analysis reports the data from the perspective of Bayesian statistics, the advantages of which were discussed in Section 3.

Bayesian analysis

A statistical analysis based on the principals of p-value significance testing only allows the rejection of a null hypothesis, and not the acceptance of one. In the case of the null hypothesis (H_0): the study of advanced level mathematics has the same effect as studying other advanced school subjects on the spatial skills of the students that study it, Bayesian statistics can go further in revealing the weight of the evidence that exists to support this, versus the alternative hypothesis (H_1): that there does exist a developmental effect of studying mathematics on spatial skills.

The statistical software program, JASP (Wagenmakers and Jove, 2016), was used to calculate an independent t-test Bayes factor for the difference in spatial reasoning gains between the mathematics and non-mathematics groups. A Cauchy prior width of 0.707 was used, as recommended for use in psychological experiments as a common effect size (Wagenmakers and Jove, 2016).

Figure 6.5 shows the Bayes factor robustness check, indicating that the current data provided strong evidence for the null hypothesis. The Bayes factor, based on the prior width and current data, is $BF_{01} = 11.596$. This can be interpreted as the data being 11.6 times more likely under the null hypothesis (that there would be no difference between the mathematics and non-mathematics groups in terms of their gain in spatial reasoning scores over the course of one year's study) than the alternative hypothesis (that there would be a difference).

The wide, and ultra wide prior dots indicate the extent of the strength of the evidence having chosen different prior widths. The value of 0.707 that was chosen for this analysis, along with the data collected, provides strong evidence for the null hypothesis. Even when the effect size for the alternative hypothesis is defined in the analysis as a distribution closer to zero (represented by a smaller prior width), these data provide evidence in favour of the null and against the alternative hypothesis. All Cauchy prior widths above 0.2 result in Bayes factors considered at least moderate evidence (see Figure 6.5).

The JASP software also produces a sequential analysis of the data (see Figure 6.6) which shows how the Bayes factor changes as the number of participants

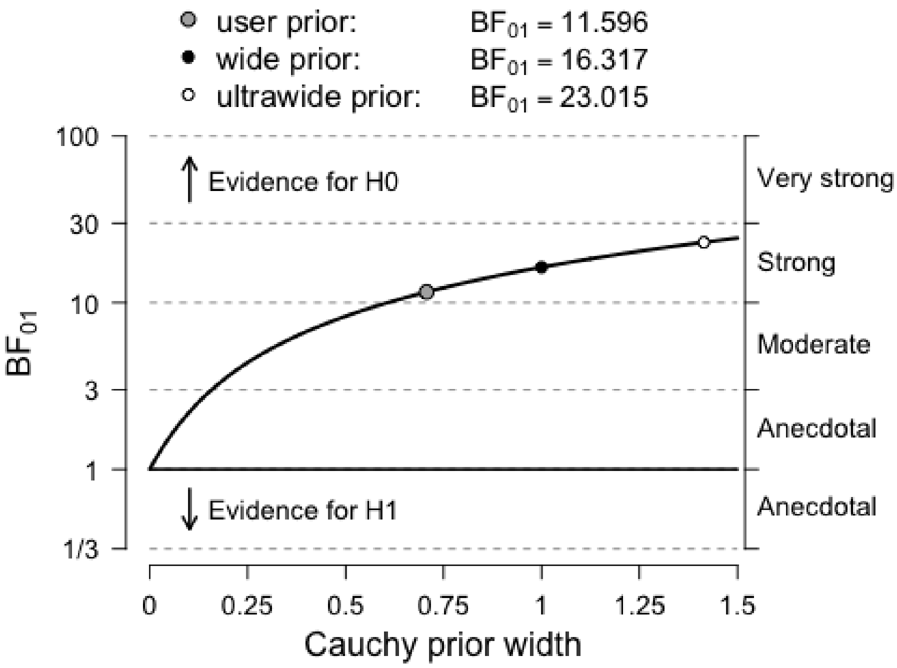


Figure 6.5: Bayes factor robustness check plot — Study Three

increases. It can be seen from Figure 6.6 that the Bayes factor becomes stable after approximately N=400.

6.4.4 Analysis Two: linear regressions

In addition to a comparison of the mathematics and non-mathematics groups, it was also possible to analyse the predictive value of a range of different school subjects in terms of the participants' spatial reasoning scores before and after a year's worth of advanced mathematical study. Three regressions were run on the data, with different dependent variables: spatial reasoning skills at Time 1, spatial reasoning skills at Time 2, and gain in spatial reasoning skills between Time 1 and Time 2. With relation to the subject groups defined in Table 6.1, the participants were coded as 1 if they took one or more of the sub-courses at AS-level, and 0 if they did not. Table 6.4 in the previous section shows the number of students that took each subject group.

First, a multiple regression was performed with spatial score at Time 1 as the dependent variable and each subject group as distinct predictor variables. Table 6.7 displays the results of this regression. It can be seen that a num-

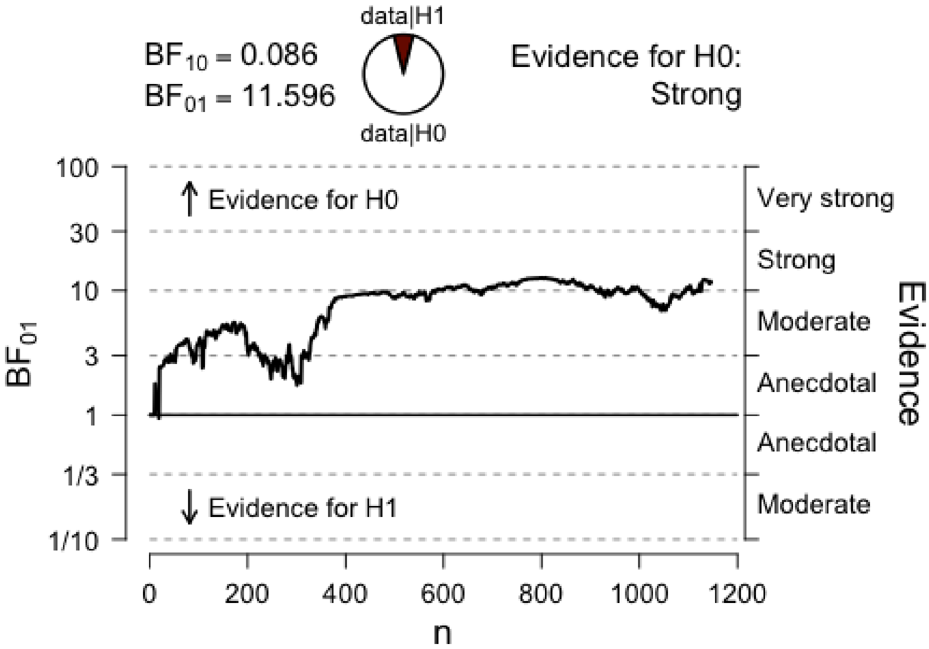


Figure 6.6: Sequential analysis of the data to show how the Bayes factor changes as the number of participants increases

ber of subject groups were significant predictors of students' spatial scores at Time 1. Of these predictors, many can be disregarded as non-significant once Bonferroni corrected: creative art ($p = .090$), computing science ($p = .0.420$), law ($p = .525$), librarian & information science ($p = .585$) and social, economic and political science ($p = ..645$). However, humanities, mathematical sciences, and physical sciences are revealed as more convincingly significant predictors of spatial ability (all remain at $p < .001$ after Bonferroni correction).

Table 6.8 shows a similar multiple regression that was performed using spatial score at Time 2 as the dependent variable. At Time 2, in addition to those subject areas identified at Time 1, creative arts, biological sciences and English significantly predicted spatial scores. The Betas for both biological sciences and English were negative, meaning that the students that took these subjects scored lower on spatial reasoning than those that did not. Humanities, mathematical sciences and physical sciences remain the strongest and most convincing significant predictors of higher spatial reasoning scores.

The fact that mathematical sciences is a strong predictor of spatial reasoning in students at both Time 1 and Time 2 provides evidence for the filtering theory: that mathematicians perform better on spatial tasks than non-mathematicians. In order to investigate whether there was any evidence for the alternative formal discipline theory: a developmental difference between mathematician and non-mathematicians, a third multiple regression was run with spatial score gains (calculated as Time 2 - Time 1) as the dependent variable. Table 6.9 displays this analysis.

This third multiple regression was not a significantly predictive model ($R^2 = .013$, $p = .415$). In terms of predicting the gains in spatial scores over a year of advanced study, knowing the subject group that a student took does not add any knowledge over and above that of the average gain score for the students. The results of both the multiple regressions, and of analysis one are discussed in the next section.

6.5 Discussion

This study aimed to provide additional evidence for the existence, or not, of a formal discipline value of studying advanced mathematics, in terms of performance on a spatial reasoning task. Students completed the task before and after an academic year's worth of advanced study in order for any gains to be analysed alongside which subjects they chose to study. Two analyses were performed with the data, which are now discussed in turn.

6.5.1 Analysis One

The first analysis considered the participants split into two groups, those that had chosen to study advanced mathematics, and those that had not. A measure of spatial reasoning was taken for each group before and after one year's study of advanced mathematics. A comparison of the gains in spatial reasoning over this time between the groups allowed the two hypotheses of this thesis to be tested. As with Study One, a main effect of group was found, with mathematicians scoring higher on the spatial reasoning task than non-mathematicians. In the discussion of the results of Study One, it was suggested that this main effect could be a result of unequal exposures to mathematics prior to the data being collected. Up to the point that the participants were tested on their spatial reasoning skills, they had all completed a number of years of compulsory mathematics education, and therefore could be considered to have been exposed to a very similar amount of mathematics. However, this may not be the case for a number of reasons. It is possible that some of the participants engaged in mathematical activities outside of the school system, for example extra hours of tutoring, or even completing mathematical tasks and puzzles in their spare time. Also, an argument exists that those who chose to take AS/A level mathematics, for whatever reason, engaged more meaningfully with the content of the GCSE syllabus, therefore possibly allowing for more formal discipline potential from the study of the subject. Alternatively, if this group had, in fact, engaged more meaningfully with the mathematics syllabus before A level, this may have been due to their higher spatial skills. This argument is more in line with the idea of individuals possessing different innate abilities to perform well on spatial reasoning tasks. If some individuals are born with a higher potential for skills in spatial reasoning, they might find success in mathematics easier and studying of the subject more enjoyable, and therefore be more likely to choose to study the subject in post-compulsory education.

The results also revealed a main effect of time, with participants scoring higher on the spatial reasoning task at Time 2. This effect was also found in the results of Study One, with the post-A level students scoring higher than the pre-A level students on all but one of the spatial skills measures. In Study One, this effect was difficult to interpret due to the cross-sectional nature of the study making it difficult to rule out the possibility of the differences being due to the cohorts being dissimilar academically. In the current study, a comparison was made within the same cohort of students at different time points, and therefore these confounds were eliminated. The presence of a main effect of time does, therefore, show that spatial reasoning scores increased over the period of one year's advanced study, suggesting a certain amount of neural plasticity in students of this age. However, if the increase in spatial reasoning scores was due to an effect of formal discipline from studying advanced mathematics, it would be expected that the students that did not study mathematics would not improve, or would improve significantly less. Conversely, the data revealed no differences in gains for the two groups. This finding mirrored that of Thorndike (1924a,b) and Wesman (1945) who found no advantage to mathematical study over any other school subject in terms of its effect on students' reasoning skills. A Bayesian analysis of the data allowed a calculation of the strength of the evidence for the null hypothesis: that there was no difference between the gains on the spatial reasoning task performance between the mathematicians and the non-mathematicians. The data provided strong evidence for this null hypothesis: that there are no spatial reasoning developmental differences between mathematicians and non-mathematicians.

6.5.2 Analysis Two

The second analysis of the data of the current study aimed to investigate the predictive value of a range of school subjects on students' spatial reasoning scores at the two time points, and on gains over time. Multiple regressions revealed that enrolment on humanities, mathematical sciences and physical sciences courses significantly predicted scores at Time 1, and at Time 2. The result for mathematical sciences is unsurprising when considering the large amount of literature linking spatial reasoning task performance and mathematical achievement discussed in Chapter 2. The result also mirrors the findings from Analysis One of this study, and of Study One: that mathematicians perform better on spatial tasks.

The significance of humanities and physical sciences has some similarity to previous research. Although the students did not differ at Time 1, Lehman and Nisbett (1990) found improvements in conditional and inductive reasoning in social science students, and gains in logical reasoning for natural sciences students (which included some of the same sub-categories as the physical sciences category of the current study) and humanities students. The students studied by Lehman and Nisbett were a university cohort, rather than the pre-university cohort of the current study, and therefore the content of the subject courses may not be comparable. Also, the reasoning tasks employed by Lehman and Nisbett were considerably different to the spatial reasoning task used in the current study, and so cannot be expected to measure a large amount of common features. However, it is noteworthy that links were found between reasoning of any kind, and these subjects. Although mathematics students were not studied specifically by Lehman and Nisbett (1990), they did find that the number of mathematics modules that a student was enrolled in, and their gains in scores on the reasoning tasks, were positively correlated, and this was hypothesised to explain the gains seen for the natural sciences students. It is very possible that a similar phenomena is present in the current study: that the students taking physical sciences were also taking mathematics, and that this contributed to the predictive value at both time points. The significance of the humanities subjects is harder to explain, and perhaps requires further investigation outside of this thesis.

The regression model for gains in performance was not significantly predictive, meaning that no knowledge could be gained about the way in which a students' spatial reasoning skills developed over a year of advanced study by knowing which subjects they chose to study. This provides evidence of no developmental differences between students that chose to study different school subjects, again reflecting the findings of Thorndike (1924a,b).

6.6 Conclusions

The data collected from Study One and Study Two provided strong evidence for a filtering effect of studying mathematics. The mathematicians performed better on every spatial measure, at every time point, for every group studied. In the case of the second research question: whether or not there exists a developmental difference between the groups, Study One was inconclusive. Study Two, however, provided strong evidence for this not being the case, and that, instead, some filtering effect existed prior to the study of advanced mathematics. Chapter 8 discusses these findings in the context of the issues relating to mathematics education and the literature introduced in Chapters 1 and 2. The next chapter describes a small experimental study which aimed to further investigate the nature of the relationship between spatial reasoning and mathematics.

Table 6.7: Predictors of spatial reasoning scores at Time 1

Variable	B	Stand. Beta	p	95% CI
Constant	**21.314**		**< .001**	**[20.3, 22.4]**
Biological Sciences	-.315	-.034	.261	[-0.86, 0.23]
Business & Administrative Studies	-.521	-.043	.156	[-1.24, 0.20]
Computing Science	**-1.09**	**-.065**	**.028**	**[-2.06, -0.121]**
Creative Arts and Design	**0.955**	**.087**	**.006**	**[0.28, 1.63]**
Electronics & Technology	-1.236	-.040	.157	[-2.95, 0.48]
English	-.492	-.050	.113	[-1.10, 0.12]
Humanities	**1.230**	**.117**	**< .001**	**[0.60, 1.87]**
Languages	.982	.052	.059	[-0.04, 2.00]
Law	**-.866**	**-.062**	**.035**	**[-1.67, 0.06]**
Librarianship & Information Science	**-1.295**	**-.059**	**.039**	**[-2.52, 0.07]**
Mathematical Sciences	**2.130**	**.232**	**< .001**	**[1.51, 2.75]**
Physical Science	**1.529**	**.164**	**< .001**	**[0.85, 2.21]**
Social, Economic & Political Science	**-.566**	**-.061**	**.043**	**[-1.11, -0.02]**
Sport Science	-0.024	-.001	.961	[-1.00, 0.95]
R^2	.172			
F	**16.70**		**< .001**	

Table 6.8: Predictors of spatial reasoning scores at Time 2

Variable	B	Stand. Beta	p	95% CI
Constant	**25.088**		**< .001**	**[23.9, 26.3]**
Biological Sciences	**-.985**	**-.096**	**.002**	**[-1.61, -0.36]**
Business & Administrative Studies	-.744	-0.055	.075	[-1.56, 0.08]
Computing Science	-.467	-.025	.406	[-1.57, 0.64]
Creative Arts and Design	**1.162**	**.095**	**.003**	**[0.39, 1.93]**
Electronics & Technology	-1.067	-.031	.283	[-3.02, 0.88]
English	**-1.040**	**-.094**	**.003**	**[-1.73, -0.35]**
Humanities	**1.254**	**.106**	**.001**	**[0.53, 1.98]**
Languages	.930	.044	.115	[-0.23, 2.09]
Law	-.625	-.040	-181	[-1.54, 0.29]
Librarianship Information Science	**-1.50**	**-.061**	**.036**	**[-2.90, -0.10]**
Mathematical Sciences	**2.284**	**.222**	**< .001**	**[1.58, 2.99]**
Physical Sciences	**1.334**	**.128**	**.001**	**[0.55, 2.11]**
Social. Economical & Politics Sciences	-.479	-.046	.132	[-1.10, 0.14]
Sport Science	.396	.021	.438	[-0.71, 1.50]
R^2	.147			
F	**13.851**		**< .001**	

Table 6.9: Predictors of spatial reasoning score gains

Variable	B	Stand. Beta	p	95% CI
Constant	**3.775**		$<$ **.001**	**[2.67, 4.88]**
Biological Sciences	**-.670**	**-.077**	**.022**	**[-1.24, -0.10]**
Business & Administrative Studies	-.223	-.019	.561	[-0.97, 0.53]
Computing Science	.623	.039	.228	[-0.39, 1.64]
Creative Arts and Design	.207	.020	.565	[-0.50, 0.91]
Electronics & Technology	.169	.006	.853	[-1.62, 1.96]
English	-.548	-.058	.091	[-1.18, 0.09]
Humanities	.024	.002	.943	[-0.64, 0.69]
Languages	-.052	-.003	.924	[-1.11, 1.01]
Law	.241	.081	.574	[-0.60, 1.08]
Librarianship & Information Science	-.204	-.010	.755	[-1.49, 1.08]
Mathematical Science	.154	.018	.639	[-0.49, 0.80]
Physical Science	-.195	-.022	.593	[-0.91, 0.52]
Social, Economic & Political Science	.087	.010	.766	[-0.49, 0.66]
Sport Science	.420	.026	.417	[-0.56, 1.44]
R^2	.013			
F	1.035		.415	

Chapter 7

Study Three — Does priming participants with mathematical reasoning improve their spatial reasoning?

7.1 Introduction and theory

The fact that mathematicians display higher levels of spatial skills in comparison to non-mathematicians has been well established through analysis of the data for Study One and Two of this thesis. The main purpose of Study Three was to more fully explore the nature of this effect. A possible account for the differences seen between mathematicians and non-mathematicians is that the mathematicians' spatial reasoning skills are being affected by the fact that they are faced with mathematical content on a day-to-day basis and are therefore experiencing a short-term, 'frame of mind' effect that may not translate to a genuine difference in their cognitive construct. Non-conscious phenomena of this sort is referred to as a 'priming effect' and can have surprisingly large effects on peoples' behaviour, as discussed in the following section.

For the current study, university engineering students were asked to complete a mathematical reasoning task followed immediately by a spatial reasoning task. By asking them to complete a mathematical reasoning task, their mental disposition will have been, at least temporarily, affected. The subsequent spatial reasoning task enabled any effects of short-term mathematical mental disposition on spatial reasoning behaviour to be observed. The following section will describe current literature regarding 'priming effects'.

7.1.1 Literature on priming effects

'Priming' is a type of unconscious memory, or perception, in which a person's behaviour is affected by recent previous experiences. One of the earliest experimental examples of this semantic organisation of memory was a study in which participants were asked to respond 'yes' if two strings of letters formed words, or 'no' if one, or both were a non-word (Meyer and Schvaneveldt, 1971). In the 'yes' condition, participants responded more quickly when the words were associated, e.g. NURSE and DOCTOR. Meyer and Schvaneveldt argued that this effect was due to the second word being easier to retrieve and recognise for the participants, having been 'primed' by the first word.

The discovery of this priming phenomenon inspired much research interest and, although cognitive psychologists could learn a lot about the organisation of memories and retrieval techniques, social psychologists were particularly interested in 'social priming' — how an individual's behaviour could be influenced by priming, further than pressing a 'yes' or 'no' button in a laboratory. The most publicised example of social priming is a study by Bargh et al. (1996) in which participants who were primed with words associated with old age took longer to walk down a corridor when leaving the experiment room than those who had been primed with neutral words. Previous research had shown a number of social priming effects on attitudes, for example on aggressive, or hostile, behaviour (Bargh and Pietromonaco, 1982; Carver et al., 1983; Srull and Wyer, 1979) and personality judgement formation (Higgins et al., 1977). Bargh et al. wanted to show that priming effects could be completely unconscious, that the behaviour outcomes could be unrelated to the situation in which they were displayed, and that they were not limited to social perception. Thirty students were randomly assigned to an elderly primed condition, or a neutral primed condition. Each participant performed a scrambled-sentence task, and were given the impression by the experimenter that this was a language proficiency test. Participants in the elderly primed condition were presented with words related to elderly stereotypes, e.g. WRINKLE, GREY and the neutral primed participants were presented with neutral words such as CLEAN, PRIVATE. A second experimenter then covertly recorded the time that it took for the participants to walk down the corridor to leave the experiment. The elderly primed students took, on average, about 1 second longer to reach the end of the 9.75m corridor. None of the participants, when interviewed afterwards, voiced any belief that the words had had any impact on their behaviour.

Within the same journal article, Bargh et al. (1996) reported a study in which subliminal images of faces were incorporated into an unrelated, tedious, visual task. Participants that were shown African American faces, as opposed to Caucasian faces, were more likely to display hostile behaviour when informed of a computer error that resulted in them having to start the experiment from the beginning. Those individuals that displayed high levels hostile behaviour did not consciously report high levels of racist attitudes towards African Americans. The findings of these studies were taken as proof that priming could have major effects on people's social behaviour and attitudes without any conscious awareness and much social theory incorporated this evidence for the following decades.

More recently, however, the replicability of these studies has been brought into question. Doyen et al. (2012) published a replication of the Bargh et al. (1996) study which failed to find the same results. Doyen et al. timed the participants leaving the experiment room using movement lasers, rather than a second experimenter, eliminating expectation bias. A second experiment in which an experimenter and stopwatch were relied upon, and in which the experimenter was aware of which participants were expected to walk slower, did reproduce Bargh et al.'s results. The set of social priming studies that Bargh et al. had built his career on, and that had been so instrumental in the formation of many social theories, was suspected as being misleading. With them the reputation of social psychology as a serious science was shaken (Bartlett, 2013).

In light of this, it is appropriate to be cautious about drawing inferences from the evidence on priming effects on human behaviour. There is, however, a much more solid evidence base for changing people's behaviour in decision making without their conscious knowledge through processes such as anchoring, both in laboratory conditions, and in the field. Tversky and Kahneman (1974) were some of the first psychologists to write about the way that humans make judgement and decisions in uncertain situations. They suggested that people rely on heuristic evidence to inform their decisions, and this can be manipulated. If certain pieces of evidence are made specifically available to the decision maker, particularly if evidence is scarce, their judgements are likely to be influenced by this. A robust example of anchoring in effect is a study in which students at the University of California were asked either "Is the Mississippi River longer or shorter than 2,000 miles?" or "Is the Mississippi River longer or shorter than 70 miles?". When asked to then estimate the length of the Mississippi River, students who were asked the first question estimated much higher (Jacowitz and Kahneman, 1995). The effects of anchoring are strong and well researched, and are often used to manipulate decision making, for example in advertising.

Although a lot of the research into social priming should be considered cautiously, there does remain evidence that a person's frame of mind and judgement can be influenced by their recent exposure to particular events. Considering this, it could be argued that the differences in spatial reasoning behaviour seen between the maths and non-maths groups in Study One and Two were due to a temporary 'frame of mind' that occurs because the participants in the mathematics group were in a context related to mathematics (their school or university). The following section will discuss the literature surrounding priming, mathematics, and spatial reasoning particularly.

7.1.1.1 Priming on spatial reasoning

A surface literature search of priming effects and spatial reasoning will quickly find a number of social psychology studies citing an effect of gender stereotyping. Considering the gender differences that have been found in many cognitive psychology studies (as discussed in previous chapters, e.g. Chapter 2), the idea that individuals might be affected by priming in this way does not seem incomprehensible. For example, Ortner and Sieverding (2008) claim to have found that adult women who are primed with male stereotypes displayed higher spatial abilities, and McGlone and Aronson (2006) reported similar results for females that were primed with their college identity as opposed to their gender identity. However, the Ortner and Sieverding study found priming effects with significance levels of only $p = .03$, and did not employ a control group, making it difficult to draw any definitive conclusions from the findings. It seem likely that reliable replication of these gender-priming/spatial findings needs to be achieved before they can be claimed to be robust.

In addition to this interest in gender priming and spatial skills, there is much research interest in the effect of music on spatial reasoning. Rauscher et al. (1993), first observed this effect through a study in which participants spent 10 minutes either listening to Mozart, listening to relaxation instructions, or sitting in silence. Immediately afterwards, they completed a number of spatial reasoning tasks, including a matrix task very similar to that used in Study One and Two of this thesis. Rauscher et al. found that the participants in the Mozart condition scored significantly higher on the abstract spatial reasoning tasks than either the relaxation or silence controls. Although in this experiment the effects did not last longer than 10-15 minutes, Rauscher et al. (1997) later did find evidence for longer lasting effects. These studies sparked an interest in the effects that music could have on spatial, and mathematical, performance due to the potential of enhancing these skills in an academic setting. Coined the 'Mozart Effect', this technique is thought to 're-train' the ear, allowing the brain to re-order cognitive processes and, as a result, enable individuals to solve problems more efficiently and effectively. Exactly how this is happening is not clear, although Steele et al. (1997) ruled out the suggestion of a working memory mediation. Although much of the early research into the Mozart Effect has, as with the social priming research, failed to be replicated since (see Stough et al. (1994) and Wilson and Brown (1997) as two of many examples), the Mozart Effect still proves popular in some research circles. See Bangerter and Heath (2004) for a full account of the history of the Mozart Effect, and Pietschnig et al. (2010) for a meta-analysis which found a small average effect size of $d = 0.37$. It is unsurprising to the majority of cognitive psychologists that the Mozart Effect has, over time, failed to live up to the potential test-boosting value that it promised on first discovery. In Chapter 1.2, the idea of 'transfer' was discussed and, in particular, the lack of evidence for the feasibility of far transfer between two elements that are relatively distant in their composition, such as music and spatial skills. Mathematics, on the other hand, has very established links in the literature with spatial reasoning and the transfer between the two would be seen as a lot less far, although still not 'near' as defined by many psychologists, e.g. Thorndike and Woodworth (1901) who failed to find effects of training on estimating the area of rectangles on a task of estimating the area of triangles.

If a certain amount of transfer between training and performance tasks is considered possible, and if having a higher level of spatial skills is a 'frame of mind' that can be temporarily induced in people, it stands to reason that priming with mathematical reasoning may have this effect. Study Three aims to test this hypothesis as an alternative explanation for any differences that have been found on performance on spatial reasoning tasks between mathematicians and non-mathematicians. This study does not aim to further knowledge in terms of the formal discipline value of advanced mathematics, but strengthens the understanding of the cognitive mechanisms that underlie the relationship between mathematics and spatial reasoning. If performance on a spatial reasoning task is unaffected by priming on mathematical reasoning, this will strengthen the argument that any effect that mathematical study does have on spatial skills is through long-lasting changes in cognitive development, and not just a 'frame of mind' effect.

In the current study, undergraduate students were primed with mathematical reasoning through a challenging paper-and-pen task before completing a spatial reasoning task. A matrix reasoning task was used for the same reasons as outlined in Section 6, as well as the fact that this task was used successfully in both Study One and Two of this thesis. The students' performance on the spatial reasoning task was compared to that of an active control group that had spent an equal amount of time working on an equally challenging, but non-mathematical task prior to completing the spatial reasoning task. It was deemed methodologically important to ensure that the control group be an active control group to be certain that any differences between the experimental group was due to the mathematical content, and not general cognitive stimulation. The following sections describe the way in which the study was executed, the results, and what conclusions can be drawn from the data.

7.2 Methods

7.2.1 Participants

One hundred and fifty four undergraduate engineers (122 male, 29 female, 3 not reported. Mean age = 20 years, 6 months) were recruited from Loughborough University, Leicestershire, UK. A variety of engineering courses were being studied by the students, of which all were in their second year. Table 7.1 lists the number of students taking the reported courses.

Type of engineering course	Number of students
Product design	53
Engineering management	7
Manufacturing engineering	21
Materials engineering	43
Sports technology	28
Not reported	2

Table 7.1: Engineering courses being taken

7.2.2 Design

The study was undertaken during lecture time allocated to a common module, 'Statistics for Engineers'. As part of this module, the students were required to prepare a piece of coursework that illustrated the skills that they have learnt in data analysis and statistical testing over the year. The data that was was collected for this thesis was also used as the data set for the students' coursework, and related coursework questions were written by the regular course lecturer. Students were told in advance that this lecture session would involve collecting data to be analysed for their coursework and were informed that attendance in the session was not compulsory.

The students were randomly allocated to either the experimental group (N=75), or the control group (N=79). This was done by placing paper tasks (with identical first pages) on to the desks prior to the students arriving and seating them as they entered the room. The students were tested in two groups, in accordance with the regular times for their lectures for 'Statistics for Engineers', at 4pm and 5pm on a Friday afternoon.

7.2.3 Measures

Each participant completed a pen-and-paper booklet. The experimental group were primed with mathematical reasoning, followed by a spatial reasoning task, and the control group completed a non-mathematical, but similarly challenging task followed by the same spatial reasoning task. Both the mathematical and the control priming tasks were designed with enough questions so that none of the participants would complete the entire task in the time allocated (15 minutes). These priming tasks can be found in Appendix 9.3.

7.2.3.1 Mathematical reasoning priming task

The mathematical priming task consisted of 16 items, taken from Swan (2005). Figure 7.1 shows an example of the items used. The items were designed to compel the students to consider thorough mathematical justifications for their answers, and to devise examples, or counterexamples, to defend their reasoning. The students were therefore not only performing mathematical operations, but also engaging in mathematical thinking and reasoning at a richer level than simply performing mathematical operations.

The more digits a number has, the larger is its value.

- Always true
- Sometimes true
- Never true

Please explain your answer:

Figure 7.1: An example of the mathematical reasoning task items: Study Three

7.2.3.2 Control priming task

The control priming task consisted of 34 items in which the participants were asked to choose the correct word to complete the sentence. Figure 7.2 shows an example of the control task. This was chosen to be a challenging but non-mathematical activity.

Tick the answer that you think is grammatically correct.

It's way past my bedtime and I'm really tired. I __________ go to bed.

- Should
- Ought
- Could

Figure 7.2: An example of the grammatical reasoning task: Study Three

After 15 minutes of working on the priming tasks, all participants were asked to turn to the section of their booklets that contained the spatial reasoning task.

7.2.3.3 Spatial reasoning task

The spatial reasoning task consisted of 18 matrix reasoning questions, in which the participants were asked to indicate the missing piece that completed the pattern correctly. This task was very similar to the task used in Study One and Study Two and an example of this type of item can be seen in Figure 7.3. The 18 items were taken from Ravens Standard Progressive Matrices (Plus) (Raven et al., 2000), suitable for the age group being tested.

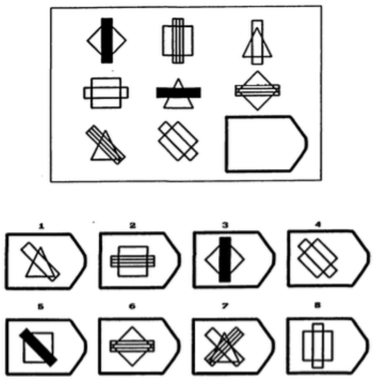


Figure 7.3: Example of the matrix reasoning items: Study Three

The participants were also asked to self-report their previous mathematical achievement as their A-level result, or equivalent, and their end of year university course mark (from their first year of study).

7.3 Results

7.3.1 Descriptive statistics

The experimental and control groups did not differ significantly in terms of the number of students that were taking any of the particular engineering courses. The groups also did not differ in terms of gender mix, age, previous mathematical achievement, overall prior achievement, or end of first year course mark (all $ps > .1$).

Table 7.2 shows the descriptive statistics for the two groups.

Ideally, the analysis would consider gender, due to the large amount of literature linking gender, mathematics and spatial reasoning. Unfortunately, as with the post-A-level group in Study Two, the sample population was so predominantly male that statistical comparisons would be hard to interpret. This gender inbalance is to be expected from a class of engineering students in a UK university.

	Exp. (N=76)	Control (N=78)
Gender balance [male;female;not reported]	58;16;2	64;13;1
Age (years)	20.52 (.82)	20.49 (1.17)
Prior maths achievement	6.39 (.88)	6.12 (.94)
Average of all A level grades	6.20 (.68)	6.12 (.66)
End of first year degree score (%)	63.34 (7.26)	64.02 (7.54)
Degree type		
Product design	26	27
Engineering management	3	4
Manufacturing engineering	11	10
Materials engineering	20	23
Sports technology	14	14

Table 7.2: Mean (SD) descriptive statistics for the two groups. For prior maths score, 6 represents a grade B at A-level, 7 represents a grade A)

7.3.2 Main analysis

Correlations

Significant correlations were found between end of first year examination course mark and previous mathematical achievement ($r = .183, p = .048$) and overall prior achievement ($r = .211, p = .014$) which can be explained through a general 'academic achievement' factor. There was also a marginally significant correlation found between previous mathematical achievement and scores on the spatial reasoning task ($r = .174, p = .055$). This correlation would be expected to be significant with a properly powered study designed to measure this.

Between group analysis

The mean spatial reasoning scores for the experimental group were compared to the control group through an independent t-test. There was no significant difference between the mean spatial reasoning scores for the experimental group (0.51, SD=0.17) and the control group (0.54, SD=0.16). No between-subjects effect of experimental condition was found ($t(152) = 1.00, p = .318$).

Figure 7.4 shows the relative spatial reasoning scores of the two groups, from which it can be seen that, in fact, the control group scored slightly higher than the experimental group.

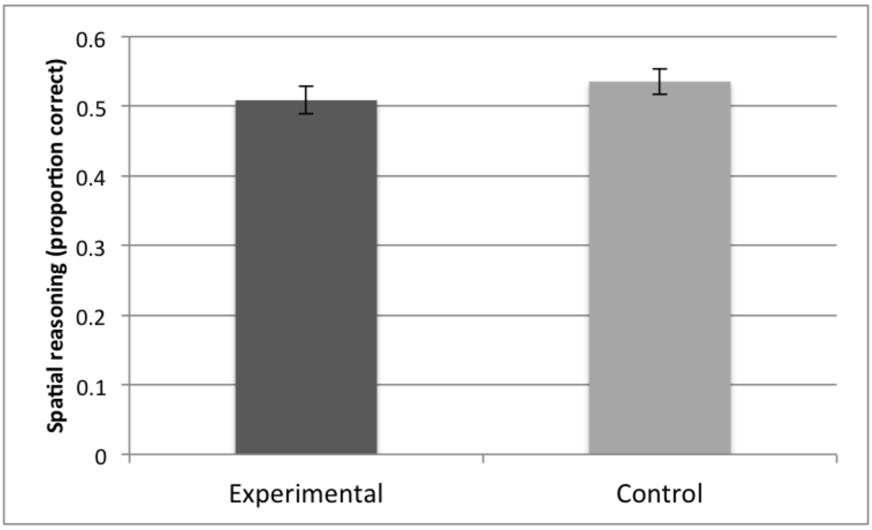


Figure 7.4: Mean accuracy scores for the spatial reasoning task - Study Three

Bayesian analysis

The above analysis, based on the principals of p-value significance testing provides evidence to suggest that the null hypothesis should not be rejected. In order to assess the strength of evidence in the data for the null hypothesis being true, a Bayesian analysis is more appropriate.

The statistical software JASP (Wagenmakers and Jove, 2016) was used to calculate an independent t-test Bayes factor for the difference in spatial reasoning scores between the experimental and control groups. The Bayes factor analysis compares evidence for the null hypothesis (that there is no difference between the spatial reasoning scores for the mathematically primed students and the control group), with evidence for the alternative hypothesis (that that is a difference). A Cauchy prior width of 0.707 was used, as this is the recommended common effect size for psychological experiments Wagenmakers and Jove (2016), and also used in previous studies of this thesis. This prior width, in combination with the current data, produced a Bayes factor of BF_{01}=3.629, indicating that the data were 3.63 times more likely to occur in a situation in which the null hypothesis was trues, as opposed to the alternative. Figure 7.5 shows the robustness check plot for this analysis. The prior width of 0.707, along with the data collected, provide moderate evidence for the null hypothesis.

A post-hoc power analysis, calculated to detect an effect size of 0.37[1], and

[1]The effect size was taken from a meta analysis of studies of the Mozart Effect, discussed in the introduction of this chapter.

considering the sample sizes used for the current study, revealed a power of 0.63. Although this value is a little below the accepted value of 0.80 for experiments, when taken alongside the fact that the effect was, in fact, in the opposite direction than predicted, and that the Bayesian analysis provided moderate evidence for the null hypothesis, it can be quite sensibly concluded that there was no priming effect present in this study.

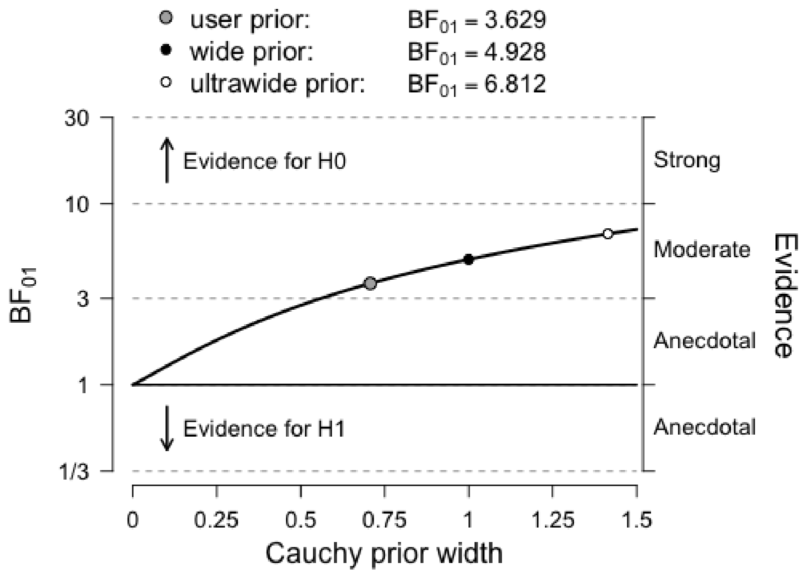


Figure 7.5: Bayes factor robustness check plot: Study Three

7.4 Discussion

This study aimed to investigate the nature of the influence that mathematical study had on individual differences in spatial reasoning. A group of undergraduate engineering students were primed on mathematical reasoning before completing a measure of spatial reasoning. Marginal correlations were found between spatial reasoning skills and mathematical ability, as measured by past academic achievement, which confirms that the participants that took part in this study possessed similar links between these two constructs as have been found in much of the literature discussed. However, no evidence was found for a priming effect of mathematical reasoning on spatial skill; inducing a mathematical thinking frame of mind did not increase the participants' ability to solve spatial problems more efficiently. The results of this study suggest that the differences in spatial reasoning observed between mathematicians and non-mathematicians in Study One and Two is not due to a mathematical 'frame of mind', created by recent interaction with mathematical content. Instead, the fact that mathematicians perform better on spatial tasks must be due to a more embedded cognitive difference between those that chose to study advanced mathematics, and those that do not.

An interesting finding of the current study was that there was no advantage to the students whatsoever of being primed with mathematical reasoning and, in fact, although not significant, the control group performed better in the spatial reasoning task. There is a small amount of literature that suggests that priming undergraduates with arithmetic problems does, in fact, reduce their performance on an algebra task (McNeil et al., 2010). Although the priming and outcome tasks of the current study and those employed by McNeil et al. were designed to evoke and measure difference cognitive aspects, the comparison is noteworthy due to its counterintuitive nature. McNeil et al. hypothesised that the solving of arithmetic problems hindered the solving of algebraic equations because it primed the participants to think in an operational way, which was not advantageous. In the current study, the mathematical priming task was designed to encourage the participants to think in a way that required well thought out mathematical reasoning, with the consideration that this would have the potential to transfer to spatial reasoning. The fact that this was not the case supports the argument that the transfer of skills from practice in one context to a non-identical task is not possible (Thorndike and Woodworth, 1901). The following chapter presents the overall conclusions of the three studies that make up this thesis.

Chapter 8

Overall conclusions

It is a widely held belief that studying mathematics poses some benefit to the students that engage with the subject, more than the learning of the subject content itself. The fact that very little evidence exists to support this theory of formal discipline for advanced mathematical study has not deterred it from being used, in part, to justify expansions to the mathematics curriculum, with mathematics now counting as double weighted in school accountability measures (Department for Education, 2016b), and students being required to continue mathematics study to the age of 18 if they have not achieved what the government defines as a good pass (Department for Education, 2014b). Employers and universities also mirror this emphasis, requiring school-leavers to hold a certain level of mathematics qualification, even when the job or further education course will require very little mathematics (Dudley, 2010). This thesis set out to add to the literature of the formal discipline value of the study of advanced mathematics, and therefore to establish in what ways the increased focus on mathematics in the education system is beneficial to the students who study the courses.

Recent literature, based on a large amount of longitudinal data, has determined that an individual's success in a STEM career can be predicted by their spatial skills at a younger age, over and above the more commonly measured mathematical and verbal skills. The fostering of spatial skills through education has therefore become a central focus of much educational research (Wai et al., 2009, 2010). Studies One and Two of this thesis explored the potential that advanced mathematical study had on the development of spatial skills. The literature discussed in Chapters 1 and 2 established the possibility of some transfer of skills from the study of academic subjects to more general cognitive abilities (Blade and Watson, 1955; Inglis and Simpson, 2007; Lehman et al., 1988; Lehman and Nisbett, 1990; Sorby, 1999), that spatial skills were closely linked to mathematics achievement (Göbel et al., 2001; Hubber, 2016; Newcombe et al., 2015), and have the potential to be trained (Bruce and Hawes, 2015; Terlecki and Newcombe, 2008; Uttal et al., 2013a). A combination of these areas of research literature suggests the plausibility of training spatial skills through the study of advanced mathematics. However, past research directly associated with investigating the formal discipline value of school subjects has found no significant effect of mathematics (Thorndike, 1924a,b; Wesman, 1945). More recently, some suggestion that mathematical study has a developmental effect on the way in which students approach some reasoning problems has been found (Attridge and Inglis, 2013). Study One and Two of this thesis aimed to provide evidence to further knowledge of the way in which advanced mathematical study and development of spatial skills were interrelated. Study Three further investigated the relationship between mathematics and spatial reasoning.

8.1 Overview of findings and implications

The first two studies of this thesis measured a number of spatial skill constructs of students that were either pre- or post-advanced study. For all of the spatial measures, the mathematicians performed better than the non-mathematicians, supporting much of the literature. Study One was susceptible to a number of features that made the results difficult to interpret, and conclusions about whether or not there were developmental differences between the groups were not certain. There were larger differences of performance on the spatial tasks between the mathematicians and non-mathematicians for the post-A level group than the pre-A level group, but none of the interactions between group and education level were significant. A further Bayesian analysis found that, for one of the spatial measures (the matrix reasoning task), the ANOVA model that included an interaction was the most likely, but only by a small amount compared with a model with only the main effects of group and education level. Study Two provided a much more clear picture of the situation: mathematicians performed better on the spatial task at Time 1 (before a year's worth of advanced mathematical study), but did not improve any more than students that studied other academic subjects by Time 2. The differences in the gains between the mathematics and non-mathematics groups were almost nonexistent, with a calculated Bayes factor revealing that the data represented strong evidence for the null hypothesis. Study Three tested the possibility that the higher levels of spatial skills displayed by the mathematicians was due to an effect of them having a temporarily altered state of mind due to being immersed in mathematical activities on a day-to-day basis. This hypothesis was not supported, suggesting that the differences between mathematicians and non-mathematicians were more deeply embedded cognitively.

The combination of results from Studies One and Two presented clear evidence that mathematicians possess higher levels of spatial skills than non-mathematicians, but that this advantage is not due to the study of advanced mathematics itself, and Study Three confirmed that the differences were not because of some 'frame of mind' effect. These findings do not support the formal discipline theory for which a number of recent studies have claimed, e.g. Attridge and Inglis (2013); Inglis and Simpson (2007); Lehman and Nisbett (1990). The most stark difference between these studies and the studies of this thesis is the type of reasoning that was the focus. An explanation for this is that the study of post-compulsory advanced mathematics does have some formal discipline value for certain types of reasoning behaviour, but not all. What is important for educationalists to establish is which of these reasoning types is most beneficial to which students. Students that wish to pursue careers in STEM will be advantaged by developing high levels of spatial skill (Wai et al., 2009), but an advanced level mathematics course might not be the most effective way of doing this. A course of graphical sketching, however, may be more effective, as a number of studies have shown this to improve 3D visualisation skills (Sorby, 1999, 2009), and the fact that both the mathematicians and the non-mathematicians in Study Two did improve their scores on the spatial reasoning task over the period of one year does suggest some malleability of this construct. However, if students wished to improve their statistical and methodological reasoning behaviour, they would be best to enrol on a psychology course (Lehman et al., 1988), and on an advanced mathematics course to enhance their logical reasoning behaviour (Attridge and Inglis, 2013).

The reason that the mathematicians displayed higher levels of spatial skills throughout the studies of this thesis must be due to some filtering effect which results in the individuals that possess these skills to be more inclined to choose to study advanced mathematics. At what point this cognitive difference occurred is open to debate. Compulsory education in the UK requires all students to cover the same syllabus of mathematics up to GCSE[1], and therefore any effect

[1] Although many mathematics qualification are split into different assessment tiers, which will determine some of the content that the students are taught, all participants involved in the studies of this thesis sat the higher tier examination.

that their mathematics education had had on reasoning, spatial or otherwise, should be equal across all students at this point. The fact that spatial reasoning skills did differ between mathematicians and non-mathematicians at Time 1 for Study Two, and marginally at pre-A level in Study One, contrasts with other studies of formal discipline that find no difference in the reasoning skills of participants prior to intervention (Attridge and Inglis, 2013; Inglis and Simpson, 2009; Lehman and Nisbett, 1990). This suggests that spatial reasoning develops in a different way, or at a different rate, to other types of reasoning. Possibly, spatial reasoning is completely unaffected by external factors, and is entirely innate, although this exclusively genetic explanation does not hold much weight in the literature (Tosto et al., 2014), and it would seem implausible that an individual's A level choices could be predicted from birth. However, another cognitive construct, such as executive functioning, or attention, could mediate both spatial skills and mathematics achievement from a very early age, and this is why the two are so interlinked. This explanation is, in part, supported by studies that have found performance on mathematics tasks and spatial tasks to be mediated by working memory (Kaufman, 2007; Tolar et al., 2009).

Another possibility is that spatial reasoning development depends on other experiences that nurture these skills at an earlier point in life, such as construction play[2] (Nath and Szucs, 2014), or playing video games (Feng et al., 2007). The filtering of individuals with higher spatial skills into advanced mathematical study might then be due to some motivation that also derives from these activities, for example a thought process such as "I spend a lot of time playing with Lego" ... "I would like to have a career in architecture" ... "I need to excel in mathematics to achieve that". On the other hand, there may be some other factor, such as the interests and hobbies of their parents, that influences both their amount of time spent engaging with activities that train spatial skills, and their interest in mathematics. A third possibility is that mathematical study does, in fact, have a formal discipline effect on the development of spatial skills, but prior to the end of a student's compulsory education. This theory requires the assumption that the individuals that choose to study advanced mathematics have had a different experience of mathematics to those that do not choose the subject, prior to making their post-compulsory education choices. This assumption is not inconceivable at all. Many students are likely to engage with extra-curricular activities, such as puzzles, and strategic games, that involve some aspects of mathematics. In fact, these activities might be more likely than lesson activities to be of the kind that would encourage the development rea-

[2]For example playing with Lego building blocks

soning skills. In addition, some students may have, for some reason, engaged more meaningfully with the content of the compulsory mathematics syllabus, maybe because they related to the teacher of the subject, or because they were taught in a class of other students that enjoyed the subject. A more engaged student will be more likely to develop a deeper understanding of mathematical topics and the links between them, as opposed to purely procedural learning. This type of 'arousing mindfulness' learning has more potential for far transfer to other skills, such as spatial reasoning (Perkins and Salomon, 1992). All of these external factors that have the potential to affect a student's engagement in, exposure to, and motivation for mathematical and spatial activities are inescapably intertwined, and what seems likely is that the effects of all act in a loop of influences, in conjunction with some more innate abilities, with the result of filtering individuals with higher spatial skills into advanced mathematics education.

Ultimately, the early identification, and fostering, of spatial skills through education would result in more students continuing in to STEM careers (Wai et al., 2009; Uttal et al., 2013b), something that governments are keen to achieve. The recent verification of the importance of spatial skills in STEM has motivated some changes to governments' education policies, for example in Canada where mathematics education is becoming more focused on the promotion of spatial skills (Ministry of Education, 2014). Since the motivation for this thesis was established, the UK government have reformed compulsory education, making a number of changes to the mathematics GCSE which include a greater focus on problem-solving skills to foster higher-level thinking and understanding (Department for Education, 2016a). It is possible that these changes may lead to a more effective nurturing of the type of 'high road' skills (Perkins and Salomon, 1992) that have the potential to be transferred to other reasoning skills that can be of benefit to a student in later life.

8.2 Future work

This thesis has found no evidence to support the theory of formal discipline in terms of advanced mathematical study and the development of spatial skills. Further investigation into why mathematicians display higher levels of spatial skills prior to advanced study should focus on some of the possible explanations that were put forward in the previous discussion section. Researching the effects of mathematics education prior to the end of compulsory education is not an easy task, as there does not exist an obvious comparison group of students who will not have been exposed to the same amount of mathematics education, unless this is confounded with other variables, such as absence from school, which could happen for a number of reasons, and would be expected to affect any measurable dependent variable indirectly. One possible solution to this would be a cross-national study of students in England, and students exposed to a curriculum that started mathematics teaching at an earlier or later age, allowing students of the same developmental age to be compared. Another version of this design would be to compare children of very similar ages within the same education system, but in different academic year groups, some having experienced an additional year of mathematics teaching. Studies of this kind would help to shed light on whether or not the study of mathematics has some formal discipline value at an earlier stage of development. Investigation of which particular aspects of the mathematics curriculum might promote the development of spatial skills would be an interesting area of research. A possible longitudinal measurement of spatial skills of classes of students that are studying different topics at different times of the academic year could provide some data for this.

Chapter 9

Appendices

9.1 Appendix A: Study One

9.1.1 Study One paper task booklet (excluding the RPM items)

Task Booklet (1)

This booklet contains four tasks for you to complete.

Each of the tasks is timed so please follow the instructions carefully.

Complete the following in block capitals:

First name ……………………..

Surname ……………………..

Date of Birth ……………………..

Sex M/F

A-level subjects studied and grades (indicate AS):

…………………………….. ……..

…………………………….. ……..

…………………………….. ……..

…………………………….. ……..

…………………………….. ……..

Degree course: ………………………………………………………………………………..

GCSE grades

Maths ………..

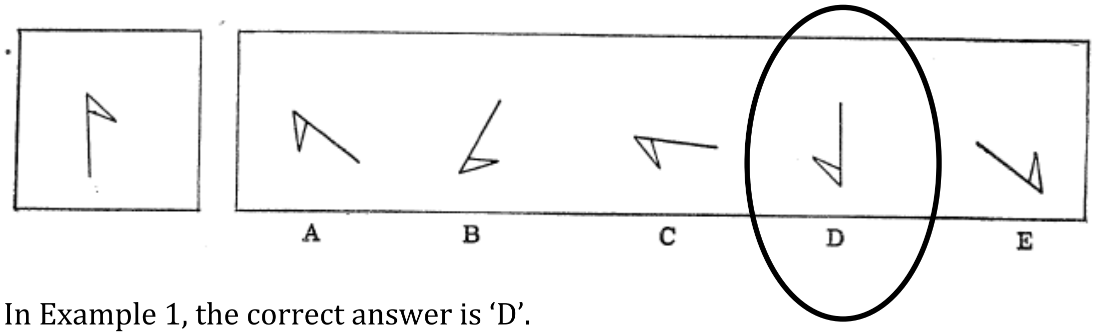
English ……….
.
Science ………..

Please wait before turning the next page

2d Rotation (1)

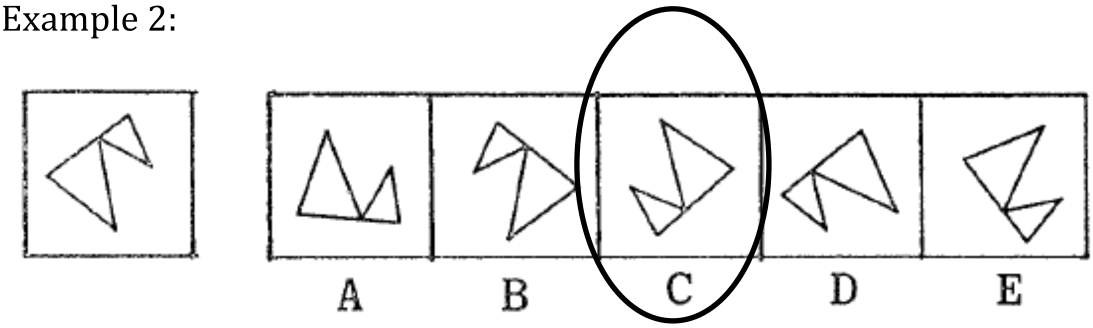
This task requires you to choose the drawing from the right that exactly matches the drawing on the left if you turn it around.

Example 1:



In Example 1, the correct answer is ‘D’.

Example 2:



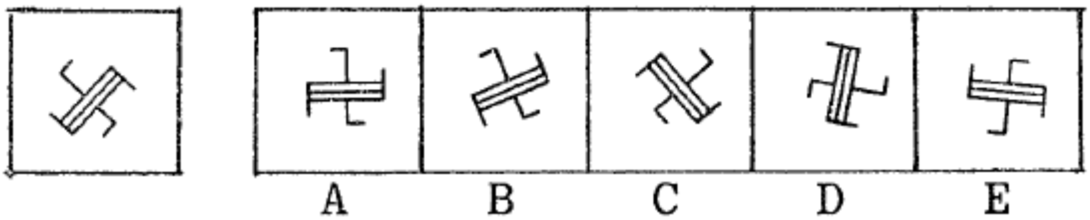
In Example 2, the correct answer is ‘C’.

If you have any questions, ask me now.

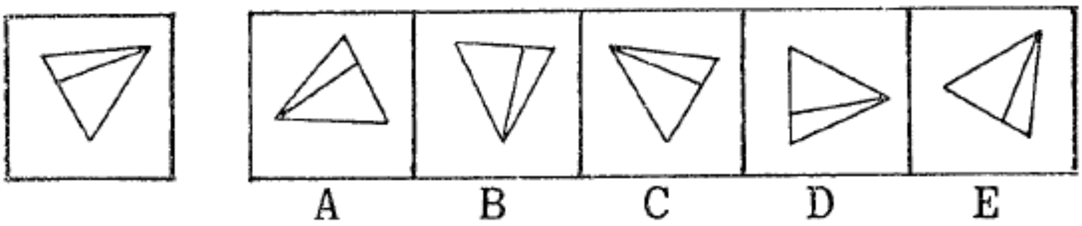
Please wait to turn over the page.

You have 5 minutes to complete this section. There are 10 questions

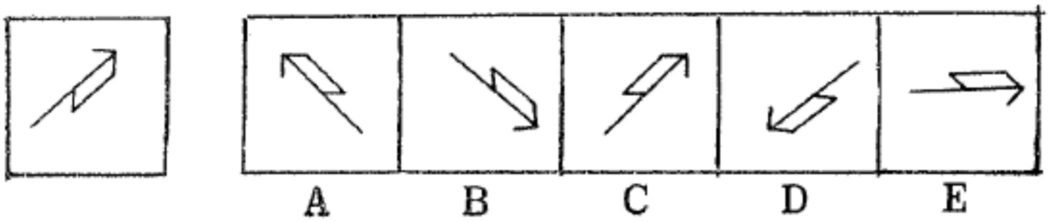
1) Circle the picture that exactly matches the one on the left.



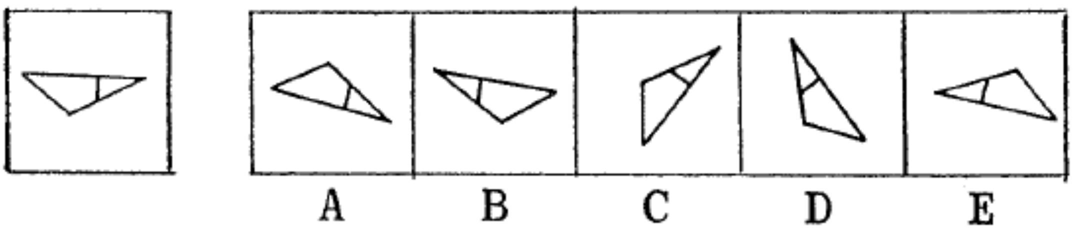
2) Circle the picture that exactly matches the one on the left.



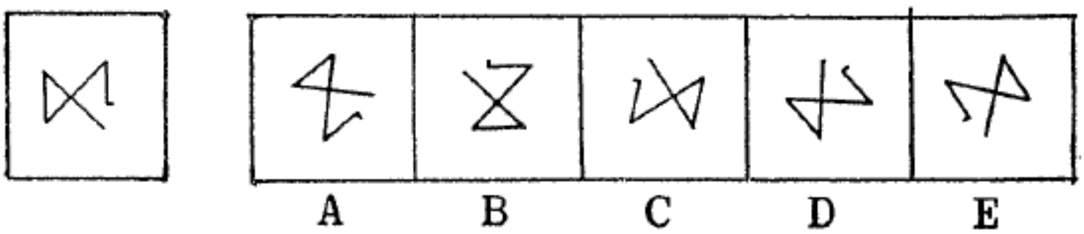
3) Circle the picture that exactly matches the one on the left.



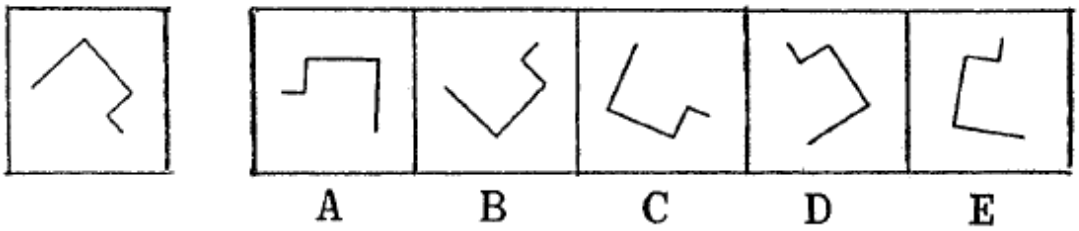
4) Circle the picture that exactly matches the one on the left.



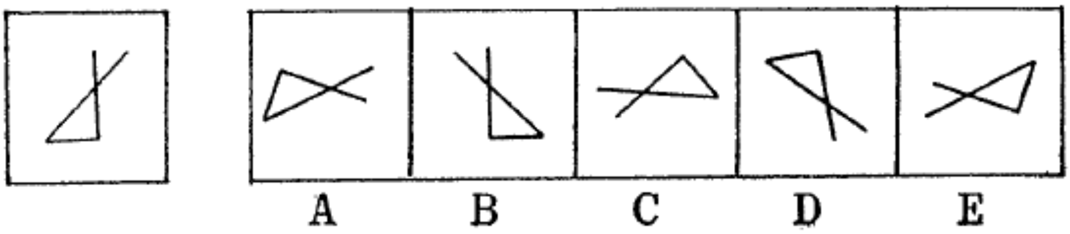
5) Circle the picture that exactly matches the one on the left.



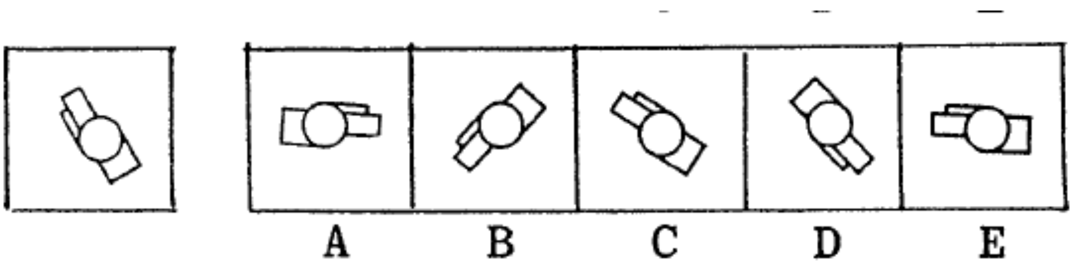
6) Circle the picture that exactly matches the one on the left.



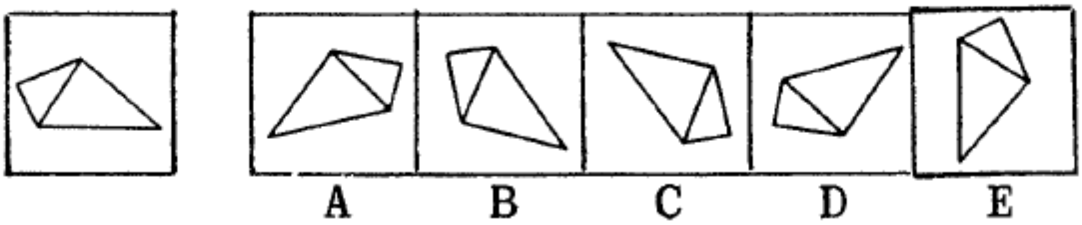
7) Circle the picture that exactly matches the one on the left.



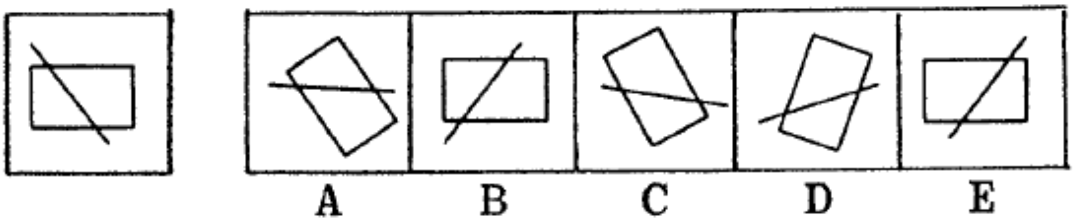
8) Circle the picture that exactly matches the one on the left.



9) Circle the picture that exactly matches the one on the left.



10) Circle the picture that exactly matches the one on the left.

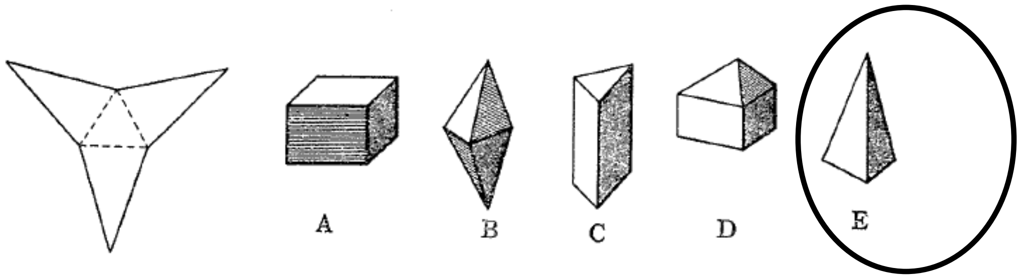


That is the end of the 2D-rotation task **– please wait to turn the page**

3D Visualisation (1)

This task requires you to choose the object on the right that can be made using the net on the left.

Example 1:



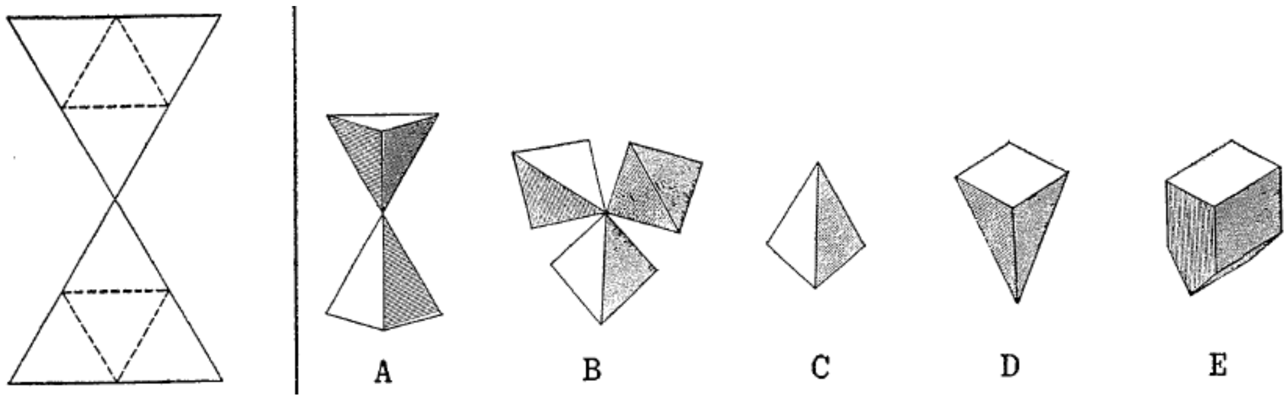
In Example 1, 'E' is the correct answer

If you have any questions, ask me now.

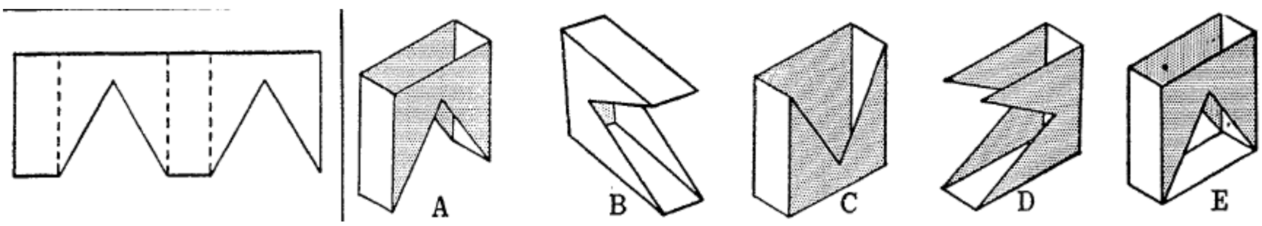
Please wait to turn over the page.

You have 5 minutes to complete this section. There are 10 questions

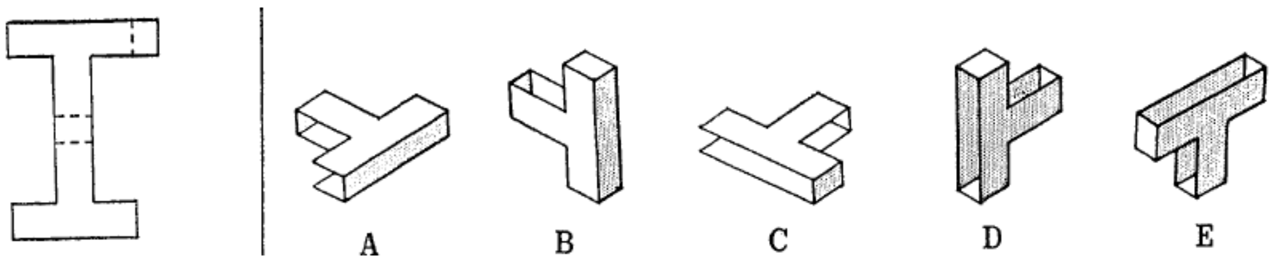
1) Circle the object that can be made with the net



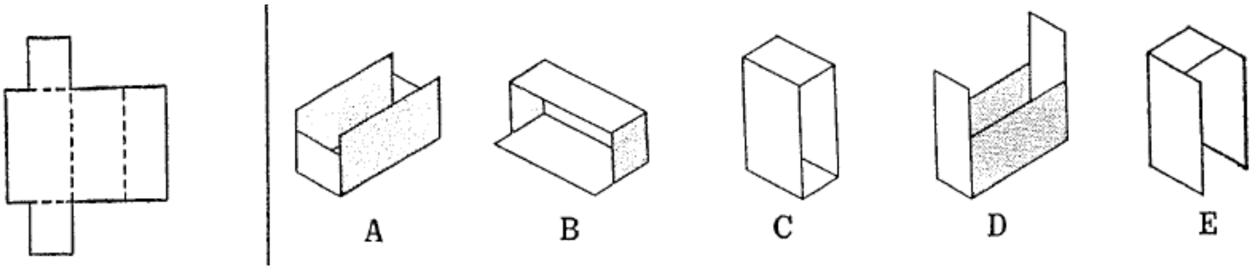
2) Circle the object that can be made with the net



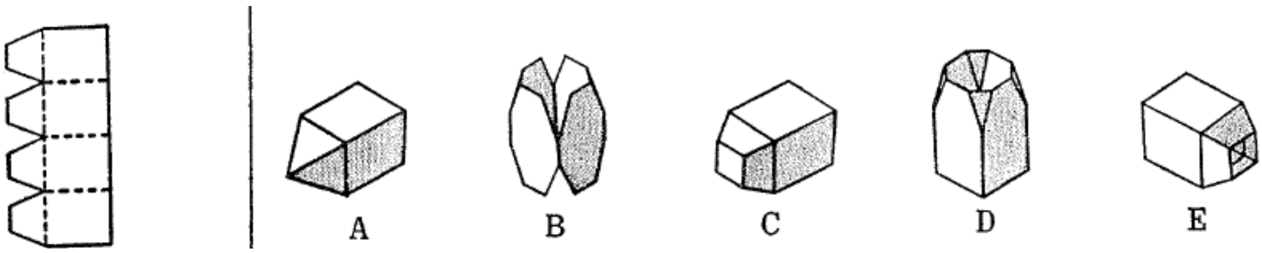
3) Circle the object that can be made with the net



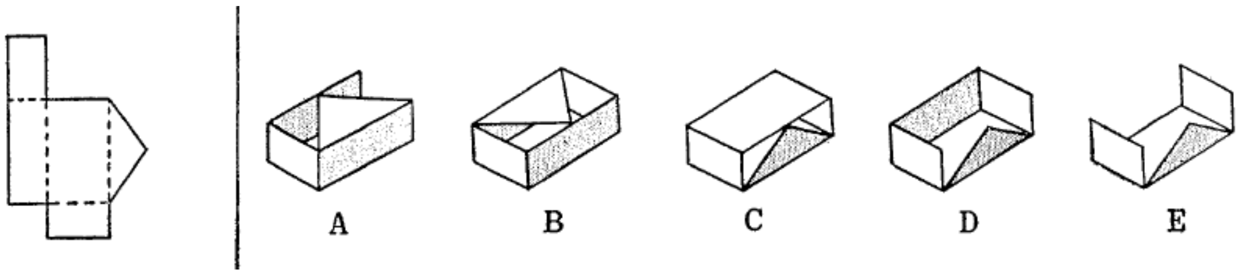
4) Circle the object that can be made with the net



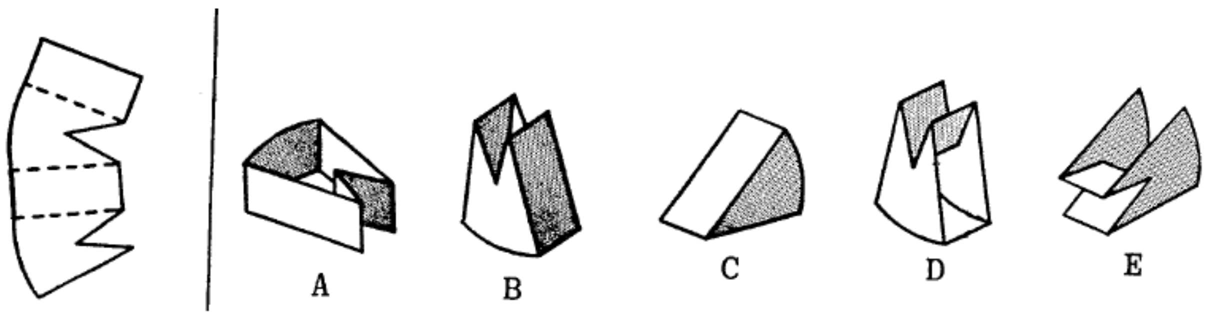
5) Circle the object that can be made with the net



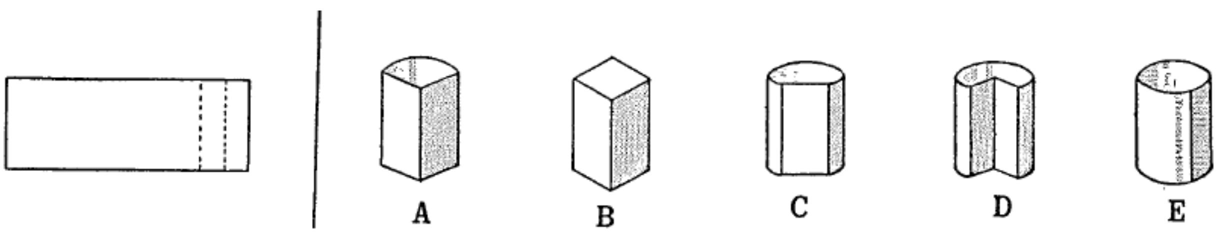
6) Circle the object that can be made with the net



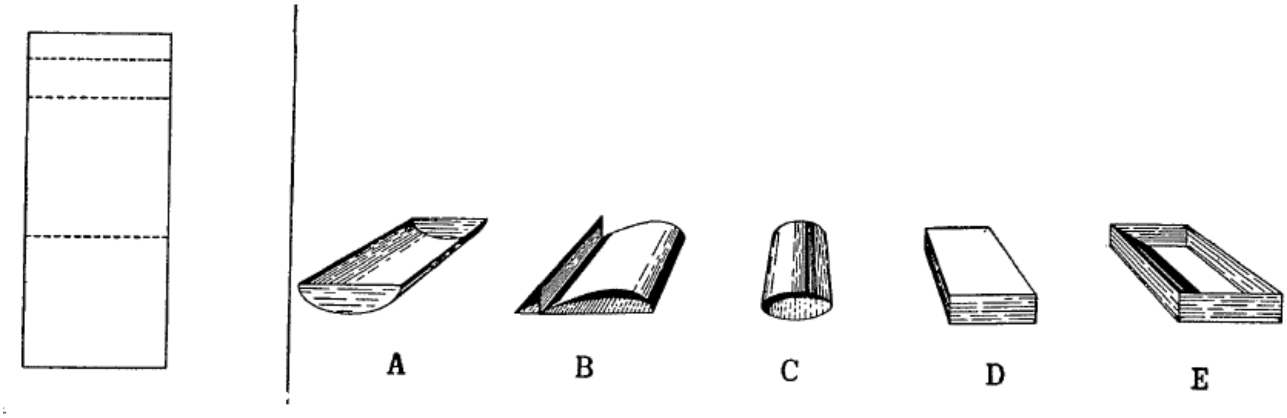
7) Circle the object that can be made with the net



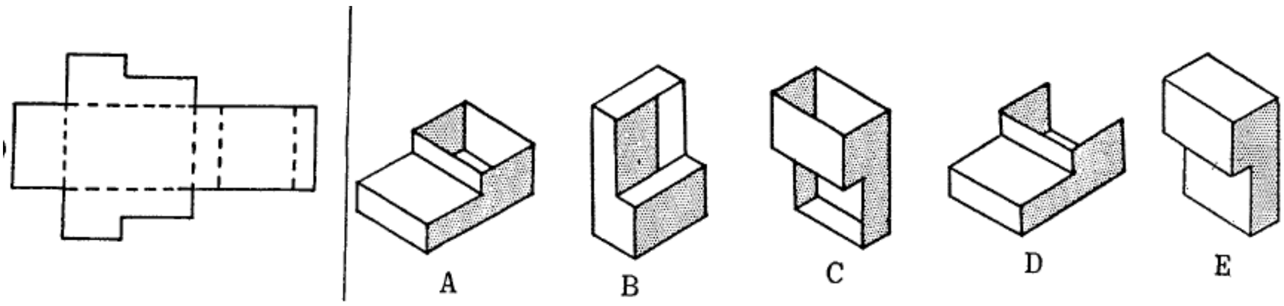
8) Circle the object that can be made with the net



9) Circle the object that can be made with the net



10) Circle the object that can be made with the net



That is the end of the 3D visualisation task –

please wait to turn the page

Maths Fluency (1)

This task requires you to answer some simple maths questions. You will have 3 minutes to answer as many as you can.

Example 1:

$$\begin{array}{r} 2 \\ +\ 3 \\ \hline 5 \end{array}$$

The answer to Example 1 is '5'.

If you have any questions, ask me now.

Please wait before turning the page.

You have 3 minutes for complete as many of these as you can. Please start at the top left and go from left to right. There are two pages altogether.

1 − 1	0 + 3	2 + 2	4 − 2	2 + 1	3 − 3	0 + 0	3 − 0	2 − 1	2 + 4
5 + 0	3 − 1	1 + 6	4 + 4	5 − 0	1 + 1	6 − 1	3 + 5	4 − 1	5 − 2
3 − 2	5 + 1	6 − 3	2 − 2	7 + 1	4 − 4	1 + 8	4 − 3	7 + 2	4 + 1
2 + 5	8 − 1	5 − 4	3 + 3	10 − 2	3 + 6	7 − 2	2 + 8	3 + 1	9 −4
6 − 2	4 + 6	9 + 3	8 − 6	7 + 5	10 − 10	2 + 6	5 − 3	6 − 6	3 + 4
5 + 5	8 − 3	5 − 1	8 + 0	7 − 4	1 + 9	10 − 6	8 + 4	6 + 8	9 − 9
1 × 1	4 + 5	7 + 7	2 × 3	10 − 5	1 × 2	3 + 2	8 − 8	9 + 5	5 × 1
3 + 7	1 × 4	9 − 2	9 + 8	3 × 3	6 − 5	2 × 2	9 − 3	3 × 1	6 + 6

Start at the top left and work from left to right.

1	8	2	7	8	7	2	9	5	10
× 0	+ 7	× 5	− 7	× 1	+ 4	× 6	− 6	× 3	− 1
1	5	8	4	7	8	3	7	9	4
× 9	− 5	× 5	+ 0	× 1	− 2	× 9	− 0	+ 9	× 2
3	7	7	4	8	4	9	3	6	0
+ 8	× 9	− 5	× 4	+ 8	× 7	− 8	× 4	− 4	× 6
9	3	9	8	10	6	10	6	9	2
− 1	× 7	+ 4	× 3	− 9	× 5	− 3	+ 7	× 9	× 8
4	10	5	7	8	9	2	7	10	7
− 0	− 4	× 9	× 7	− 5	× 6	+ 9	× 6	− 8	× 8
9	4	10	4	9	2	6	8	5	8
+ 6	× 8	− 7	× 6	− 7	× 9	× 6	− 4	× 7	+ 5
5	7	6	4	9	7	5	8	0	7
× 5	− 1	× 1	× 0	× 8	− 6	× 4	× 6	× 3	− 3
2	6	8	9	7	6	0	8	4	10
× 7	+ 5	× 8	− 5	+ 9	× 3	× 8	− 7	× 9	− 0

That is the end of the maths fluency task - **please wait to turn the page**

9.2 Appendix B: Study Two

9.2.1 Study Two paper task booklet (excluding the RPM items)

Reasoning and Critical Thinking Task (1)

Information and Informed Consent

This study is about reasoning skills of students studying post-16 subjects. It will involve you answering questions that are designed to measure your reasoning and critical thinking once now, and again at the end of the school year.

If at any point, you wish to withdraw from the study, you may do.
If at any point in the future, you wish to withdraw your results from the study, you can contact me by email on s.m.humphries@lboro.ac.uk
Thank you for taking part.

I agree to take part in this study of reasoning and critical thinking

Signed ………………………………… (participant)

Reasoning and Critical Thinking (1)

---

*Please fill out all of the following information **in block capitals**:*

1) First name: …………………Surname …………………… Student Number:…………….

   College or School: ………………………………………….. Form/Tutor group ……………..

   Date of Birth: ……/……/…… Sex: male/female

   Native Language: …………………………..

2) Please list **all** of the subjects that you are taking this year and at which level (please indicate if these subjects are part of the International Baccalaureate programme) e.g. *Geography AS-level; Mathematics IB*

   ____

   ____

   ____

   ____

   ____

---

3) What were your grades for:

   Maths GCSE ……………………..

   English GCSE (Lit and Lang) …………………….

   Science GCSE ……………………..

   (If you were awarded separate science grades, give your highest grade)

4) Do you have or suspect having any of the following?

	Suspect	**Diagnosed**	**No**
Dyslexia			
Dyscalculia			
Attention deficit disorder			
Neurological disorders			
Learning difficulties			

9.3 Appendix C: Study Three

9.3.1 Study Three mathematics priming task booklet (excluding the RPM items)

Reasoning Study

This is a study about reasoning skills of undergraduate engineers. The purpose of this study is to look at the relationship between different types of reasoning.

The data collected from this task will be stored anonymously and will be used by students at Loughborough University as part of their Statistical Methods module. If you wish to withdraw your data at any point, you are free to and should do so using the following e-mail address: **s.m.humphries@lboro.ac.uk**.

If you do not wish to participate in the study at all, please let the research know now.

On the following pages you will be asked to complete two reasoning tasks. Please complete them to the best of your ability and try not to leave any questions blank. If you don't know the answer to a question, please guess.

If you are happy to continue with the study, please fill out all of the following information.

Date of birth: ____________________________

Gender: **male/female**

First language: ____________________________

Degree course ____________________________ **1st year average exam score:** ___________

A-level results (or alternative qualifications):

____ ___________
____ ___________
____ ___________
____ ___________
____ ___________

Signature: ____

Please do not turn over the page until instructed to do so.

Version 1.1

Mathematical Reasoning

1.1

On the following pages, you will see a number of mathematical statements. Your task is to decide whether each statement is always, sometimes, or never true and to justify your reasoning.

If you consider a statement to be **always true**, then try to explain how you know it is always true. If you think a statement is **sometimes true**, then try to describe all the cases in which it is true and all the cases in which it is false. If you think a statement is **never true**, then again explain how you can be sure.

There are 16 items in the Mathematical Reasoning section. Tick the answer that you believe to be correct. You should aim to spend approximately 30 mins on this section and then move on to the Abstract Reasoning section.

1. If you add n consecutive numbers together the result is divisible by n.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

07

---

2. $3x^2 = (3x)^2$

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

03

---

3. Pentagons have fewer right angles than triangles.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

09

---

4. It is possible to divide a triangle into 6 equal areas by folding.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

01

---

5. When you add two numbers, you get the same answer as when you multiply them.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

06

6. The square of a number is greater than that number.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

04

7. You see tyre tracks of a bicycle in the mud. You can deduce from these which direction the bicycle was travelling in.

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

05

---

8. $(x-2)^2 = x^2 - 4x$

   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

12

---

9. Quadrilaterals tesselate.
   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

08

---

10. A shape with a finite area has a finite perimeter.
   - Always true
   - Sometimes true
   - Never true

   Please explain your answer:

010

---

11. If you double a number, you get an even number.

- Always true
- Sometimes true
- Never true

Please explain your answer:

14

---

12. If you are told the values for the perimeter and the area, it is possible to draw the shape.

- Always true
- Sometimes true
- Never true

Please explain your answer:

13

---

13. $\sqrt{ab} > \frac{a+b}{2}$

- Always true
- Sometimes true
- Never true

Please explain your answer:

02

14. If you multiply two odd numbers, you get an odd number.

- Always true
- Sometimes true
- Never true

Please explain your answer:

15

15. The more digits a number has, the larger is its value.

    - Always true
    - Sometimes true
    - Never true

    Please explain your answer:

11

---

16. When you cut a piece off a shape, you reduce the area and the perimeter.

    - Always true
    - Sometimes true
    - Never true

    Please explain your answer:

16

9.3.2 Study Three control task booklet

Reasoning Study

This is a study about reasoning skills of undergraduate engineers. The purpose of this study is to look at the relationship between different types of reasoning.

The data collected from this task will be stored anonymously and will be used by students at Loughborough University as part of their Statistical Methods module. If you wish to withdraw your data at any point, you are free to do so by contacting the following e-mail address: **s.m.humphries@lboro.ac.uk**.

On the following pages you will be asked to complete two reasoning tasks. Please complete them to the best of your ability and try not to leave any questions blank. If you don't know the answer to a question, please guess.

If you are happy to continue with the study, please fill out all of the following information.

Student number: ___________________________

Gender: **male/female**

First language: ___________________________

January exam score average: ___________

Signature: ____

Please do not turn over the page until instructed to do so.

Version 1.1

Grammatical Reasoning: Part One

1.1

On the following pages, you will see a number of statements with a word or phrase missing. Your task is to choose the grammatically correct word or phrase to fill in the blank from a list of options. Tick the circle next to your answer.

There are 40 short items in Part One and 6 longer items in Part Two. You should aim to spend approximately 30 mins on these Grammatical Reasoning sections and then move onto the Abstract Reasoning section.

1. Tick the answer that you think is grammatically correct.

   ___________ shall I say is calling?

   - ◯ Who
   - ◯ Whom
   - ◯ Whose
   - ◯ Who's

   01

2. Tick the answer that you think is grammatically correct.

   The boy ___________ thew the ball was blond.

   - ◯ Himself
   - ◯ That
   - ◯ Which
   - ◯ Who

   02

3. Tick the answer that you think is grammatically correct.

   The cat has ___________ the canary.

   - ◯ Eat
   - ◯ Eaten
   - ◯ Ate
   - ◯ Eated

   03

4. Tick the answer that you think is grammatically correct.

   Chairs __________ don't have cushions are uncomfortable to sit on.

   - ◯ That
   - ◯ Which
   - ◯ Whose
   - ◯ Where

   04

---

5. Tick the answer that you think is grammatically correct.

   Uncle David is really __________ man.

   - ◯ An old sweet
   - ◯ A sweet, old
   - ◯ A sweet old

   05

---

6. Tick the answer that you think is grammatically correct.

   The bus is usually on time. It __________ to be here any time now.

   - ◯ Might
   - ◯ Has
   - ◯ Ought

   06

---

7. Tick the answer that you think is grammatically correct.

   It's way past my bedtime and I'm really tired. I __________ go to bed.

   - ◯ Should
   - ◯ Ought
   - ◯ Could

   07

---

8. Tick the answer that you think is grammatically correct.

   ___________ are no excuses this time Madison!

   ◯ There
   ◯ Their
   ◯ They're

   08

---

9. Tick the answer that you think is grammatically correct.

   The climate of New Zealand can be a pleasure for you if ___________ don't mind a little rain.

   ◯ We
   ◯ He
   ◯ You

   09

---

10. Tick the answer that you think is grammatically correct.

    Everyone was home for the holidays. What could make for ___________ Christmas than that?

    ◯ A merryer
    ◯ The merriest
    ◯ A merrier

    10

---

11. Tick the answer that you think is grammatically correct.

    If it ___________, I would stay home and study.

    ◯ Rains
    ◯ Will rain
    ◯ Rained
    ◯ Both a and c

    11

---

12. Tick the answer that you think is grammatically correct.

I am terribly afraid of heights. If I __________ that tall tree in the front yard, I would die.

- ◯ Climbed
- ◯ Climb
- ◯ Both a and c

12

---

13. Tick the answer that you think is grammatically correct.

He was not thinking well __________ that occasion.

- ◯ At
- ◯ In
- ◯ On
- ◯ When

13

---

14. Tick the answer that you think is grammatically correct.

The other boys or Dave __________ to blame.

- ◯ Is
- ◯ Are
- ◯ Were
- ◯ Will

14

---

15. Tick the answer that you think is grammatically correct.

Bill graduated from college last spring. If he __________, I think his mother would have told him to leave the house.

- ◯ Was not graduated
- ◯ Is graduating
- ◯ Had not graduated

15

---

16. Tick the answer that you think is grammatically correct.

They grew up in __________ house in Mexico City.

- ◯ A comfortable, little
- ◯ A little, comfortable
- ◯ A comfortable little

16

---

17. Tick the answer that you think is grammatically correct.

Professor Smith, we've finished our work for today. __________ we leave now?

- ◯ May
- ◯ Can
- ◯ Must

17

---

18. Tick the answer that you think is grammatically correct.

The child responded to his mother's demands __________ throwing a tantrum.

- ◯ With
- ◯ By
- ◯ From

18

---

19. Tick the answer that you think is grammatically correct.

Choose the proper sentence structure:

- ◯ If you look through Angelo's telescope, you can see Saturn's ring.
- ◯ Look through Angelo's telescope, you can see Saturn's ring.
- ◯ You can see Saturn's ring, look through Angelo's telescope.

19

---

20. Tick the answer that you think is grammatically correct.

It's been snowing __________ Christmas morning.

- ◯ Since
- ◯ For
- ◯ Until

20

---

21. Tick the answer that you think is grammatically correct.

My cold is definitely ___________ this morning.

- ◯ Worse
- ◯ Worst
- ◯ Worser

21

---

22. Tick the answer that you think is grammatically correct.

I knew what model car it was, but I wasn't sure about ___________ colour.

- ◯ Its
- ◯ It's

22

---

23. Tick the answer that you think is grammatically correct.

Half the students ___________ against the tuition strike.

- ◯ Is
- ◯ Are

23

---

24. Tick the answer that you think is grammatically correct.

I'll be ready to leave ___________ about twenty minutes.

- ◯ In
- ◯ On
- ◯ At

24

---

25. Tick the answer that you think is grammatically correct.

Those are probably the ___________ curtains in the store.

- ◯ Fanciest
- ◯ Most fanciest

25

---

26. Tick the answer that you think is grammatically correct.

I'm really lost, ___________ showing me how to get out of here?

- ◯ Would you mind
- ◯ Would you be
- ◯ Must you be

26

---

27. Tick the answer that you think is grammatically correct.

It is very important that all employees ___________ in their proper uniforms before 6:30am.

- ◯ Are dressed
- ◯ Will be dressed
- ◯ Be dressed

27

---

28. Tick the answer that you think is grammatically correct.

My best friend lives ___________ Boretz Road.

- ◯ In
- ◯ On
- ◯ At

28

---

29. Tick the answer that you think is grammatically correct.

This will just be between you and ___________.

- ◯ Myself
- ◯ I
- ◯ Me
- ◯ Mine

29

---

30. Choose the correct punctuation.

- ◯ Our solar system has nine major planets, only one is known to have intelligent life.
- ◯ Our solar system has nine major planets only one is known to have intelligent life.
- ◯ Our solar system has nine major planets; only one is known to have intelligent life.

30

---

31. Tick the answer that you think is grammatically correct.

Some of the votes ___________ to have been miscounted

◯ Seem
◯ Seems 31

---

32. Tick the answer that you think is grammatically correct.

It seems to me that we've had ___________ assignments in English this term.

◯ Much
◯ Many 32

---

33. Tick the answer that you think is grammatically correct.

That ice is dangerously thin now. You ___________ go ice-skating today.

◯ Mustn't
◯ Might not
◯ Would mind not to 33

---

34. Tick the answer that you think is grammatically correct.

I wish I ___________ better today.

◯ Feel
◯ Felt 34

---

35. Tick the answer that you think is grammatically correct.

No one has offered to let us use ___________ home for the department meeting.

◯ Her
◯ Their
◯ His or her 35

---

36. Tick the answer that you think is grammatically correct.

Charles and ___________ are attending the conference.

- ◯ Me
- ◯ I
- ◯ Myself
- ◯ Mine

36

---

37. Tick the answer that you think is grammatically correct.

Mircosoft announced ___________ releasing a new product next week.

- ◯ It is
- ◯ They are
- ◯ Itself
- ◯ She is

37

---

38. Tick the answer that you think is grammatically correct.

Which of the following sentences contains a verb agreement error?

- ◯ These are a collection of valuable nineteenth-century manuscripts.
- ◯ The professor that teaches nineteenth-century manuscripts.
- ◯ The only dog that the buyers want are Dalmatians.
- ◯ All of the above
- ◯ None of the above

38

---

39. Tick the answer that you think is grammatically correct.

It I ever find my glasses, I think I'll have ___________ replaced.

- ◯ It
- ◯ Them

39

---

40. Tick the answer that you think is grammatically correct.

You seem to be having trouble there. ___________ I help you?

- ◯ Would
- ◯ Will
- ◯ Shall

40

Grammatical Reasoning: Part Two

1

This section will ask you to complete some questions concerning grammar. There are 6 items in this section.

1. Combine the following two sentences into one effective sentence containing only one independent clause:
   Chicago is a capital of Illinois ___________ it is the third most populated city in America.

01

2. Combine the following two sentences into one effective sentence containing only one independent clause:

   Some factories have been torn down ___________ they have been converted to artists' studios.

02

3. Rewrite the following sentence to achieve a more concise statement:

   At this point in time we can't ascertain the reason as to why the screen door was left open. 03

---

4. Rewrite the following sentence to achieve a more concise statement:

   My cousin who is employed as a nutritionists at the University of Florida, recommends the daily intake of mega doses of Vitamin C. 04

---

5. There are 10 errors in the following paragraph. Correct each error.

   In the article, Gavzer state that the traditional American families often includes “two parents, a father who works, and a mother who raise her two or three children at home”. This is also true for tradition Japanese families. Japanese men want their wives to stay at home and take care of their children while they are out working very hard to support their families. Therefore it can be said that families in America and Japan bases their beliefs of a tradition family on the same points. But I also found some differences. Americans seem to date more people before marriage than Japanese people does. Although an American experience many dates, this do not make it any easier to marry the right person. Everybody have a hard time picking the right person for a husband or wife. In Japan there is many networks that can arrange marriages for men and women. When the right person is found, the marriage follow.

05

---

6. There are five errors in the following paragraph. Correct each error.

   This is about an Indian family. The parents decided to come to America with the intention of getting jobs and giving their children a better education. Before they came to America they had sold most of their property in their country. They thought they can earn three of four times more money than what they were earning in India. When they first arrived in America, they don’t know anybody in the country. The family stayed in a hotel until they find a place to live. As soon as they move to an apartment, they started to apply for jobs that were related to their fields, but they didn’t succeed. At first, they were unsuccessful because they don’t speak English well and their degrees in engineering were not valid in the state they were living in. Their pride and self dignity were hurt and too many doors were closed to their success.

06

9.4 Appendix D: UK education system – explanation of terms

Key Stage 1	The first two years of official primary schooling (ages 5 to 7 years old).
Key Stage 2	The last four years of primary schooling (ages 7 to 11 years old).
Key Stage 3	The first three years of secondary education (ages 11 to 14 years old).
Key Stage 4	The last two years of secondary education (ages 14 to 16 years old). Most commonly, students sit GCSE examinations at the end of these years.
Key Stage 5	An optional extra two years of secondary education (usually ages 16 to 18 years old). Most commonly, students sit A level examinations at the end of these years.
GCSE	(General Certificate of Secondary Education) A regulated and internationally recognised qualification usually sat at the end of compulsory education. Offered in a wide range of subjects of which students often take around ten.
Tiering	The assessment of a number of GCSEs (including mathematics) are tiered with the foundation tier aimed at the lower end of the ability scale, and the higher at the most able.
A levels	(Advanced level General Certificate of Education) A regulated and internationally recognised qualification usually completed before university education. At the time of the data collection for this thesis, the A level was split into two parts: AS level completed in the first year, and A level in the second. Students usually sit AS/A levels in three or four subjects.

Table 9.1: Explanation of UK education levels and main examinations

References

Andersson, U. (2007). The contribution of working memory to children's mathematical word problem solving. *Applied Cognitive Psychology*, 21(9):1201–1216.

Angelakis, E., Stathopoulou, S., Frymiare, J. L., Green, D. L., Lubar, J. F., and Kounios, J. (2007). EEG neurofeedback: a brief overview and an example of peak alpha frequency training for cognitive enhancement in the elderly. *The Clinical Neuropsychologist*, 21(1):110–129.

Attridge, N. and Inglis, M. (2013). Advanced mathematical study and the development of conditional reasoning skills. *PloS one*, 8(7):1–8.

Baddeley, A. D. (1992). Working memory. *Science*, 255(5044):556–559.

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, 4(11):417–423.

Baddeley, A. D. and Hitch, G. (1974). Working memory. *Psychology of learning and motivation*, 8:47–89.

Bangerter, A. and Heath, C. (2004). The Mozart effect: Tracking the evolution of a scientific legend. *British Journal of Social Psychology*, 43:605–623.

Bargh, J. A., Chen, M., and Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype activation on action. *Journal of Personality and Social Psychology*, 71(2):230–244.

Bargh, J. A. and Pietromonaco, P. (1982). Automatic information processing and social perception: The influence of trait information presented outside of conscious awareness on impression formation. *Journal of Personality and Social Psychology*, 43(3):437–449.

Baron-Cohen, S. (2003). *The essential difference: The truth about the male and female brian.* Basic Books, New York.

Bartlett, T. (2013). Power of suggestion . *The Chronicle of Higher Education*.

BBC (2013). BBC News - Teenagers have to keep studying English and maths.

Benson, E. (2003). Intelligent intelligence testing. *Monitor on Psychology*, 34(2):48–58.

Binet, A. and Simon, T. (1916). *The development of intelligence in children: The Binet-Simon scale*. Williams & Wilkins Company.

Blade, M. F. and Watson, W. S. (1955). Increase in spatial visualization test scores during engineering study. *Psychological Monographs: General and Applied*, 69(12):1–13.

Boake, C. (2002). From the Binet–Simon to the Wechsler–Bellevue: Tracing the history of intelligence testing. *Journal of Clinical and Experimental Neuropsychology*, 24(3):383–405.

Boaler, J., Wiliam, D., and Brown, M. (2000). Students' experiences of ability grouping-disaffection, polarisation and the construction of failure. *British educational research journal*, 26(5):631–648.

Bors, D. A. and Vigneau, F. (2003). The effect of practice on Raven's Advanced Progressive Matrices. *Learning and Individual Differences*, 13(4):291–312.

Bouchard, Jr, T. J. (1998). Genetic and environmental influences on adult intelligence and special mental abilities. *Human Biology*, 70:257–279.

Brooks, E. (1883). *Readings in the History of Mathematics Education*. National Council of Teachers of Mathematics, Washington, DC.

Brooks, F. D. (1924). The transfer of training in relation to intelligence. *Journal of Educational Psychology*, 15(7):413–422.

Broyler, C. R., Thorndike, E. L., and Woodyard, E. (1927). A second study of mental discipline in high school studies. *Journal of Educational Psychology*, 18(6):377–404.

Bruce, C. D. and Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: what is it? Why does it matter? And what can we do about it? *ZDM Mathematics Education*, 47:331–343.

Burnett, S. A. and Lane, D. M. (1980). Effects of academic instruction on spatial visualization. *Intelligence*, 4(3):233–242.

Burton, A. M., White, D., and McNeill, A. (2010). The Glasgow face matching test. *Behavior Research Methods*, 42(1):286–291.

Buschkuehl, M. and Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss Medical Weekly*, 140(19-20):266–272.

Carpenter, P. A., Just, M. A., and Shell, P. (1990). What one intelligence test measures - a theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review*, 97(3):404–431.

Carroll, J. B. (1997). The three-stratum theory of cognitive abilities. In Flanagan, D. P. and Harrison, P. L., editors, *Contemporary intellectual assessment: Theories, tests and issues*, pages 122–130. Guilford Press, New York.

Carver, C. S., Ganellen, R. J., and Froming, W. J. (1983). Modeling: An analysis in terms of category accessibility. *Journal of Experimental Social Psychology*, 19:403–421.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54(1):1–22.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology*, 27(5):703–722.

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

Chooi, W.-T. and Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, 40(6):531–542.

Clements, D. H. and Gullo, D. F. (1984). Effects of computer programming on young children's cognition. *Journal of Educational Psychology*, 76(6):1051–1058.

Clouston, S. A., Kuh, D., Herd, P., Elliott, J., Richards, M., and Hofer, S. M. (2012). Benefits of educational attainment on adult fluid cognition: international evidence from three birth cohorts. *International Journal of Epidemiology*, 41:1729–1736.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences* . Lawrence Erlbaum Associates Publishers.

Colom, R., Escorial, S., and Rebollo, I. (2004). Sex differences on the Progressive Matrices are influenced by sex differences on spatial ability. *Personality and Individual Differences*, 37(6):1289–1293.

Colom, R. and Flores-Mendoza, C. E. (2007). Intelligence predicts scholastic achievement irrespective of SES factors: Evidence from Brazil. *Intelligence*, 35(3):243–251.

Coluccia, E. and Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of environmental psychology*, 24:329–340.

Conway, A. R., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., and Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, 12(5):769–786.

Cragg, L. and Gilmore, C. (2014). Trends in Neuroscience and Education. *Trends in Neuroscience and Education*, 3(2):63–68.

Dahl, R. E. (2004). Adolescent brain development: A period of vunerabilities and oppotunities. In *Annual New York Academy of Sciences*, pages 1–22, Pittsburgh, Pennsylavania.

Dark, V. J. and Benbow, C. P. (1990). Enhanced problem translation and short-term memory: Components of mathematical talent. *Journal of Educational Psychology*, 82(3):420–429.

Deary, I. J. (2001). *Intelligence: A very short introduction*. Oxford University Press, Oxford, UK.

Deary, I. J., Gow, A. J., Pattie, A., and Starr, J. M. (2012). Cohort Profile: The Lothian Birth Cohorts of 1921 and 1936. *International Journal of Epidemiology*, 41(6):1576–1584.

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

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

Department for Education, U. K. (2013). Mathematics programmes of study: Key stage 3. Technical report, London.

Department for Education, U. K. (2014a). Mathematics: AS and A level content. Technical report, London, UK.

Department for Education, U. K. (2014b). School leaving age.

Department for Education, U. K. (2016a). Mathematics: GCSE subject content and assessment objectives. Technical report, London, UK.

Department for Education, U. K. (2016b). Progress 8 measure in 2016, 2017, and 2018. Technical report, London.

Dorsey, M. F. and Hopkins, L. T. (1930). The influence of attitude upon transfer. *Journal of Educational Psychology*, 21(6):410–417.

Downey, D. B., Von Hippel, P. T., and Broh, B. A. (2004). Are schools the great equalizer? Cognitive inequality during the summer months and the school year. *American Sociological Review*, 69(5):613–635.

Doyen, S., Klein, O., Pichon, C. L., and Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind. *PloS one*, 7(1):1–7.

Dudley, U. (2010). What Is Mathematics For? *Notices of the AMS*, 57(5):608–613.

Education Funding Agency (2014). 16 to 19 funding: maths and English condition of funding.

Evans, J. S. B., Newstead, S., and Byrne, R. M. J. (1993). *Human Reasoning: The psychology of deduction.* Lawrence Erlbaum Associates Publishers, Hove, UK.

Evans, J. S. B. T., Clibbens, J., and Rood, B. (1995). Bias in Conditional Inference: Implications for Mental Models and Mental Logic. *The Quarterly Journal of Experimental Psychology Section A*, 48(3):644–670.

Facione, P. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic*, 20(1):61–84.

Feng, J., Spence, I., and Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, 18(10):850–855.

Feurzeig, W., Horwitz, P., and Nickerson, R. (1981). Microcomputer in education. Technical Report 4798, Cambridge, MA.

Fong, G. T., Krantz, D. H., and Nisbett, R. E. (1986). The effects of statistical training on thinking about everyday problems. *Cognitive Psychology*, 18(3):253–292.

Forbus, K., Usher, J., Lovett, A., Lockwood, K., and Wetzel, J. (2011). CogSketch: Sketch Understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science*, 3(4):648–666.

Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives*, 19(4):25–42.

Frick, A., Daum, M. M., Walser, S., and Mast, F. W. (2009). Motor Processes in Children's Mental Rotation. *Journal of Cognition and Development*, 10(1-2):18–40.

Frick, A., Hansen, M. A., and Newcombe, N. S. (2013). Cognitive Development. *Cognitive Development*, 28(4):386–399.

Fry, A. F. and Hale, S. (1996). Processing speed, working memort, and fluid intelligence: Evidence for developmental cascade. *Psychological Science*, 7(4):237–241.

Gadske, R. E. (1940). *Demonstrative Geometry as a Means for Improving Critical Thinking* . PhD thesis, Northwestern University.

Galton, F. (1883). *Inquires into human faculty and its development.* The Macmillan Company.

Geary, D. C., Saults, S. J., Liu, F., and Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology*, 77:337–353.

Göbel, S., Walsh, V., and Rushworth, M. F. S. (2001). The Mental Number Line and the Human Angular Gyrus. *NeuroImage*, 14(6):1278–1289.

Goffaux, V. and Rossion, B. (2006). Faces are spatial — holistic face perception is supported by low spatial frequencies. *Journal of Experimental Psychology: Human Perception and Performance*, 32(4):1023–1039.

Gottfredson, L. S. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, 29(3):379–410.

Gottfredson, L. S. (2004). Intelligence: is it the epidemiologists' elusive fundamental cause of social class inequalities in health? *Journal of Personality and Social Psychology*, 86(1):174–199.

Gottfredson, L. S. and Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science*, 13(1):1–4.

Gould, S. J. (2006). *The Mismeasure of Man (Revised & Expanded)*. W. W. Norton & Company.

Guildford, J. P. and Zimmerman, W. S. (1948). The Guildford-Zimmerman aptitude survey. *Journal of Applied Psychology*, 32(1):24.

Guttman, R., Epstein, E., Amir, M., and Guttman, L. (1990). A structural theory of spatial abilities. *Applied Psychological Measurement*, 14(3):217–236.

Handscombe, J. (2013). The comparative difficulty of Higher Mathematics on the International Baccalaureate. *Teaching Mathematics and its Applications*, 32(3):112–122.

Hartung, M. L. (1942). Chapter VII: Teaching of Mathematics in Senior High School and Junior College. *Review of Educational Research*, 12:425–434.

Hegarty, M. and Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology*, 91(4):684–689.

Henderson, E. N. (1911). Formal Discipline. In *Cyclopedia of Education*, pages 642–647. The Macmillan Company.

Higgins, E. T., Rholes, W. S., and Jones, C. R. (1977). Category accessibility and impression formation. *Journal of Experimental Social Psychology*, 13:141–154.

Hirschfeld, G., Thielsch, M. T., and Zernikow, B. (2013). Reliabilities of mental rotation tasks: limits to the assessment of individual differences. *BioMed Research International*, 2013(1–3):1–7.

Hitch, G., Halliday, S., Schaafstal, A., and Schraagen, J. (1988). Visual working memory in young children. *Memory & Cognition*, 16(2):120–132.

Holmes, J., Adams, J. W., and Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology*, 20(2):272–289.

Hubber, P. (2016). *Understanding the role of visuo-spatial working memory in adult mathematics*. PhD thesis, The University of Nottingham.

Hubber, P. J., Gilmore, C., and Cragg, L. (2014). The roles of the central executive and visuospatial storage in mental arithmetic: A comparison across strategies. *The Quarterly Journal of Experimental Psychology*, 67(5):936–954.

Hutson, P. W. (1935). The need for reorientation of mathematics in the secondary school. *The Mathematics Teacher*, 28(3):145–153.

Inglis, M. and Simpson, A. (2004). Mathematicians and the selection task. In *International Group for the Psychology of Mathematics Education*, pages 89–96.

Inglis, M. and Simpson, A. (2007). Conditional inference and advanced mathematical study. *Educational Studies in Mathematics*, 67(3):187–204.

Inglis, M. and Simpson, A. (2009). Conditional inference and advanced mathematical study: further evidence. *Educational Studies in Mathematics*, 72(2):185–198.

Institute of Education (2010). Institute of education archives subject guide. *University of London*.

Jackson, S. and Griggs, R. (1988). Education and the selection task. *Bulletin of the Psychonometric Society*, 26(4):327–330.

Jacowitz, K. E. and Kahneman, D. (1995). Measures of anchoring in estimation tasks. *Personality and Social Psychology Bulletin*, 21(11):1161–1166.

Jaeggi, S., Studer-Luethi, B., Buschkuehl, M., Su, Y.-F., Jonides, J., and Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning — Implications for training and transfer. *Intelligence*, 38(6):625–635.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, 105(19):6829–6833.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., and Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, 108(25):10081–10086.

Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., and Heil, M. (2013). Mental rotation performance in primary school age children: Are there gender differences in chronometric tests? *Cognitive Development*, 28(1):51–62.

Jarosz, A. F. and Wiley, J. (2014). What are the odds? A practical guide to computing and reporting Bayes factors. *The Journal of Problem Solving*, 7(1):2–9.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39(1):1–123.

Jensen, A. R. (1980). *Bias in mental testing.* Free Press, New York.

Kail, R. (1986). The impact of extended practice on rate of mental rotation. *Journal of Experimental Child Psychology*, 42:378–391.

Kane, M. J., Hambrick, D. Z., and Conway, A. R. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005). 131(1):66–71.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence*, 35(3):211–223.

Kirkham, N. Z., Slemmer, J. A., and Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition*, 82:B35–B42.

Klingberg, T. (2010). Training and plasticity of working memory. *Trends in Cognitive Sciences*, 14:317–324.

Klingberg, T., Forssberg, H., and Westerberg, H. (2002a). Increased brain activity in frontal and parietal cortex underlies the development of visuospatial working memory capacity during childhood. *Journal of Cognitive Neuroscience*, 14(1):1–10.

Klingberg, T., Forssberg, H., and Westerberg, H. (2002b). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology*, 24(6):781–791.

Kunda, M., McGreggor, K., and Goel, A. (2009). Addressing the Raven's Progressive Matrices test of "general" intelligence. *Association for the Advancement of Artificial Intelligence.*

Kyllonen, P. C. and Christal, R. E. (1990). Reasoning ability is (little more than) working-memory capacity?! *Intelligence*, 14(4):389–433.

Lachance, J. A. and Mazzocco, M. M. (2006). A longitudinal analysis of sex differences in math and spatial skills in primary school age children. *Learning and Individual Differences*, 16(3):195–216.

Lehman, D. R., Lempert, R. O., and Nisbett, R. E. (1988). The effects of graduate training on reasoning: Formal discipline and thinking about everyday-life events. *American Psychologist*, 43(6):431–442.

Lehman, D. R. and Nisbett, R. E. (1990). A longitudinal study of the effects of undergraduate training on reasoning. *Developmental Psychology*, 26(6):952–960.

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

LGA (2015). Achievement and retention in post 16 education. Technical report.

Lim, T. K. (1994). Gender-related differences in intelligence: Application of confirmatory factor analysis. *Intelligence*, 19(2):179–192.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences*, 49(4):344–351.

Lynn, R. and Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, 32(5):481–498.

Mackintosh, N. J. and Bennett, E. S. (2005). What do Raven's Matrices measure? An analysis in terms of sex differences. *Intelligence*, 33:663–674.

Manktelow, K. (2012). Reasoning with Propositions. In *Thinking and Reasoning: An introduction to the psychology of reasoning, judgement and decision making*, pages 57–71. Psychology Press Ltd, Hove.

McGlone, M. S. and Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27:486–493.

McNeil, N. M., Rittle-Johnson, B., Hattikudur, S., and Petersen, L. A. (2010). Continuity in representation between children and adults: arithmetic knowledge hinders undergraduates' algebraic problem solving. *Journal of Cognition and Development*, 11(4):437–457.

Medina, A. C. and Sorby, S. A. (1998). Identifying gender differences in the 3-D visualization skills of engineering students in Brazil and in the United States. In *International Conference for Enigneering Education*, pages 1–8, Rio de Janeiro, Brazil.

Melby-Lervåg, M. and Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology*, 49(2):270–291.

Meyer, D. E. and Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2):227–234.

Ministry of Education, O. (2014). Paying attention to spatial reasoning K-12: Support document for 'Paying attention to mathematics education'. Ontario Ministry of Education.

Moody, D. E. (2009). Can intelligence be increased by training on a task of working memory? *Intelligence*, 37(4):327–328.

Myerson, J., Hale, S., Rhee, S. H., and Jenkins, L. (1999). Selective interference with verbal and spatial working memory in young and older adults. *Journal of Gerontology*, 54B(3):161–164.

Nath, S. and Szucs, D. (2014). Learning and Instruction. *Learning and instruction*, 32(C):73–80.

Neisser, U., Boodoo, G., Bouchard, Jr, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., and Sternberg, R. J. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, 51(2):77–101.

Newcombe, N. S., Levine, S. C., and Mix, K. S. (2015). Thinking about quantity: the intertwined development of spatial and numerical cognition. *Wiley Periodicals. Inc.*, 6:491–505.

Newstead, S. (2003). Peter Wason (1924-2003). *Thinking & Reasoning*, 9(3):177–184.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., and Turkheimer, E. (2012). Intelligence: new findings and theoretical developments. *American Psychologist*, 67(2):130–159.

Ofsted (2016). Raising standards and improving lives.

Okagaki, L. and Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, 15:33–59.

Ortner, T. M. and Sieverding, M. (2008). Where are the gender differences? Male priming boosts spatial skills in women. *Sex roles*, 59:274–281.

Papert, S. (1998). Child power: Keys to the new learning of the digital century. In *Colin Cherry Memorial Lecture on Communication*, Imperial College London.

Passolunghi, M. C. and Siegel, L. S. (2004). Working memory and access to numerical information in children with disability in mathematics. *Journal of Experimental Child Psychology*, 88(4):348–367.

Pea, R. D. and Kurland, D. M. (1984). On the cognitive effects of learning computer programming. *New Ideas in Psychology*, 2(2):137–168.

Pearson (2007). Raven's Standard Progressive Matrices (SPM) - Evidence of reliability and validity. TalentLens, USA.

Peirce, J. W. (2008). Psychopy2.

Perkins, D. N. and Salomon, G. (1992). Transfer of learning. In *International Encyclopedia of Education*, pages 2–13. Pergamon Press, Oxford, England.

Peterson, D. (2016). The baby factory: difficult research objects, disciplinary standards, and the production of statistical significance. *Socius: Sociological Research for a Dynamic World*, 2:1–10.

Pickering, S. J. (2001). The development of visuo-spatial working memory. *Memory*, 9(4-6):423–432.

Pietschnig, J., Voracek, M., and Formann, A. K. (2010). Intelligence. *Intelligence*, 38(3):314–323.

Plato (2008). *The Republic (375BC) - Modern translation.* Gutenberg.

Quinn, P. C. and Liben, L. S. (2008). A sex difference in nental rotation in young infants. *Psychological Science*, 19(11):1067–1070.

Rabipour, S. and Raz, A. (2012). Training the brain: Fact and fad in cognitive and behavioral remediation. *Brain and Cognition*, 79(2):159–179.

Raghubar, K. P., Barnes, M. A., and Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences*, 20(2):110–122.

Rauscher, F. H., Shaw, G. L., and Ky, C. N. (1993). Music and spatial task performance. *Nature*, 365(6447):611.

Rauscher, F. H., Shaw, G. L., Levine, L. J., Wright, E. L., Dennis, W. R., and Newcomb, R. L. (1997). Music training causes long-term enhancement of preschool children's spatial-temporal reasoning. *Neurological Research*, 19:2–8.

Raven, J., Raven, J. C., and Court, J. H. (2000). *Raven Manual: Section 3.*

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D. E., Hambrick, D. Z., Kane, M. J., and Engle, R. W. (2012). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, 142(2):359–379.

Ritchie, S. J. (2015). *Intelligence - All that matters.* John Murray Learning, UK.

Ritchie, S. J., Bates, T. C., and Deary, I. J. (2015). Is education associated with improvements in general cognitive ability, or in specific skills? *Developmental Psychology*, 51(5):573–582.

Ritchie, S. J., Bates, T. C., Derr, G., Starr, J. M., and Deary, I. J. (2013). Education is associated with higher later life IQ scores, but not with faster cognitive processing speed. *Psychology and Aging*, 28(2):515–521.

Rowe, G., Hasher, L., and Turcotte, J. (2008). Age differences in visuospatial working memory. *Psychology and Aging*, 23(1):79–84.

Salthouse, T. A., Babcock, R. L., Skovronek, E., Mitchell, D. R. D., and Palmon, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, 26(1):128–136.

Schmidt, F. L. (2000). Select on intelligence. In Locke, E. A., editor, *Handbook of principles of organizational behvior*, pages 3–18. Wiley, Chichester, UK.

Schmidt, F. L. and Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124(2):262–274.

Schmidt, G. P. (1958). The Liberal Arts college: A chapter in American cultural history. *The Catholic Historical Review*, 44(1):65–67.

SharpBrains (2013). Executive summary: Infographic on the digital brain health market 2012-2020. Technical report.

Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, 93(3):604–614.

Smith, A. F. M. (2004). Making Mathematics Count.

Smith, E. E., Jonides, J., and Koeppe, R. A. (1996). Dissociating verbal and spatial working memory using PET. *Cerebal Cortex*, 6:11–20.

Sorby, S. A. (1999). Developing 3-D spatial visualization skills. *Engineering Design Graphics Journal*, 63:21–32.

Sorby, S. A. (2009). Educational Research in Developing 3D Spatial Skills for Engineering Students. *International Journal of Science Education*, 31(3):459–480.

Spearman, C. (1927). *The Abilities of Man.* The Macmillan Company, New York.

Spearman, C. E. (1904). "General intelligence" objectively determined and measured. *The American Journal of Psychology*, 15(2):201–292.

Spelke, E. S. (2005). Sex differences in intrinsic aptitude for mathematics and science?: a critical review. *American Psychologist*, 60(9):950.

Srull, T. K. and Wyer, R. S. (1979). The role of category accessibility in the interpretation of information about persons: Some determinants and implications. *Journal of Personality and Social Psychology*, 37(10):1660–1672.

Stanic, G. A. (1986). Mental discipline theory and mathematics education. *For the Learning of Mathematics*, 6(1):39–47.

Steele, K. M., Ball, T. N., and Runk, R. (1997). Listening to Mozart does not enhance backwards digit span performance. *Perceptual and Motor Skills*, 84:1179–1184.

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, 105(19):6791–6792.

Stough, C., Kerkin, B., Bates, T. C., and Mangan, G. (1994). Music and spatial IQ. *Personality and Individual Differences*, 17(5):695.

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, 35(5):401–426.

Swan, M. (2005). Improving Learning in Mathematics. Department for Education and Skills Standards Unit.

Terlecki, M. S. and Newcombe, N. S. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology*, 22:996–1013.

Thorndike, E. L. (1922). Instruments for measuring the disciplinary values of studies. *Journal for Educational Research*, 5(4):269–279.

Thorndike, E. L. (1924a). Mental discipline in high school studies. *Journal of Educational Psychology*, 15(1):1–22.

Thorndike, E. L. (1924b). Mental discipline in high school studies - Continued from January. *Journal of Educational Psychology*, 15(2):83–98.

Thorndike, E. L. and Woodworth, R. S. (1901). The influence of improvement in one mental function upon the efficiency of other functions - The estimation of magnitudes. *Psychological Review*, 8(4):384–395.

Tolar, T. D., Lederberg, A. R., and Fletcher, J. M. (2009). A structural model of algebra achievement: computational fluency and spatial visualisation as mediators of the effect of working memory on algebra achievement. *Educational Psychology*, 29(2):239–266.

Tosto, M. G., Hanscombe, K. B., Haworth, C. M. A., Davis, O. S. P., Petrill, S. A., Dale, P. S., Malykh, S., Plomin, R., and Kovas, Y. (2014). Why do spatial abilities predict mathematical performance? *Developmental Science*, 17(3):462–470.

Tversky, A. and Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185(4157):1124–1131.

Ulmer, G. (1939). Teaching geometry to cultivate reflective thinking: An experimental study with 1239 high school pupils. *The Journal of Experimental Education*, 8(1):18–25.

USGS (2016). Statistical interpretation.

Uttal, D. H. and Cohen, C. A. (2012). Spatial thinking and STEM education: When, why and how? In *Psychology of learning and motivation*, pages 148–178. Elsevier Inc.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., and Newcombe, N. S. (2013a). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin*, 139(2):352–402.

Uttal, D. H., Miller, D. I., and Newcombe, N. S. (2013b). Exploring and Enhancing Spatial Thinking Links to Achievement in Science, Technology, Engineering, and Mathematics? *Current Directions in Psychological Science*, 22(5):367–373.

van der Ven, A. H. G. S. and Ellis, J. L. (2000). A Rasch analysis of Raven's standard progressive matrices. *Personality and Individual Di erence*, 29:45–64.

van Garderen, D. (2006). Spatial visualization, visual imagery, and mathematical problem solving of students with varying abilities. *Journal of Learning Disabilities*, 39(6):496–506.

Vandenberg, S. G. and Kuse, A. R. (1978). Mental roations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47:599–604.

Venkatesan, S. (2014). Celebrating a century on form boards with special reference to Seguin Form Board as measure of intelligence in children. *Global Journal of Interdisciplinary Social Sciences*, 3(6):43–51.

Vernon, P. E. (1969). *Intelligence and cultural environment*. Methuen, London.

Vorderman, C., Budd, C., Dunne, R., Hart, M., and Porkess, R. (2011). A world-class mathematics education for all our young people. Technical report.

Voyer, D., Voyer, S. D., and Saint-Aubin, J. (2016). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review*, pages 1–28.

Wagenmakers, E.-J. and Jove, J. (2016). JASP.

Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4):817–835.

Wai, J., Lubinski, D., Benbow, C. P., and Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology*, 102(4):860–871.

Wei, W., Yuan, H., Chen, C., and Zhou, X. (2011). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology*, 82(1):157–181.

Weisert, J. J. (1939). Foreign languages as mental discipline: a survey. *The German Quarterly*, 12(2):61–70.

Wesman, A. G. (1945). *Transfer of Training from School Subjects to Intelligence*. PhD thesis, AMS Press, New York.

Wiedenbauer, G. and Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and instruction*, 18(1):30–41.

Wilson, T. L. and Brown, T. L. (1997). Reexamination of the effect of Mozart's music on spatial-task performance. *The Journal of Psychology*, 131(4):365–370.

Wolf, A. (2011). Review of vocational education - The Wolf report. Technical report.

Wood, G., Willmes, K., Nuerk, H.-C., and Fischer, M. H. (2008). On the cognitive link between space and number: a meta-analysis of the SNARC effect. *Psychology Science Quarterly*, 50(4):489–525.

Woodcock, R. W., McGrew, K. S., and Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. Riverside, Itasca, IL.

Woodworth, R. S. and Thorndike, E. L. (1901). The influence of improvement in one mental function upon the efficiency of other functions.(I). *Psychological Review*, 8(3):247–261.

Zald, D. H. and Iacono, W. G. (1998). The development of spatial working memory abilities. *Developmental neuropsychology*, 14(4):563–578.