Gender Differences in Learning with Instructional Animations

Author:
Wong, Pui Shan

Publication Date:
2016

DOI:
https://doi.org/10.26190/unsworks/2992

Gender Differences in Learning with Instructional Animations

Pui Shan Wong

A thesis in fulfilment of the requirements for the degree of

Doctor of Philosophy

School of Education

Faculty of Arts and Social Sciences

March 2016

PLEASE TYPE

THE UNIVERSITY OF NEW SOUTH WALES
Thesis/Dissertation Sheet

Surname or Family name: **Wong**

First name: **Pui Shan** Other name/s: **Mona**

Abbreviation for degree as given in the University calendar: **PhD**

School: **School of Education** Faculty: **Faculty of Arts and Social Sciences**

Title: **Gender differences in learning with instructional animations**

Abstract 350 words maximum: (PLEASE TYPE)

Research into instructional animations has produced inconsistent results. Some studies have found animations superior to static pictures, while others have found no differences, or even advantages in favour of statics. These mixed results have often puzzled researchers because animations have a greater flexibility than static pictures in depicting physical and temporal changes. Several factors have been proposed to explain the lack of a clear pattern, including poorly designed studies with a number of biases, and failure to consider moderating factors such as gender and spatial ability. However, one explanation on why animations can be ineffective is that, they are a cause of transient information, which requires extra working memory resources to process and therefore hinders learning. However, there is a special case for learning human movement tasks where animations do not create additional cognitive load because humans have evolved to learn about movement through observations.

This thesis continues the research into comparing animations and statics using a human movement task (building Lego shapes). It also examines the impact of gender and spatial ability (and the relationship between the two) on learning from both animations and statics. Furthermore, some embodied cognition effects are investigated in the form of gesturing and observing hands manipulating the shapes.

Four experiments were conducted with university students having an equal number of males and females, randomly assigned to the designated conditions. Participants were required to observe a Lego construction, and then rebuild it. Spatial ability was measured and statistically controlled in the analysis. Results (Experiments 1, 2 & 3) did not show any animation advantages; however, there was a consistent gender-presentation format interaction effect where females performed better after observing animations while males performed better with static pictures. Furthermore, in contrast to the gesturing research, gesturing was found to be impedimental to learning, as was the observation of hands (Experiment 4). Furthermore, the best predictors of male performance were objective measures of spatial ability, whereas for females self-rating measures were best.

Overall it was found that gender plays a significant role in animation research as do a number of other moderating factors.

Declaration relating to disposition of project thesis/dissertation

………………………………………	………………………………………	30 Mar 2016
Signature	Witness	Date

FOR OFFICE USE ONLY Date of completion of requirements for Award:

THIS SHEET IS TO BE GLUED TO THE INSIDE FRONT COVER OF THE THESIS

ORIGINALITY STATEMENT

‘I hereby declare that this submission is my own work and to the best of my knowledge it contains no materials previously published or written by another person, or substantial proportions of material which have been accepted for the award of any other degree or diploma at UNSW or any other educational institution, except where due acknowledgement is made in the thesis. Any contribution made to the research by others, with whom I have worked at UNSW or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except to the extent that assistance from others in the project's design and conception or in style, presentation and linguistic expression is acknowledged.’

Date …………30 Mar 2016……………………..............

COPYRIGHT STATEMENT

Date ………………30 Mar 2016……………..........................

AUTHENTICITY STATEMENT

Signed …………………………………………….........................

Date ………………30 Mar 2016……………..........................

Acknowledgements

First and foremost I would like to express the deepest appreciation to my supervisor, Emeritus Professor Paul Ayres. I would not have made this journey through without your continuous supports, guidance, encouragements (and your British protections!). Thank you for being the best supervisor that one could ask for. I am very grateful and proud to have you as my supervisor, and I truly hope that I can make you proud in return.

Besides, I would like to thank my second supervisor, Professor Slava Kalyuga. Your attitudes are inspiring and your warmth is always encouraging. Thank you for looking out for me.

My sincere thanks also go to Professor Fred Paas who provided valuable advices and insights onto the experiments. It is my honour and pleasure to work with one of the most productive researchers in the field.

A huge love and thank to my family for supporting my dream. Mum and Dad, I know it is never easy supporting a child studying aboard, but you never hesitated giving and would never doubt my decisions. Jerry, thank you for taking care of the family on behalf of your selfish sister. To my fiancé Ricky, I will never forget how we overcame obstacles and maintained our long-distance relationship all these years. You embraced all my nonsense and your love give me wings to fly. I am blessed having you in my life.

To my friends Anne-Marie, Cris, Naoko, Kevin, Adrian, Joanne, Olga, Ellen, Roshi, Israt, Nur…etc., you all coloured my life in Australia. Thank you for all your spiritual and physical support (especially for all the snack supply which keeps me going!). A large part of my thesis is made out of your sweets and coffee!

Last but not the least; thank all my cognians (especially R-post mates and guam) for your loves and supports from over the ocean. I am proud to be a cognian, as always.

Table of Contents

List of Figures

List of Tables

List of Appendices

Abstract

Research into instructional animations has produced inconsistent results. Some studies have found animations superior to static pictures, while others have found no differences, or even advantages in favour of statics. These mixed results have often puzzled researchers because animations have a greater flexibility than static pictures in depicting physical and temporal changes. Several factors have been proposed to explain the lack of a clear pattern, including poorly designed studies with a number of biases, and failure to consider moderating factors such as gender and spatial ability. However, one explanation on why animations can be ineffective is that, they are a cause of transient information, which requires extra working memory resources to process and therefore hinders learning. However, there is a special case for learning human movement tasks where animations do not create additional cognitive load because humans have evolved to learn about movement through observations.

This thesis continues the research into comparing animations and statics using a human movement task (building Lego shapes). It also examines the impact of gender and spatial ability (and the relationship between the two) on learning from both animations and statics. Furthermore, some embodied cognition effects are investigated in the form of gesturing and observing hands manipulating the shapes.

Four experiments were conducted with university students having an equal number of males and females, randomly assigned to the designated conditions. Participants were required to observe a Lego construction, and then rebuild it. Spatial ability was measured and statistically controlled in the analysis. Results (Experiments 1, 2 & 3) did not show any animation advantages; however, there was a consistent gender-presentation format interaction effect where females performed better after observing animations while males performed better with static pictures. Furthermore, in contrast to

the gesturing research, gesturing was found to be impedimental to learning, as was the observation of hands (Experiment 4). Furthermore, the best predictors of male performance were objective measures of spatial ability, whereas for females self-rating measures were best.

Overall it was found that gender plays a significant role in animation research as do a number of other moderating factors.

Preface

Dynamic visualisations, animations, are often used in teaching (e.g., Bétrancourt, Dillenbourg, & Clavien, 2008; Rieber, 1991a). One of the potential advantages of instructional animations is their flexibility in depicting dynamic or temporal changes. The close match between showing an animation of a dynamic learning task and the actual task, have led many educational researchers to predict that animations should be superior to statics (see Bétrancourt & Tversky, 2000; Chandler, 2004; Hegarty, 1992).

However, research has shown that animations are only superior to equivalent static pictures in some cases (e.g. Ayres, Marcus, Chan, & Qian, 2009; Bétrancourt et al., 2008; Höffler & Leutner, 2007). In some studies, animations have been found to be equivalent to (e.g. Morrison & Tversky, 2001) or even inferior to equivalent static pictures (e.g. Castro-Alonso, Ayres, & Paas, 2014b; Mayer, Hegarty, Mayer, & Campbell, 2005). Some researchers (e.g. Bétrancourt & Tversky, 2000; Tversky, Morrison, & Bétrancourt, 2002) have indicated that the inconclusive results might be due to inconsistencies in the experimental design and/or material design. Other researchers (e.g. Ayres & Paas, 2007b; van Gog, Paas, Marcus, Ayres, & Sweller, 2009) proposed that the transient nature of animations might create an unfavourable environment for learning under certain conditions. From a cognitive load theory perspective, it is believed that the dynamic nature of animations produces transient information, which is difficult to process and hinders learning effectiveness (Ayres & Paas, 2007b; Castro-Alonso, Ayres, & Paas, 2015b; van Gog et al., 2009).

Recently, animations have been found to be consistently more effective than static diagrams in facilitating the learning of human movement skills (e.g. Ayres & Paas, 2007b; Castro-Alonso et al., 2014b; Marcus, Cleary, Wong, & Ayres, 2013). Based on Geary’s (2007) concept of biologically primary knowledge, the ability to learn through observation and to simulate movements have evolved in human beings (see Castro-Alonso, Ayres, & Paas, 2014a; Paas & Sweller, 2012). Animations, which can depict changes and continuous movements, are believed to create an excellent environment for observations and simulations. Consequently for human movement tasks, dealing with transient information is less problematical, and studies have shown that animations can be superior to equivalent static pictures.

Besides the issues associated with learning topics, there are also some indications that gender and spatial ability are also factors that might affect learning from animations (Ayres, 2015; Ayres et al., 2015). Although not conclusive, some evidence has emerged that females learn better with animations than static pictures while males learn better (or equally) with static pictures than animations (e.g., Cowards, Crooks, Flores, & Dao, 2012; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006). Spatial ability is essential for learners to understand visual information and construct mental representations to assist in the learning process. It is often assumed that because females are weaker in spatial ability than males (see Guillem & Mograss, 2005; Halpern, 2012, pp. 128–145; Linn & Petersen, 1985; Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995), and animations help learners with lower spatial ability, females benefit from animations because they have lower spatial ability. However, there is a lack of direct investigations confirming this link.

The main aim of this thesis was to continue the research into animations and static presentations with a human movement task (constructing a Lego shape); but also to investigate how gender influences the effectiveness of both types of presentations. It also aimed to explore the relationship between spatial ability and gender when learning from visual presentations. Furthermore two embodied cognition factors (gesturing and observing hands manipulating the constructions) were explored to see if they could improve the learning from animations and statics. All investigations were completed in order to improve the instructional efficiency of animations and statics.

Thesis structure

The first three chapters provide extensive reviews on the major factors involved in the sphere of study. Chapter 1 outlines the basic principles of human cognitive architecture and the associated learning mechanisms. It is then followed by some basic concepts of evolutionary educational psychology, which provides some grounds and insights on how learning abilities have evolved. Finally this chapter outlines the main theoretical framework of this thesis- cognitive load theory and its most relevant effects.

Chapter 2 explores the literature on instructional animations, another major element of this thesis. It provides an overview on the existing literature on instructional animations and how they compare to static pictures, with evidence for and against the effectiveness of instructional animations. This chapter also explore some reasons why instructional animations are not always successful. In particular, the role played by embodied cognition, which taps into natural learning abilities is discussed. The chapter finishes with some key factors influencing animation designs.

Chapter 3 examines the role played by gender and spatial ability in instructional animations and statics. It outlines the main empirical evidence regarding gender differences in learning from visual presentations.

Following on from Chapters 1-3, Chapter 4 describes the development of the hypotheses and provides a brief description of each experiment. Based on the literature review and the main aims of the study, four hypotheses were developed. Chapter 5 provides the details (methodologies, procedures, analyses and results) of the four experiments completed to test the hypotheses. Finally Chapter 6 provides a summary of the results and how they match the hypotheses, as well as a general discussion of the findings, along with the study’s limitations, implications, and future directions.

Chapter 1: Human Cognitive Architecture and Cognitive Load Theory

Cognitive load theory is an educational psychology theory focusing on instructional material design. It was developed based on the existing knowledge of human cognitive architecture. As Sweller, Ayres, and Kalyuga (2011, p. iii) remarked: ‘without knowledge of human cognitive processes, instructional design is blind’, suggesting that it is wasteful discussing instructional material design before understanding how our human brains work. Hence, in this chapter, human cognitive architecture will be discussed first, followed by evolutionary educational psychology in order to understand how our learning abilities have evolved. Finally the main ideas of cognitive load theory will be discussed followed by some of the main effects that the theory has generated.

1.1 Human Cognitive Architecture

1.1.1 The modal model

R. C. Atkinson and Shiffrin (1968) developed a functional model of human memory known as the *multiple store* or *modal model* (Figure 1.1), illustrating how information is perceived, processed and stored. The model comprised of three components: sensory register, short-term memory (now called working memory, WM) and long-term memory (LTM). In this model, information or a stimulus initially enters the sensory register as a form of environmental input, often in large quantities. However, only a small amount which humans attend to enters working memory, and the remaining unattended information decays (Broadbent, 1958; Treisman, 1960). After entering working memory, the information will be processed, and potentially organised and encoded into long-term memory for permanent storage. While processing the information, previously stored information will be retrieved from long-term memory back into the working memory to assist in the organisation and understanding of the new information.

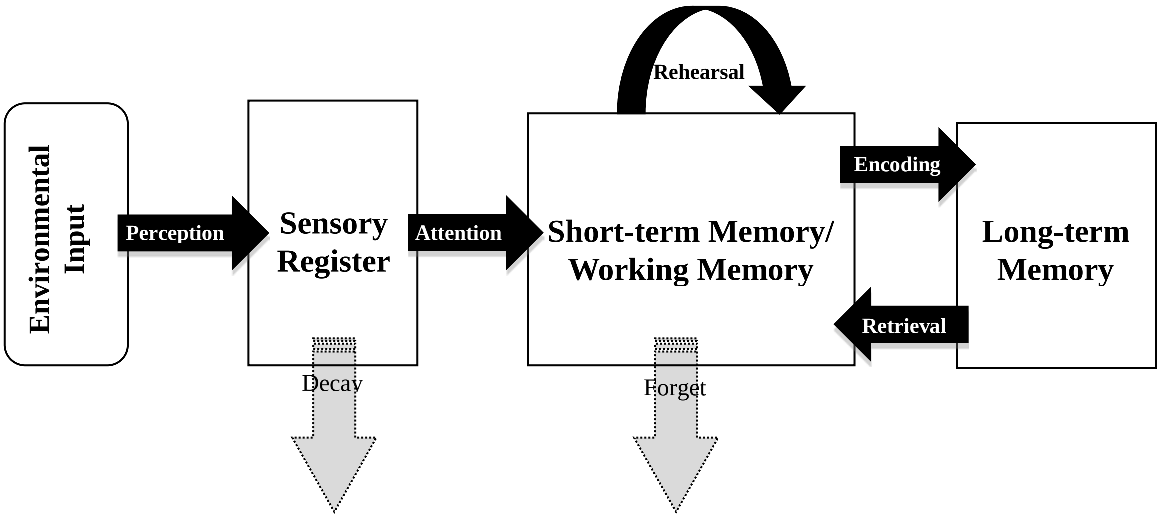


Figure 1.1. Redrawn Modal Model (for Original See R. C. Atkinson & Shiffrin, 1968, p. 113)

In the development of cognitive load theory, the role of sensory memory was considered less important than the other two components; and hence in this thesis only working memory and long-term memory will be discussed in further detail. However, it is noted that sensory memory plays crucial roles in attention and perception, which are also critical factors in learning.

1.1.1.1 Working memory

Historically, working memory (WM) was initially known as short-term memory suggesting it was only a temporary storage of information. This term was then gradually replaced by *working memory* as it became clearer from the research that this component of the memory system served as both a temporary storage (*memory*) and an active processing system (*working*), where information is processed consciously (see

Baddeley, 1992). This system is known to be limited in both capacity and duration (Sweller, van Merriënboer, & Paas, 1998).

Working memory can hold only seven ± two chunks of information (Miller, 1956) for about two to three seconds (Peterson & Peterson, 1959), and this capacity decreases to about four if the information requires processing (see also Cowan, 2010). That means only a limited amount of information can be processed for a limited time unless an individual rehearses the information to keep it active in the working memory.

Among a number of theories related to working memory, Baddeley and Hitch's (1974) multi-component working memory model (Figure 1.2, updated in 2000) has been one of the most widely used and discussed models. It illustrates how information in the working memory is encoded and linked with stored information in the long-term memory.

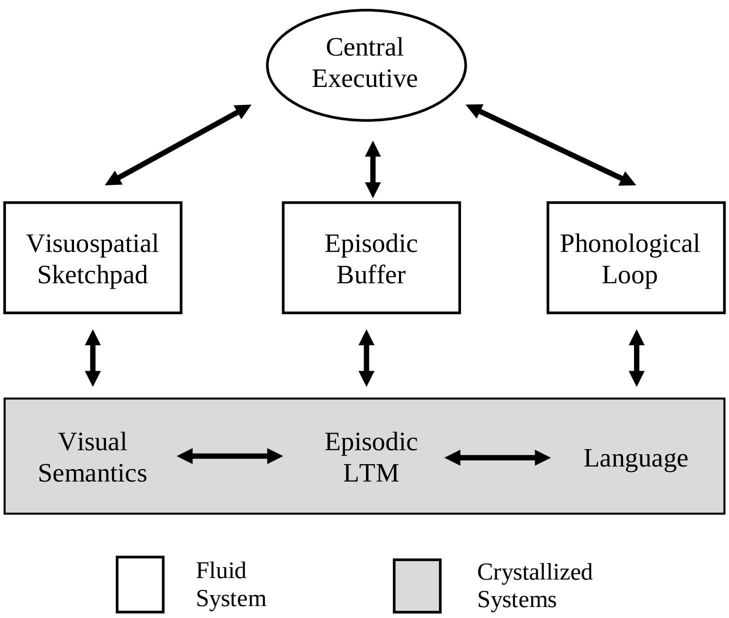


Figure 1.2. Baddeley and Hitch's Multi-component Working Memory Model (Adopted from Baddeley, 2002, p.93)

In the updated Baddeley model (2000), working memory is divided into three sub-systems (middle layer in Figure 1.2): visuospatial sketchpad, episodic buffer and phonological loop. Each of them is responsible for a specific type of information: visuospatial sketchpad is responsible for visual information such as graphs and pictures (both two- and three-dimensional objects); the phonological loop deals with auditory materials such as language/ speech; and the episodic buffer handles memory about personal past events in multi-dimensional code, and provides a temporary interface between the other two (slave) working memory systems and long-term memory. Information entering working memory will be processed independently in the sub-system based on the materials' nature. Consolidated knowledge stored in the crystallised system (bottom layer in Figure 1.2) can be retrieved back into the subsystems and combined with novel information. The process for binding the information from a number of sources into coherent episodes is governed by the central executive.

The original model of Baddeley and Hitch (1974) did not contain the episodic buffer, consisting only of the two slave systems (the phonological loop & visuospatial sketchpad) and the central executive. The episodic buffer was added to give a more holistic model of memory and to fill some missing gaps, such as dealing with episodic memories. It is notable, that the development of cognitive load theory was based exclusively on the 1974 model with little or no reference to the episodic buffer.

1.1.1.2 Long-term memory

In contrast to the working memory, long-term memory (LTM) has no known limits in both duration and capacity (Cowan, 2008; Shiffrin & Atkinson, 1969). Shiffrin and Atkinson (1969, p. 180) described long-term memory as a 'permanent repository for information'. In other words, once information is encoded and stored into the long-term memory, it is never lost. According to Shiffrin and Atkinson (1969), the encoding process (storage process) involves *transfer, placement* and *image-production.* The stored information will later be retrieved back into the working memory through *retrieval process* when needed. Moreover, the retrieval process requires *search* and *recovery*. Decrements in performance over time can also reduce the search effectiveness and lead to a failure of the retrieval process. Such occasions are known as ‘forgetting’.

Schema formation and automation.

With a seemingly infinite capacity in the LTM, Shiffrin and Atkinson (1969) believed that the search process was unlikely to be random. The information needed to be organised in a way that can enhance effective search. The most cited cognitive construct to achieve this organization is known as a *schema* (*schemata* or *schemas* in plural form). Pioneer researchers who contributed much to the early understanding of the roles played by schemas were Piaget and Bartlett.

Piaget (1926) and Bartlett (1932) first introduced the term schema into the psychology field with a belief that elements of information are organised and categorised in the way that will be used (Sweller et al., 1998). Piaget referred to a schema as a unit of knowledge, which is a result of a number of elements being grouped together according to some theme by the process of *assimilation* or *accommodation* (see Piaget, 1926). Individual elements in lower level schemas can be combined into higher level schemas, creating more complex schemas as learning becomes more sophisticated.

Bartlett’s (1932) experiments showed some further light on how knowledge was constructed. His results suggested that new information was not remembered completely; rather it was constructed based on not only prior knowledge and episodic memory about events, but also cultural knowledge. In other words, humans often filled in the gaps when remembering information according to their knowledge and beliefs.

From a cognitive load theory perspective, learning is considered a change in long-term memory (Sweller et al., 2011), and consequently the construction of schemas plays an extremely important role in the learning process. Schemas do not only store learnt information in long-term memory in a helpful way; but can also reduce the burden on working memory when retrieved from long-term memory (Sweller et al., 1998; see also Kalyuga, Chandler, & Sweller, 1999). When information is stored in the form of a schema, multiple elements of information are linked together, which can be treated as a single element when held in working memory, thus overcoming the limitations of working memory for specific tasks related to that information.

Another important property of schemas is *schema automation*. This is a process achieved through either problem explorations or continuous practice (Kotovsky, Hayes, & Simon, 1985; Shiffrin & Schneider, 1977). Automation of schemas allows easy access to knowledge which can be applied (including the execution of tasks) almost effortlessly. Kotovsky et al. (1985) conducted a series of experiments using a well-known problem-solving task- the ‘Tower of Hanoi’- to examine the relationship between the number of problem states and processing load (working memory load). Results revealed that experts showed a higher familiarity of and a better ability to utilise the problem rules to make fewer moves. Such behaviour demonstrated that automation of rules was helpful for planning during problem solving. Microanalysis of participants’ behaviour (i.e. pattern of moves and changes) also showed that automation could overcome memory limitations. Such results were in line with those from Shiffrin and Schneider (1977) who found that repetitive training significantly improved the reaction time for signal detection and the number of correct responses. These two studies provided evidence that accessing and using automated knowledge requires less working memory resources for processing and understanding new information (Sweller et al., 2011). Sweller (1994) suggests that a key function of learning is to store automated schemas in long-term memory.

Sweller (2003) claims that the role of long-term memory is more than just an information store recognising or recalling information; instead, is an integral component of cognitive activity at all levels. This claim was heavily influenced by the research of de Groot (1965) who demonstrated that the difference between chess masters and less-skilled players was not in the ability to conduct more problem-solving searches, but in the ability to memorise board configuration. Experts simply had better memories for real-game chessboard configurations and what to do next. Tellingly when experts were compared with novices on random chess configurations no difference was found between them (Chase & Simon, 1973) . Sweller et al. (1998) argued that chess masters have learnt the basic moves associated with each configuration through years of practice, and have them stored in long-term memory. Unlike less-skilled players who need to search for good moves using their limited working memory, chess masters can make use of their permanent knowledge.

The chess study of De Groot demonstrated that memory of board configurations was critical to expert performance. Results from other studies (e.g. Sweller & Cooper, 1985) also reported similar results demonstrating that long-term memory plays a critical role in task performance, and is the major factor distinguishing novice from expert problem solvers.

1.2 Evolutionary Educational Psychology

The classical elements of human cognitive architecture, described above, have played a major role in the development of cognitive load theory. In more recent times, these elements have been linked by Sweller (Sweller, 2008; Sweller & Sweller, 2006) with evolutionary educational psychology as described next.

Evolutionary educational psychology provides insights into how human have evolved to learn. The main guiding principle is governed by the theory of evolution as proposed by Darwin (Sweller et al., 2011). The prime ideas of evolution are the laws of variation, natural selection and sexual selection. *Variation* is a random process occurring when certain characteristics vary the chance of survival by transferring to the next generation. Such characteristics may or may not be beneficial to the chance of survival. However, survivors who are able to produce offspring, allow the successful gene to be carried on. *Natural selection*, unlike the variation process, which is random, is a process shaped by the environment. It shapes an individual's traits to fit into the environment at that time. If individuals are not competitive enough and cannot be fit into the ecology, they will be eliminated. *Sexual selection* is the social process that combines the effects of intrasexual competition and intersexual selection. In combination, all three selections produce strong and successful individuals of that specie, able to survive through ecological, social and physical challenges, and produce offspring (Darwin, 1859; see also Geary, 2002).

In relation to education and learning, it is believed that 'natural selection and sexual selection are the mechanisms that have driven the evolution of brain, cognition, and g [general intelligent]' (Geary, 2005, p. 23). Based on these ideas, Geary (2002) developed a framework that outlines 'the relation between universal social and cognitive adaptations and academic learning' (p. 318), featuring the concepts of primary and secondary knowledge (Geary, 2005, 2007, 2008).

1.2.1 Biologically primary knowledge

Learning, sometimes, can be instinctive to humans. Some knowledge can be acquired effortlessly and unconsciously without any explicit instructions, Geary (2005) categorised these types of knowledge or skills as biologically primary knowledge. Speaking the first language and recognising faces are examples of biological primary skills (see also Sweller et al., 2011). Each biologically primary skill is modular – they are independent from each other. It is believed that the ability to acquire these skills is biologically driven through evolution and is necessary for intellectual development.

Motor skills learning/ observational learning

Besides face recognitions and first language acquisition, human movement can also be considered as a form of biologically primary knowledge (Geary, 2005, 2007, 2008). Geary argued that all organisms (including humans) have evolved to attend to and process movement patterns of prey or predators. It is necessary to be able to learn and use movement quickly so not to be naturally eliminated. In the case of humans, supporting evidence has been found in anatomic research where humans have been shown to have developed a good sized motor and somatosensory cortex responsible for movement and bodily sensations (Sanes & Donoghue, 2000; Ungerleider, Doyon, & Karni, 2002). As such, human movement (an important factor in this thesis) can be considered a form of biologically primary knowledge (Geary, 2005). Humans have evolved to learn about many forms of human movements, quickly and easily; often though observation without explicit instructions.

1.2.2 Biologically secondary knowledge

In contrast, knowledge which is not instinctive and requires effortful and explicit instructions falls into the category of biologically secondary knowledge. Most of the knowledge taught in educational institutions falls into this category (Sweller et al., 2011). To distinguish biologically secondary knowledge from primary knowledge, Sweller et al. (2011) give writing as an example. An individual can learn to speak by immersing in the environment, but being surrounded by people who can write does not help in learning to write. Although an individual can easily learn motor movements and muscle coordination, more explicit instructions and practice are required to learn to write. Geary (2007, 2008) believed that biologically secondary knowledge was also modular; however this module is unified by sharing amongst all secondary knowledge (see also Sweller et al., 2011, p. 7). In other words, biologically secondary skills are transferable.

1.2.3 Principles of natural information processing

Based on the work of Geary, Sweller (2004) theorised five fundamental principles that link evolution to human cognition. It is argued that the human cognitive system has evolved in a way which is analogous to the theory of evolution by natural selection (Sweller, 2004; Sweller & Sweller, 2006). These five principles form the biological bases of the cognitive load theory. A summary of the five principles from a cognitive perspective is provided Table 1.1 and detailed descriptions follow.

Table 1.1.
Overview of the Natural Information Processing System Principles (Adopted from Sweller & Sweller, 2006, p. 436)

Principles	Cognitive case	Evolutionary case	Function
Information store principle	Long-term memory	Genome	Store information for indefinite periods
Borrowing and reorganizing principle	Transfer information to long-term memory	Transfer information to a genome	Permit the rapid building of an information store
Randomness as genesis principle	Create novel ideas	Create novel genetic codes	Create novel information
Narrow limits of change principle	Working memory	Epigenetic system handling environmental information	Input environmental information to the information store
Environmental organizing and linking principle	Long-term working memory	Epigenetic system handling genetic information	Use information stored in the information store

Information store principle. Evolutionarily, it requires a huge storage facility to store a large amount of information in order to deal with the multifaceted situations that occur for survival purposes. This huge storage area, in human cognitive architecture, is known as long-term memory. Long-term memory not only stores knowledge but also influences how novel information is processed.

Borrowing and reorganizing principle. Based on stored information, humans are able to evaluate if the environment is threatening and produce corresponding actions to increase their chance of survival. In order to build such a database, Sweller and Sweller (2006) suggest that almost all of the semantic information stored in an individual's long-term memory has been borrowed from another's long-term memory through various forms of instruction. For early humans, such borrowing was achieved by imitating other humans. Supporting evidences can be found in neuroscience research, which indicates that humans are born to learn via observations (see also section 2.5.4 about embodied cognition and the mirror neuron system). The newly borrowed information is then combined with stored information to form a new construct (i.e. schema), providing ownership of one's own knowledge.

Randomness as genesis principle. This is a principle illustrating how human generate new or novel knowledge, through a strategy of random generation and test of effectiveness. The survival environment is ever changing which creates different challenges to survive. Constrained by only existing information, generated solutions can sometimes lead to dead ends. With the aim to increase the chance of survival and successful reproduction in a particular environment, humans have evolved the ability to generate new knowledge. In the natural environment, randomness happens in gene mutations (variation) where adaptive genes that can pass tests of the environment will be transferred to offspring. Such mechanisms (random generate and test for effectiveness) are also found in human cognition, where general problem-solving strategies such as means-ends analysis have been developed as part of primary biological knowledge (Sweller et al., 2011, pp. 32-38).

Narrow limits of change principle. A combination of the previous three principles can result in an information explosion, which can take a long time to process. In biological evolution, one minor change can lead to significant advantages. However, if a large number of changes occur at once, the offspring is unlikely to be successful (Strachan & Read, 2004; cited Sweller & Sweller, 2006). Too much information is also destructive to our information processing. If a large amount of information needs to be processed at one time, an individual might not be able to respond rapidly during emergency situations with fatal results. Hence, it is necessary to have a mechanism limiting the changes. Working memory, because of its very limited nature, provides an intermediary limiting information flow as part of human cognitive architecture. It reduces the information load by streamlining the amount of information that can enter working memory for active processing (Sweller et al., 2011).

Environmental organizing and linking principle. This principle deals with the executive system governing the information processing and storing. Previous principles illustrated how information is borrowed or created and stored in an unlimited store in the form of schemas. Schema formation does not happen randomly but according to environment factors. This principle explains why schemas in long-term memory are inactive until triggered by relevant environmental stimulus. A major advantage is that an appropriate response can be activated/ produced correspondingly in a timely manner. When information stored as schemas in long-term memory is retrieved, working memory is relatively increased.

1.3 Cognitive Load Theory

Cognitive load theory was first developed by John Sweller and colleagues in the 1980's. Simon and Simon (1978; cited in Ayres & Sweller, 1990) found that problem solvers tended to use means-ends analysis (a heuristic search that find the differences between current and goal state, see Simon & Newell, 1971) when presented with novel problems. However, means-ends analysis imposes a very heavy cognitive load onto problem solvers, as they are required to simultaneously consider the current state, goal state, and the differences and the relations between them. Initial studies focused on providing explanations and alternatives (such as worked examples and goal-free problems) on why problem solving was not a good way to learn (Sweller, 1988). Through these studies the theory was built on the properties of human cognitive architecture, especially the roles played by working memory and long-term memory. In more recent times, considerations of evolutionary cognition have been added to the theory.

The difference between biologically primary knowledge and biologically secondary knowledge is an important distinction from a cognitive load theory perspective. It provides educators with some ‘instructional ramifications’ (Sweller et al., 2011, p. 5). Cognitive load theory is now considered to apply specifically to biologically secondary knowledge, which to learn is effortful and requires direct instructions. In contrast, primary knowledge is learnt instinctively and does not need instructions (Paas & Sweller, 2012). Recent developments of cognitive load theory assume that the role of instruction is to build a large information store (LTM) so that the environmental organising and linking principle can occur. As a result, a primary focus of cognitive load theory is to provide guidelines that enable instruction to be designed that saves effort in processing new information in working memory and facilitates schema construction as a consequence.

1.3.1 Cognitive load

As discussed in section 1.1.1, new information is processed in a limited working memory. When learning something new, relevant information is integrated with prior knowledge and then stored in long-term memory as new knowledge in the form of schemas. Processing information in working memory requires working memory resources to be assigned. Any load imposed onto the human cognitive system during learning is considered to be cognitive load. Cognitive load affects the amount of resource allocation. If more resources are required to deal with the load created, then fewer resources are available for schema construction. Two main sources of cognitive load are intrinsic cognitive load and extraneous cognitive load. These are described in the later section, but first element interactivity is discussed, as it is an important construct in cognitive load theory.

1.3.1.1 Element interactivity

An element is the simplest form of a single learning item that needs to be learnt or processed, or has been learnt or processed (Chandler & Sweller, 1996; Sweller et al., 2011). Element interactivity (Hasler, Kersten, & Sweller, 2007; Sweller et al., 2011; Sweller & Chandler, 1994; Sweller et al., 1998) can be determined by counting the number of interacting elements that a person has to process simultaneously in working memory when dealing with new information. Sometimes the elements can be learnt or processed in isolation from each other, for example learning an alphabet, where element interactivity is considered low. However, if a large number of elements need to be processed simultaneously (e.g. solving a complex mathematics formula), then element interactivity is considered high, and the learning process becomes more demanding on working memory resources. Element interactivity explains why some new materials are more difficult to learn than others (Sweller & Chandler, 1994). However, after a new schema is created consisting originally of a number of sub-elements that were previously treated individually, retrieval can be treated as one element in the working memory. Hence, element interactivity depends on a learner's prior knowledge (Sweller et al., 2011). Individual differences in prior knowledge mean that the element interactivity experienced by individual learners varies accordingly.

1.3.1.2 Intrinsic cognitive load

Intrinsic cognitive load (Sweller, 2010; Sweller et al., 1998) is the load imposed by the intrinsic nature of the information, and is determined by levels of element interactivity. In other words, intrinsic load is dependent upon the basic structure of the materials over which instructors have no controls (Sweller et al., 1998). For materials that are high in element interactivity, a high working memory load will be imposed during the learning process, resulting in a high intrinsic cognitive load. As a consequence, if a learning task has very high element interactivity, an individual's working memory capacity may become overloaded and learning is severely limited. In contrast, materials low in element interactivity generate low intrinsic cognitive load, and are easier to learn.

Sweller et al. (2011) emphasise that intrinsic cognitive load needs to be distinguished from levels of task difficulty. While element interactivity deals with the number of elements that *interact* with each other, task difficulty relates to the *number of items* to be processed. A task low in element interactivity can still be difficult if there are a large number of items. Sweller (1994) and Sweller et al. (2011) used second language learning as an example. When learning a second language, each word to be learnt in a foreign vocabulary is low in element interactivity; however, the large bank of vocabulary makes learning a second language a difficult task.

Germane cognitive load. To learn, working memory processes must be applied to the learning process (schema formation). Consequently, a specific form of cognitive load is required. Historically, Sweller et al. (1998) referred to this load as germane cognitive load (see also Sweller, 2010), and was considered independent of intrinsic cognitive load. However, a recent reconceptualization by Sweller et al. (2011) argues that this load should be considered as resources devoted to information that is germane to learning, and hence should be referred to as germane resources instead of germane cognitive load. Although notably many researchers are yet to adopt this definition, germane resources are the resources allocated by the learner to deal with intrinsic cognitive load. Consequently, germane resources are linked directly to intrinsic cognitive load. In this thesis, the term germane cognitive load is used in order to preserve the meaning from the cited authors.

1.3.1.3 Extraneous cognitive load

In contrast to intrinsic cognitive load, extraneous cognitive load is an unnecessary load caused by poorly designed instructional materials that overlook working memory limitations (Sweller et al., 2011; Sweller et al., 1998). Poorly designed materials divert students' cognitive resources away from schema acquisition and the concepts to be learnt into dealing with the demands of the instructions. Examples include learning through problem-solving and split-attention materials (discussed in more detail in Section 1.3.4). The level of extraneous cognitive load is decided by the number of interacting elements embedded in the instructions (Sweller et al., 2011). If learners have to deal with high extraneous load, fewer cognitive resources will be available for schema acquisitions and automation as a result. Hence extraneous cognitive load is considered as a negative load that hinders learning. Moreover, since this load is imposed through the instructional material representation, instructors have the power to alter this load through altering their instructional methods.

1.3.1.4 Additivity of intrinsic and extraneous cognitive load

By definition, intrinsic and extraneous cognitive load are two separate loads imposed by two independent sources. However, from learners' perspective, it may be difficult to clearly distinguish the source of loads. The resources devoted to dealing with these two types of loads come from the same working memory pool and are undifferentiated from each other (Sweller et al., 2011). Nevertheless, in order to learn most effectively, the total cognitive load generated is an important consideration because of the working memory limitations. If either or both intrinsic and extraneous load requires many working memory resources, learning may be inhibited. Previously, total cognitive load (Sweller et al., 1998) was calculated by adding together intrinsic, germane, and extraneous load, but because of the recent argument that germane and intrinsic are not independent of each other, adding intrinsic and extraneous cognitive load together determines the total cognitive load generated (Sweller et al., 2011).

1.3.2 Cognitive load measures

As cognitive load theory developed it became important to measure cognitive load. Table 1.2 summarises a number of methods to measure cognitive loads. In general, measurements can be categorised into direct and indirect measures.

Indirect measures were first investigated in the 1980s. With the assumption that there was a correlation between performance (including strategy employed, errors committed and performance accuracy) and cognitive processing loads, it was believed that measuring performance provided good indicators of the total cognitive load.

However, some cognitive load theorists believed that measuring performance was not enough in cases where additional demands, such as through the instructional materials, are imposed (Paas, van Merriënboer, & Adam, 1994). Performance did not necessarily reflect the amount of effort invested or difficulty experienced. Hence, Paas (1992) developed a subjective 9-point mental effort rating scale to capture an index of cognitive load. Subjective measures assume that learners are able to reflect on their own cognitive processes, and there is a correlation between the subject measures and actual cognitive load. Initially the rating scale was created to measure mental effort and task difficulty, and has been successfully employed for two decades in cognitive load theory research (see Sweller et al., 2011; van Gog & Paas, 2008).

Table 1.2.
Summary of Cognitive Load Measurements

	Measurement	**Reference**
Indirect	Performance accuracy	Chandler and Sweller (1991); Sweller and Chandler (1994)
	Number of productions (strategies used)	Ayres and Sweller (1990); Sweller (1988);
	Categorisation of solutions and acquisition time	Sweller (1994)
	Number of errors	Ayres and Sweller (1990)
	Error rate	Ayres (2001)
Direct	*Subjective measurement:*	
	Paas' mental effort rating	Paas (1992), Paas and van Merriënboer (1994a), Paas, van-Gerven, and Wouters (2007)
	Task difficulty rating	Ayres and Youssef (2008); Marcus, Cooper, and Sweller (1996)
	Leppink's 13-item cognitive load measure	Leppink, Paas, van Gog, van der Vleuten, and van Merriënboer (2014), Leppink and van den Heuvel (2015)
	Objective measurement:	
	Heart rate variability (HRV)	Althaus, Mulder, Mulder, van Roon, and Minderaa (1998),
	Event-related brain potentials (ERP)	Trejo, Kramer, and Arnold (1994)
	Pupillary dilation (TEPRs)	Beatty and Lucero-Wagoner (2000), Klingner, Kumar, and Hanrahan (2008), van Gerven, Paas, van Merriënboer, Hendriks, and Schmidt (2003)
	Blink rate	Holland and Tarlow (1972)
	Electroencephalography (EEG)	Antonenko, Paas, Grabner, and van Gog (2010)
	Functional magnetic resonance imaging (fMRI)	Paas, Ayres, and Pachman (2008; cited Sweller et al., 2011), Whelan (2007)
	Eye tracking	Schwonke, Berthold, and Renkl (2009), Hyönä (2010), Mayer (2010), Ozcelik, Arslan-Ari, and Cagiltay (2010), van Gog and Scheiter (2010)

The Paas (1992) 9-point mental effort rating scale and other derivatives are 1-item measures, and provide only a total cognitive load index. They fail to identify the different type of load (intrinsic or extraneous). Both experimentally and theoretically, it has been considered helpful to measure the different types of load. This goal has been somewhat elusive, but recently Leppink, Paas, van der Vleuten, van Gog, and van Merriënboer (2013) developed a 10-item cognitive load questionnaire (and it became a

14-item questionnaire later, see Leppink et al., 2014; Leppink & van den Heuvel, 2015) that show promise in measuring both intrinsic and extraneous load separately.

Psychophysiological techniques assume that physiological responses are sensitive to changes in cognitive functioning. However, initial attempts to use physiological measures have not been promising as they have been unable to find subtle changes in cognitive load compared with subjective measures (see Althaus et al., 1998; Paas, Tuovinen, Tabbers, & van-Gerven, 2003; Paas & van Merriënboer, 1994a). More recent studies have used more updated technologies such Electroencephalography (EEG, see Antonenko et al., 2010) for greater effect. Also an increasing number of researchers (e.g., Hyönä, 2010; Mayer, 2010; Ozcelik et al., 2010; Schwonke et al., 2009; van Gog & Scheiter, 2010) have employed eye-tracking methods to measure learners' attention and cognitive load spent on tasks. It is believed that eye-tracking data can provide details on how learners interact and process the learning materials (for details, see van Gog & Scheiter, 2010), and be useful in understanding various cognitive load effects. Despite the potential of physiological methods, especially in providing on-line measures of cognitive load, more data is needed to verify the accuracy and effectiveness of this method. Furthermore, subjective measures have the added advantage of being very easy to administer without interventions and considerably cheaper than costly equipment (Paas & van Merriënboer, 1993).

1.3.3 Instructional efficiency

Based on Paas's 9-point mental effort rating, Paas and van Merriënboer (1993, 1994a) developed an efficiency measure (E) by combining performance and mental effort measures (see also Paas et al., 2003). Efficiency provides a measure of the cognitive cost of learning, which can serve as an indicator of schema acquisition and automation. High performance and low mental effort would result in having high instructional efficiency ($E > 0$), whereas low performance and high mental effort would result in having low instructional efficiency ($E < 0$). Efficiency provides an additional measure to the effectiveness of instructional interventions.

$$E = \frac{Z_{Performance} - Z_{Mental\ Effort}}{\sqrt{2}}$$

1.3.4 Cognitive load effects

As described above, cognitive load theory started with investigations into why problem solving was not effective and looking for alternative instructional strategies. Over decades of experimental examinations and demonstrations, researchers have generated a number of cognitive load effects which can guide effective instructional procedures. Since the major interest of this thesis is on instructional animations which will be discussed thoroughly in Chapter 2, only some out of a large pool of effects will be discussed to briefly outline the main concepts and issues. This section will be further divided into three subsections: effects dealing with extraneous cognitive load, effects dealing with intrinsic load and effects promoting germane load.

1.3.4.1 Strategies to reduce extraneous cognitive load

Worked example effect

Extraneous cognitive load is caused by the instructional designer, or teacher, using poorly designed materials or promoting inefficient learning strategies. For example, trying to learn through discovery methods like problem solving has shown to be highly inefficient (see Kirschner, Sweller, & Clark, 2006). As an alternative worked example has been found to facilitate more learning compared with problem solving. This advantage has been called the worked example effect (Sweller et al., 2011).

Sweller et al. (2011) argued that the worked example effect flows directly from the cognitive architecture. A worked example, which contains a problem state and step-by-step solution to that problem, enable learners to learn the key aspects about the problem and the solution (Atkinson, Derry, Renkl, & Wortham, 2000; Renkl, 2014; Sweller et al., 2011, pp. 99-100). Flowing from the borrowing and reorganising principle discussed above, learners (especially novices) can rapidly be equipped with domain specific schemas (Sweller et al., 2011) through borrowing the key aspect from the worked example in order to solve other problems in an efficient manner. As a consequence, worked examples reduce learners' extraneous load and facilitates learning. This strategy has been shown to be beneficial in many subject areas such as mathematics (Cooper & Sweller, 1987; Sweller & Cooper, 1985), music (Owens & Sweller, 2008), second languages learning (Diao, Chandler, & Sweller, 2007), literature (Kyun, Kalyuga, & Sweller, 2013; Oksa, Kalyuga, & Chandler, 2009), arts appreciation (Rourke & Sweller, 2009) and essay writing (Hübner, Nückles, & Renkl, 2010).

More recently, researchers have applied worked-examples in designing instructional multimedia materials and again proved their effectiveness (Nievelstein, van Gog, van Dijck, & Boshuizen, 2013; Renkl, 2002; Schworm & Renkl, 2006; Spanjers, Wouters, van Gog, & van Merriënboer, 2011; see also Renkl, 2014). However, in some cases, worked examples have found to be relatively ineffective, for example when split-attention materials are used (e.g. Tarmizi & Sweller, 1988).

Split-attention effect

Split-attention occurs when learners are required to split their attention across at least two sources of information in order to learn from the materials (Sweller et al., 2011). Such separation of information can occur in physical form and/or temporal form (particularly in instructional animations) (Ayres & Sweller, 2005, 2014; Mayer, Deleeuw, & Ayres, 2007). Split-attention materials create extraneous load by requiring learners to perform mental integration, and in consequence reduce learning effectiveness (Ayres & Sweller, 2005). The research on the split-attention effect started with an investigation into worked example effect where Tarmizi and Sweller (1988) found that worked examples were not effective for geometry problems because traditional solutions were positioned seperatedly below the diagrams. Moreover, split-attention effect could be reduced by inserting the solution steps into the diagram (integration of materials) at the relevant points. In such a way, extraneous search processes were reduced, and worked examples in integrated format became effective compared to non-integrated materials and conventional problem-solving. This effect has been replicated many times (Ayres & Sweller, 2014) leading to the conclusion that extraneous cognitive load can be reduced by avoiding split-attention materials.

The modality effect

The modality effect occurs when spoken text plus diagrams/pictures leads to superior learning than written text plus diagrams/pictures. Sweller et al. (2011) described the modality effect as an alternative way dealing with the split-attention. The basis for this effect is found in Paivio’s dual coding theory (for details, see Clark & Paivio, 1991; Paivio, 1971, 1991) and Mayer’s model of multimedia learning (Mayer, 2001, 2008, 2014a), which show how different forms of information are processed by different channels (visual and auditory) and integrated into a unified representation in working memory. Whereas visual text and pictures are initially processed in the visual channel, replacing visual text with auditory text can reduce the total load imposed on the visual channel, and relatively expand working memory overall by using both channels (Sweller et al., 2011). The modality effect has been found many times (e.g. Hasler et al., 2007; Jeung, Chandler, & Sweller, 1997; Kühl, Scheiter, Gerjets, & Edelmann, 2011; Mousavi, Low, & Sweller, 1995; see also Ginns, 2006). This effect is particularly important to instructional multimedia, as using multiple sources of materials in different mediums is a major characteristic. Making good use of the modality effect by offloading part of the explanatory text from visual form to auditory form can enhance learning.

The redundancy effect

The modality effect may not occur if redundant materials are used. With many instructional animation/multimedia presentations, it is common to find materials using both narrated and written forms of identical text simultaneously. However, Kalyuga, Chandler, and Sweller (2004) demonstrated that using both spoken and written forms concurrently resulted in less efficient learning than the spoken form alone. Gerjets, Scheiter, Opfermann, Hesse, and Eysink (2009) also obtained a similar result where written-text only instruction was superior to spoken plus written text instruction. These findings are examples of the redundancy effect (Kalyuga & Sweller, 2014). When the information that can be understood in isolation is presented using multiple sources, at least one source will be redundant (Sweller et al., 2011). Having to process redundant information requires extra mental resources and causes extraneous cognitive load (Kalyuga & Sweller, 2014; Sweller et al., 2011).

Expertise reversal effect

Whereas redundancy can be caused by replicating information, providing learners with instruction that is unnecessary for the level of expertise can also be redundant. The expertise reversal effect occurs when information beneficial to novice learners becomes redundant for more knowledgeable learners (Sweller et al., 2011). Based on the worked-example effect and the modality effect, detailed descriptions of visualisations with step-by-step narrated guidance should be beneficial to learning from instructional animations. However, Kalyuga, Chandler, and Sweller (2000) found that these advantages only occurred with novice learners. When learners became more knowledgeable, the advantage of work-examples and modality effects disappeared (see also Kalyuga, 2014). Later Kalyuga, Chandler, and Sweller (2001) found a similar redundancy effect with discovery learning, which was not effective for novices, but effective for more knowledgeable learners. Spanjers et al. (2011) also demonstrated an expertise reversal effect with instructional animation where segmented animation was more efficient than continuous animation for novices but not for knowledgeable learners. These results suggest that detailed instructions are most beneficial for initial schema acquisition. After learners become more knowledgeable, the detailed information becomes redundant and increases learners' extraneous cognitive load when they try using it.

1.3.4.2 Strategies to reduce intrinsic cognitive load

Much of the research into cognitive load theory has focused on reducing extraneous cognitive load. However, some research has been devoted to reducing intrinsic cognitive load. As intrinsic cognitive load is imposed by the materials' nature, intrinsic cognitive load is fixed relative to the learner unless instructors change the learning task itself (to alter element interactivity) or through additional learning (facilitate schema acquisition).

Pre-training (effect)

Providing pre-training is one way to change the learners' knowledge base. As discussed in Section 1.3.1.1, element interactivity is dependent upon a learner's prior knowledge. Based on the organising and linking principle discussed in Section 1.2.3, the more prior knowledge (in form of schema) there is in LTM, the easier it is to process new information. Hence, equipping learners with relevant schema by providing pre-training can reduce intrinsic load and leave more cognitive load for acquiring new schemas (Sweller et al., 2011). Successful research examples of pre-training include providing some basic information on how simple brakes work to learn about the overall brake systems (Mayer, Mathias, & Wetzell, 2002) and initial training on spread-sheets, when learning mathematics through spread-sheets (Clarke, Ayres, & Sweller, 2005).

Isolated elements effect

Besides changing the knowledge base, reducing element interactivity is an alternative strategy to reduce intrinsic load (Sweller et al., 2011). This can be achieved by initially segmenting the interacting elements into isolated elements, and then, sequencing the materials over time until full interacting elements are experienced. Results from Pollock, Chandler, and Sweller (2002) showed that participants in an isolated-elements group outperformed those in interacting-elements group using the same learning material. Furthermore, the effect is stronger in tasks that are high in element interactivity. Moreover, the isolated element effect is only effective for novice learners, as more knowledgeable learners will more likely demonstrate an expertise reversal effect (Ayres, 2006).

1.3.4.3 Strategies to promote germane cognitive load

Whereas the previous two sections describe methods to reduce cognitive load, some strategies have focused on promoting germane cognitive load. Germane cognitive load, as discussed, are the cognitive resources devoted to dealing with intrinsic cognitive load to generate schema acquisition (learning). One example is variation. It has been shown that when students are presented with a varied problem set, they need to deal with deeper structures and relations between the elements. Hence, variations can engage learners into deeper processing and to build more flexible and well-connected schemas (Paas & van Merriënboer, 1994b; Sweller et al., 1998).

Self-explanations (Renkl & Atkinson, 2003) and redirecting attention during learning (van Merriënboer, Schuurman, de Croock, & Paas, 2002) can also promote germane cognitive load. Since better understanding can lead to more well-structured schemas (Owens & Sweller, 2008), self-explanation provide deeper understanding of the connected elements (Sweller et al., 2011). Similarly, redirecting learners' attention allows learners to focus only on relevant materials. It can reduce extraneous cognitive load and consequently leave more working memory capacity available to promote germane cognitive load.

1.4 Chapter Summary

Learning is a human cognitive activity and occurs within the limitations of human cognitive architecture. To understand how humans learn, it is important to understand how the human cognitive architecture functions. With knowledge from cognitive psychology, it is known that learning occurs when information in various forms is processed in working memory and then stored in long-term memory. Working memory is known to have a limited capacity and duration. Humans cannot hold unlimited amounts of information for an infinite duration. Although, according to Baddeley and Hitch's working memory model, various forms (e.g. visual and audio) of information are processed differently, none of these processes can exceed the working memory limitations. Nonetheless, long-term memory where a vast amount of knowledge is stored is unlimited in capacity. This 'warehouse' stores information in the form of schemas, which guides the way as an individual interprets new information and helps facilitate information processing in working memory.

Evolutionary educational psychology provides insights on how learning happens within the boundaries of human cognitive architecture. Making reference to Geary's work, some knowledge (biologically primary knowledge) is innate to humans while some others (biologically secondary knowledge) require much effort in learning them. This knowledge division, together with Sweller's five principles of natural information processing, underpin the cognitive load theory, which provides instructional recommendations to instructors. This theory distinguishes the different sources and types of cognitive loads (intrinsic, extraneous, and germane) imposed on working memory during learning.

The major goal of the cognitive load theory is to facilitate schema acquisition and automation by reducing intrinsic and extraneous cognitive load, and increasing germane cognitive load. To do this a number of cognitive theory effects were identified and discussed, with the following implications. Providing detailed solutions (worked examples) are effective for guiding novice learners to build schematic knowledge. Also explanatory/comprehensive information should be integrated so as to reduce extraneous cognitive load in performing mental integrations. Furthermore, instead of overloading the visual channel, it is recommended to offload part of explanatory text stored in the visual channel (written form) onto the auditory channel (spoken form) to reduce extraneous load. However, identical information should not be presented in both channels wasting limited mental resources processing redundant information.

In addition, learners' prior knowledge and experience play an important role in the instruction strategy, hence sequencing instructional guidance based on learners' expertise is recommended. To reduce intrinsic load, it is recommended to segment high-element-interactivity materials into low-element-interactivity materials, and provide pre-training to novices. Lastly, to promote germane cognitive load, providing varied problem sets (*variation*) enable students to focus more on deeper processing and to build better schemas. In addition, self-explanations and redirecting students' attention are alternatives to promoting germane cognitive load. Instructors are encouraged to employ these strategies to foster stronger schema acquisition.

Chapter 2: Instructional Animations and Embodied Cognitions

Animated visualisations are commonly used to aid instructions; however, experimental studies have provided inconsistent results and left instructional animation somewhat as an enigma (Ayres & Paas, 2007b). The aim of this thesis is to contribute to this research field by investigating factors that influence animation research results. There will be three major segments in this chapter. The first segment (Section 2.1-2.2) provides a general review on animation definitions, and their effectiveness with a major focus on dynamic vs. static comparison. The second segment (Section 2.3-2.4) illustrates some counter-examples showing animations are not as effective as claimed. Reasons for the failure will then be provided. The final segment (Section 2.6) examines key factors in designing effective animations. These investigations will mainly be situated in the framework of cognitive load theory.

2.1 Definition and Characteristics of Animations

Animations are considered to be visualisations that are comprised of a series of frames (static pictures). Each frame contains incremental changes appearing as an alternation of the previous frame. By showing all the frames in sequence at high speed, human brains automatically fill the gap between each frame, and a perception of movement is produced. This gives animations a capacity to depict temporal changes directly which is the greatest distinction between animations and static pictures (Bétrancourt & Tversky, 2000; Castro-Alonso et al., 2014a, p. 552; Lowe, 2003).

Lowe (2003) identified three main types of change depicted in graphic entities in animations: *transformation* (change of form e.g. size, shape, colour and texture), *translations* (change of position that can be perceived with respect to the border of the display), and *transition* (change of presence i.e. appearing or disappearing from the display). Combinations of these changes can depict apparent motion (e.g. kinetic motion in physics, see Beichner, 1996) or movements for observational learning (e.g. observing teaching videos can help pre-service teachers to link theory and practice, see Blomberg, Sherin, Renkl, Glogger, & Seidel, 2014; Seidel, Blomberg, & Renkl, 2013). They can also show very small changes such as micro-steps during geometrical transformations (Thompson & Riding, 1990), imperceptible mechanisms at the molecular level (Barnea & Dori, 1999; Stern, Barnea, & Shauli, 2008; Yarden & Yarden, 2011), abstract concepts like statistics models (Wender & Muehlboeck, 2003), and temporal concepts or rules containing temporal progression, movement or spatial relations (Bétrancourt & Tversky, 2000; Wender & Muehlboeck, 2003).

In current research, the terms *animations* (e.g. Castro-Alonso, Ayres, & Paas, 2015a; ChanLin, 1998; Lowe & Schnotz, 2014; Mayer & Anderson, 1991), *videos* (e.g.Arguel & Jamet, 2009; Blomberg et al., 2014; Derry, Sherin, & Sherin, 2014; Schwan & Riempp, 2004), *animated diagrams*/ *demonstrations*/ *representations*/ *models* (e.g. Jones & Scaife, 2000; Kalyuga, 2008; Lowe, 2004; Lusk & Atkinson, 2007; Wender & Muehlboeck, 2003; Wouters, Paas, & van Merriënboer, 2008), and *dynamic pictures*/ *graphics*/ *representations*/ *displays* (e.g. Ainsworth & VanLabeke, 2004; Akinlofa, Holt, & Elyan, 2014; Boucheix & Schneider, 2009; Castro-Alonso et al., 2015b; Lowe, 2004), are often used interchangeably. Therefore in this thesis the term animation will be used generally to represent all these different categories. However, specific categories will be described when necessary to provide important details of the research materials reported in this literature review.

2.2 Initial Learning Advantages Predicted for Using Instructional Animations

Generally visualisations (both animations and static pictures) have an advantage in displaying spatial information. Larkin and Simon (1987) compared the use of textual and graphic representations (*sentential representation* and *diagrammatic representation*) at a theoretical level. They argued that diagrams offer advantages for recognition and search processes when solving problems in content areas where information is naturally organised in spatial ways (e.g. topological and geometric relations). They can help the viewer to detect essential elements unambiguously on a diagram and reduce the need to search for multiple information elements relating to a single idea, as this information is grouped into a single location (see also Rieber, 1991a). Animations, comprising of a set of static pictures, carry the same advantages.

With insight from Johnson-Laird's theory of mental models (1983; cited in de Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005), de Beni et al. argued that visualisations (both animations and static pictures) are useful in showing the complex relationship between elements embedded in text, and help construct internal representations (see also Bétrancourt & Tversky, 2000; Hegarty, 1992). From a cognitive load theory perspective, any such help in constructing metal representations reduces a learner's cognitive load and allows more resources to be available for learning.

In comparison to static pictures, animations have more flexibility to be congruent to dynamic learning tasks because of their superior capacity in depicting changes directly (Rieber, 1990b; see also congruence principle by Tversky et al., 2002). The close match between showing an animation of a dynamic learning task and the actual task, as well as sophisticated developments in technology, have led many educationalists to predict that animations should be superior to statics, especially when dynamic change is involved (Chandler, 2004). Indeed empirical research has shown that instructional animations have advantages in learning certain subject domains, such as sciences, motor-related tasks and abstract concepts, as the following sections will report.

2.2.1 Animations vs. statics in learning sciences

There is supporting evidence showing animations are more effective than statics in learning scientific or mathematics concepts. For example, Wender and Muehlboeck (2003) compared the effectiveness of animations and static pictures in teaching statistics at a university introductory level. As in many studies of this nature to ensure comparability between interventions, the static pictures were obtained by taking screen shots from the animation condition. The results showed that students in the animation group significantly outperformed those in the static picture group. In addition, Bétrancourt et al. (2008), and Rebetez, Bétrancourt, Sangin, and Dillenbourg (2009) also found animation advantages with materials teaching formation of lightning and astronomy to university students respectively. Similarly, Pfeiffer, Scheiter, Kühl, and Gemballa (2011) found that animations were superior to key pictures when teaching university students zoology (recognising fish species). However, such difference disappeared after real-word experiment (visiting aquarium) was implemented.

Besides adult learners, Rieber conducted a series of experiments comparing animations with static pictures in learning Newton’s laws of motion, and the results demonstrated that animations: 1) were superior to static picture for young children’s learning (see, Rieber, 1990b, 1991b), 2) promoted incidental learning as it depicts more information such as variation of resultant speeds (see, Rieber, 1991b), 3) encouraged stronger motivation than static pictures (Rieber, 1991a). Thompson and Riding (1990) also showed a similar trend in that adding animation depicting a continuous shear geometric transformation yielded significantly better learning outcomes than depicting a set of discrete flashing pictures of the same process. Participants were students aged eleven to fourteen. Furthermore, Korakakis, Pavlatou, Palyvos, and Spyrellis (2009) compared the use of 3D pictures, 3D animation and interactive 3D animation in teaching various science topics to year-eight students. Although results showed no significant group differences in performance, animation was a more efficient tool, as the static picture group spent significantly more time and studied significantly less content than those in the two animation groups.

2.2.2 Animations vs. statics in learning motor-related tasks

Besides scientific concepts, evidence has been found that animations can be an advantage in teaching human movement related tasks. A number of studies have demonstrated the effectiveness of animations for learning motor skills, especially those involving hand manipulation tasks.

For example, Wong et al. (2009) conducted three experiments examining the role of animation in teaching origami (paper folding). The first experiment examined the effectiveness of narrated animations and narrated static pictures in teaching paper Viking helmet constructions to primary students. The animation, which was developed using an Adobe flash programme, showed a computerised paper illustrating the steps. For static pictures, there were two versions: single-static and double-static. The single-static version showed only one picture at a time whilst the double-static version showed both the initial and final stage of each step. User controls were added to replay each step. The results showed that participants in the animation group spent less time in the learning and revision phases, yet more of them were able to complete the construction, in comparison with the two static picture groups. The second and third experiment examined teaching paper basket construction to younger primary students using animations and static pictures without narrations and user controls. The results again showed animations had advantage in performance scores (i.e. the number of correct steps) in both experiments.

Later Ayres et al. (2009) conducted two experiments and obtained similar results. Instead of using computerised materials, which were less realistic, Ayres et al. used more realistic video-based materials. A video depicting human hands tying three separate knots sequentially (in the first experiment) and two videos disassembling two key-ring puzzles (in the second experiment) were employed in the animation conditions. Pictures captured from the video depicting each step were employed in the static picture conditions. No interactivities or narrations were embedded in any of the materials. The results from the first experiment revealed that participants in the animation group completed significantly more correct steps in significantly less time than those in the static picture group. In the second experiment, results showed that participants in the animation group outperformed those in the static picture group for completing the motor task, but also on recognition tasks identifying the previous or next move.

Similar results were obtained by Akinlofa, Holt, and Elyan (2013) who used a task based on building and altering a Lego model truck. Similar to the design of Wong et al. (2009), each picture depicted the initial and final stages of each step. In addition, text describing the process was added physically close to the pictures. In the animation condition, a long video-clip was segmented into 22 short clips with relevant text describing the process. The results showed that participants in the animation group committed significantly fewer errors and completed the tasks in less time than those in the static picture group.

Similarly, Castro-Alonso et al. (2015a) found animations to be superior to equivalent static pictures when learning Lego construction tasks. Two experiments were conducted with university level students under various environments (i.e. physical vs. virtual building environment, hands-observations vs. no-hands-observations, and different level of completeness.)

Garland and Sánchez (2013) also found overall animation benefits in learning knot tying from video-based material depicted from both face-to-face views and over-the-shoulder views. Results from eighty undergraduate student participants demonstrated that animations were superior to static pictures in both learning time and proportional correct steps (i.e. number of correct attempts divided by the number of total attempts). In addition, in regards to the viewing angles, the over-the-shoulder view was more beneficial than the face-to-face view in both learning time and correct steps made, suggesting that the angle of observation is an important factor.

Besides fine motor movement tasks, Arguel and Jamet (2009) found the same animation advantage in gross movement tasks (i.e. completing first-aid procedures). In their study, various types of first-aid techniques including suffocation treatments, compression and pressure bandaging, identifying pressure points, and applying tourniquets were examined in two experiments. Seventy-two undergraduate students were allocated into a picture group, an animation (video) group and a picture plus animation group. Results from the first experiment showed participants learning with animations achieved better results than those with static pictures. Moreover, those in picture plus animation group achieved the best result amongst all three groups. The second experiment used the same materials to examine the relation between the completeness of the materials and animations. The result showed presenting only key elements brought significantly better performance than presenting the complete information.

Furthermore, Michas and Berry (2000) replicated the animation advantage in first-aid (i.e. bandage a hand) skill when comparing materials in different conditions: text-only, line-drawings, text plus line-drawing, video, static-video (pictures captured from videos). Results showed that participants in the video condition and text plus line drawing condition (no significant difference between these two) significantly outperformed those in the other conditions in both motor-retention (includes accuracy and fluency) and recognition tasks.

2.2.3 Animations vs. statics in other learning areas

Asides from sciences and motor tasks, animations have also been found to be an advantage in other areas like writing a computer algorithm. For example, Byrne, Catrambone, and Stasko (1999) found an animation advantage in teaching computer algorithm to university students. Blake (1977) also showed that animation showing actual chess movements was superior to both a static picture condition and an animated-arrow condition showing a moving arrow indicating the path.

2.2.4 Animations vs. other instructional strategies

Besides comparisons with static pictures, animations have also been found to be superior to traditional classroom teaching (textbook plus verbal explanations or laboratory demonstrations) when teaching year-seven chemistry (Çığrık & Ergül, 2009), year-five elementary science (Dalacosta, Kamariotaki-Paparrigopoulou, Palyvos, & Spyrellis, 2009) and university introductory chemistry (Sanger, Brecheisen, & Hynek, 2001; Sanger, Phelps, & Fienhold, 2000). Furthermore, Lewi and Barron (2009) found animations reduced the completion time in solving worked examples about image editing using Photoshop among university level participants in comparison to without-animations.

2.3 Conflicting Results – Animations Not Superior to Static Pictures

The dynamic property makes animations distinctive from static pictures and as reported above there are many studies that report clear advantages to animations; however, an animation advantage is not always found and sometimes there can be a disadvantage (ChanLin, 1998; see also Ayres & Paas, 2007b).

For example, Rieber's *Newton's law of motion* studies that were discussed above yielded consistent animation advantage when teaching to elementary students, however, this advantage disappeared for adult learners (Rieber, 1996; Rieber, Boyce, & Assad, 1990). Also, in an earlier study in 1989 Rieber found no significant difference between animations with static pictures when teaching the same topic to elementary students.

Morrison and Tversky (2001) found no significant advantage when examining the effectiveness of three types of visualisations (i.e. text-only, static picture plus text, and animation plus text) in teaching social movements to university students. Participants in both animation and static picture groups yielded higher scores than those in text-only condition but no differences were found between them. This result was in line with those from Watson, Butterfield, Curran, and Craig (2010) who compared animations with a text-only condition and a picture plus text condition in teaching how to assemble an abstract and novel device. Again, an animation advantage was found when comparing with the text-only condition, but not compared with the pictures plus text condition. Furthermore, although Bétrancourt et al. (2008), discussed above, found a performance advantage for animation in teaching lightning formation, participants in the animation condition actually spent significantly more time studying the material than those in the picture condition. Consequently, this study suggests that animations may produce higher tests scores but at a cost of low instructional efficiency.

In addition to the above studies revealing no differences between animations and static pictures, some studies have even found that static pictures are actually superior to animations. For example, Mayer et al. (2005) conducted four experiments comparing the effectiveness of static picture (with text) and animations (with narrations) using various learning topics (cause and effect processes) including lightning formation, toilet tank mechanism, ocean wave formation, and machinery of car brakes. Results from all the four experiments consistently indicated that participants in the static picture groups scored higher on tests than those in the animation groups.

Scheiter, Gerjets, and Catrambone (2006) obtained a similar result in teaching probability to university participants under four conditions: text-only, static picture, animation, and imagery (i.e. require participants to mentally imagine the content of text material). Participants in the animation group spent the most learning time but achieved the worst scores amongst all four conditions. In comparing only static with dynamic displays, participants who chose to use animation sparsely achieved the highest score in novel problems; however those who chose to use animation frequently scored the lowest in both isomorphic problems (questions that are close to instructional examples) and novel problems. Participants having frequent use of static picture scored the highest on isomorphic problems. These results suggested that frequent use of animation is impedimental.

In addition, Jones and Scaife (2000) compared the effectiveness of animation and static picture in teaching physiology of the human heart (included mechanisms, labels and blood flow). Quantitative results showed that there was no difference in performance between the animation and static picture groups. Moreover, qualitative results found that participants from the animation group appeared to be overconfident which inhibited their effort at comprehensive learning. Also, the dynamic information in the animations (e.g. the flowing blood, the moving heart muscles and valves) prevented participants from paying appropriate attention to the information. Similarly, Koroghlanian and Klein (2004) found no difference in performance between animation group and static picture group, but students in the animation group spent significantly more time, suggesting that animation was less efficient.

Castro-Alonso et al. (2014b) conducted two experiments comparing animation and static pictures in memorising abstract symbol patterns. In the first experiment, results showed no differences between animation and static pictures. But in the second experiment where more symbols were included, static pictures were found to be superior to animation.

2.4 Why Animations Fail to Show an Advantage Compared to Static Pictures

2.4.1 Inconsistency amongst studies

As described above results showing the effectiveness of instructional animations are mixed. Rieber (1989) justified that the inconsistency amongst her studies was from i) material difficulty, ii) visual load (see also summary, Rieber, 1990a). Bétrancourt and Tversky (2000) reviewed 17 studies and suggested one of reasons for the inconsistency lies in the diversity of the research field itself. Such diversity may come from the definition and the function of animations, and the study objectives.

As discussed in Section 2.2 in describing the many definitions of animations, Ploetzner and Lowe (2012) argue that inconsistent definitions and characteristics of computer animations lead to inconsistent conclusions in the research. These authors proposed that future animation studies should be categorised according to their objective and characteristics.

Regarding learning objectives, Boucheix and Schneider (2009) found that participants in an animation group scored the lowest in two types of recall tasks but scored the highest in a functional mental model task (i.e. open questions integrating the knowledge from two recall tasks). This result suggests that static pictures are more effective in demonstrating tasks that required attention to specific details, whilst animations have an advantage in depicting dynamic tasks and delivering a more global picture of dynamic functions. This argument is consistent with the study of Williamson and Abraham (1995) where students in animation groups outperformed only in conceptual understanding but not in course achievement.

Morrison, Tversky, and Bétrancourt (2000) examined the learning materials from several research studies and concluded that there was often a discrepancy in the amount of information presented in static picture and animations, indicating a lack of equivalence. Tversky et al. (2002) subdivided the type of inconsistencies into the content and procedures displayed in animation and static picture comparisons. Some information, such as blood flow direction in learning biology (Large, Beheshti, Breuleux, & Renaud, 1996; see also Morrison et al., 2000; Tversky et al., 2002) or the prior knowledge required in learning Newton’s Law (Rieber, 1990b, 1991a, 1991b; see also Morrison et al., 2000; Tversky et al., 2002) were given in animation conditions but not for statics. Consequently, Tversky et al. (2002) argued that most of the successes of animation seemed to be due to the extra information conveyed rather than the animated information.

Results from Michas and Berry (2000), discussed above, also add support towards the equivalence issue. Despite the fact that animations were shown to be superior to static pictures, Michas and Berry (2000) also found a close gap between the video and text plus line-drawing conditions, suggesting that animations are able to deliver additional information that static pictures are not able to deliver by themselves, but static pictures can also be as effective as animation if additional text is included.

In addition, animations are often found to be interactive compared to statics thus introducing another moderating variable (e.g. Moreno & Mayer, 2007). Interactivity, or learner-control, is a feature that can benefit learners (see section 2.5), and it should be separated when examining the effectiveness of animations (see also Bétrancourt & Tversky, 2000; Tversky et al., 2002), especially if statics do not include this feature.

Furthermore, instead of comparing the use of animation directly with equivalent visualisation (either static picture, board-and-chalk or real-model demonstration), some studies have investigated the effectiveness of animations by using them as an add-on to other existing strategies such as learning from text (Sperling, Seyedmonir, Aleksic, & Meadows, 2003) or with normal classroom instructions (Sanger et al., 2000; Williamson & Abraham, 1995). Such combinations make it difficult to assess the direct contribution of the animations.

Also, some studies such as by Sanger and Greenbowe (1997), claim an animation advantage, by comparing their results with another researchers' study. Even though the same materials were adopted, the statistical validity of such comparisons is low.

The experimental design problems in comparing statics and animations are well documented in the meta-analysis of Höffler and Leutner (2007). 25 out of 57 animation studies were excluded from the meta-analysis due to 1) not comparing animations with equivalent static pictures, 2) mixing both types of visualisation, or 3) containing interactivity within animations. In other words, almost half of the existing studies identified had methodological issues, thus adding to the uncertainty surrounding animation-static comparisons.

2.4.2 Learning task characteristics

Besides issues connected to presentation characteristics, learning task characteristics can also play a role. As discussed above, visualisations (static pictures or animations) offer an advantage in an area where information is naturally organised in spatial ways. However, text offers advantages in areas where information is organised in the form of natural language (e.g. temporal or logical sequence), which is harder for pictures (static or dynamic) to represent (Larkin & Simon, 1987). Both presentation formats are weak in delivering logical sequence in natural language, which may explain why most of the studies in Section 2.3 (describing animations advantages) are not language based.

Despite some methodological issues and conflicting results, some indications of what type of learning content is best suited to animations have started to emerge recently. Although the Höffler and Leutner (2007) meta-analysis excluded many studies, those included revealed that 71% of the comparisons indicated a positive effect of animation over static pictures with a small effect size $d = .37$. In addition, although there were only five visualisations involving procedural-motor knowledge in the analysis, their mean effect size ($d = 1.06$) greatly exceeds the mean effect size of visualisation involving other knowledge types. This result suggests that there is superior learning when procedural-motor knowledge is involved. The findings of the meta-analysis and other recent studies into learning about tasks that involve human movement have led a number of researchers to argue that instructional animations can successfully tap into human innate ability to learn procedural motor skills through observations (see also Paas & Sweller, 2012). Their argument also includes a reason (*the transient information effect*) why animations may not be effective for non-human movement tasks, which is discussed in the next section.

2.4.3 The transient information effect

Sweller et al. (2011, p. 220) defined the transient information effect as ‘a loss of learning due to information disappearing before the learner has time to adequately process it or link it with new information’. Because of the dynamic nature in animations, information on one frame will soon roll over to another frame and can be lost quickly. In other words the information on each frame is not persistent or permanent (see also Ayres & Paas, 2007a, 2007b). Learners, when learning with animations, are required to interpret a larger amount of information within a designated time limit (in comparison to static picture which is more permanent), to attend and identify the changes, to hold and process the information in their working memory, and to integrate the new information with the old information simultaneously (Ayres & Paas, 2007b; van Gog et al., 2009; see also Castro-Alonso et al., 2015b). If the information is low in element interactivity (discussed in Chapter 1), learners may still be able to process and link the new information on the current frame with those in their existing schema. However, if the information is high in element interactivity, learners might need to spend excessive cognitive load to process, or even fail to process, the information before it disappears. In other words, transient nature may hinder learning from animations. Evidence of transient effects has also been found comparing spoken text (a fundamental form of transient information) with the more permanent written text (see Singh, Marcus, & Ayres, 2012). Sweller et al. (2011, p. 222) argue that this effect might be a key factor in explaining why instructional animations have not consistently revealed positive results.

2.4.4 Embodied cognition

Whereas the transient information effect provides a plausible explanation on why animation can be inferior to static pictures, it does not explain why motor-related tasks appear to be a special case and benefit from animations. The study by Wong, Leahy, Marcus, and Sweller (2012) that examined the effectiveness of animations in human movements learning (i.e. origami, paper folding task) and the role of transient information provides an insight. Consistent with other research discussed above, there was an overall animation advantage for this motor skills task. There was also an interaction between animation and segment length, where animations were superior in the short-segment condition, but not for the long-segment condition. For long segments, the effects of transient information were exaggerated leading to no advantage over statics.

Wong et al. (2012) argued, like other researchers (Castro Alonso et al., 2014a; Paas & Sweller, 2012; Sweller et al., 2011), that humans have an innate ability to learn via observations. Tasks that depict movements create an ideal observational learning environment which allows animations to be effective. Although, animations create transient environments, because humans have evolved to learn motor tasks fairly easily (Paas & Sweller, 2012), the normal cognitive load increases associated with animations do not occur because of these evolutionary advantages. In the case of the Wong et al. (2012) study, for short segments there were little transient effects but for long segments, transient information was a much greater disadvantage. To give a physiological explanation on how humans are able to learn from observations quite effortlessly, van Gog et al. (2009) proposed that the mirror neuron system plays a significant role.

2.4.4.1 Mirror neuron system

The mirror neuron system was first identified by di Pellegrino, Fadiga, Fogassi, Gallese, and Rizzolatti (1992) when studying goal-related movements in monkey brains (see also Rizzolatti, 2005; Rizzolatti & Sinigaglia, 2010). It was found that the motor system, which is responsible for generating movements, also has a mirroring capacity. In other words, the brain perceived the same meaning when the monkey observed an action and when the monkey repeated the action. The indirect evidence provided by data from monkeys has provided convincing evidence that the human motor system has the same capacity. However, it has been questionable whether the activation is from mirror mechanism, or simply reflects a preparation mechanism (Rizzolatti & Sinigaglia, 2010). With more advanced technology, the mirror mechanism in humans has been supported (see Mukamel, Ekstrom, Kaplan, Iacoboni, & Fried, 2010). One important feature of the system is that it does not simply repeat the action, instead, it is goal-related involving understanding of the action intention (see Rizzolatti & Craighero, 2004; Vander Wyk, Hudac, Carter, Sobel, & Pelphrey, 2009).

The mirror-neuron system provides a physiological explanation of how humans imitate body actions through observations, and provides an important implication: observational learning is a human innate ability and can be highly beneficial to teaching and learning (Paas & Sweller, 2012; van Gog et al., 2009). Watching an animation (in various forms such as videos or cartoons) can activate the mirror-neuron system and lead to learning without increases in cognitive load caused by the transitory effects of animations. Humans have evolved to deal with such information.

Paas and Sweller (2012) refer to the phenomenon to learn motor skills from animations as the *human movement effect*, and is one example of ‘using biologically primary knowledge to facilitate the acquisition of biologically secondary knowledge’ (see Chapter 1). It is argued that the brain and the body have evolved to learn through observation and be adept at spreading the load in doing so (Clark, 2008).

2.4.4.2 Hand observation – hand movement effect

It is notable that many of the human movement studies have included tasks that require hand manipulations such as making origami figures (Wong et al., 2009). As part of the current study it was investigated whether it is necessarily to actually observe the hands or not in the animations. For example, Marcus et al. (2013) examined hand-observations in learning knot-tying under three conditions: with-hand animations (observing the hands tying the knots), without-hand animations (the hands blanked out), and without-hand static pictures. There was an overall animation advantage in learning how to tie the knots. Although the with-hand animation was found to be superior to the without-hand static picture on measures of cognitive load (lower) and efficiency (higher), there was no difference between the with-hand animation and the without-hand animation.

Another example is the study by Castro-Alonso et al. (2015a), who investigated hand-observations in learning Lego constructions using animation and static pictures. Although there was support for an animation advantage, participants who did not observe hands moving the Lego bricks in the presentations outperformed those who did observe hands. Furthermore the study by Castro-Alonso et al. (2014b), discussed above, explored the role of hand-observations in memorising abstract symbols patterns. The result from the first experiment showed that there was an interaction effect between presentations (animations vs. static picture) and hand-observations (with-hands vs. without-hands). The with-hands condition was significantly better than the without-hands condition for static picture, whereas there were no difference for the animation condition. The second experiment, where the materials were more complex, results also demonstrated an interaction effect between presentation and hand-observation, where learning without-hands was more efficient than learning with-hands in the animation condition but not the static picture condition. Drawing the two results together, it was suggested that showing hands in the material was beneficial for static pictures but not animations, where negative effects were observed.

In both the Castro-Alonso et al. (2015a) and Castro-Alonso et al. (2014b) studies, as well as the Marcus et al. (2013) study, showing hands in the animation did not lead to an advantage, but sometimes a disadvantage. From a cognitive load theory point of view it is possible that showing the hands are redundant, as the animations may contain sufficient information already. If redundant material is processed, then extraneous cognitive load will be generated, with the potential for a loss of learning (Kalyuga & Sweller, 2014).

2.4.4.3 The role of hands in embodied cognition

The hands play an important role in embodied cognition and their use can facilitate learning and the completion of tasks. For example Quinn and Ralston (1986) showed that hand movement played a role in spatial coding: consistent movements can enhance spatial location memory while inconsistent movement disrupted the performance. In conversations, Roth (2000) showed that pointing allowed students to express themselves more effectively. Goldin-Meadow, Nusbaum, Kelly, and Wagner (2001) also found that gestures lightened the load during math-explanation tasks and thus freed up more working memory space for a following word-memorisation task. Similarly, when solving spatial problems, gestures has been shown to facilitate spatial reasoning (Chu & Kita, 2011; Matlen, Atit, Göksun, Rau, & Ptouchkina, 2012). In such cases making gestures seems to activate the embodied cognition system. Embodied cognition is based on the notion that cognitive processes develop from goal-directed interactions between organisms and their environment (Barsalou, 1999, 2008; Glenberg, 1997) Research shows that visual and motor processes in the brain are active during the performance of cognitive tasks, suggesting that cognitive and sensorimotor processes are closely intertwined. Observing or making gestures leads to enhanced encoding and therefore richer cognitive representations are formed increasing task accuracy and efficiency, with lower demands on working memory.

From a cognitive load perspective, Macken and Ginns (2014) examined cognitive load when using gesture in learning about the human heart from text and pictures. There was no significant difference in participants' reported intrinsic load and extraneous load, but participants in a gesture group significantly outperformed students in a non-gesture group in both tests of terminology and comprehension, suggesting that gestures improved performance without adding further cognitive load.

2.5 Other Key Factors in Designing Effective Animations

The main aim of cognitive load theory is to produce effective instructional designs. A fundamental principle is to avoid extraneous cognitive load, whose creation is likely to reduce learning. There are a number of strategies identified in Chapter 1 that can avoid this unhelpful load such as using worked examples and avoiding split-attention (see Sweller et al., 2011). As discussed in this chapter, instructional animations may also be a source of extraneous load as they often contain transitory information. To reduce the transient information effect there are a number of well-researched strategies that are effective (Ayres & Paas, 2007a, 2007b), such as segmentation (dividing the learning materials into smaller parts, see Mayer & Chandler, 2001), cueing (see Kühl, Scheiter, & Gerjets, 2012) and learner-control (see Hasler et al., 2007; Kühl, Eitel, Damnik, & Körndle, 2014). However, these strategies may not be needed when human movement tasks are the learning topics. Nevertheless, there are also some other issues that need to be considered when designing effective animations.

2.5.1 Realism vs. schematic

The meta-analysis by Höffler and Leutner (2007), discussed above, showed that animations are more likely to be superior to equivalent static pictures when they are video-based and highly realistic. However, Tversky et al. (2002) argued that animation should be less realistic but simple and clear enough for observers to perceive and understand detailed changes. Scheiter, Gerjets, Huk, Imhof, and Kammerer (2009) also argued that a less realistic presentation (schematic visualisation) enhanced learning outcomes. Nevertheless, Höffler and Leutner also suggested that the realistic advantage might be partly caused by the nature of the learning task itself. The more realistic animations are mostly video-based which are also representational (i.e. the learning content is displayed explicitly) which has shown to be an effective feature. Whereas, many of the less realistic animations are not representational but have also yeided postive results. Excluding the influence of representational factors, Höffler and Leutner have also argued that animation should be schematic and to exclude unneccesary elements that might raise extraneous cognitive load.

2.5.2 Congruency and apprehension

With respect to the visual design, researchers (e.g. Bétrancourt, 2005; Large, 1996; Mayer & Sims, 1994; Tversky et al., 2002) recommend that animation should be congruent with the learning content, but apprehensive for learners to easily understand. From a visualisation viewpoint, being congruent is to represent the content closely with the task nature (e.g. depicting the actual blood flow direction when teaching about heart mechanisms), thus saving the learner from performing additional mental visualisations. However, if the task is very complex (e.g. simultaneously depicting muscle contraction and relaxation, blood flow, valve moving, haemoglobin and white blood cell flowing…etc), it may be very difficult to understand the meaning of the visual representations. Thus unnecessary details should be excluded. Consistent with the realism principle described in section 2.6.1, realistic visualisations are beneficial, but an important feature is to be constructive and be apprehensive for learners in order to understand and to build effective internal representations (Bétrancourt, 2005; Tversky et al., 2002).

2.6 Chapter Summary

In summary, animations have a capacity to depict movements and change, and thus it is often believed that animations should be superior to equivalent static pictures as instructional materials. Supporting evidence has been found in various learning areas including sciences, motor-related tasks, and other tasks involved temporal movements or abstract spatial concepts. However, there are also some counterexamples in the existing literature demonstrating that animations are not more effective than static pictures, and even sometimes detrimental to learning.

To explain these inconsistent results, researchers have provided several explanations such as methodological flaws, differences in material designs, presentation characteristic (transient nature) and task characteristics. One recent argument to support why animations seem to facilitate learning in human movement domains is based on embodied cognition, and taps into human innate learning ability. It suggests that animations, usually in combination with the observation of hands, may activate human mirror-neurons and consequently foster learning with less mental effort.

Besides embodied cognition effects, there are some other key factors in designing effective instructional animations such as using segmentation, learner control, using schematic but clear presentations, high congruency and more apprehensive materials. In general, the key is to not overwhelm learners with too much or too complex information, and ensure learners have both enough time and working memory resources available to thoroughly process the content.

Chapter 3: Gender Differences & Spatial Ability

3.1 Introduction

A main aim of this thesis was to investigate gender differences when learning from instructional animations. Gender differences in learning have been studied quite extensively, but similar to general animation studies, the results are often inconsistent. Some researchers have found significant gender differences in performance when learning key curriculum areas such as Language (Hyde & Linn, 1988; Kramer, Delis, & Daniel, 1988; Vogel, 1990), Mathematics and Science (Halpern et al., 2007; Hyde, Fennema, & Lamon, 1990; Hyde & Mertz, 2009; V. E. Lee & Burkam, 1996; Raymond & Benbow, 1986), Statistics (Schram, 1996; Woehlike & Leitner, 1980), Engineering (Legewie & DiPrete, 2014), and Geoscience (Orion & Kali, 2005); whilst other researchers have found no differences (Jee et al., 2013; Neuschmidt, Barth, & Hastedt, 2008).

As far back as the 1980's some researchers concluded that the gender issue had diminished due to environmental factors such as parent support (Raymond & Benbow, 1986), or the disappearance of sexual stereotypes in society (Boswell, 1985, p. 197). Yet, there is little consensus explaining why gender inconsistencies still exist today.

To identify influences underpinning potential gender differences in learning a number of factors have been investigated. Some studies have concluded that males and females process information differently. For example, Guillem and Mograss (2005) found that males tended to process information faster but at a more superficial level; whereas females tended to process information in a more thorough manner. This may explains why females report more difficulties than males when answering multiple-choice tests (Murphy, 1982; Smith & Miller, 2005), or prioritise their learning materials

(Speth & Brown, 1990). Other factors have included differences in preferences for interface designs (see Passig & Levin, 2000) and learning traits (see Terzis & Economides, 2011).

However, despite many potential factors influencing how males and females interact with multimedia instructions, differences in spatial ability has often been proposed as having a major impact (e.g. Barnea & Dori, 1999; Bonanno & Kommers, 2005; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006), and consequently is discussed in detail in this chapter.

This chapter is divided into two sections. The first section (3.1) examines the research detailing gender differences when learning in multimedia environments. The second section (3.2 – 3.6) examines the interactions between gender, spatial ability and learning environments.

3.2 Gender Differences in Multimedia Environments

3.2.1 Gender differences comparing statics with animations

An early study by Jacek (1997) examined whether instructional animations benefit males and females to the same degree in learning about physical science concepts. Jacek found that females improved significantly with animations compared to static pictures, whereas males showed no significant difference between formats. Furthermore, self-reported questionnaire data revealed that females in an animation group reported higher confidence doing science than those in a less animated group.

Later Yezierski and Birk (2006) conducted a similar study investigating firstly the effectiveness of computer animations in helping students to overcome misconceptions in chemistry, and secondly whether the animation worked equally for males and females. A large sized sample with 350 males and 369 females were recruited from middle school, high school and university students. The results showed that an embedded animation intervention significantly improved students understanding overall. Furthermore, they found a significant gender effect. Initially males outperformed females in a pre-test, but this difference disappeared after the intervention in a post-test. Their results suggested that learning with animation closed the gender gap (that existed at the beginning of the study) by boosting females learning performance.

Sánchez and Wiley (2010) also found a similar pattern in a study which investigated gender differences in learning science with animations. An interaction effect between gender and presentation type (animation vs. static picture) was not significant on sentence verification tasks (i.e. memory task), but was significant on a comprehension-measure assessing the concepts. Males outperformed females in the static picture condition but not in the animation condition. Furthermore, questionnaires revealed a consistent pattern in that males showed significantly more interest in learning than females in the static picture condition, but not the animation condition. Although there was no explicit analysis testing the difference between animation and static pictures for females, overall evidence suggested that animation boosted females' performance and thus narrowed the gap between males and females in learning science concepts.

Consistent with these findings, Cowards et al. (2012) found that university females significantly outperformed males in learning about vision science in an animation condition, but no difference was found with static pictures.

However, in contrast with the results above, Griffin, MacEachren, Hardisty, Steiner, and Li (2006) found that males learnt significantly better with animated maps than with static small-multiple maps; while for females no differences were found between formats. Furthermore, ChanLin (2001) examined gender differences comparing animations, graphs, and texts when learning descriptive knowledge and procedural knowledge in physic problem solving with eighth- and ninth- grade students. It was found that the visual format affected both descriptive and procedural learning amongst females but not males. Females were found to benefit equally from both animations and graphs for descriptive learning, but benefited slightly more from static pictures for procedural learning.

3.2.2 Gender differences with other multimedia materials

As well as static-animation comparisons, a number of studies have also examined gender effects with more broad-based multimedia learning materials. For example, Flores, Cowards, and Crooks (2010), investigated gender differences in learning a biology topic using different multimedia conditions. A gender-presentation format interaction was found where males performed better when material was presented in a dual module (visual plus auditory) while females performed better with materials presented in a single module (visual only). This result for females is contrary to extensive research by Mayer (2014b) who found that the dual mode is generally superior to a single mode (termed the multimedia effect by Mayer). Flores et al. (2010) concluded that females are more effective in processing verbal information, spending less mental effort in verbal-information processing which freed up more cognitive resources to processing spatial information. From a cognitive load theory perspective this explanation is plausible, because additional diagrams would be redundant (see Kalyuga & Sweller, 2014) leading to reduced learning if sufficient knowledge could be gained from the verbal information alone.

Riding and Grimley (1999) compared a reading-text plus reading-graph format (written group) with a narrated-text plus reading graph format (audio group). According to cognitive load theory a modality effect (see Chapter 1; see also Low & Sweller, 2014) should occur, where learners in the audio group outperform learners in the written group. However, the results showed that the expected modality effect only occurred according to gender and cognitive styles. For females those categorised as 'Wholist-Managers and Analytic-Verbalisers' experienced a multimedia effect, with the reverse modality effect for 'Wholist-Verbalisers and Analytic-Imagers', while the opposite effects were experienced by males according to cognitive styles.

Togo and Hood (1992) found an interaction between the data presentation format (tables vs. graphs) and gender, where females learnt significantly better with the tabular format than the graphic format, but males learnt slightly better (but not significantly) with a graphic format.

Barnea and Dori (1999) found that females learning with an animated molecular model on a computer improved significantly more than males on tasks related to 3D structures, but no gender differences were found in real-life demonstrations that were not computer based.

Males and females have also been found to have different spatial strategies. For example, Sandstrom, Kaufman, and Huettel (1998) demonstrated that females relied predominantly on landmark information, while males used both landmark and geometric information in virtual environment navigation tasks. Law, Pellegrho, and Hunt (1993) found that males were more sensitive to dynamic spatial reasoning (relative velocity judgement) than females, but not in static spatial reasoning (relative distance judgement). Gender differences have been found in game preferences. Bonanno and Kommers (2005) found males preferred games that demanded a higher visuospatial ability (including all localisation, orientation, mental rotation), target-directed motor skills and a greater reaction speed, whereas females prefer games that capitalised on perceptual speed, fine motor skills, and sequenced hand movements.

Other studies have investigated interactions between gender and cognitive styles Researchers (e.g., Chang & Yang, 2010; Cowards et al., 2012; Grimley, 2007; Pöhnl & Bogner, 2012, 2013; Riding & Grimley, 1999) have argued that males and females show different information processing characteristics because they have different cognitive styles. For example, Pöhnl and Bogner (2012) studied gender effects in learning from a computer or a textbook with the assumption that males and females process information in a different manner (i.e. cognitive style). Their results revealed that females in textbook-based learning reported a lower mental effort and thus higher mental efficiency than males, but no gender difference was found in computer-based learning. Although the cognitive style was not measured, the authors believed that there was a gender difference in schema storing and retrieving which contributed to the difference in information processing.

Qualitative result from Passig and Levin (2000) revealed there were differences in gender preferences for the interface design among kindergarten classes. Young girls prefer having text, colours, pictures and moderate-speed movements; whereas young boys prefer having user-controls, sharp moves and plenty of movements on the screen. Similarly, questionnaires from Terzis and Economides (2011) revealed that males would change their tendency in using computer based assessments based on its playfulness, usefulness, content and social influence; but females were mainly influenced by its playfulness, ease of use, content and goal expectancy. These results demonstrated males and females tended to have different learning traits and preferences towards learning with multimedia in interactions with computers.

Gender differences have also been found according to the gender of the learning agents used in multimedia presentations. Hoogerheide, Loyens, and van Gog (2016) investigated the interactions between a model's gender (the agent that performs the demonstrations in the learning materials) and the observers' gender. Hoogerheide et al. found a significant interaction for mental effort where males reported it was less effortful to study from a male model than females, but no differences were found for female students.

It has also been proposed that males and females have different preferences. For example, Bonanno and Kommers (2005) surveyed 367 students finding gender difference in game preferences. Males preferred games that demanded a higher visuospatial ability (spatial localisation, spatial orientation and mental rotation), target-directed motor skills and a greater reaction speed; whereas females preferred games that capitalised on perceptual speed, fine motor skills, and sequenced hand movements.

3.3 Gender Effects and Spatial Ability

Although cognitive styles and preferences have been proposed as possible explanations for gender differences in learning from multimedia materials, many researchers (e.g. Barnea & Dori, 1999; Bonanno & Kommers, 2005; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006) have argued that spatial ability is the major contributing factor. The following sections discuss spatial effects in more detail, starting with definitions of spatial ability.

3.3.1 Classification of spatial ability

In general, spatial ability is the ability to represent, generate, and recall symbolic/non-linguistic information (see Linn & Petersen, 1985; see also Coluccia &

Louse, 2004). Although definitions vary, one widely adopted definition emerged from the meta-analysis conducted by Linn and Petersen (1985), who categorised spatial ability into three components: *spatial perception*, *mental rotation*, and *spatial visualisation*. Firstly, *spatial perception* is the ability to determine the spatial relationships with respect to the orientation of oneself. Secondly, *mental rotation* is the ability to mentally rotate a two- or three-dimensional figure in an accurate and rapid manner. Thirdly, *spatial visualisation* involves multistep manipulations of spatial information, which often requires both spatial perception and mental rotation. Spatial visualisation tests often require participants to work out the spatial information, rotate the figure and keep track of multistep operations quickly.

3.3.2 Spatial orientation

Besides spatial ability as defined by Linn & Petersen (1985), there is another similar but distinctive concept namely *spatial orientation ability*, which is often confused with spatial ability (as claimed by Coluccia & Louse, 2004). Spatial orientation ability always involves movements, spatial environment and acquisition of knowledge about the environment (Coluccia & Louse, 2004; Martens & Antonenko, 2012). This skill is particularly important when learning map reading, way-finding in the real environment, or learning with virtual/ simulated environments. In regards to measurement, spatial orientation (just like spatial abilities) is a complex skill used to locate oneself with respect to a reference point or coordinate.

3.3.3 Visuospatial working memory

While some researchers approach spatial performance as a result of a general spatial ability, others specifically identify spatial memory as a main contributor (e.g.

Carroll, 1974). Referring to Baddeley and Hitch's working memory model (see Chapter 1), the visuospatial sketchpad within working memory is responsible for processing spatial information. Based on this model, some researchers (e.g. Darling, Della Sala, Logie, & Cantagallo, 2006; de Beni et al., 2005; Logie, 1995) have argued that spatial performance is highly related to the visuospatial working memory capacity. Quinn and colleagues (McConnell & Quinn, 2000, 2004; Quinn, 1994, 2008; Quinn & McConnell, 1996; Quinn & Ralston, 1986) have conducted studies to identify the actual mechanisms involved in spatial processing by examining the roles played by the various working memory components. Results consistently showed that spatial processing was dissociated from visual and movement processing (see also Smyth & Pendleton, 1989). However, Baddeley (1988, cited Quinn, 1994) has argued that the visuospatial sketchpad is sensitive to both visual and spatial coding with the nature of sensitivity depending on the nature of the task.

Another proposed role for visuospatial working memory in learning spatial information lays in the association between the visuospatial working memory and verbal working memory. Since the two type of working memories operate closely together under the central executive, some researchers have hypothesised that spatial information is partly memorised using verbal mechanism (Choi & L'Hirondelle, 2005; de Beni et al., 2005; Duff & Hampson, 2001; Pazzaglia & Cornoldi, 1999). de Beni et al. (2005) used a dual-task paradigm (i.e. requires participants to perform two tasks that require the same cognitive component) to examine the involvement of both verbal working memory (phonologic loop) and visuospatial working memory (visuospatial sketchpad) when comprehending and memorising spatial text. Results showed that recalling non-spatial text and recalling spatial text were both interfered with by the concurrent task, but depending on the type of text. Recalling non-spatial text was only impaired by the concurrent verbal task, but recalling spatial text was impaired by both the concurrent spatial task and the concurrent verbal task.

It has also been argued that people with a better visuospatial memory can create better visuospatial representations (Fernandez, 2000; cited in Deyzac, Logie, & Denis, 2006) or mental imagery (Kosslyn, 1980; cited in Logie, Zucco, & Baddeley, 1990), which help them to memorise more spatial information during learning. Such claims are consistent with the process of mental animation which is needed to animate a series of static pictures (Hegarty, 2014), which is discussed in Section 3.6.

Although studies into visuospatial working memory and spatial performance have not produced a common model, evidence has emerged that visual spatial memory has an important role to play in spatial ability.

3.3.4 Spatial measurement

In considering the role played by spatial ability in learning it has been important to use reliable measures of spatial ability. Based on the different cognitive processes involved, a number of different psychometric measures have been used. Table 3.1 categorises some commonly used spatial measurements based on the definitions discussed in the previous section. Different researchers have used a variety of categorisations and terms for spatial ability in their original papers. However, they are categorised in the table according to the outlined definitions rather than individual author categories.

Table 3.1.
Summary of Spatial Measurements for Respective Spatial Factor

Factor	Measurement	Reference
Spatial perception	Thurstone's Hand Test and Flag Test	Michael, Guilford, Fruchter, and Zimmerman (1957)[1]
	Piaget's Water level Task	Thomas and Jamison (1975), Willemsen and Reynolds (1973), Geiringer and Hyde (1976), Wittig and Allen (1984)
Mental rotation	Ekstrom et al's Card Rotations Test	Ekstrom, French, Harman, and Dermen (1976), Lohman and Nichols (1990),
	Shepard-Metzler 3D Mental Rotations Test	Shepard and Metzler (1971), Shepard and Judd (1975), Peters and Battista (2008), Vandenberg and Kuse (1978), Voyer et al. (1995)
	Thurstone's Figure Test	Lohman and Nichols (1990)
	The Purdue Visualization of Rotations (ROT)	Bodner and Guay (1997)
	Ekstrom et al's Cube Comparisons Test	Ekstrom et al. (1976)
Spatial visualisation	Ekstrom et al's Paper Folding Test	Michael et al. (1957), McGee (1979), , Lohman and Nichols (1990),
	Ekstrom et al's Form Board	Michael et al. (1957), Lohman (1988), Lohman and Nichols (1990),
	Thurstone's Cube Comparisons	Lohman (1988), Campos (2012)
	Punched Holes	French (1963), Michael et al. (1957),
Spatial orientation	Guilford-Zimmerman test of Spatial Orientation	McGee (1979), Lohman (1988)
	Thurstone's Space Task	Willis and Schaie (1986)
Visuospatial working memory	Corsi Block-Tapping Test	Della Sala, Gray, Baddeley, Allamano, and Wilson (1999), Vecchi and Richardson (2001), Vandierendonck, Kemps, Fastame, and Szmalec (2004), Busch, Farrell, Lisdahl-Medina, and Krikorian (2005),
	Visual Patterns Test	Della Sala, Gray, Baddeley, and Wilson (1997), Della Sala et al. (1999)
	Brief Visuospatial Memory Test	Benedict, Schretlen, Groninger, Dobraski, and Shpritz (1996)
	Operation Span Task/ Automated Operation Span Task (AOSPAN)	Unsworth, Heitz, Schrock, and Engle (2005), Kilic and Yildirim (2010), Redick et al. (2012)
	Silverman-Eals Tests of Object and Location Memory	Silverman and Eals (1992), Carroll (1993), Choi and L'Hirondelle (2005), Postma, Izendoorn, and de Haan (1998)

[1] The authors used the term Kinesthetic Imagery (the ability to determine left and right).

3.4 Gender differences in spatial ability/ memory

3.4.1 Gender difference in spatial ability and spatial orientation

Researchers have tried to link gender differences on many learning topics with individual spatial ability (see Baenninger & Newcombe, 1995; Casey, Nuttall, Pezaris, & Benbow, 1995; Geary, 1996, 2004; Pearson & Ferguson, 1989). Research generally supports the findings that males outperform females in spatial ability related tasks (see Guillem & Mograss, 2005; Halpern, 2012, pp. 128–145; Linn & Petersen, 1985; Masters & Sanders, 1993; Voyer et al., 1995). Hyde (1981), Linn and Petersen (1985), and Masters and Sanders (1993), conducted three separate meta-analyses covering studies from 1955 to 1997 on the gender differences in various spatial ability tests. Hyde (1981) reported a large gender difference in both spatial ability and numeric ability where males displayed advantages, although the effect sizes were both small. In Linn and Petersen's study (1985), they found that males consistently performed better in both spatial perception and mental rotation tasks. These results were in line with the report from Masters and Sanders (1993), who found a stable gender difference on the Mental Rotations test (developed by Peters et al., 1995). All three studies concluded that males scored significantly higher than females.

Yet, some gender studies are inconsistent (e.g. Galea & Kimura, 1993; Lawton, Charleston, & Zieles, 1996; see also Coluccia & Louse, 2004). Coluccia and Louse (2004) conducted a systematic review exploring gender differences in spatial orientations only. Their results revealed that males consistently performed better than (or equal with) females in simulated and real environments. However, in a map environment (i.e. when information was presented on a map) males performed only slightly (and insignificantly) better and sometimes even significantly worse than females. In particular, females were found to be consistently better at *recalling landmarks, street names,* and *reporting route strategies* (the opposite of orientation strategies) in map navigation. Coluccia and Louse (2004) argued that the inconsistent results for gender differences in spatial ability were due to confusion in spatial definitions. Their review found that different competences were often grouped together under the same sub-division of spatial ability, thus leading to conflicting outcomes.

3.4.2 Gender differences in visuospatial working memory and spatial processing

Similar to spatial ability studies, gender differences in visuospatial working memory have also produced mixed findings. Some researchers (e.g. Loring-Meier & Halpern, 1999; Sánchez & Wiley, 2010) argue that males consistently demonstrate larger visuospatial memory capacity and thus exhibit better visuospatial performance than females. Loring-Meier and Halpern (1999) conducted four experiments examining different processes in visuospatial working memory (i.e. image generation, image maintenance, image scanning, and image rotation) finding that males outperformed females in all tasks in reaction times, but no difference in task accuracy. The greatest effect was found on mental rotation tasks, which requires a particularly high demand on visuospatial working memory to both store and perform the transformation simultaneously.

Whereas males have been found to perform better at object transformation tasks, females perform better at tasks that require memory of locations (see e.g. Choi & L'Hirondelle, 2005; Eals & Silverman, 1994; McBurney, Gaulin, Devineni, & Adams, 1997; Silverman, Choi, & Peters, 2007; Silverman & Eals, 1992). The ability to recall specific locations of individual objects is referred to as object location memory (Choi &

L'Hirondelle, 2005). It requires the reconstruction of the positions of previously studied objects (see Postma et al., 1998).

Choi and L'Hirondelle (2005) examined memory location further by investigating potential correlations between language and location memory. Their results showed that the gender difference (females are superior to males) in the concrete condition (recognising the location of familiar objects) disappeared in the abstract condition (recognising the location of abstract objects). The authors argued that verbal memory exerted an indirect influence onto object location memory, but no correlation between verbal memory and object location memory was found. Furthermore, McGivern et al. (1997) found that females recognised both significantly more abstract shapes and nameable objects than males.

Other researchers have argued that there is a gender difference in spatial memory initiated from different spatial processing. McGivern and colleagues (see McGivern et al., 1997; McGivern et al., 1998) showed there was a gender effect in recognition memory which was sensitive to the stimuli's nature. Females showed significant advantage over males in recognising female-oriented objects and random objects, but no differences were found when the objects were male-oriented. Combined with the research of Guillem and Mograss (2005) who found that males tended to process information which they found to be relevant but females tended to process everything, it suggests that gender differences start at the recognition level.

Males and females have also been found to have different spatial strategies. For example, Sandstrom et al. (1998) demonstrated that females relied predominantly on landmark information, while males used both landmark and geometric information in virtual environment navigation tasks. Law et al. (1993) found that males were better at dynamic spatial reasoning (relative velocity judgement) than females, but not for static spatial reasoning (relative distance judgement). The authors claimed the similarities of the two tasks (i.e. relative velocity judgement and relative distance judgement) were very high and thus concluded that the difference was caused by the nature of the judgement and the information type.

Evidence from Neuroscience

Gender differences in spatial processing have also been found using neuro-scientific methods. Human spatial learning has been found to rely heavily on the hippocampus and parahippocampus sections of the brain (Cornwell, Johnson, Holroyd, Carver, & Grillon, 2008), which are also important for episodic memory (Robin et al., 2015). Magnetic resonance imaging (MRI) evidence has found gender differences in basic visuospatial processes when judging line orientation (e.g. Clements-Stephens, Rimrodt, & Cutting, 2009; Frings et al., 2006). There was a significantly greater activation in the right lingual gyrus and cerebellum in males than in females. These two regions have been found to be associated with motor planning/sequencing and with encoding/decoding spatial memories respectively. Consequently, the findings suggest that males utilize the spatial regions to a greater degree than females (see Clements-Stephens et al., 2009). In contrast, hippocampal activation was found predominantly left lateralised in females, but more right lateralised in males, suggesting that females may use greater verbal processes when encoding and recognising object locations in a 3D environment (see Frings et al., 2006). Furthermore, both Speck et al. (2000) and Schöning et al. (2007) found similar results using functional magnetic resonance imaging (fMRI).

Jordan, Wüstenberg, Heinze, Peters, and Jäncke (2002), and Weiss et al. (2003), compared brain activation patterns (using an fMRI) in males and females who performed at a similar standard in the Mental Rotations test. Results showed that the mental rotation conditions did not evoke any different cortical activation patterns, but gender did. Females showed stronger activations in the interior temporal gyrus, which is known to involve spatial memory, objective identification, objective categorisation and object memory; whereas males showed stronger activations in the motor area (manipulating object rotation). However, no significantly differences in activation in the left front and left temporal area (verbal area) of the brain were found.

Overall these studies examining brain patterns suggest that gender differences occur in the functional organisation of the brain.

3.5 Sexual selection – Evolution theory

Both behavioural (psychological) and physiological studies have shown a gender gap in spatial ability; however, the underpinning reasons have seldom been explored in such research. The following section examines some possible causes.

3.5.1 Sexual selection

Evolution theory provides some insights. Sexual selection, an evolutionary concept articulated by Darwin (1871), plays an important role in the mental differences between males and females (p. 326). It is the mechanisms that have driven the ‘evolution of brain, (and) cognition’ (Geary, 2005, p. 23). Based on sexual selection, and supported by neurobiological research, Jacobs (1996) argued that the gender difference in learning is a result of intrinsic difference in the human brain.

In focusing on classroom learning, Geary (1995, p. 291) argued that the gender difference in 3-dimensional spatial ability was a result of sexual selection, ‘directly related to intra-male competition and courtship of females’. Males are innately better at locomotion in order to ‘find’ the females. Moreover, since humans’ neurocognitive systems that support habitat representations are evolved in 3- dimensional but not 2-dimensional worlds, or a more static form (e.g. object locations), the gender difference is expected to be smaller or non-existent when processing 2-dimensional information.

Hunter-Gatherer theory of Spatial Sex Differences

Silverman and Eals (1992) have proposed a theory called a Hunter-Gatherer theory of spatial sex differences based on evolution theory. They believe that the critical factor in the natural selection for spatial dimorphism was the sexual division of labour between hunting and gathering during human evolution. The types of spatial capabilities that showed male-bias (e.g. mental rotations, maze learning…etc) comprised of abilities to 'orient oneself in relation to objects or places in view of conceptualised across distances, and to perform the mental transformations necessary to maintain accurate orientation during movement' (Silverman & Eals, 1992, p. 534). Such capability was important for the pursuit of animals for food. Similarly, female-biased spatial capabilities included the ability to recognise and recall spatial configuration of objects and object location in relation to other objects. These capabilities enhanced females to rapidly learn and remember the content of object arrays (i.e. edible plant vs. inedible plant; the growing season…etc) and are important to successful foraging. A number of studies discussed in Section 3.4.2 reported these spatial differences.

Sex hormones

In addition to experimental results, some other studies have examined the organisational effects of sex hormones. Silverman and Phillips (1993) investigated female performance on spatial tasks at different menstrual stages. Both within-subjects and between-subjects results indicated that females performed better when their estrogen levels (one of the female hormones) were low (see also Gaulin, Silverman, Phillips, & Reiber, 1997).

Medical scientists have also found that testosterone (a male hormone) intake could improve spatial memory/ability of healthy women (Postma et al., 2000), older men (Cherrier et al., 2001) and transsexuals (van Goozen, Slabbekoorn, Gooren, Sanders, & Cohen-Kettenis, 2002). Schöning et al. (2007) showed the importance of sex hormones by directly measuring the sex hormone in males and females, and comparing them with both spatial task performance and fMRI scans. They found a strong correlation between the testosterone level and activation of the left inferior parietal cortex (a key region for mental rotation tasks) in males but not females. These medical results explain the correlation between age and spatial ability (because the level of testosterone declines with age). They also provide support for sexual selection in evolutionary theory that males and females are biologically different in their spatial ability.

3.6 Spatial Ability and Instructional Animations and Static Pictures

Spatial ability is an essential skill to navigate both real and virtual environments. When it comes to classroom teaching and learning, spatial ability is often regard as an essential skills to understand scientific and abstract concepts such as orthographic projection (e.g. Pillay, 1994), mathematical problems (e.g. Hegarty & Kozhevnikov, 1999), mechanical systems (e.g. Boucheix & Schneider, 2009; Hegarty, Kriz, & Cate, 2003; Hegarty & Waller, 2005). Researchers (e.g. Baenninger & Newcombe, 1995; Casey et al., 1995; Geary, 1996, 2004; Pearson & Ferguson, 1989) have argued that males outperform females in maths and science because they have a stronger spatial ability that enables them to mentally visualise and understand the math and science concepts, and then relate them with the real world more easily.

In regards to instructional visualisations, when static pictures are used to display kinematic or dynamic processes, the learner must mentally animate the processes in order to fully understand them. Motion must be inferred. Studies have found that spatial ability is highly correlated with mental animation (see Hegarty et al., 2003; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Hegarty & Sims, 1994; Hegarty & Waller, 2005; Narayanan & Hegarty, 2002). Such studies have often featured learning about mechanical systems where cause and effect has to be understood. Results have consistently indicated that learners with higher spatial ability make fewer mistakes on retention tasks (i.e. better recall) and performed better on transfer tasks (i.e. deeper understandings) in both animation and static picture groups. Hegarty and colleagues concluded that high spatial ability learners are able to build better mental models, thus forming better conceptual knowledge and understanding from the visual presentations.

The same argument has been applied to instructional animations, which are able to depict real-time changes requiring fewer processing resources to animate mental images than static presentations- the latter being more demanding for those with lower spatial ability. Hence some researchers have argued that lower spatial ability learners (i.e. females) should benefit more from animation than from static pictures (Jacek, 1997). Evidence for this assumption has been supported by Sánchez & Wiley's study (2010) mentioned above, who found that animation closed the gender gap by successfully boosting females participants performance. In their study, females had lower spatial ability.

Although there is some evidence suggesting that the gender effect is influenced by spatial ability, there has been a lack of direct investigation of the interactions between gender, spatial ability and presentation types (i.e. animation vs. static pictures). Some researchers have drawn their conclusions based on the argument that females generally have lower spatial ability than males rather than include this variable directly in their own studies (e.g. Jacek, 1997; Yezierski & Birk, 2006). On some occasions, the participants' spatial ability and learning performance have been measured, but analysed separately (e.g. Falvo & Suits, 2009; Sánchez & Wiley, 2010).

There are also studies that have found gender effects regardless of the learners' spatial ability. For example Falvo and Suits (2009) found that participants with higher spatial ability outperformed those with low spatial ability when learning with animations, but females outperformed males even though females had lower-spatial ability group than males.

A general linear regression model reported by Garg, Norman, and Sperotable (2001) showed that participants' spatial ability and the material view (key view which showed only back and front view vs. multiple view which showed back, front and 10 degree rotated view) significantly affected the learning process, whereas age, sex, handedness and experience of using computers, did not. The meta-analysis results of Höffler (2010) also showed that spatial ability and its sub-dimensions significantly affected the visualisation process. There was an overall effect demonstrating a medium advantage for high spatial ability learners when learning with both animations and static pictures. These results are consistent with result from Lee and Shin (2012) who studied the relations between participants' spatial ability and the learning of a procedural task (printer cartridge replacement) from static and animations. The result showed that participants with low spatial ability committed fewer errors when learning with animation than static pictures. However, for participants with high spatial ability, there were no differences found between animations and static picture group.

Although some evidence suggests that high spatial ability learners benefit more from visualisation than low spatial ability learners, there is also conflicting evidence suggesting that low spatial ability learners can benefit from animations. For example, a Rieber (1990b) review on animations found a difference between adult learners and children. In adult learning, animation did not appear to facilitate visualisation/ imagination effects (named imaging in Rieber's paper), however, children appeared to benefit from animations. If adults have higher spatial ability than children it can be concluded that learners with the lower spatial ability (children) benefited more from animations. Also, Wender and Muehlboeck (2003) found that low spatial ability participants tended to profit more from animations than static pictures when learning statistics at university level (although the interaction effect was not significant and high spatial ability participants still scored significantly better). Similarly, Koroghlanian and Klein (2004) found a similar result when examining the use of animation in teaching biology The pre-test showed that students with high spatial ability outperformed those with low spatial ability. However, the difference disappeared in a post-test. When only considering the post-test result, Koroghlanian and Klein (2004) concluded that the result did not support previous finding that animations benefits low spatial ability learners. Moreover, the result seemingly suggested that animations once again mind the performance gap between high and low spatial ability learners.

The Höffler (2010) meta-analysis reported that low spatial ability learners were significantly supported by animations compared to static pictures, as well as by 3D pictures instead of 2D pictures. These results are in line with the ability-as-compensator hypothesis (see Höffler & Leutner, 2011; Mayer & Sims, 1994) which proposes that animations may be more effective for low spatial ability learners than static pictures (Blake, 1977; Höffler & Leutner, 2011; Rieber, 1989).

3.7 Chapter Summary

A gender difference in learning is a controversial topic, which has been studied extensively but still no consensus has been found. In regards to multimedia learning, some evidence has emerged that males and females learn differently with different presentation modules (dual modules vs. single module), strategies (wholistic vs. analytic), depth-of-process (superficial vs. thorough approach), and preferences such as colours, text, and movement speeds for example.

In relations to animations and static pictures, the research outlined in this chapter, although at times inconsistent, suggests that female learners can benefit more from instructional animations than males. Studies have shown that learning with animations helps closing the gap (males are superior to females) that is often found when learning science subjects in particular. In contrast when learning from static pictures the evidence suggests that generally the gap still remains, as males benefit more from static pictures.

One potential cause of the inconsistent findings associated with gender and visual presentations is spatial ability. Spatial ability has been identified as an important factor when learning with visualisations (both animations and static pictures). It helps learners to process information from visualisations and to construct mental representations that are essential for understanding. Although it has been shown that high spatial ability learners have an advantage in learning many curriculum topics, the research evidence also shows that low spatial ability learners benefit more from animations than static pictures. As females generally have lower spatial ability than males it is often concluded that animations are a better tool for females to learn about spatial related tasks.

However, more definite conclusions are difficult to make because of a lack of studies that have directly examined the relationships between gender, spatial ability and animations, simultaneously. The research also reports some issues in defining spatial ability (e.g. spatial ability, spatial orientation or visuospatial working memory) and wide variations in the tasks and learners used in studies, making identification of moderating variables illusive. In addition, some evidence based on neuroscience and evolution theory, has shown that females outperform males in visuospatial location memory, which may give them an edge in related tasks. All of which make it difficult to define the boundary conditions for identifying the effectiveness of instructional animations. Hence, a direct examination of gender, spatial ability and animations was carried out in the empirical study of this thesis.

Chapter 4: Hypothesis Development and Study Overview

As discussed in Chapter 2, instructional animations are theoretically beneficial to learning. However, research has produced inconsistent results. From a cognitive load perspective, it is believed that the transient information effect is one of the major causes for this ineffectiveness (Ayres & Paas, 2007a, 2007b; van Gog et al., 2009). Other factors may also be involved as other studies (e.g. Bétrancourt & Tversky, 2000; Höffler & Leutner, 2007; Morrison et al., 2000) have suggested that the inconsistent results are caused by variations in research materials, unreliable comparison methods, or characteristics of learning tasks.

Nevertheless studies into instructional animations have provided some insights into the types of conditions where animations may be effective. However, there are still some gaps in this research area, and more empirical results are needed to clearly identify these conditions. Consequently, a major aim of this thesis was to continue the research into comparing animations with statics. Specifically a human movement task was chosen (building a Lego construction) to also enable further investigation of the human movement effect (Paas & Sweller, 2012; Sweller, Ayres, & Kalyuga, 2011).

One of the gaps in the research area is the impact of various learners' characteristics. In cognitive load theory studies, individual differences are mostly confined to an individual's expertise level (i.e. level of domain specific knowledge), and more recently, spatial ability. Gender is seldom considered and consequently few studies have been completed within this theoretical framework in relation to gender and instructional animations. As discussed in Chapter 3, some researchers outside cognitive load theory (e.g. Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006) have found significant gender differences in learning with animations, concluding that animation helps females more than males, and forming a link with lower spatial ability.

However, there is insufficient empirical evidence supporting this link. Hence, the second aim of this thesis was to examine gender differences when comparing animations with statics, and to investigate the role played by spatial ability.

A third aim of the study was to examine the effects of embodied cognition in this domain (Paas & Sweller, 2012; van Gog et al., 2009). As discussed in Chapter 2, a number of experimental studies have been conducted using hand manipulative tasks (e.g. Akinlofa et al., 2013; Ayres et al., 2009; Castro-Alonso et al., 2015a; Wong et al., 2012; Wong et al., 2009). These studies consistently showed that animations are superior to static pictures. However, whether it is necessary to view the hands in such tasks is still an open question. Hence, this factor was explored. In addition, as part of greater embodied cognition effects, gesturing has been found to be a helpful learning strategy (Paas & Sweller, 2012). Because little research has been conducted into the use of gesturing with animations this factor was also explored.

In order to investigate these three main aims, four experiments were completed using Lego materials (a human movement task). Four hypotheses were tested, which are described below.

4.1 Hypothesis Development

4.1.1 Hypothesis 1: Animations will be superior instructional materials to static pictures

With tasks that involve human movement, animations have been found to be consistently superior to equivalent static pictures (e.g. Akinlofa et al., 2013; Ayres et al., 2009; Castro-Alonso et al., 2014a, 2015a; Höffler & Leutner, 2007; Paas & Sweller, 2012; Wong et al., 2009). According to (Ayres & Paas, 2007a, 2007b) animations can generate extraneous load due to their transient nature. However, because tasks involving human movement activate the mirror-neuron system extraneous load is reduced (Paas & Sweller, 2012; van Gog et al., 2009). As discussed in Chapter 2, learning such tasks through observation involves biologically primary skills and creates no, or minimal effort, to learn. Learners can observe and simulate human movements from animations without the usual cognitive costs associated with transient information.

In this thesis, Lego materials were employed because they are human movement tasks involving the hands, and empirical evidence from Castro-Alonso et al. (2015a) showed that animations were superior to statics in this domain. Hence it was predicted that for Lego tasks:

Hypothesis 1: Animations will be superior instructional materials to static pictures

4.1.2 Hypothesis 2: There will be a gender-presentation format interaction

Gender-related research with instructional animations has, similar to general animation research, produced inconsistent results. Some studies have shown that females learning with instructional animations have significantly outperform those learning with static pictures (e.g. Cowards et al., 2012). Sometimes there is no difference between males and females with animation presentations, and sometimes evidence has emerged that males performed significantly better with statics than females (e.g. Sánchez & Wiley, 2010). Results have also shown that using animations have boosted females learning and closed the gender gap (e.g. Jacek, 1997; Yezierski & Birk, 2006). Although results vary, the overall weight of evidence suggests that animations are more helpful for females than males, whom are less likely to be disadvantaged with static pictures.

As described in Chapter 3, the conclusion that animations may help low spatial ability learners (in this case females) in particular, is drawn from the general consensus that females are weaker in spatial ability. However, females have been found to be better than males at object location memory - the ability to recall specific locations of individual objects (see e.g. Choi & L'Hirondelle, 2005; Eals & Silverman, 1994; McBurney et al., 1997; Silverman et al., 2007; Silverman & Eals, 1992). Because the Lego tasks in this study have high object location memory demands, females may have an advantage in these experiments. Taken together that females may have higher object location memory and also benefit from animations, it was predicted that females would outperform males in the animation format, but not necessarily in the static format. Hence the following interaction was predicted for the Lego tasks:

Hypothesis 2: There will be a gender-presentation format interaction.

4.1.3 Hypothesis 3: Observing hands will lead to superior learning than not observing hands

In regards to animation providing an advantage for learning human movement tasks, one explanation lies in the field of embodied cognition. As discussed above, and also in Chapter 2, observing human movements activate the human mirror neuron system, offloading the effects of transient information when learning. Because many of the tasks that have found advantages for animations have involved hand manipulations, of interest to researchers has been the question: is it necessary or not to view the hands performing the tasks in the presentations? Comparing the study of Marcus et al. (2013),Castro-Alonso et al. (2015a) and Castro-Alonso et al. (2014b), there were gradual changes in the level of hand involvements from high to low. Knot-tying from Marcus et al. (2013) is a procedural motor task which cannot be completed without involving hands by nature. However, remembering abstract symbol positions from Castro-Alonso et al. (2014b) does not involve the hands. Whilst the Lego construction from Castro-Alonso et al. (2015a) is in the middle of the scale: the hands are involved in moving and manipulating the bricks, but not in remembering the positions. Furthermore, the study of Wong et al. (2009), discussed in Chapter 2 employed computerised materials without showing hands. This study showed that learning how to make origami figures from an animation with an unrealistic environment without hand demonstrations was still effective. Hence it is unclear to what extent hand-observation can affect the learning of motor tasks; especially memorising the locations of objects in a motor task. Notably in both the Castro-Alonso et al. (2015a) and Castro-Alonso et al. (2014b) studies, as well as the Marcus et al. (2013) study, showing hands in the animation did not lead to an advantage, but sometimes a disadvantage. However, generally based on the human movement effect and embodied cognition, it is believed that observing hands movements can activate our mirror neuron system, making it easier to learn human movement tasks. Hence, despite some evidence to the contrary, for the given Lego tasks it was predicted that:

Hypothesis 3: Observing hands will lead to superior learning than not observing hands

4.1.4 Hypothesis 4: Gesturing will lead to superior learning than no gesturing

The final prediction is also based on an embodied cognition effect. Neuro-scientific evidence indicates that observing and executing gestures share a similar mechanism in the mirror neuron system. Although, there are a number of studies conducted on observing hand movement when learning with animations, few have examined the execution of gestures when learning with animations at the same time

(Ouwehand, van Gog, & Paas, 2015; Post, van Gog, Paas, & Zwaan, 2013). Hence, there is some uncertainty about the role played by gesturing when learning from animations, and under what conditions gesturing may be useful (or harmful) to learning with animations.

As discussed in Chapter 2, evidence shows that the use of gesture can influence learning and problem solving (Barsalou, 1999, 2008; Chu & Kita, 2011; Glenberg, 1997; Goldin-Meadow et al., 2001; Matlen et al., 2012; Roth, 2000). In regards to solving spatial problems, gesturing also plays a role in facilitating spatial reasoning (Chu & Kita, 2011; Matlen et al., 2012), capturing essential spatial information (Wagner, Nusbaum, & Goldin-Meadow, 2004) and encoding/decoding non-verbal spatial information (Morsella & Krauss, 2004; So, Shum, & Wong, 2015). Hence, because of the positive results associated with gesturing it was predicted that for the given Lego tasks:

Hypothesis 4: Gesturing will lead to superior learning than no gesturing.

4.2 The Role of Spatial Ability

As discussed in Chapter 3, spatial ability plays an important role in both learning with static and animated visualisations. When static pictures are used to display kinematic or dynamic processes, learners are required to mentally animate the information (i.e. transforming static motions into dynamic motions) in order to understand the information and consequently construct a mental model. Such processing is highly correlated with spatial abilities (see Hegarty et al., 2003; Hegarty et al., 2006; Hegarty & Sims, 1994; Hegarty & Waller, 2005; Narayanan & Hegarty, 2002). The same mechanisms also apply to instructional animations except learners are not required to mentally animate the presentations as they are already animated. Consequently animations can help learners with lower spatial ability by making it easier to construct a mental model without additional processing, thus saving their cognitive load for learning.

Based on a general assumption that females have a lower spatial ability (in particular spatial rotation ability) than males (see Guillem & Mograss, 2005; Halpern, 2012, pp. 128–145; Hyde, 1981; Linn & Petersen, 1985; Masters & Sanders, 1993; Voyer et al., 1995), it has been argued that any advantage for females from instructional animations is due to their lower spatial ability, as animations generally benefit learners with low spatial ability. In the Lego tasks used in the present study, mental rotation ability is important for the transfer task (which will be introduced in the next section). Moreover, spatial location memory also plays an important role in memorising the Lego structure (locations of each brick), which will be the major learning task. Hence spatial ability will be an important factor potentially moderating the learning performance of the different conditions. Consequently, the role played by spatial ability was investigated in all the experiments.

4.3 Experiments Overview

The experimental materials were adopted from Castro-Alonso et al. (2015a) who found there were significant animation advantages with Lego tasks. These results were consistent with those from the literature claiming that animations were superior to equivalent static pictures when learning about human movement related tasks. However, in the Castro et al. study there were significantly more female participants than male participants. Although the participants' spatial ability was controlled for, it was questionable whether gender played a role in the advantage afforded the instructional animations. Hence, this thesis used similar materials, not only because they have a human movement element, but also because the results of the Castro et al. study needs further investigations regarding a possible gender bias, and can be used as a direct comparison. As previously stated, many of the previous results into animation studies may have been biased by including uncontrolled moderating factors. If gender is one such factor then it is important to balance males and females in animation studies. Consequently, an equal number of male and female participants were recruited in this study.

The study consisted of four experiments following a similar procedure of a spatial ability testing section, a practice section and the main experimental section. In the spatial ability testing section, all participants from various conditions received standardised spatial ability assessments (more details will be discussed in respective methodology sections in Chapter 5). Afterwards a short practice section was completed, which varied slightly according to the assigned treatment conditions. The major purpose of the practice was to ensure participants were familiar with the main experiment procedures and conditions. In the main experimental section, there were three phases: a learning phase, a retention test phase, and a transfer phase. During the learning phase, participants were required to watch the assigned presentations (a Lego shape construction) and to memorise as much information as they could within a fixed time. Then participants moved onto the retention test phase where they were required to reconstruct the viewed model onto an assigned platform. This cycle of view the presentation and complete the retention test was repeated once. Finally, in the transfer task, participants were required to perform a transformation (rebuild a rotated version of the model). During the main experiment sections, accuracy scores in the two retention tasks and the transfer task, a self-rated mental effort measure and instructional efficiency (by calculation of accuracy and mental effort, see Chapter 1) were collected for each experiment.

Experiment 1 was a pilot study to examine if the proposed materials produced significant results and the spatial ability test (card rotational test, CRT) was meaningful. *Hypothesis 1* and *Hypothesis 2* were tested by using a 2 (gender: males vs. females) x 2 (presentation-format: animations vs. static pictures) factorial design. For the presentations, the Lego shape was built without any hands being observed. Testing was conducted on a physical platform with real Lego bricks.

Experiment 2 repeated Experiment 1 with a larger sample size. The same hypotheses (*Hypothesis 1* and *Hypothesis 2*) were tested. The learning materials from Experiment 1 were slightly modified in order for hands to be observed within the presentations. In other words, participants were required to watch a video or a set of static pictures showing a hand moving the bricks.

Experiment 3 replicated Experiment 1 and Experiment 2 testing the same hypotheses (*Hypothesis 1* and *Hypothesis 2*). Identical learning materials from Experiment 2 were reemployed. However, the performing platform was modified. Instead of rebuilding the Lego shape using physical bricks, participants were required to rebuild on a computer using computerised bricks.

Experiment 4 tested *Hypothesis 3* and *Hypothesis 4* using only animation materials from the previous experiments. Two new variables (hand-observations: with-hands vs. without-hands) and gesturing (encouraged gesture vs. prohibited gesture) were tested. Because, no static presentation was included, differences between males and females were investigated for each condition of gesturing and hand observations. Hence, Experiment 4 comprised of a 2x2x2 factorial design. The computerised performing platform from Experiment 3 was readopted.

Chapter 5: Empirical Studies

5.1 Experiment 1

The first experiment compared the learning outcomes when imitating an object manipulative task after being modelled either in a static or animated presentation. In addition, gender effects were also investigated. Two of the central hypotheses of this thesis were investigated.

Hypothesis 1: Animations will be superior instructional materials to static pictures
Hypothesis 2: There will be a gender-presentation format interaction.

The experiment used a 2×2 design. There were two factors: 1) type of presentation (animation vs. static), and 2) gender (male vs. female). These two factors constituted four experimental groups: males learning with animation, males learning with static pictures, females learning with animation, and females learning with static pictures. As this experiment was designed to gauge the effectiveness of the materials used to test the study hypotheses, a small sample was employed. Furthermore the presentations did not show hands performing the tasks, and hence embodied cognition effects were not investigated here.

5.1.1 Method

5.1.1.1 Participants

The sample consisted of 26 university students (13 male and 13 female) aged between 19 and 34 (M = 23.9, SD = 3.84). They were current undergraduate (N = 12) and postgraduate (N = 14) students from a large Sydney university drawn from various faculties including Arts & Social Sciences, Business, Engineering, Medicine and Science. They were randomly allocated according to the groupings shown in Table 5.1.

Table 5.1.
Groupings of Participants in Experiment 1

	Animation	Static Picture	Total
Male	6	7	13
Female	7	6	13
Total	13	13	26

5.1.1.2 Materials

A half-page survey (see Appendix 1) was designed to capture some background information, self-rated spatial ability and a subjective measure of cognitive load (questions a-d). Answers were provided according to the experimental sequence requirements.

Survey of background knowledge. The background information survey was given to participants on commencement of the study and assessed their gender, their university programme and year that students were enrolled in, and their level of study and handedness (right or left).

Subject assessment of spatial ability. Two questions on self-rated spatial ability were used to examine participants' confidence levels towards their own spatial ability. According to the categorisation from Linn and Petersen (1985), spatial ability could be sub-categorised into spatial perception, mental rotation and spatial visualisation. However, only the mental rotation subcategory was measured in addition to general spatial ability because it was more closely related to the transfer task. These two questions were presented to participants together with the survey of background knowledge and were completed at the commencement of the experiment.

1. How would you rate your *mental rotation* ability (ie. ability to rotate or flip 2-D or 3-D shape mentally)?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

2. In general, how would you rate your *overall spatial ability*?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

Objective assessment of spatial ability. As well as self-rated measures of spatial ability the Card Rotation Test (CRT, see Appendix 2) was also given to objectively assess participants' spatial ability. The CRT was developed by Ekstrom, French, Harman, and Dermen (1976) examining spatial rotation ability (see Chapter 3). The test was originally composed of two sub-sections (named S1 and S2). However, only part one (Card Rotation Test-S1) was used here to shorten the procedure and time, hence saving the participants' cognitive resources for the main experiment. The test comprised of an instructional page and two pages of questions. Each problem consisted of an irregular shaped card on the left, and eight other drawings of the same card on the right (see Figure 5.1 below). Some cards on the right were merely rotated (marked 'S', see below) and some of them were turned over (marked 'D', see below). Participants were required to decide whether each card on the right was rotated or turned over compared to the card on the left. There were in total ten problems on each page and three minutes were given to complete each page. In order to let every participant be clear about the time, the experimenter prepared a digital counter timer. Participants were invited to start the timer when they were ready to start the test.

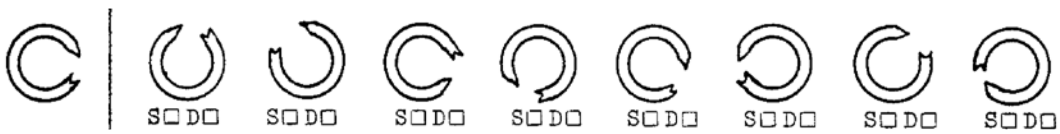


Figure 5.1. Sample of CRT in Experiment 1

Learning Materials. The learning materials consisted of a practice task and the main learning task based on the LEGO tasked designed by Castro-Alonso et al. (2015a).

Practice task. First a video depicting 4 steps of LEGO® construction from an aerial view was made, which included showing the relevant hand actions placing the LEGO® bricks. The video showed human hands (in green gloves) manipulating one Lego brick at a time, and placing them on a LEGO® platform one-by-one to form a shape. The animations' display sizes were adjusted to 160 x 160 pixels based on the LEGO® platform size. It showed the platform centred in the screen filled with white background. The video was filmed by a Canon S95 camera in .MOV format (size 1280 x 720 pixels; 23 frames per second). The muted version was then edited and exported with Adobe Premiere Pro CS3 (Adobe, 2007) to Adobe Flash Video format (.flv).

For each brick, 6 seconds of the animation (5-seconds for moving the brick and 1-second for showing a placed brick) was allotted to showing the brick being placed. Initially, the empty platform was shown for 1-second, and finally the completed shape was shown for 1 second after all the bricks were placed. In total, the animation lasted 26 seconds.

The video was then modified to create a without-hands animation condition (see Figure 5.2, left) to be used in this experiment. The hands were digitally removed using Adobe Premiere Pro CS3 with all other information retained. In order to create an identical environment in both the animation condition and static picture condition, the video display size in the animation condition was intentional kept at 160 x 160 pixels instead of increasing it to the screen-size.

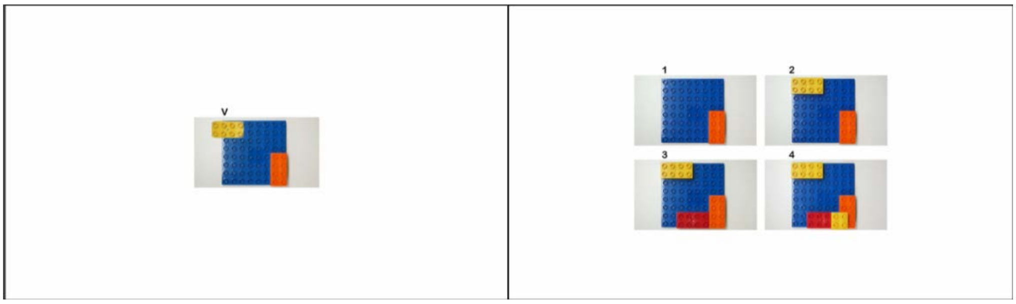


Figure 5.2. Practice Task Material in Experiment 1 (Left: a Frame from the Animation Condition; Right: the Static Picture Condition)

For the static conditions, 4 key frames of each brick position were derived directly from the video in the animation condition. All the key frames were presented simultaneously on the computer (see Figure 5.2, right). They had the same size and depicted the same information as in the corresponding video. Each static sequence was presented on the computer and lasted the same time as the animation (26 seconds). Each frame was sequentially numbered on the top left, in order to assist the learner with the order of the building task. This number was maintained in the animation condition. For both animation and static conditions, the learning materials were embedded into a flash (.fla) file with screen size 3100 × 810 pixels, and participants had no control over the pacing of the learning materials.

Main instructional materials. In the animation condition, a video depicting 15 steps of LEGO® construction from an aerial view was filmed with a digital Sony Handycam in PAL standard (size 768 x 576 pixels; 25 frames per second) without audio. The video showed human hands (in green gloves) manipulating one LEGO® brick at a time, and placing them on a LEGO® platform one-by-one. The video was then edited and exported with Adobe Premiere Pro CS3 (Adobe, 2007) to Adobe Flash Video format (flv) and was compressed to 10,000 kbps via the codec On2 VP6. The size of the output animation was adjusted to a LEGO® platform's size of approximately 200 x 200 pixels leaving the LEGO® platform placed at the centre of the screen filled with white background.

For each brick 6 seconds of the animation was allotted to showing the brick being placed. Initially, the empty platform was shown for 1-second, and finally the completed shape was shown for 1 second after all the bricks were placed. In total, the animation lasted 92 seconds. For the use of this experiment, the hands were then removed digitally with all other information retained (see Figure 5.3, left).

In order to create a close-to-identical environment in the animation and static picture conditions, the video display size in the animation condition was intentional adjusted to be 200 x 200 pixel (same size as static picture) instead of filling the whole screen.

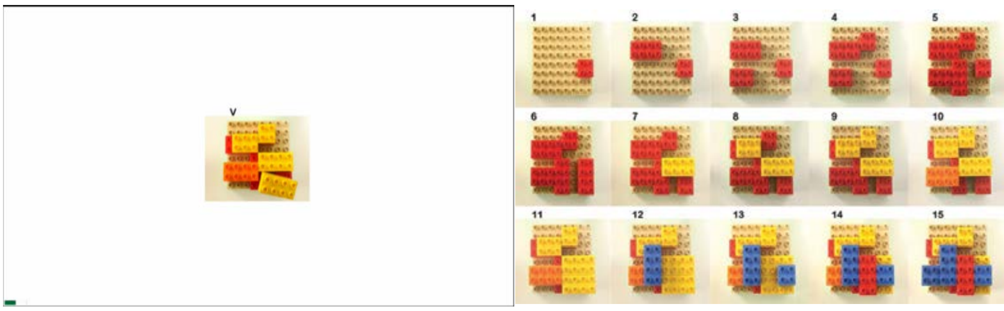


Figure 5.3. Main Instructional Task Material in Experiment 1 (Left: a Frame from the Animation Condition; Right: the Static Picture Condition)

For the static condition, 15 key frames showing each brick being placed was extracted from the animation. They had the same size and depicted the same information as in the corresponding video (see Figure 5.3, right). All 15 key frames were presented at the same time on the computer and lasted for the same viewing time as the animation (92seconds).

Having the same layout and rationale as in the practice task, the frames were sequentially numbered on the top left. This number was maintained in the animation condition. For both animation and static conditions, the learning materials were again embedded into the same flash (.fla) file with screen size 3100 × 810 pixels as in the practice task, and again, participants had no control over the pacing of the learning materials.

Cognitive Load Measure. A subjective measure of cognitive load was designed based on the instrument constructed by Paas and van Merrienboer (1994b), and Paas, van Merriënboer, and Adam (1994), with the assumption that participants could introspect on their cognitive processes. The participants were asked how much mental effort they spent in completing the task right after they completed each task. In order not to confuse the participants their ratings on the survey, the tasks were referred to as the practice task, task 1, task 2 and task 3 (according to the sequence) respectively. The responses were given on a 9-point Likert scale rating spaced as little, fair, and heavy (see Table 5.2).

Table 5.2.
Cognitive Load Tasks Used in Experiment 1

How much mental effort did you spend in completing…

	Little			Fair			Heavy		
a. Practice Task	☐	☐	☐	☐	☐	☐	☐	☐	☐
b. Task 1	☐	☐	☐	☐	☐	☐	☐	☐	☐
c. Task 2	☐	☐	☐	☐	☐	☐	☐	☐	☐
d. Task 3	☐	☐	☐	☐	☐	☐	☐	☐	☐

5.1.1.3 Testing environment

Practice task. The same set of Lego Duplo bricks used in filming the practice task material was employed (see Figure 5.4). The bricks were presented in a vertical manner and arranged according to the order that they appeared in the learning material starting from left to right. After participants finished watching the learning material, real bricks together with a square platform, the base to build onto, were brought to the participants. This platform had a fixed orientation identical to the learning presentation. Participants were then required to build the shape viewed in the presentation.

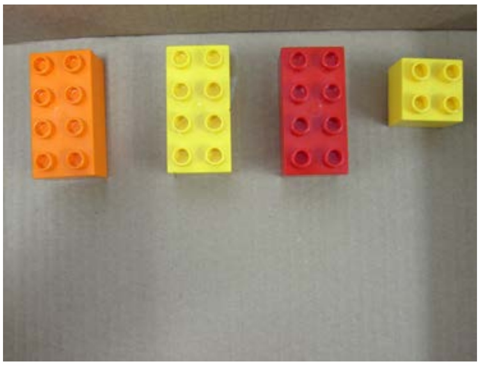


Figure 5.4. Testing Environment Used in Experiment 1 Practice Task

Completion Task. The same set of Lego Duplo bricks used in filming the main learning material was used for testing (see Figure 5.5). The actual bricks given to the participants were arranged in a vertical position, according to the order that they appeared in the learning materials starting from top left to bottom right. A brown square building platform was provided on their work desk for participants to build the required shape on. Again, this platform had a fixed orientation identical to the learning presentation and participants were required to build the shape viewed during the presentation.

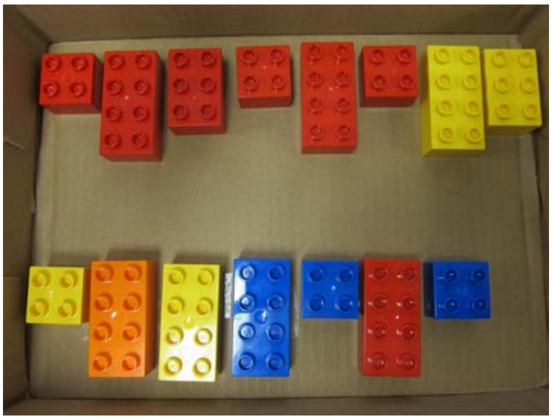


Figure 5.5. Testing Environment Used in Experiment 1 Completion Task

Transfer Task. The first six Lego Duplo bricks used in the completion task were reused (see Figure 5.6). All 6 bricks were placed vertically and were arranged according to the order in the completion task starting from left to right. After the bricks and the platform were brought to the participants, the brown square platform was fixed onto the desk using double-sided tape so it could not be moved. Participants were required to rebuild the bottom layer (i.e. the first six bricks) viewed during the presentation with a 90-degree rotation.

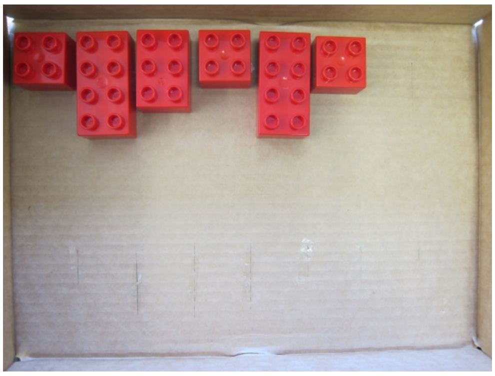


Figure 5.6. Testing Environment in Experiment 1 Transfer Task

5.1.2 Procedure

The experiment sessions were conducted in a quiet room. Each session lasted about 40 minutes, and there was only one student in each session. All the instructions and the learning materials (except the questionnaires and spatial assessments which were in paper form) were embedded into one flash file (experiment programme hereby; size: 3100 × 810 pixels) to ensure every student received the same instructions and to reduce procedural bias.

All participants were invited to sit in front of a computer with the experiment programme ready to start. They first signed the experimental consent form (Appendix 3) indicating their consent in volunteering for the study. Then they completed the background survey followed by the Card Rotational Test (CRT). Participants were given sufficient time to read the instruction page of the test, and three minutes to complete each page of the test. They were allowed to take a break upon the completion of each page. When they were ready to start the test, they were invited to start the 3-minutes countdown timer pre-set as described above.

After spatial assessments, they watched the assigned (according to their experimental condition) practice task materials on the computer. Participants were told to memorise the sequence and the position of every brick shown on the screen. Then they were given a real set of bricks together with the platform and were required to rebuild the shape in the same order and sequence as shown in the presentation. The practice task was a warm-up exercise to help the participants understand the instructions, and become familiar with their specific condition. During the building process, they were required to press the space bar on the computer after each brick position was confirmed. After confirmation, no change was allowed. The employment of the space bar was used to remind participants that no changes could be made. Once the construction was completed, they were required to rate their mental effort spent on the practice task (*question a* in Appendix 1).

After the practice task, they watched the assigned learning task materials for the first time (labelled the 1st attempt). Following the same procedure as in the practice task, participants were required to memorise the position of the Lego bricks shown on the computer and then to rebuild the shape with a real set of bricks. Immediately after they finished the task, they were required to rate their mental effort (*question b* in Appendix 1). This procedure was then repeated by watching the learning task material for the second time constructing the shape (the 2nd attempt) and rating mental effort (*question c* in Appendix 1). In both the 1st attempt and the 2nd attempt of the completion task, the same learning material was employed, i.e. participants watched the same material twice. With the aim of ensuring that the participants put their effort into memorising the material for each attempt independently, participants were told they were going to watch ‘another’ piece of learning material, rather than a repeat presentation.

Immediately after the 2nd attempt of the completion task and cognitive load rating, students started the transfer task. They were required to rebuild only the bottom layer (i.e. the first 6 bricks) of the completion task, as if the platform was rotated 90 degree clockwise. Up to this point, no more material would be shown before building. They were explicitly told not to rotate their head nor the platform, instead the rotation needed to be completed mentally. They were then given a real set of bricks for the bottom layer (6 bricks) together with the platform, and required to re-build the rotated layer. Upon completion, they were required to rate their mental effort for the transfer task (*question d* in Appendix 1).

5.1.2.1 Grading rubrics

1st & 2nd completion attempts. The learning effectiveness of this manipulation task was evaluated based on the accuracy of the brick positions placed in both attempts.

Even though the participants were told to memorise both sequence and the position of the bricks, it was found that the vast majority of participants failed to rebuild the construction in the correct sequence – instead, only patterns were recalled. If a scoring rubric was adopted that simply gave 1 mark for a correct position and sequence of a brick, and no marks assigned for correct misplaced sequences, then scores would have been very low. Hence, in these grading rubrics, only brick placement patterns were taken into consideration: as long as they preserved the same ‘pattern’, marks would be given. Each level was scored separately – if a brick was placed in a correct position but at a different level, no marks would be given. The maximum score was 15 and the minimum was 0. The followings are the patterns for which marks were given.

Pattern: disordered. Figure 5.7 demonstrated some identical bricks placed in a wrong sequence. The graph on the left represents the target answer and it was worth three marks. One mark would be given to each brick in the correct position regardless of the sequence. For answers on the right, three marks would also be given as they represented a correct placement, even though the order of placement was incorrect.

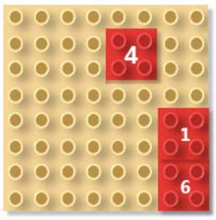
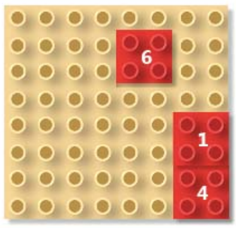
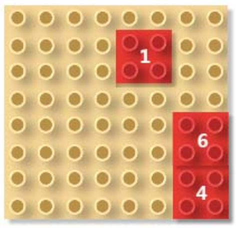
Target	*Answer*	
		
3 marks	3 marks	3 marks

Figure 5.7. Sample of a Disordered Construction with Scores

Pattern: rotated. If two or more adjacent bricks were allocated in a correct position but in a rotated manner (see Figure 5.8 right), 1 mark would be deducted from the total marks. However, if there was only one brick but placed in a rotated manner onto the correct spot, no mark would be given.

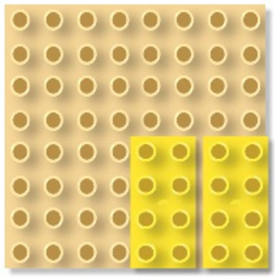
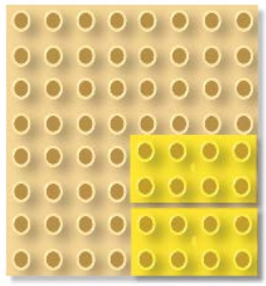
Target	*Answer*
	
2 marks	1 mark

Figure 5.8. Sample of a Rotated Construction and Scores

Pattern: switched. If the bricks were placed in a correct orientation and the whole pattern was in a correct allocation, with only the bricks switched like in Figure 5.9 (right), 1 mark would be deducted.

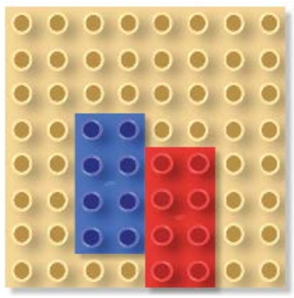
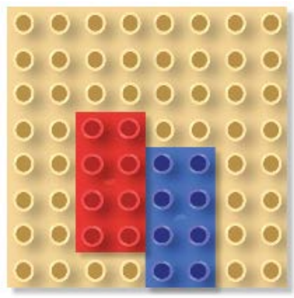
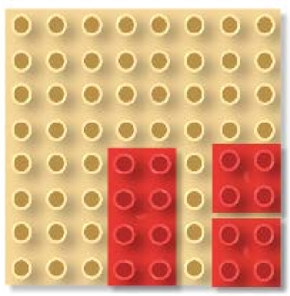
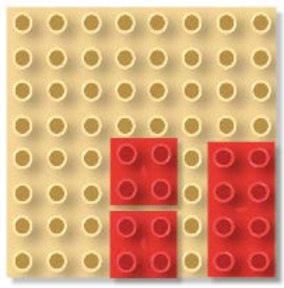
Target	*Answer*
	
2 marks	1 mark
Target	*Answer*
	
3 marks	2 marks

Figure 5.9. Sample of a Switched Construction and Scores

Pattern: translated. In a translated pattern, both orientation and configuration are correctly recalled, but the whole structure was displaced (as shown in Figure 5.10), 1 mark was deducted.

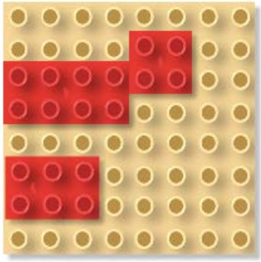
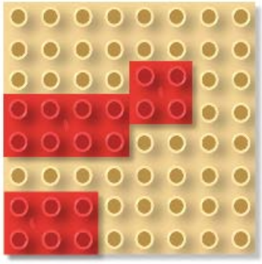
Target	*Answer*
	
3 marks	2 marks

Figure 5.10. Sample of a Translated Construction and Scores

Transfer task. In the transfer task, the participants were required to re-build only the bottom layer from the learning material in the same configuration but in clockwise-manner (see Figure 5.11, left). Some marks would be given to various patterns (disorder, rotated, switched and translated) constructed as long as they meet the criteria stated above. Furthermore, if participants were able to recall the correct configuration as well as correct position, but rebuild them in a wrong orientation (like in Figure 5.11, right), 6 marks would be given. If participants were able to recall the correct configuration, sequence and to rebuild in a correct orientation, 1 extra mark would be given. The maximum score for the transfer task was 7 and the minimum was 0.

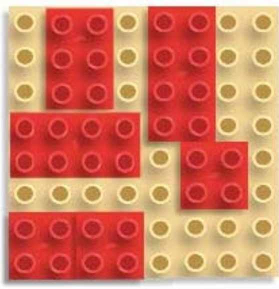
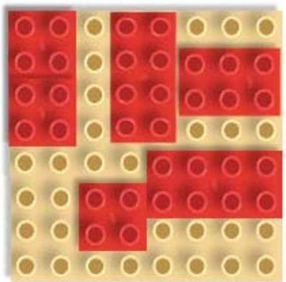
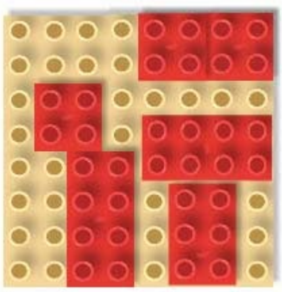
Target	*Answer*	
		
7 marks	6 marks	6 marks

Figure 5.11. Sample of Different Orientations in the Transfer Task and Scores

5.1.3 Results

5.1.3.1 Test of spatial differences

To test for gender spatial differences and possible pre-existing differences between the groups, the spatial measures (CRT score) collected was analysed using a 2 × 2 ANOVA [type of presentation (animation vs. static picture) and gender (males vs. females)]. The spatial measure (CRT scores) was significant for gender $F(1, 22) = 5.93$, $p = .023$, $\eta_{p2} = .21$, $MSE = 1288.93$, where males ($M = 112.46$, $SD = 29.29$) had significantly higher scores than females ($M = 75.92$, $SD = 45.71$). There was also a close-to-significant ($p < .10$) effect for the presentation type $F(1, 22) = 3.94$, $p = .060$, $\eta_{p2} = .15$, but no interaction, $F(1, 22) =1.50$, $p = .234$, $\eta_{p2} = .06$. Due to pre-existing significant differences, the CRT spatial measure was used as a covariate in all further analyses.

5.1.3.2 Test performance

Table 5.3 reports the mean (and SD) scores for each task under the different learning conditions and gender types.

Table 5.3.
Mean (SD) of Performance Scores in Experiment 1

		Animation	Static picture	Total
Male	1st attempt	5.00 (3.78)	7.57 (1.67)	6.38 (3.02)
	2nd attempt	7.86 (2.67)	10.43 (2.71)	8.77 (3.51)
	Transfer	3.17 (2.42)	5.14 (1.86)	4.23 (2.29)
Female	1st attempt	5.36 (2.88)	4.25 (2.46)	4.85 (2.65)
	2nd attempt	7.86 (2.67)	6.42 (3.09)	7.19 (2.85)
	Transfer	1.64 (2.39)	3.58 (1.93)	2.54 (2.33)
Total	1st attempt	5.19 (3.19)	6.04 (2.63)	5.62 (2.89)
	2nd attempt	7.38 (3.00)	8.58 (3.46)	7.98 (3.23)
	Transfer	2.34 (2.44)	4.42 (1.99)	3.38 (2.42)

Two-way ANCOVAs showed that there was no significant main effect for presentation-format for the 1st attempt ($F < 1$, ns), 2nd attempt ($F < 1$, ns), or the transfer task, $F(1, 21) = 2.14$, $MSE = 4.08$, $p = .158$, $\eta p^2 = .09$. In addition, there was no significant main effect for gender on the 1st attempt, the 2nd attempt, or the transfer task (all $F < 1$, ns). There was a close-to-significant ($p < .10$) interaction for the 1st attempt, $F(1, 21) = 3.98$, $p = .059$, $\eta p^2 = .16$, and a significant interaction for the 2nd attempt, $F(1, 21) = 4.71$, $p = .042$, $\eta p^2 = .18$, but there was no significant interaction for the transfer task ($F < 1$, ns).

Simple effects follow-up tests indicated that for the animation format, there was no significant difference between males ($M = 4.23$, $SE = 1.51$) and females ($M = 6.02$, $SE = 1.37$) for the 1st attempt, $F(1, 10) = .65$, $p = .441$, $\eta_p^2 = .06$, $MSE = 10.77$. However for the static format, males ($M = 7.49$, $SE = .83$) scored significantly higher than females ($M = 4.35$, $SE = .90$); $F(1, 10) = 6.37$, $p = .030$, $\eta_p^2 = .39$, $MSE = 4.59$. Similarly, for the 2nd attempt, there was no significant difference between males ($M = 6.06$, $SE = 1.38$) and females ($M = 8.52$, $SE = 1.26$) in the animation format, $F(1, 10) = 1.44$, $p = .258$, $\eta_p^2 = .13$, $MSE = 9.11$. However for the static format, males ($M = 10.59$, $SE = 1.14$) scored significantly higher than females ($M = 6.23$, $SE = 1.24$), $F(1, 10) = 6.48$, $p = .029$, $\eta_p^2 = .39$, $MSE = 8.73$.

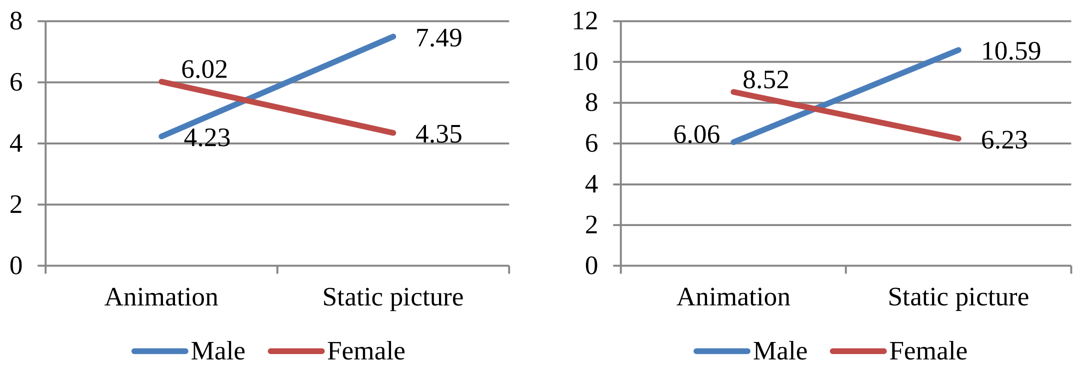


Figure 5.12. Interaction Effects for Performance in the 1st Attempt (Left) and the 2nd Attempt (Right) in Experiment 1

In summary there were no significant main effects for animation or gender on performance scores. However, there were significant interactions indicating that males outperformed females when learning with static pictures, but no difference when learning from animations.

5.1.3.3 Cognitive load measures

Participants were asked to reflect on their cognitive load spent on upon completion of each task. Table 5.4 below illustrates the mean (and SD) scores of the reported cognitive load spent on each task (the practice task was not reported). The higher score represents the more cognitive load spent on the task.

Table 5.4.
Mean (SD) of Cognitive Load Spent in Experiment 1

		Animation	Static picture	Total
Male	1st attempt	8.33 (.82)	6.86 (1.22)	7.54 (1.27)
	2nd attempt	7.83 (1.17)	4.57 (1.27)	6.08 (2.06)
	Transfer	6.50 (1.38)	4.43 (1.27)	5.38 (1.66)
Female	1st attempt	7.71 (1.50)	6.83 (1.94)	7.31 (1.70)
	2nd attempt	7.00 (2.24)	5.50 (2.26)	6.31 (2.28)
	Transfer	8.14 (1.46)	4.67 (2.42)	6.54 (2.60)
Total	1st attempt	8.00 (1.23)	6.85 (1.52)	7.42 (1.47)
	2nd attempt	7.38 (1.81)	5.00 (1.78)	6.19 (2.14)
	Transfer	7.38 (1.61)	4.54 (1.81)	5.96 (2.21)

Two-way ANCOVAs were conducted on each of the 3 mental effort ratings. Results indicated there was a close to significant ($p < .10$) main effect for presentation format for the 1st attempt, $F(1,21) = 3.23$, $p = .087$, $\eta_p 2 = .13$, $MSE = 2.11$; where the participants in the animation group ($M = 7.99$, $SE = .42$) rated cognitive load significantly higher than in the static picture group ($M = 6.88$, $SE = .42$). There was significance for the 2nd attempt, $F(1, 21) = 7.58$, $p = .012$, $\eta_p 2 = .27$, $MSE = 3.34$; where again the animation group ($M = 7.31$, $SE = .53$) reported higher cognitive load than the static picture group (M = 5.16, SE = .53). There was also a significant presentation effect for the transfer task, $F(1, 21) = 13.51$, $p = .001$, $\eta_p2 = .39$, $MSE = 2.91$; where the animation group (M = 7.28, SE = .49) reported higher cognitive load than the static picture group (M = 4.60. SE = .50).

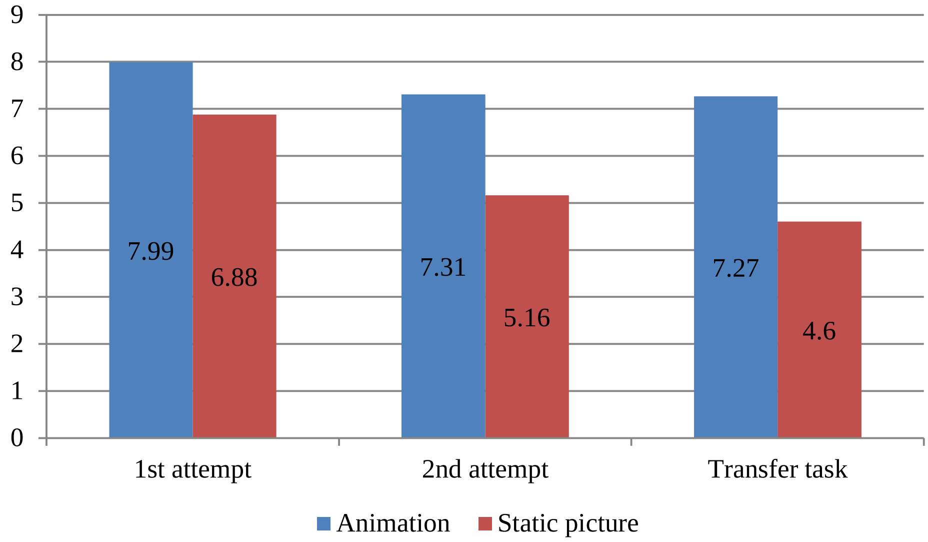


Figure 5.13. Main Presentation-format Effect for Cognitive Load Spent in Different Attempts in Experiment 1

There were no significant gender effects for the 1st attempt, 2nd attempt (both F < 1, ns), or for the transfer task, $F(1, 21) = 1.19$, $p = .288$, $\eta_p2 = .05$. Furthermore, there were no significant interaction effects for the 1st attempt (F < 1, ns), for the 2nd attempt, $F(1, 21) = 1.90$, $p = .183$, $\eta_p2 = .08$, or for the transfer task (F < 1, ns).

5.1.3.4 Learning efficiency

As described in *Chapter 1,* instructional efficiency compares task performance with the amount of mental effort made in completing the task. Hence performance and the cognitive load reported above were combined according to the following equation.

$$E = \frac{Z_{performance} - Z_{mental\ effort}}{\sqrt{2}}$$

Table 5.5.
Mean (SD) of Learning Efficiency in Experiment 1

		Animation	Static picture	Total
Male	1st attempt	-.54 (1.19)	.77 (.79)	.16 (1.16)
	2nd attempt	-.57 (1.09)	1.37 (.72)	.47 (1.33)
	Transfer	-.15 (1.02)	1.06 (.77)	.50 (1.06)
Female	1st attempt	-.15 (1.13)	.02 (1.14)	-.07 (1.09)
	2nd attempt	-.06 (1.16)	.28 (.98)	.10 (1.05)
	Transfer	-1.11 (1.14)	.53 (.76)	-.35 (1.27)
Total	1st attempt	-.33 (1.12)	.42 (1.00)	.05 (1.11)
	2nd attempt	-.30 (1.11)	.87 (.99)	.29 (1.19)
	Transfer	-.67 (1.15)	.82 (.78)	.08 (1.23)

Table 5.5 reports the group mean (and SD) efficiency scores on each task. Two-way ANCOVAs were conducted for each learning efficiency calculation. Results indicated there was no significant main effect for presentation format for the 1st attempt, $F(1,21) = 1.59$, $p = .221$, $\eta_p2 = .07$, $MSE = 1.14$. However, there was a significant main presentation effect for the 2nd attempt, $F(1, 21) = 5.46$, $p = .029$, $\eta_p2 = .21$, $MSE = 1.00$; where the static format ($M = .75$, $SE = .29$) led to significantly higher efficiency than the animated format ($M = -.25$, $SE = .29$). There was a significant effect for the transfer task, $F(1, 21) = 9.35$, $p = .006$, $\eta_p2 = .31$, $MSE = .83$; where again the static format ($M = .67$, $SE = .27$) showed significantly higher efficiency than the animated format ($M = -.52$, $SE = .26$).

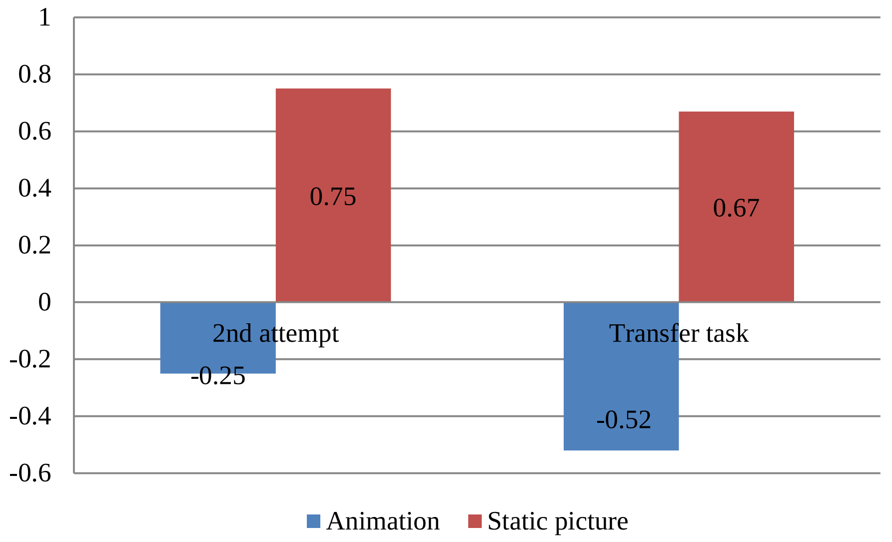


Figure 5.14. Main Presentation-format Effect for Learning Efficiency in the 2nd Attempts and Transfer Task in Experiment 1

There were no significant gender effects for the 1st attempt or for the 2nd attempt (both $F < 1$, ns), or for the transfer task, $F(1, 21) = 1.27$, $p = .272$, $\eta_p 2 = .06$. Furthermore, there were no significant interaction effect for the 1st attempt, $F(1, 21) = 2.35$, $p = .140$, $\eta_p 2 = .10$, or for the transfer task ($F < 1$).

Nonetheless, there was a significant interaction effect for the 2nd attempt, $F(1, 21) = 4.78$, $p = .040$, $\eta_p 2 = .19$. Follow-up tests indicated that for the animation format, there was no significant gender difference, F (1, 10) = 2.20, $p = .169$, $\eta_p^2 = .18$, $MSE = 1.18$; between males ($M = -.88$, $SE = .50$) and females ($M = .21$, $SE = .46$). However, for the static format, $F(1, 10) = 5.50$, $p = .041$, $\eta_p^2 = .36$, $MSE = .76$, there was a higher learning efficiency for males ($M = 1.42$, $SE = .36$) than for females ($M = .23$, $SE = .36$).

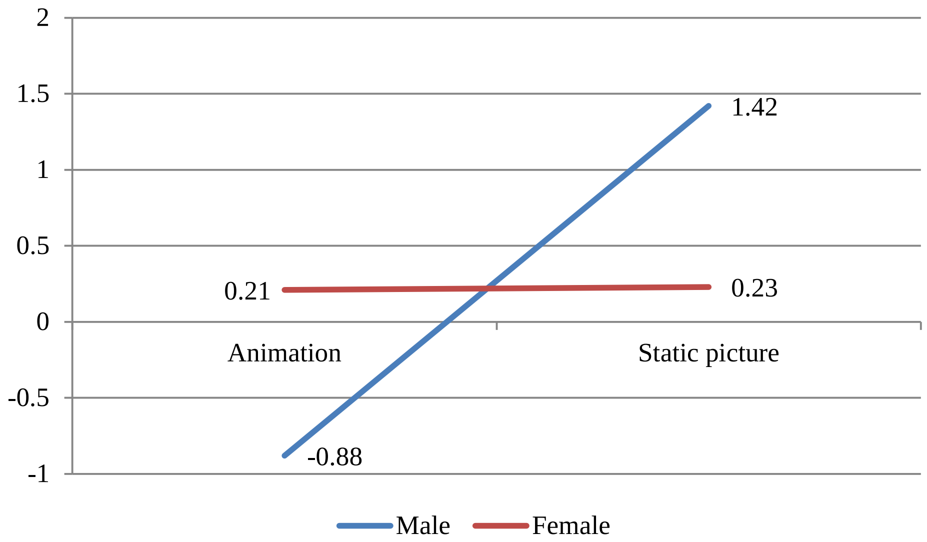


Figure 5.15. Interaction Effects for the Learning Efficiency in the 2nd Attempt in Experiment 1.

5.1.4 Summary of main analysis

In Experiment 1, the use of the CRT test successfully captured the gender difference in spatial ability, and thus ANCOVAs were used throughout the entire experiment. Hypothesis 1 predicted animations would be superior instructional materials to static pictures. The results indicated that there were no significant differences between animations and static pictures in performance scores. However, there was a close-to-significant presentation-format effect for the cognitive load spent on the 1st attempt, and significant effects on the 2nd attempt and transfer task, where participants in the animation group spent more cognitive load than those in the static picture group. Also, the significant presentation-format effect for the learning efficiency on the 2nd attempt and the transfer task supported the same trend, suggesting that animation was inferior to static pictures on these tasks. Hence Hypothesis 1 was rejected.

Hypothesis 2 predicted a gender–presentation format interaction and supporting evidence was found. On performance score there was a significant interaction effect between gender and presentation-format on the 2^{nd} attempt. Simple effects tests indicated that males outperformed females in the static format, but no significant gender difference in the animated format. Such an interaction effect was also found for the learning efficiency on the 2^{nd} attempt.

5.1.5 Predictors of gender performance

In addition to the main analysis answering hypotheses, regression analysis was conducted to explore the potential relationship between gender, spatial ability and animations. This analysis was for exploratory purposes hoping to gain more insights for future directions.

5.1.5.1 Test of subjective spatial measures

Independent *t*-tests were conducted to compare the mean difference of self-rated mental rotational ability and of self-rated overall spatial ability between males and females. Result showed there were no differences in the self-rated mental rotational ability between males (M = 3.54, SD = .97) and females (M = 3.23, SD = .83), $t(24)$ = .87, p = .393, d = .34. Also, there were no differences in the self-rated overall spatial ability between males (M = 3.38, SD = 1.04) and females (M = 3.15, SD = .90), $t(24)$ = .60, p = .55, d = .24.

5.1.5.2 Regression analysis

Correlation coefficients between the three test scores, the CRT score, and the two self-rating measures were calculated for both genders individually (see Table 5.6 and Table 5.7).

Table 5.6.
Pearson Correlation Matrix for Males in Experiment 1(N = 13)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	.382				
Overall spatial ability	.160	.851**			
Task 1 score	.009	-.062	.068		
Task 2 score	-.079	.322	.447	.683*	
Transfer task score	.208	.485	.378	.592*	.674*

Table 5.7.
Pearson Correlation Matrix for Females in Experiment 1(N = 13)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	.353				
Overall spatial ability	.457	. 952**			
Task 1 score	.360	.509	.501		
Task 2 score	.180	.666*	.737*	.383	
Transfer task score	.694**	.382	.454	.524	.432

For each test measure the CRT was expected to be a significant predictor, as was the static-animation condition and therefore both were entered into the regression model. As can be seen from Table 5.7, both self-rating measures were significantly correlated with Task 2 scores for females and therefore they were also entered into the regression model using the enter method.

Regression results for males. For Task 1 (first attempt), Task 2 (second attempt), or Task 3 (transfer) no significant predictors for males were found.

Regression results for females. For Task 1 a close to significant model (p = .081) was found which included only the predictor self-perception of spatial ability (B = 1.48, SE = 2.60, β = .50, R^2 = .25). For Task 2, a single factor significant model was found (p = .004) with self-perception of spatial ability the only predictor (B = 2.33, SE = .65, β = .74, R^2 = .54). For Task 3 a single factor significant model was found (p = .009) with the CRT as the only predictor (B = .035, SE = .01, β = .69, R^2 = .48).

In summary, there were no significant predictors found for male performance. However for females, self-perception of overall spatial ability successfully predicted the performance in retention tasks (Task 1 and Task 2); whilst CRT, the objective assessment of mental rotation ability predicted the transfer task performance.

5.1.6 Assessment of materials

An aim of this experiment was to see if the materials were appropriate to test the given hypotheses. As a number of significant results were found in this small sample confidence was raised that the materials were suitable and therefore they were used again with further modifications in later experiments with scaled up sample sizes.

5.2 Experiment 2

The results found in Experiment 1 indicated no advantage to the animation condition. Moreover, the animation group reported higher cognitive load and lower learning efficiency on some tasks, suggesting that animations were less effective. Nonetheless, the results supported the second hypothesis: there were gender-presentation format interactions, suggesting that males learnt better with static pictures than females, but with no gender differences with animations.

With revised materials and a large sample size, Experiment 2 continued to examine the gender and presentation format interaction effect using a different learning condition. The learning material in the Experiment 1 depicted the Lego construction (manipulative task) without showing hand manipulations. However, Castro-Alonso et al. (2014b) showed there was an interaction between hand-observations and presentation-format with these non-manipulative LEGO tasks; where static pictures were more effective with hands observed, but animation was more effective when hands were not observed. Hence, the main aim of this experiment was to investigate whether the observation of hands would interact differently with gender in this domain. The same hypotheses as Experiment 1 were again tested.

Hypothesis 1: Animations will be superior instructional materials to static pictures
Hypothesis 2: There will be a gender–presentation format interaction.

Using the same groupings as in Experiment 1, the experimental setting was based on a two by two factorial design. There were two independent variables: 1) type of presentation (animation vs. static), and 2) gender (male vs. female), constituting four experimental groups (males learning with animation, males learning with static pictures, females learning with animation, and females learning with static pictures).

5.2.1 Method

5.2.1.1 Participants

The sample contained 59 university students (30 male and 29 female) aged between 17 and 40 (*M* = 22.5, *SD* = 5.29). They were current undergraduate (*N* = 46) and postgraduate (*N* = 13) students at UNSW from various faculties including Arts and Social Sciences, Business, Engineering, Medicine and Science. They were then randomly assigned according to the groupings shown in Table 5.8.

Table 5.8.
Groupings of Participants in Experiment 2

	Animation	Static Picture	Total
Male	16	14	30
Female	14	15	29
Total	30	29	59

5.2.1.2 Materials

The same participants' survey, self-rated spatial ability questions and the spatial ability assessment (CRT) from Experiment 1 were again used (see Section 5.1.1.2).

Learning Materials. The learning materials consisted of a practice task and the main instructional materials.

Practice task. The same practice task in Experiment 1 (without digitally removing the hand manipulations, see Section 5.1.1.2) was again used. Figure 5.16 shows a screen-cap of the animation (left) and the static picture condition (right).

Figure 5.16. Practice Task Materials in Experiment 2 (Left: Animation Condition; Right: Static Picture Condition)

Main instructional materials. The learning materials from the main learning task in Experiment 1 (before digitally removing the hand manipulations) were employed (see Section 5.1.1.2). Figure 5.17 shows a screen-cap of the animation (left) and the static picture condition (right).

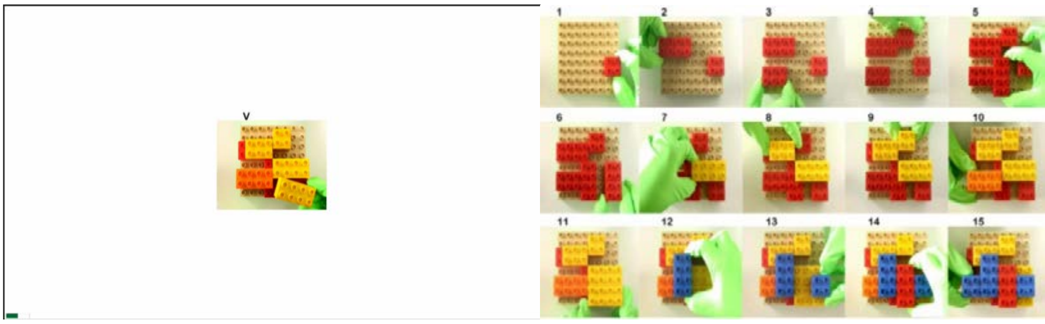


Figure 5.17. Main Instructional Materials in Experiment 2 (Left: Animation Condition; Right: Static Picture Condition)

Subjective measure of cognitive load. The same cognitive load measure from Experiment 1 was re-employed (see Section 5.1.1.2)

5.2.1.3 Testing environment and grading rubric

The same tests (1st attempt, 2nd attempt, & Transfer), testing environment (physical placements of the Lego bricks (see Section 5.1.1.3) and grading rubric (see Section 5.1.2.1) from Experiment 1 was re-employed.

5.2.2 Procedure

The same procedure in Experiment 1 was used again (see Section 5.1.2)

5.2.3 Results

5.2.3.1 Test of spatial differences

A test for initial spatial differences amongst the groups was conducted before conducting the main analysis. The spatial measures (CRT score) collected before the acquisition phase were analysed using a 2 × 2 ANOVA [type of presentation (animation vs. static picture) and gender (males vs. females)]. The spatial measure (CRT scores) was not significant for gender (male = 107.87, female = 101.62), presentation type, or interaction (all $F < 1$, ns).

Hence, the CRT spatial measure was not used as a covariate in this experiment. 2 (animation vs. static picture) × 2 (male vs. female) ANOVAs were used to investigate the two given hypotheses of this experiment.

5.2.3.2 Test performance

Table 5.9.
Mean (SD) of Performance Scores in Experiment 2

		Animation	Static picture	Total
Male	1st attempt	5.43 (3.52)	6.14 (3.44)	5.77 (3.44)
	2nd attempt	7.22 (4.55)	9.82 (4.51)	8.43 (4.64)
	Transfer	3.88 (2.60)	4.07 (2.75)	3.97 (2.63)
Female	1st attempt	7.00 (2.50)	5.43 (3.08)	6.19 (2.88)
	2nd attempt	10.75 (2.83)	8.80 (4.10)	9.74 (3.62)
	Transfer	5.18 (1.58)	3.63 (2.64)	4.38 (2.29)
Total	1st attempt	6.17 (3.13)	5.78 (3.22)	5.98 (3.16)
	2nd attempt	8.87 (4.19)	9.29 (4.25)	9.08 (4.19)
	Transfer	4.48 (2.25)	3.85 (2.65)	4.17 (2.46)

Table 5.9 illustrates the mean (and SD) scores for each task under the different learning conditions. A two-way ANOVA showed that there was no significant main animation effect for the 1st attempt, 2nd attempt (both $F < 1$, ns), or the transfer task, $F(1, 55) = 1.12$, $p = .295$, $\eta_{p2} = .02$, $MSE = 6.00$. In addition, there was no significant main gender effect for the 1st attempt (F < 1, ns), the 2nd attempt, $F(1, 55) = 1.39$, $p = .243$, $\eta_{p2} = .03$; or for the transfer task, ($F < 1$, ns). There was no significant interaction for the 1st attempt, $F(1, 55) = 1.89$, $p = .175$, $\eta_{p2} = .03$, or the transfer task $F(1, 55) = 1.86$, $p = .178$, $\eta_{p2} = .03$. However, there was a significant interaction effect for the 2nd attempt, $F(1, 55) = 4.59$, $p = .037$, $\eta_{p2} = .08$. Follow-up simple effect tests (independent sample t-test) indicated that for the animation format, females ($M = 10.75$, $SD = 2.83$) scored significantly higher than males ($M = 7.22$, $SD = 4.55$); $t(28) = -2.51$, $p = .018$, $d = .93$. However for the static format, there was no significant gender difference, $t(27) = .64$, $p = .528$, $d = .24$, between males ($M = 9.82$, $SD = 4.51$) and females ($M = 8.80$, $SD = 4.10$).

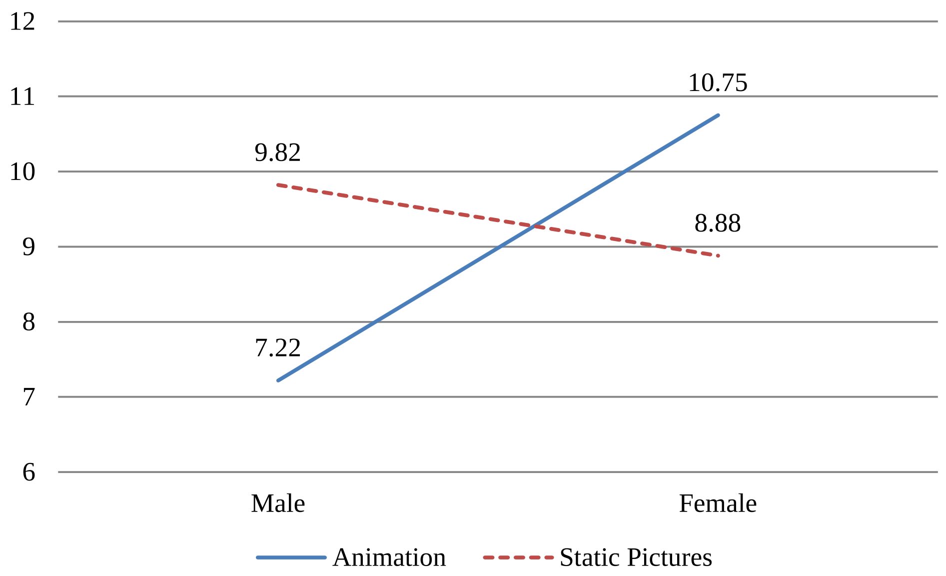


Figure 5.18. Interaction Effect for Performance Score in the 2nd Attempt in Experiment 2

In summary, there was no significant main effect for presentation format or gender. However, there was a significant interaction effect where females outperformed males when learning with animations, but no difference for static pictures (2^{nd} attempt).

5.2.3.3 Cognitive load measures

Participants were asked to reflect on the cognitive load spent on the task upon completion of each task. Table 5.10 illustrates the mean (and SD) score of the reported cognitive load spent in each task. The higher score represents the more cognitive load spent on the task.

Table 5.10.
Mean (SD) of Cognitive Load Spent in Experiment 2

		Animation	Static picture	Total
Male	1^{st} attempt	7.88 (1.50)	6.79 (1.93)	7.37 (1.77)
	2^{nd} attempt	7.25 (1.65)	6.57 (1.83)	6.93 (1.74)
	Transfer	5.81 (2.23)	5.14 (2.48)	5.5 (2.33)
Female	1^{st} attempt	8.00 (.88)	7.13 (1.25)	7.55 (1.15)
	2^{nd} attempt	7.50 (1.16)	6.73 (1.44)	7.10 (1.35)
	Transfer	6.93 (1.82)	6.47 (2.42)	6.69 (2.12)
Total	1^{st} attempt	7.93 (1.23)	6.97 (1.59)	7.46 (1.49)
	2^{nd} attempt	7.37 (1.43)	6.66 (1.44)	7.02 (1.55)
	Transfer	6.33 (2.09)	5.83 (2.49)	6.08 (2.29)

A two-way ANOVA indicated a significant main effect for presentation format for the 1^{st} attempt, $F(1, 55) = 6.80$, $p = .012$, $\eta_p 2 = .11$, $MSE = 2.07$; where the participants in animation group ($M = 7.93$, $SD = 1.23$) rated cognitive load significantly higher than in the static picture group ($M = 6.97$, $SD = 1.59$). Also, there was a close to significant effect ($p < .10$) for the 2^{nd} attempt, $F(1, 55) = 3.23$, $p = .079$, $\eta_p 2 = .06$, $MSE = 2.38$; where again the animation group ($M = 7.37$, $SD = 1.43$) reported higher cognitive load than static picture group ($M = 6.66$, $SD = 1.61$). There was no significance for the transfer task ($F < 1$, ns).

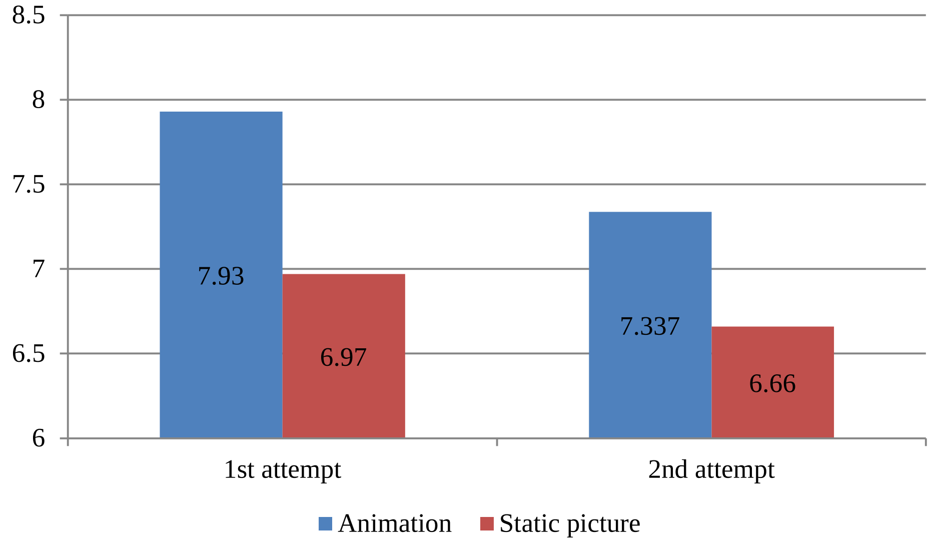


Figure 5.19. Main Presentation-format Effects for Cognitive Load Spent in the 1st and 2nd Attempt in Experiment 2

There were no significant gender effects for the first attempt and the second attempt (both $F < 1$, ns). However, there was a significant gender effect for the transfer task, $F(1, 55) = 4.32$, $p = .04$, $\eta_p 2 = .07$, where females reported spending significantly higher cognitive load ($M = 6.69$, $SD = 2.12$) than males ($M = 5.47$, $SD = 2.33$). There were no significant interaction effects for the 1st attempt, 2nd attempt, or the transfer task (all $F < 1$, ns).

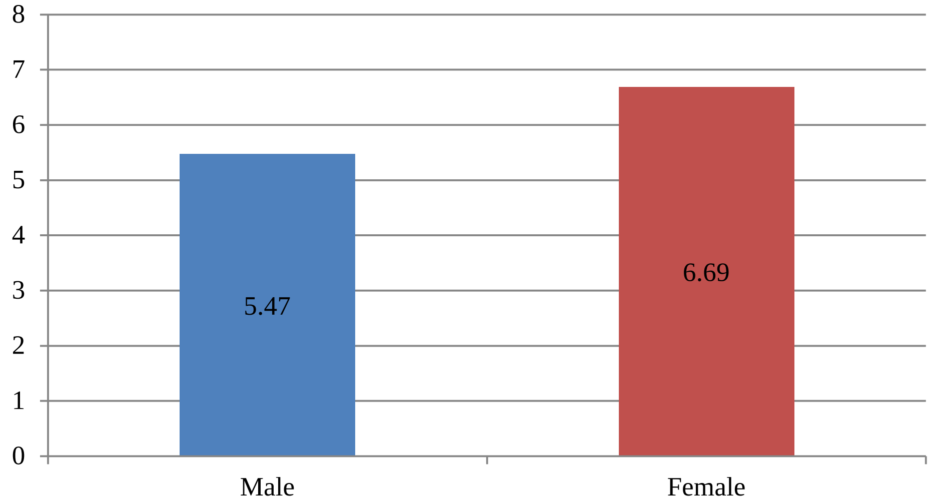


Figure 5.20. Main Gender Effects for Cognitive Load Spent in Transfer Task in Experiment 2

5.2.3.4 Learning efficiency

Table 5.11 below shows the means and SDs for learning efficiency in each task.

Table 5.11.
Mean (SD) of Learning Efficiency in Experiment 2

		Animation	Static picture	Total
Male	1st attempt	-.32 (1.24)	.35 (1.47)	-.003 (1.37)
	2nd attempt	-.42 (1.40)	.33 (1.39)	-.07 (1.42)
	Transfer	-.001 (1.34)	.26 (1.40)	.12 (1.35)
Female	1st attempt	-.03 (.76)	.03 (1.04)	.004 (.90)
	2nd attempt	.06 (.82)	.08 (1.20)	.07 (1.01)
	Transfer	.03 (.94)	-.27 (1.15)	-.13 (1.05)
Total	1st attempt	.18 (1.03)	.19 (1.25)	.00 (1.15)
	2nd attempt	-.20 (1.17)	.20 (1.28)	.00 (1.23)
	Transfer	.01 (1.15)	-.01 (1.28)	

Results showed that there was no significant main animation effect for the 1st attempt, $F(1, 55) = 1.49$, $p = .228$, $\eta_p^2 = .03$, $MSE = 1.34$, for the 2nd attempt, $F(1, 55) = 1.44$, $p = .236$, $\eta_p^2 = .03$, $MSE = 1.52$, or for the transfer task ($F < 1$, ns). Also, there were no significant gender effects for the first attempt 1st attempt, 2nd attempt, or the transfer task, (all $F < 1$, ns). Likewise, there were no significant interactions for the 1st attempt, $F(1, 55) = 1.04$, $p = .313$, $\eta_p^2 = .02$, for the 2nd attempt, $F(1, 55) = 1.29$, $p = .262$, $\eta_p^2 = .02$, or for the transfer task ($F < 1$, ns).

5.2.4 Summary of main analysis

Summarising the results, in contrast with Experiment 1, the CRT did not identify a gender difference in spatial ability. Hypothesis 1 predicted that animations would be superior instructional materials to static pictures. The results showed that there were no significant differences between animations and static pictures in performance scores or instructional efficiency. However, there was a significant presentation-format effect for the cognitive load spent on the 1st attempt, plus a close-to-significant effect on the 2nd attempt, where in both attempts participants in the animation condition spent more cognitive load than the static picture condition. Hence, Hypothesis 1 was rejected. Hypothesis 2 predicted a gender–presentation format interaction and supporting evidence was found. On performance score there was a significant interaction effect between gender and presentation-format in the 2nd attempt. Simple effects tests indicated that females significantly outperformed males in the animation condition but no gender difference was found in the static picture condition.

5.2.5 Predictors of gender performance

Similar with Experiment 1, regression analysis was conducted following the main analysis for exploratory purposes.

5.2.5.1 Test of subjective spatial measures

Independent T-tests were conducted to compare the mean difference of self-rated mental rotational ability and of self-rated overall spatial ability between males and females. Result showed there were no differences in the self-rated mental rotational ability between males (M = 3.43, SD = .63) and females (M = 3.28, SD = .70), $t(57)$ = .91, p = .367, d = .23. Also, there were no differences in the self-rated overall spatial ability between males (M = 3.30, SD = .53) and females (M = 3.14, SD = .64), $t(57)$ = 1.06, p = .295, d = .27.

5.2.5.2 Regression analysis

Table 5.12 and Table 5.13 indicate the correlation coefficients between the three test scores, the CRT score, and the two self-rating measures for both genders individually. As can be seen from the tables only the CRT measure was significantly correlated with test measures, and therefore only it and the static-animation condition was entered into the regression model using the step-wise method.

Table 5.12.
Pearson Correlation Matrix for Males in Experiment 2(N = 30)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	.356				
Overall spatial ability	.192	.422*			
Task 1 score	.476**	.233	.199		
Task 2 score	.542**	.170	.244	.809**	
Transfer task score	.506**	.187	.424*	.805**	.846**

Table 5.13.
Pearson Correlation Matrix for Females in Experiment 2 (N = 29)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	.187				
Overall spatial ability	.166	.549**			
Task 1 score	.227	.079	.121		
Task 2 score	.303	-.027	.024	.675**	
Transfer task score	.172	-.034	-.098	.562**	.649**

Regression results for males. For Task 1, a single factor significant model was found (p = .008) with the CRT as the only predictor (B = .05, SE = .02, β = .48, R^2 = .23). For Task 2, a single factor significant model was found (p = .002) with the CRT as the only predictor (B = .08, SE = .02, β = .54, R^2 = .30). For Task 3 there was a 3-factor significant model (p = .04) with the CRT (B = .04 SE = .01, β = .51, R^2 = .23) and self-perception of spatial ability (*B* = 1.67, *SE* = .78, β = .34, ΔR^2 = .11) significant factors, and animation non-significant.

Regression results for females. For Tasks 1 and 2 there were no significant models. For Task 3, there was a close to significant (*p* = .068) one factor model that included the animation-static condition only (*B* = 1.55, *SE* =, .81 β = .34, R^2 = 12).

In summary, CRT successfully predicted males performance in the retention tasks (Task 1 and Task 2), and the transfer task performance could be predicted from a 3-factor model with CRT score, self-perception of spatial ability and animation. For females there was no model found for the retention tasks (Task 1 and Task 2), but the animation-static condition was found to be a significant predictor for transfer task performance.

5.3 Experiment 3

Experiments 1 and 2 found no advantage for the animated conditions. In fact, some evidence emerged that the static conditions were superior, contradicting the findings of Castro-Alonso et al. (2015a). One possible reason was that in their study they tested with a virtual environment whereas in Experiments 1 & 2 used a real-world environment. This difference of construction environment may explain the conflicting results, as other research has shown that animations are most effective when closely aligned with the conditions of the task (Morrison et al., 2000). Consequently in Experiment 3, the previous experiment was replicated using a virtual testing environment instead of real-life Lego bricks. The same two hypotheses were again tested.

Hypothesis 1: Animations will be superior instructional materials to static pictures

Hypothesis 2: There will be a gender–presentation format interaction.

The experimental setting was based on a two by two design. There were two independent factors: 1) type of presentation (animation vs. static), 2) gender (male vs. female). These two factors constituted four experimental groups: males learning with animation, males learning with static picture, females learning with animation, and females learning with static pictures.

5.3.1 Method

5.3.1.1 Participants

The sample contained 86 university students (42 male and 44 female) aged between 17 and 40 (M = 21.85, SD = 5.64). They were current undergraduate (N = 72) and postgraduate (N = 14) students at UNSW from various faculties including Arts & Social Sciences, Business, Engineering, and Science. Both genders were randomly allocated to treatment conditions according to the groupings shown in Table 5.14.

Table 5.14.
Grouping of Participants in Experiment 3

	Animation	Static Picture	Total
Male	22	20	42
Female	22	22	44
Total	44	42	86

5.3.1.2 Materials

Survey of background knowledge. The same survey from Experiment 1 and 2 was employed. For details, refer to Experiment 1 (Section 5.1.1.2).

Objective assessment of spatial ability. The same Card Rotation Test (CRT) used in Experiments 1 and 2 (see Appendix 2) was re-employed in this experiment (see Experiment 1 Section 5.1.1.2).

Learning materials. The identical learning materials (both practice task and main instructional materials) from Experiment 2 were used (see Section 5.1.1.2).

5.3.1.3 Testing environment

Practice task. An Adobe flash (.swf) file was developed using Adobe Premiere Pro CS6. The same number and colour of bricks as in the practice task learning material were depicted on the computer screen. The bricks were arranged in the exact same order as in the learning material starting from left to right in a vertical manner (see Figure 5.21). The .swf file frame-sized 800 x 600 pixels was embedded into the instruction flash file (frame size 1200 x 800) so it would appear automatically after the learning material was presented at the appropriate time frame. Participants would be required to rebuild the observed pattern onto the blue square stage (top left relative to the observer) with computerised bricks (bottom half relative to the observer). Participant would be able to perform the following operations: drag the brick to anywhere on the screen by clicking-and-holding the computer mouse button; rotate the brick double-clicking the mouse; and place the brick by releasing the mouse button. The order and numbers of operations (i.e. clicks, unclicks and double-clicks) were not limited. Moreover, moving only one brick was allowed each time in the given sequence. After the brick position was confirmed, participants were required to press the next button on the top right corner of the screen in order to start moving the next brick. However, the existing brick position would be locked and no change would be allowed afterwards.

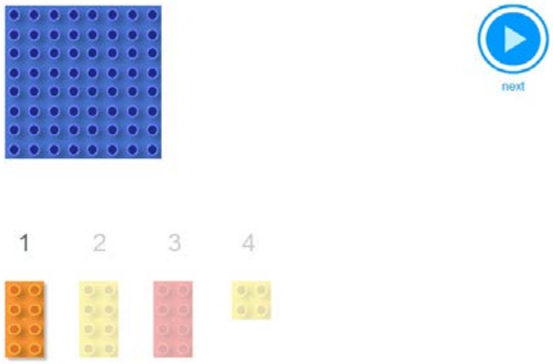


Figure 5.21. Performing Environment in Experiment 3 Practice Task

Completion Task. Similar to the practice task, an Adobe flash (.swf) file was developed using Adobe Premiere Pro CS3. The same number and colour of bricks were depicted as in the presented task (see Figure 5.22). The bricks were arranged from left to right in the exact same order as in the learning material, but in a vertical manner. The .swf file frame-sized 800 x 600 pixels was embedded into the instruction flash file (frame size 1200 x 800) so it would appear automatically after the learning material was presented at the appropriate time frame. Identical to the practice task, the bricks could be dragged with a computer mouse and be placed anywhere onto the brown square platform (top left). Bricks could be rotated by double-clicking the mouse. Furthermore, moving only one brick was allowed at a time and no change was allowed after the position was confirmed. The brick positions were confirmed and locked by clicking the next button on the top right of the screen (see Figure 5.22).

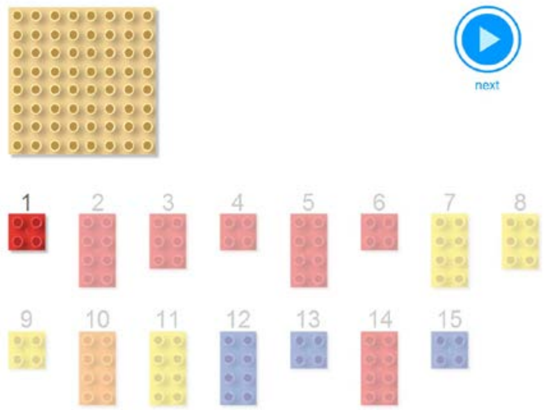


Figure 5.22. Performing Environment Used in Experiment 3 Completion Task

Transfer Task. Using the same procedures as the practice task and the completion task, an Adobe flash (.swf) file was developed using Adobe Premiere Pro CS3. All 6 bricks were placed vertically and were arranged according to the order presented in the completion task starting from left to right (see Figure 5.23). The .swf file frame-sized 800 x 600 pixels was embedded into the instruction flash file (frame size 1200 x 800) so it would appear automatically after the learning material was presented at the appropriate time frame. The bricks could be dragged with a computer mouse and be placed anywhere on the brown square stage. The brick could be rotated with a double-click. Furthermore, moving only one brick was allowed at a time and no change was allowed after the position was confirmed. The brick positions were confirmed and locked by clicking the next button (see Figure 5.23).

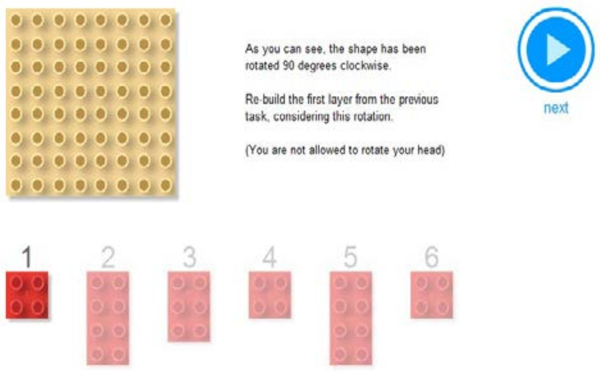


Figure 5.23. Performing Environment used in Experiment 3 Transfer Task

5.3.1.4 Grading rubrics

The same set of grading rubrics used in Experiments 1 and 2 were readopted (see Section 5.1.2.1).

5.3.2 Procedure

The procedure was similar to the previous experiments. The experimental sessions were conducted in a quiet room with only one student in each session, which lasted about 40 minutes. Participants were randomly allocated into two groups (animation vs. static picture) before the experiment sessions started. Consistent with Experiment 1, the instructions of the study as well as the learning materials were embedded into one flash programme to ensure every student received the same instructions. However, the procedure was slightly modified from Experiment 1 to provide more information on moving the bricks in the testing phase in order to enhance better clarity, especially towards to use of the computerised Lego.

All participants were firstly invited to sit in front of a computer (with experiment programme ready). They then signed the experimental consent form (Appendix 4) followed by the questionnaire (Appendix 1) related to background information and self-rated spatial ability. The CRT instrument, assessing mental rotation ability, was then given. Afterwards, in order to ensure every participant felt comfortable moving the computerised bricks and were familiar with the rules, a little demonstration and short trial-time was given before starting the practice task. Participants were told to place the bricks anywhere on the base just to get familiar with the use of the computerised bricks and manipulation. In particular, they were explicitly told to try all manipulations i.e. rotating the brick (double-click the computer mouse), dragging the brick (click-and-hold the mouse), placing the brick (release the mouse) and confirming the position (the blue next button). After participants were familiar with the environment and the rules, they completed the practice task, the main learning task, and finally the transfer task along with the respective cognitive load questions with the same procedure as used in Experiment 1.

5.3.3 Results

5.3.3.1 Test of spatial differences

Before conducting the main analysis, a test for initial spatial differences amongst the groups was conducted. The spatial measures (CRT score) collected before the acquisition phase were analysed using a 2 × 2 ANOVA [type of presentation (animation vs. static picture) and gender (males vs. females)]. There was no main effect for gender $F(1, 82) = 2.11$, $p = .151$, $\eta_p^2 = .03$, $MSE = 253.92$, presentation type $F(1, 82) = 2.90$, $p = .093$, $\eta_p^2 = .03$, or interaction $F(1, 55) = 1.11$, $p = .295$, $\eta_p^2 = .01$.Hence, the CRT spatial measure was not used as a covariate in this experiment. Two (animation vs. static picture) × two (male vs. female) ANOVAs were conducted to investigate the hypotheses.

5.3.3.2 Test performance

Table 5.15 below shows the mean (and SD) scores for each task under the different learning conditions.

Table 5.15.
Mean (SD) of Performance Scores in Experiment 3

		Animation	Static picture	Total
Male	1st attempt	6.23 (3.25)	6.63 (2.95)	6.42 (3.08)
	2nd attempt	9.84 (4.07)	11.33 (3.21)	10.55 (3.72)
	Transfer	5.00 (2.12)	5.65 (2.12)	5.31 (2.12)
Female	1st attempt	7.68 (1.82)	6.32 (2.55)	7.00 (2.30)
	2nd attempt	11.36 (2.30)	9.91 (3.48)	10.64 (3.00)
	Transfer	5.80 (1.41)	4.64 (2.13)	5.22 (1.88)
Total	1st attempt	6.96 (2.70)	6.46 (2.72)	6.72 (2.71)
	2nd attempt	10.60 (3.35)	10.58 (3.39)	10.59 (3.35)
	Transfer	5.40 (1.83)	5.12 (2.16)	5.26 (1.99)

A two-way ANOVA showed that there was no significant main animation effect for the 1st attempt, the 2nd attempt, or the transfer task (all $F < 1$, ns). Likewise, there was also no significant main gender effect for the 1st attempt, the 2nd attempt (both $F <$ 1, ns), or the transfer task $F(1, 82) = 2.30$, $p = .13$, $\eta_p^2 = .03$. Also, there was no significant interaction for the 1st attempt $F(1, 82) = 2.30$, $p = .13$, $\eta_p^2 = .03$.

However, there were significant interactions for the 2nd attempt $F(1, 82) = 4.18$, $p = .044$, $\eta_p^2 = .05$, and for the transfer task $F(1, 82) = 4.55$, $p = .036$, $\eta_p^2 = .05$. Follow-up simple effect tests (independent sample t-tests) showed there was no significant difference between males ($M = 9.84$, $SD = 4.07$) and females ($M = 11.36$, $SD = 2.30$) for the 2nd attempt in the animation format, $t(33.19) = -1.53$, $p = .136$, $d = -.46$. Similarly, there was no significant difference between males ($M = 11.33$, $SD = 3.21$) and females ($M = 9.91$, $SD = 3.48$) in the static format, $t(40) = 1.37$, $p = .179$, $d = .42$. For the transfer task, simple effect tests again indicated there was no significant difference between males ($M = 5.00$, $SD = 2.12$) and females ($M = 5.80$, $SD = 1.41$) in the animated format $t(36.55) = -1.46$, $p = .152$, $d = -.44$. Likewise, there was no significant difference between male ($M = 5.65$, $SD = 2.12$) and female ($M = 4.64$, $SD = 2.13$) in the static format, $t(40) = 1.55$, $p = .130$, $d = .48$.

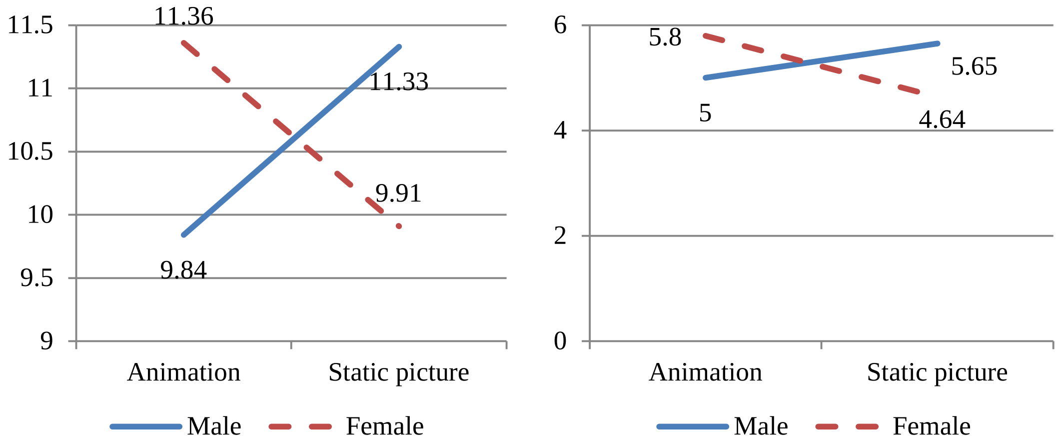


Figure 5.24. Interaction Effects for Performance Score in Experiment 3 (Left: 2nd Attempt; Right: Transfer Task)

5.3.3.3 Cognitive load measure

Table 5.16 shows the mean (and SD) scores of the self-report cognitive load spent on each task under the different learning conditions. The higher score represents the more cognitive load spent on the task.

Table 5.16.
Mean (SD) of Cognitive Load Spent in Experiment 3

		Animation	Static picture	Total
Male	1st attempt	7.77 (1.51)	7.10 (1.29)	7.45 (1.44)
	2nd attempt	6.91 (1.57)	6.25 (1.74)	6.66 (1.67)
	Transfer	6.09 (1.85)	5.15 (2.37)	5.64 (2.14)
Female	1st attempt	7.68 (1.00)	7.68 (1.09)	7.68 (1.03)
	2nd attempt	6.77 (1.11)	6.73 (1.52)	6.75 (1.31)
	Transfer	6.18 (2.26)	6.00 (2.05)	6.09 (2.13)
Total	1st attempt	7.73 (1.26)	7.40 (1.21)	7.57 (1.24)
	2nd attempt	6.84 (1.65)	6.50 (1.63)	6.67 (1.49)
	Transfer	6.14 (2.04)	5.60 (2.22)	5.87 (2.14)

Two-way ANOVAs showed that there was no significant main animation effect for the 1st attempt, $F(1, 82) = 1.59$, $p = .21$, $\eta_p^2 = .02$, $MSE = 1.53$, the 2nd attempt, $F(1, 82) = 1.19$, $p = .28$, $\eta_p^2 = .02$, $MSE = 2.24$, or for the transfer task, $F(1, 82) = 1.48$, $p = .23$, $\eta_p^2 = .02$, $MSE = 4.56$. Likewise, there was also no significant main gender effect for the 1st attempt, the 2nd attempt (both $F < 1$, ns), and for the transfer task $F(1, 82) = 1.04$, $p = .31$, $\eta_p^2 = .01$. Also, there was no significant interaction for the 1st attempt $F(1, 82) = 1.59$, $p = .21$, $\eta_p^2 = .02$, the 2nd attempt, and for the transfer task (both $F < 1$, ns).

5.3.3.4 Learning efficiency

Table 5.17 below shows the mean (and SD) scores of the self-reported learning efficiency spent on each task under the different learning conditions.

Table 5.17.
Mean (SD) of Learning Efficiency in Experiment 3

		Animation	Static picture	Total
Male	1st attempt	-.24 (1.37)	.24 (1.27)	-.01 (1.33)
	2nd attempt	-.27 (1.44)	.36 (1.30)	.02 (1.39)
	Transfer	-.17 (1.28)	.38 (1.35)	.09 (1.32)
Female	1st attempt	.19 (.79)	-.17 (.99)	.01 (.90)
	2nd attempt	.12 (.80)	-.17 (1.26)	-.03 (1.05)
	Transfer	.09 (1.13)	-.26 (1.06)	-.09 (1.10)
Total	1st attempt	-.03 (1.13)	.028 (1.13)	.00 (1.12)
	2nd attempt	-.08 (1.17)	.08 (1.29)	.00 (1.22)
	Transfer	-.04 (1.20)	.04 (1.24)	.00 (1.21)

Two-way ANOVAs showed that there was no significant main animation effect for the 1st attempt, the 2nd attempt, or for the transfer task (all $F < 1$, ns). Likewise, there was also no significant main gender effect for the 1st attempt, the 2nd attempt or the transfer task (all $F < 1$, ns).

However, there were close-to-significant ($p < .10$) interaction effects for the 1st attempt $F(1, 82) = 3.02$, $p = .086$, $\eta_p^2 = .04$, for the 2nd attempt, $F(1, 82) = 2.98$, p

= .088, η_p^2 = .04, and for the transfer task *F*(1, 82) = 2.96, *p* = .089, η_p^2 = .04. Follow-up simple effect tests (independent sample t-tests) showed that in animation format, there was no significant difference between males (*M* = -.24, *SD* = 1.37) and females (*M* = .19, *SD* = .79) for the 1st attempt, *t*(33.67) = -1.28, *p* = .209, *d* = -.38. Also, no significant difference between males (*M* = -.27, *SD* = 1.44) and females (*M* = .12, *SD* = .80) were found for the 2nd attempt, *t*(32.99) = -1.10, *p* = .280, *d* = . Likewise, no significant difference between males (*M* = -.17, *SD* = 1.28) and females (*M* = .09, *SD* = 1.13) were found for the transfer task, *t*(42) = -.70, *p* = .490, *d* = -.22. For the static format, tests demonstrated there was no significant difference between males (*M* = .24, *SD* = 1.26) and females (*M* = -.17, *SD* = .99) for the 1st attempt, *t*(40) = 1.18, *p* = .245, *d* = .36. Also, no significant difference between males (*M* = .36, *SD* = 1.30) and females (*M* = -.17, *SD* = 1.26) were found for the 2nd attempt, *t*(40) = 1.33, *p* = .191, *d* = .41. However, a close-to-significant (*p* < .10) difference between males (*M* = .38, *SD* = 1.35) and females (*M* = -.26, *SD* = 1.06) were found for the transfer task, *t*(40) = 1.72, *p* = .093, *d* = .53 where males performed better than females. Although no significant differences were found in the simple effect tests, there is a consistent pattern demonstrating males found the static format to be more effective but females found animation to be more effective.

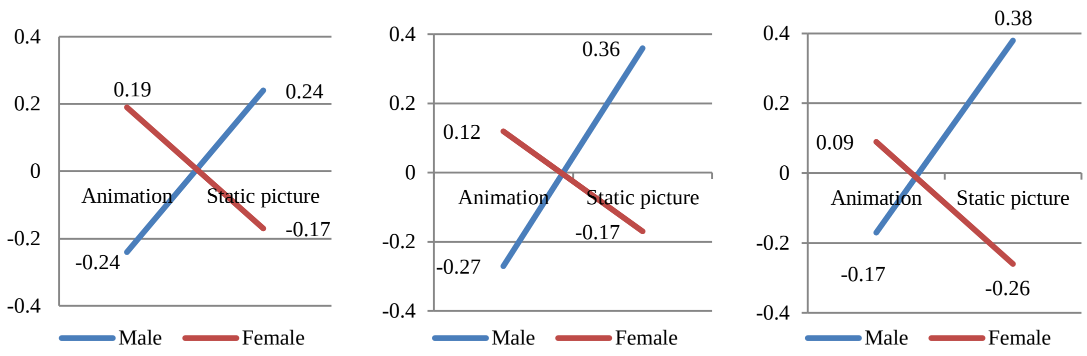


Figure 5.25. Interaction Effects for Learning Efficiency in Experiment 3 (Left: 1st Attempt; Middle: 2nd Attempt; Right: Transfer Task)

5.3.4 Summary of main analysis

Hypothesis 1 predicted an animation effect. The result was consistent with those from Experiment 1and 2, in that no overall advantage was found for the animated format. Thus hypothesis 1 is again rejected. However, considerable support was found for Hypothesis 2 in that two gender and presentation interactions were found on the second attempt and the transfer task. Furthermore interactions ($p < .10$) were found on all three learning efficiency measures. Both performance and efficiency results suggest that comparative performance between males and females was moderated by the original presentational format. Although no simple effects tests were significant, there was a consistent pattern demonstrating males found the static format to be more helpful but females found animation to be more helpful.

It was notable that the CRT (the spatial ability test) once again failed to capture any gender difference in spatial ability. The result is consistent with that from the Experiment 2, but incompatible with that from the Experiment 1.The inconsistency of the CRT results suggested a need for a different spatial ability measurement which was used in Experiment 4.

5.3.5 Predictors of gender performance

Consistent with Experiment 1 and Experiment 2, regression analysis was conducted following the main analysis for exploratory purposes.

5.3.5.1 Test of subjective spatial measures

Independent T-tests were conducted to compare the mean difference of self-rated mental rotational ability and of self-rated overall spatial ability between males and females. Result showed there were no differences in the self-rated mental rotational ability between males (M = 3.50, SD = .77) and females (M = 3.45, SD = .79), t(84)

= .27, *p* = .788, *d* = .06. Also, there were no differences in the self-rated overall spatial ability between males (*M* = 3.43, *SD* = .74) and females (*M* = 3.23, *SD* = .80), *t*(84) = 1.21, *p* = .230, *d* = .26.

5.3.5.2 Regression analysis

Table 5.18 and Table 5.19 indicate the correlation coefficients between the three test scores, the CRT score, and the two self-rating measures for both genders individually. Both self-rating measures and the CRT significantly correlated with Task 2 for males, and therefore these measures and the animation-static condition were initially entered into the regression models.

Table 5.18.
Pearson Correlation Matrix for Males in Experiment 3 (*N* = 42)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	-.213				
Overall spatial ability	-.083	.728**			
Task 1 score	-.159	.003	-.059		
Task 2 score	.339*	.395**	.317*	.711**	
Transfer task score	.239	.238	.296	.586**	.831**

Table 5.19.
Pearson Correlation Matrix for Females in Experiment 3 (N = 44)

	Correlations				
	CRT	Self-rated mental rotation	Overall spatial ability	Task 1 score	Task 2 score
CRT	1				
Self-rated mental rotation	.221				
Overall spatial ability	.168	.566**			
Task 1 score	.048	.128	.025		
Task 2 score	-.008	-.120	-.191	.640**	
Transfer task score	.149	.018	-.002	.421**	.436**

Regression results for males. For Test 1 there was no significant model. For Test 2 there was a significant 2-factor model (*p* = .007) that included self-reported mental rotation ability (*B* = 1.62, *SE* = .69, β = .34, R^2 = .16) and the CRT measure (*B* = .06, *SE* = .03, β = .27, ΔR^2 = .07, *p* = .072). However, no significant predictors were found for the transfer task.

Regression results for females. For Test 1 there was a significant one-factor model (*p* = .039) that included the animation-static condition (*B* = 1.36, *SE* = .69, β = .30, R^2 = .09). For Test 2 there was no significant model. For Test 3 there was a significant one-factor model (*p* = .047) that included the animation-static condition (*B* = 1.16, *SE* = .54, β = .31, R^2 = .09).

In summary, the self-rated mental rotation ability and CRT score successfully predicted males Task 2 performance. However, there was no other significant model for Task 1 and the transfer task. In contrast, for females, there was no model found for Task 2 performance, but animation-static condition was found to be a good predictor for Task 1 and transfer task.

5.4 Experiment 4

Experiments 1, 2 and 3 consistently yielded no overall advantage for animations but gender-presentation format interactions were observed under both real and virtual testing conditions, as well as showing hands and not showing hands conditions. In conclusion hypothesis 1 was not supported but hypothesis 2 was. In order to examine the influence of other factors on gender and animations, the static condition used previously was excluded in Experiment 4 and therefore hypotheses 1 and 2 were not tested.

Instead two new variables (hand observations and gestures) were added. As a consequence two new hypotheses outlined in Chapter 3 were tested:

Hypothesis 3: Observing hands will lead to superior learning than not observing hands
Hypothesis 4: Gesturing will lead to superior learning than no gesturing

Although no specific hypotheses were predicted (see Chapter 3) a further aim of this experiment was to investigate potential interactions between gender, hands, and gesturing. Furthermore, as indicated in the discussion of Experiment 3, the CRT was not successful in identifying gender differences and therefore two further spatial ability tests were introduced, namely the Mental Rotations Test (Peters et al., 1995) and the Corsi test (Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000).

5.4.1 Method

The experiment used a 2×2×2 design. There were three independent factors: 1) Hand observation (With-hand vs. Without-hand), 2) Gender (male vs. female), and 3) Gestures (gesturing vs. no gesturing).

5.4.1.1 Participants

The sample contained 120 university students (60 male and 60 female) aged between 17 and 46 (*M* = 24.43, *SD* = 6.24). They were current undergraduate (*N* = 70) and postgraduate (*N* = 50) students at UNSW from various faculties. They were randomly allocated to treatments according to the groupings shown in Table 5.20.

Table 5.20.
Groupings of Participants in Experiment 4

	With-hands	Without-hands
Gesturing	Male: 15	Male: 15
	Female: 15	Female: 15
No-Gesturing	Male: 15	Male: 15
	Female: 15	Female: 15

5.4.1.2 Materials

Survey of background knowledge. The survey of background knowledge from previous experiments was modified (see Appendix 5). Questions on participants' basic background information (gender, age, study programme, year of study, level of study, and handedness) and self-rated spatial ability questions were retained. In addition, two questions (question 3 and 4 below) related to Lego experience were added to capture participants' familiarity with Lego. Both questions were in a 5-point Likert scale rating listed below.

3. How much experience have you had playing with Lego (or similar) bricks?
None ☐ Some ☐ Quite a bit ☐ Much ☐ Very Much ☐

4. How would you consider your ability to build shapes with Lego (or similar) bricks?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

Furthermore, the survey also introduced two new questions on frequency of learning motor-related tasks with animation or static picture respectively to capture participant familiarity of learning with animations and static pictures. These two questions were in a 5-point Likert scale rating from *Never, Rarely, Occasionally, Frequently* to *Very Frequently.*

5. When learning motor-related task (e.g. tying knots, folding paper…etc), how often did you learn it from animation/ video?
Never ☐ Rarely ☐ Occasionally ☐ Frequently ☐ Very Frequently ☐

6. When learning motor-related task (e.g. tying knots, folding paper…etc), how often did you learn it from pictures/ books?
Never ☐ Rarely ☐ Occasionally ☐ Frequently ☐ Very Frequently ☐

Lastly, a question related to participants' preference on presentation type of instructional material (animation or static picture) before and after they participated in the experiment was added to capture the change of preference (if any).

Objective assessment of spatial ability. The CRT measure was not found to be consistently useful in previous experiments as there was a significant gender difference on this measure in Experiment 1, but no significant main effect or correlation with participants' performance in Experiments 2 and 3. Hence the CRT was replaced by the Mental Rotations Test (MRT) in this experiment. The MRT was originally designed by Vandenberg and Kuse (1978) based on the figures provided by Shepard and Metzler (1971). However, the available versions were of poor quality and thus a redrawn version of the MRT (Peters et al., 1995) was employed in this experiment with the permission from the author. Due to the agreement made with the author, detailed test questions of the MRT are not attached in this thesis.

There were four different variations of the redrawn version available and only the standard set, MRT-A, was used. In the test, there were 24 problem sets with each problem has a target figure shown on the left and four stimulus figures shown on the right (see Figure 5.26 below for an example). Two of the four stimulus figures were rotated version of the target figure and two others were different figures. Participants were required to identify and mark the rotated version as many as they could within the time limit. The test paper consisted of two pages of instructions with a practice task (3 problem sets), and four pages of test questions (6 problem sets each page). Participants were required to complete 12 problem sets (half of the total questions) in 3 minutes. They would then have a short break before starting the second half.

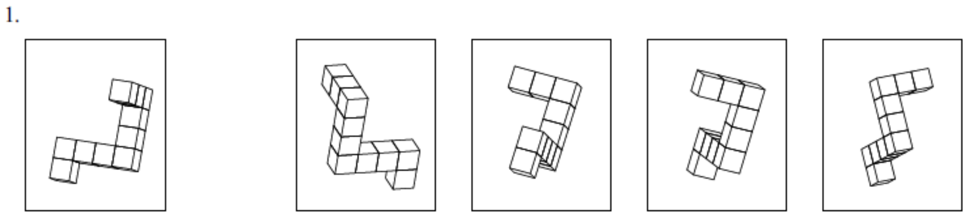


Figure 5.26. The Mental Rotation Test-A (Peters et al, 1995) used in Experiment 4

Objective assessment of visual spatial memory span. Besides the MRT, the Corsi Blocking Tapping test (referred to as The Corsi from here on) was also introduced into this experiment. The Corsi was developed by Philip Corsi in 1972 and it is a pure measurement of the visuospatial memory span (Logie, 1995; Smirni, Villardita, & Zappalá, 1983). The origin version from Corsi (1972) was built with 9 wooden block glued onto a large wooden board. It was operated by an experimenter pointing to the blocks in a definite order. Later a number of researchers computerised the physical blocks into digital blocks and different variations now exist. In this experiment, the block allocations were derived from Kessels et al. (2000) versions as they reported each and every block's coordination which made it replicable. In order to fit the Corsi, of which the stage size was 255 × 205mm (726.75 × 584.25 pixels), into the experiment's flash (.fla) file of which the screen size was 3100 × 810 pixels, the blocks' coordination was tilted to the centred without changing relative locations (see Figure 5.27 below). The digits 1 to 9 were added for a clear explanation in this thesis only. Participants would see only 9 blank white blocks.

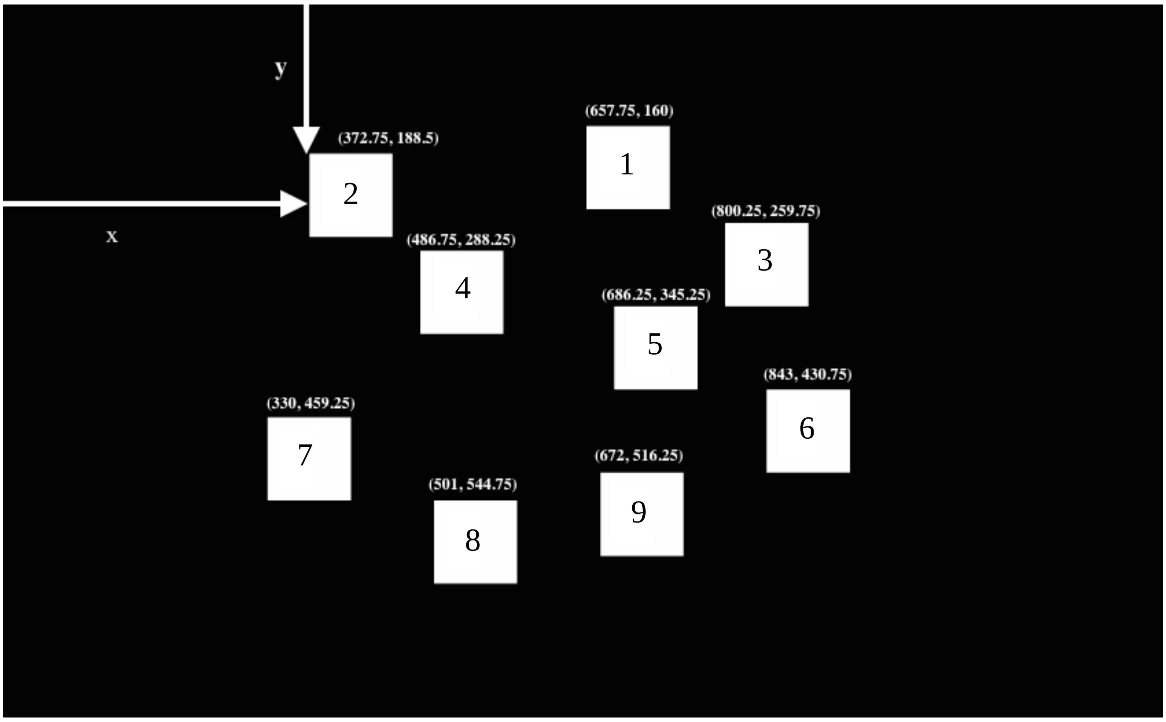


Figure 5.27. The Corsi Test Used in Experiment 4 (Modified from Kessels, Zandvoort, Postma, Kappelle, & Haan, 2000)

Instead of having an experimenter pointing to the block like in the physical version, the computer indicated the block sequence with a yellow flash. The block would flash in yellow for one second in designated order (see Table 5.21) with half a second break in between. Participants were required to memorise the sequence and then repeat back by clicking on the correspondent block in the correct order. The block sequences were adopted from Smirni et al. (1983) starting from blocks-of-five to blocks-of-nine. Each participant would have all 15 sequences and was scored on the number of correct blocks recalled in the correct presented order (Busch et al., 2005; Kaufman, 2007; Pagulayan, Busch, Medina, Bartok, & Krikorian, 2006). One mark would be given to each correct block in a correct order and the maximum score was 105.

Table 5.21.
Set of Sequence Used in The Corsi Blocking Tapping Test (Adopted from Smirni, Villardita, and Zappalá, 1983)

5–2–1–8–6	9–2–7–6–1–9	5–9–3–6–7–2–4–3
4–2–7–3–1	5–9–1–7–4–2–8	3–6–5–1–9–1–2–7
9–7–5–8–3	5–7–9–2–8–4–6	5–3–8–7–1–2–4–6–9
3–9–2–4–8–7	1–9–6–2–7–9–1	4–2–6–8–1–7–9–3–5
3–7–8–2–9–4	5–8–1–9–2–6–4–7	2–7–5–8–6–2–5–8–4

Learning Materials. The identical learning materials from Experiments 2 and 3 were used in the with-hand condition (see Section 5.2.1.2), and the identical learning materials from Experiment 1 were used in the without-hand condition (see Section 5.1.1.2).

Cognitive Load Measure. In Experiments 1, 2 and 3, the cognitive load measure was designed based on the original 1-item instrument developed by (Paas et al., 1994), but it did not capture the differences in conditions consistently. Hence a new set of cognitive load measure were used based on the 13-item-instrument constructed by Leppink et al. (2014). In the original version, there were 13 questions – 4 for intrinsic load, 4 for extraneous load and 5 for germane load. Only the eight questions about intrinsic load (question a – d) and extraneous load (question e – f) were modified and used here (see Table 5.22) as some researchers believe germane load cannot be separated from intrinsic load (see Sweller et al., 2011). The cognitive load measure was designed to examine participants' cognitive load spent on understanding the learning material (the Lego videos). A difference from the previous experiments was that cognitive load was previously measured after *each* task, but for the new questionnaire, respective cognitive loads were measured at the end of the last test (transfer). To ensure the participants answered the questions based on the Lego instructions but not the experimental instructions or spatial ability tests instructions, the experimenter verbally reminded the participants the aim of the questions before they completed this questionnaire.

Table 5.22.
Cognitive Load Measure in Experiment 4 (Adopted and Modified from Leppink et al., 2014)

6. All of the following 8 questions refer to the activity that just finished. Please take your time to read each of the questions carefully and respond to each of the questions by ticking "✓" the most appropriate box.

	Not at all the case									*Completely the case*
a) The content of the video (building a Lego shape) was very complex	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b) The problem (this particular Lego structure) covered in the video was very complex	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c) In the video, very complex concepts were involved.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d) I invested a very high mental effort in the complexity of this activity.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
e) The instructions in the video were very unclear.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
f) The instructions in the video were full of unnecessary components.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
g) The instructions in the video were, in terms of learning, very ineffective.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
h) I invested a very high mental effort in unclear and ineffective instructions in this activity.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

Gesturing conditions

Participants in both gesture and non-gesture groups were notified that their hands would be video-taped during the learning and testing phase. Thus they were all aware of the prohibition (non-gesture group) and encouragement (gesture group) of their hand movement.

Non-gesture group. For the non-gesturing group, participants were told to lay their hands flat on the table in front of them and not to move when watching the video. After the video finished, they were invited to move their hands and rebuild the observed shape onto the computer using a computer mouse.

Gesture group. Researchers (e.g. Chu & Kita, 2008; Cook, Mitchell, & Goldin-Meadow, 2008; Stevanoni & Salmon, 2005) found that free-gestures (i.e. participants are encouraged to make their own gestures without any pre-instructions) work better than instructed-gestures (i.e. the set of gesture used in the main task was instructed beforehand). Hence, in this experiment, the participants in the gesture group were encouraged to gesture in their own style while they were watching the video. They were told that they could move their hands while watching the video because it (the action of moving hands) may help them memorise the brick pattern. In order to give them a better idea of the purpose of the hand movement and to encourage them using gestures, the experimenter gave some examples such as: participants could move along with the hand (with hand condition) or with the brick (without hand condition) in the video, or try to simulate the brick pattern, or even the brick position. Moreover, the experimenter also explained that there were no limitations or restrictions on hand movements and participants have the complete freedom to decide the gesture. The exact scripts presented are recorded below.

With-hand script: While you are watching the video, your hands are encouraged to move because it may help you to memorise the brick pattern better. You can move along with the <u>*hand*</u> *in the video, or try to simulate the brick pattern, or the brick position. You can move wherever and whenever you feel like to. It is all up to you. There are no any limitations or restrictions on your hand movements. As long as you find it useful, please feel free to move.*

Without-hand script: While you are watching the video, your hands are encouraged to move because it may help you to memorise the brick pattern better. You can move along with the brick in the video, or try to simulate the brick pattern, or the brick position. You can move wherever and whenever you feel like to. It is all up to you. There are no any limitations or restrictions on your hand movements. As long as you find it useful, please feel free to move

5.4.1.3 Testing environment

Practice task, completion tasks and transfer task. The same virtual condition of the testing environment in Experiment 3 was employed (see Section 5.3.1.3).

5.4.1.4 Grading rubrics

The same set of grading rubrics from Experiment 1 was adopted (see Section 5.1.2.1).

5.4.2 Procedure

The experiment sessions were conducted in a quiet room with only one participant each time, and each session was about 40-45 minutes. Participants firstly signed the consent form and completed the questionnaire (Appendix 4) related to background information and self-rated spatial ability. Then they were invited to complete the MRT followed by the Corsi tests. Afterwards, consistent with Experiment 3, participants were given a little demonstration and short trial-time before starting the practice task. After participants were familiar with the virtual environment and the rules, they started the practice task, learning tasks followed by transfer task with the same sequence as in previous experiments. Before participants started watching each of the materials (videos), they were reminded about the use of (for gesturing group) or the limitation of (for non-gesturing group) their hand movements.

5.4.3 Results

5.4.3.1 Evaluating the cognitive load measure questionnaire

A factor analysis and reliability test (Cronbach's alpha) was conducted to examine the internal reliability of the questionnaire. Initially, a principal factor analysis was conducted on the eight items. The Kaiser-Meyer-Olkin (KMO) was .726, which verified the sample adequacy for analysis. Bartlett's test of sphericity was strongly significant, $\chi^2(28) = 288.95$, $p < .001$. Since the two-factor solution was expected and correlation between the factors were expected (based on Leppink et al., 2014), oblique rotation was performed to account for the intercorrelation of the factors. Table 5.23 presents the factor loading of each item, the Cronbach's alpha value of the factor, sub-item (corrected item-total correlation) and after deletion of the item (alpha if item deleted). The Cronbach's alpha of both intrinsic load measure (4 items, α = .83) and extraneous load measure (4 items, α = .77) were found to be highly reliable.

Table 5.23.
Reliability Test Result (Cronbach's Alpha Value) in Experiment 4

Factor/item (N=120)	Factor loading	Cronbach's alpha	Corrected item total correlation	Alpha if item deleted
Intrinsic load		.804		
Item 1	.779		.600	.765
Item 2	.841		.687	.726
Item 3	.807		.633	.752
Item 4	.757		.576	.776
Extraneous load		.748		
Item 5	.738		.530	.701
Item 6	.686		.466	.739
Item 7	.788		.598	.674
Item 8	.829		.655	.622

5.4.3.2 Tests of spatial differences

ANOVAs were conducted for the two spatial ability scores with the independent variables (gender, hands-observation, and gesture) to investigate pre-existing differences. Results showed that, for the MRT scores, there was a significant main gender effect, $F(1, 112) = 10.48$, $p = .002$, $\eta_{p2} = .09$, $MSE = 120.28$, where males (M = 20.97) had significantly higher spatial ability than females ($M = 14.48$). There were no significant main effects for hands-observation or for gestures (both $F < 1$, ns). Also, there were no interaction effects between gestures and gender, $F(1, 112) = 2.07$, $p = .153$, $\eta_{p2} = .02$, between gesture and hands-observation or between gender and hands-observation (both $F < 1$, ns). There was no three-way interaction effect between gestures, gender and hand ($F < 1$, ns).

For the Corsi test, there were no main effects for gender, $F(1, 112) = 2.57$, $p = .112$, $\eta_{p2} = .02$, $MSE = 180.85$, gestures or hand-observation (both $F < 1$, ns). Also, there were no interaction effects between gender and hands-observation, $F(1, 112) = 1.15$, $p = .286$, $\eta_{p2} = .01$, between gestures and gender or between gesture and hands-observation (both $F < 1$, ns). Similarly, there was no three-way interaction effect between gestures, gender and hand ($F < 1$, ns).

Hence, as there was a significant difference on the MRT (a gender effect) but not for the Corsi test, the MRT was used as the only covariate in the 2 (gender: male vs. female) × 2 (gesture: gesture vs. non-gesture) × 2 (hand-observation: with-hand vs. without-hand) ANCOVAs used in the main analysis.

5.4.3.3 Test performance

Table 5.24 illustrates the mean (and SD) scores for each task under the different learning conditions. The ANCOVA results revealed that there was a significant main gesture effect for the 1st attempt, $F(1, 111) = 9.37$, $p = .003$, $\eta_{p2} < .08$, $MSE = 4.87$, where participants in the non-gesture group ($M = 6.68$, $SE = .29$) outperformed those in the gesture group ($M = 5.45$, $SE = .29$). Similarly, there was a close-to-significant ($p < .10$) gesture effect for the 2nd attempt, $F(1, 111) = 3.42$, $p = .067$, $\eta_{p2} < .03$, $MSE = 10.66$, where participants in the non-gesture group ($M = 9.76$, $SE = .42$) scored higher than those in the gesture group ($M = 8.66$, $SE = .42$). However, there was no significant gesture effect for the transfer task, ($F < 1$, ns).

There was no significant main gender effect or main hand effect for the 1st attempt, 2nd attempt and the transfer task (all $F < 1$, ns).

Table 5.24.
Mean (SD) of Performance Scores in Experiment 4

		With-hands	Without-hands	Total
Gesture	1st attempt	Male = 6.20 (2.18)	Male = 5.87 (2.06)	Male = 6.03 (2.09)
		Female = 5.53 (2.26)	Female = 4.23 (1.94)	Female = 4.88 (2.17)
		Total = 5.87 (2.20)	Total = 5.05 (2.13)	Total = 5.46 (2.19)
	2nd attempt	Male = 9.70 (2.85)	Male = 8.57 (3.49)	Male = 9.13 (3.23)
		Female = 9.57 (3.67)	Female = 6.87 (3.88)	Female = 8.22 (3.96)
		Total = 9.63 (3.27)	Total = 7.72 (3.73)	Total = 8.68 (3.61)
	Transfer	Male = 4.50 (2.31)	Male = 4.60 (2.12)	Male = 4.55 (2.18)
		Female = 4.27 (2.16)	Female = 3.73 (2.68)	Female = 4.00 (2.41)
		Total = 4.38 (2.20)	Total = 4.17 (2.42)	Total = 4.28 (2.29)
No-gesture	1st attempt	Male = 6.30 (2.36)	Male = 7.13 (1.82)	Male = 6.72 (2.11)
		Female = 5.76 (2.27)	Female = 7.50 (2.90)	Female = 6.63 (2.71)
		Total = 6.03 (2.29)	Total = 7.32 (2.39)	Total = 6.68 (2.41)
	2nd attempt	Male = 9.17 (3.32)	Male = 10.50 (2.88)	Male = 9.83 (3.13)
		Female = 9.30 (3.18)	Female=10.00 (3.78)	Female = 9.65 (3.45)
		Total = 9.23 (3.19)	Total = 10.25 (3.31)	Total = 9.74 (3.27)
	Transfer	Male = 4.00 (2.51)	Male = 4.87 (2.17)	Male = 4.43 (2.30)
		Female = 3.93 (2.76)	Female = 4.77 (2.43)	Female = 4.35 (2.59)
		Total = 3.97 (2.59)	Total = 4.82 (2.22)	Total = 4.39 (2.43)
Total	1st attempt	Male = 6.25 (2.23)	Male = 6.50 (2.01)	Male = 6.38 (2.11)
		Female = 5.65 (2.23)	Female = 5.87 (2.94)	Female = 5.76 (2.59)
		Total = 5.95 (2.23)	Total = 6.18 (2.52)	Total = 6.07 (2.37)
	2nd attempt	Male = 9.43 (3.10)	Male = 9.53 (3.30)	Male = 9.48 (3.17)
		Female = 9.43 (3.38)	Female = 8.43 (4.09)	Female = 8.93 (3.75)
		Total = 9.43 (3.21)	Total = 8.98 (3.72)	Total = 9.21 (3.47)
	Transfer	Male = 4.25 (2.38)	Male = 4.73 (2.07)	Male = 4.49 (2.23)
		Female = 4.10 (2.44)	Female = 4.25 (2.57)	Female = 4.18 (2.48)
		Total = 4.18 (2.39)	Total = 4.49 (2.32)	Total = 4.33 (2.35)

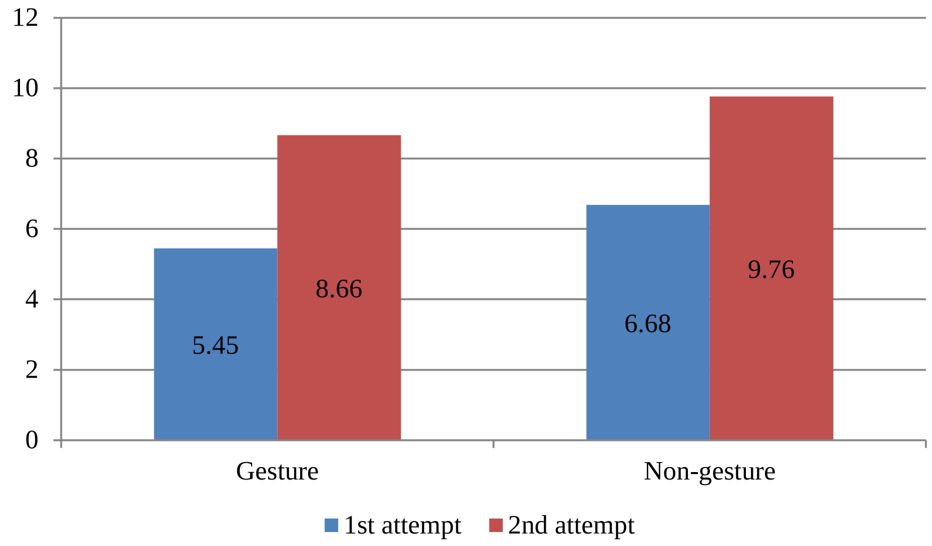


Figure 5.28. Main Gesture Effect for Performance Score in the 1st and 2nd Attempt in the Experiment 4

There was a significant interaction effect between gesture and hand for the 1st attempt, $F(1, 111) = 5.92$, $p = .017$, $\eta_{p2} < .05$, and the 2nd attempt, $F(1, 111) = 4.85$, $p = .030$, $\eta_{p2} < .04$. Simple effects tests revealed that in the without-hands condition, participants in the non-gesture group ($M = 7.30$, $SE = .41$) significantly outperformed those in the gesture group ($M = 5.07$, $SE = .41$) at the 1st attempt, $F(1, 57) = 14.56$, $p < .001$, $\eta_{p2} = .20$, $MSE = 5.11$. However there was no significant difference between the gesture group ($M = 5.81$, $SE = .39$) and the non-gesture group ($M = 6.09$, $SE = .39$) in the with-hand condition, $F(1, 57) = .26$, $p = .610$, $\eta_{p2} < .01$, $MSE = 4.61$.

Similarly in the 2nd attempt, participants in the non-gesture group ($M = 10.20$, $SE = .63$) significantly outperformed the gesture group ($M = 7.77$, $SE = .63$) in the without-hand condition, $F(1, 57) = 7.50$, $p = .008$, $\eta_{p2} = .12$, $MSE = 11.77$; but in the with-hand condition, there was no significant difference between the non-gesture group ($M = 9.33$, $SE = .56$) and the gesture group ($M = 9.54$, $SE = .56$), $F(1, 57) = .07$, $p = .787$, $\eta_{p2} < .01$, $MSE = 9.35$.

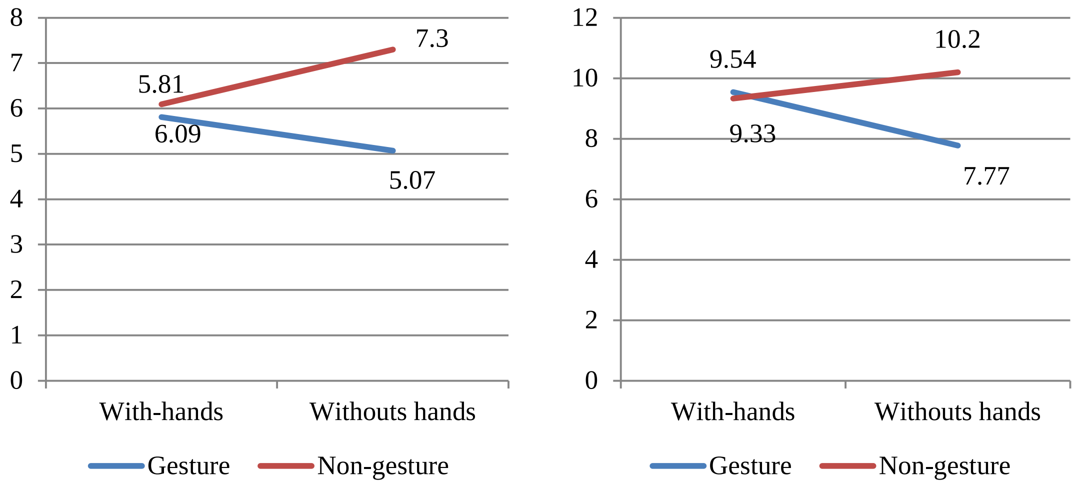


Figure 5.29. Interaction Effects between Gesture and Hand-Observations for Performance Score in the Experiment 4 (Left: 1st Attempt; Right: the 2nd Attempt)

There was no significant gesture and hand interaction for the transfer task, ($F < 1$, ns). Also, there were no gesture and gender interactions for the 1st attempt, 2nd attempt and the transfer task (all $F < 1$, ns), nor gender and hand interactions in the 1st attempt ($F < 1$, ns), 2nd attempt, $F(1, 111) = 1.08$, $p = .30$, $\eta_{p2} = .01$, $MSE = 10.80$, and the transfer task ($F < 1$, ns.). There was also no three-way interaction effect between gender, gesture and hand in the 1st attempt $F(1, 111) = 1.75$, $p = .189$, $\eta_{p2} = .02$, $MSE = 4.88$, 2nd attempt, nor the transfer task (both $F < 1$, ns).

In short, there was a significant gesture effect for performance score in the 1st attempt and a close-to-significant effect in the 2nd attempt. Participants in the non-gesture group outperformed those in gesture group. Furthermore, there was a significant gesture and hand interaction effect in both the 1st and the 2nd attempt, where the non-gesture group outperformed the gesture group in the without-hands condition but not the with-hand condition.

5.4.3.4 Cognitive load measure

8 questions measuring the cognitive load (4 questions on intrinsic load and 4 questions on extraneous load) spent on understanding the learning material were collected after participants finished attempting all tasks. The result from the 4 questions for intrinsic load (i.e. question a – d) and the 4 questions for extraneous load (i.e. question e – f) were averaged and became an averaged intrinsic load and an averaged extraneous load respectively. Table 5.25 and Table 5.26 below separately show the mean (and SD) scores of the averaged intrinsic load and averaged extraneous load spent.

Table 5.25.
Mean (SD) of the Averaged Intrinsic Load Spent in Experiment 4 (N=120)

	With-Hands	Without-Hands	Total
Gesturing	Male = 7.73 (1.87)	Male = 7.73 (2.40)	Male = 7.73 (2.11)
	Female = 7.47 (2.23)	Female = 8.07 (1.75)	Female = 7.77 (2.00)
	Total = 7.60 (2.03)	Total = 7.90 (2.07)	Total = 7.75 (2.04)
No-Gesturing	Male = 6.93 (1.98)	Male = 7.07 (2.05)	Male = 7.00 (1.98)
	Female = 7.00 (2.56)	Female = 6.87 (2.62)	Female = 6.93 (2.55)
	Total = 6.97 (2.25)	Total = 6.97 (2.31)	Total = 6.97 (2.26)
Total	Male = 7.33 (1.94)	Male = 7.40 (2.22)	Male = 7.37 (2.07)
	Female = 7.23 (2.37)	Female = 7.47 (2.27)	Female = 7.35 (2.31)
	Total = 7.28 (2.15)	Total = 7.43 (2.23)	Total = 7.36 (2.18)

Table 5.26.
Mean (SD) of the Averaged Extraneous Load Spent in Experiment 4 (N=120)

	With-Hands	Without-Hands	Total
Gesturing	Male = 2.80 (2.27)	Male = 3.87 (2.56)	Male = 3.33 (2.44)
	Female = 2.80 (2.48)	Female = 2.20 (1.74)	Female = 2.50 (2.13)
	Total = 2.80 (2.34)	Total = 3.03 (2.31)	Total = 2.92 (2.31)
No-Gesturing	Male = 2.73 (1.39)	Male = 3.60 (2.47)	Male = 3.17 (2.02)
	Female = 2.47 (1.73)	Female = 2.47 (1.77)	Female = 2.47 (1.72)
	Total = 2.60 (1.55)	Total = 3.03 (2.19)	Total = 2.82 (1.89)
Total	Male = 2.77 (1.85)	Male = 3.73 (2.48)	Male = 3.25 (2.22)
	Female = 2.63 (2.11)	Female = 2.33 (1.73)	Female = 2.48 (1.92)
	Total = 2.70 (1.97)	Total = 3.03 (2.23)	Total = 2.87 (2.10)

For intrinsic load, ANCOVA results showed that there were no significant main effects for gender, hand or gesture (all $F < 1$, ns). Similarly, there were no significant interaction between gender and hand, gender and gesture, or hand and gesture (all $F < 1$, ns). Likewise, there was no three-way interaction between gender, hand and gesture ($F < 1$, ns).

For extraneous load, ANCOVA results showed that there was a significant gender effect, $F(1, 111) = 8.25$, $p = .005$, $\eta_{p2} = .07$, $MSE = 2.01$, in which males reported significantly higher extraneous load ($M = 2.91$, $SE = .19$) than females ($M = 2.14$, $SE = .19$).

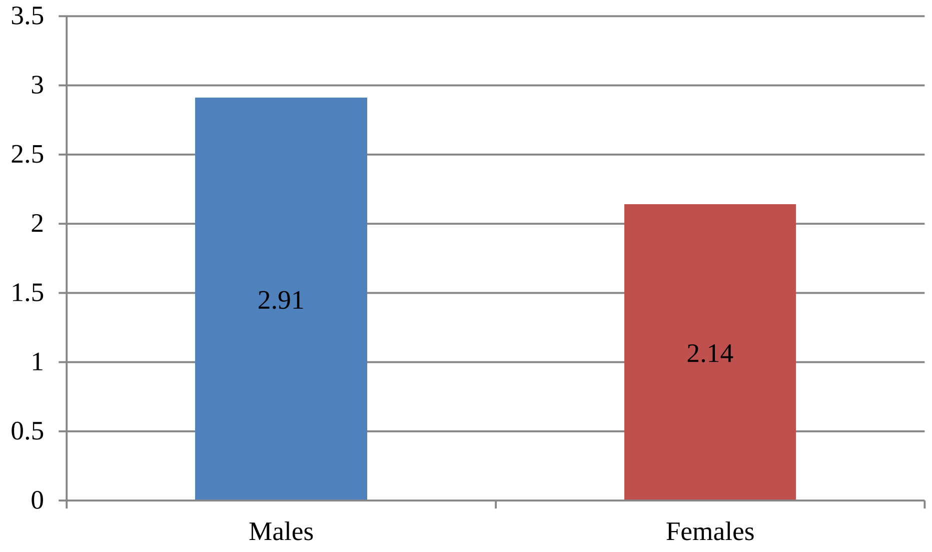


Figure 5.30. Main Gender Effect for Extraneous Load Spent in Experiment 4.

However, there were no significant main effects for hand or gesture (all $F < 1$, ns). Similarly, there were no significant interaction between gender and hand, gender and gesture, or hand and gesture (all $F < 1$, ns). Likewise, there was no three-way interaction between gender, hand and gesture ($F < 1$, ns).

5.4.3.5 Learning efficiency

In the previous experiments cognitive load was measured after each task using a single item measure. In this experiment, the multi-item measure was collected after the two retention tasks and the transfer task were completed. Therefore a total score was calculated by summing the scores for each task to give an overall performance score (formula below).

$$\text{Overall performance score} = \frac{\text{Task 1 score}}{(\text{max} = 15)} + \frac{\text{Task 2 score}}{(\text{max} = 15)} + \frac{\text{Transfer task score}}{(\text{max} = 7)}$$

This overall performance score was then combined with the two cognitive load measures (i.e. intrinsic load and extraneous load) to calculate two efficiency measures labelled: intrinsic efficiency and extraneous efficiency (see formulae below).

$$E_{intrinsic} = \frac{Z_{overall\ performance} - Z_{averaged\ intrinsic\ load}}{\sqrt{2}}$$

$$E_{extraneous} = \frac{Z_{overall\ performance} - Z_{averaged\ extraneous\ load}}{\sqrt{2}}$$

Table 5.27 and Table 5.28 show the means (and SDs) of the two efficiency measures.

Table 5.27.
Mean (SD) of Intrinsic Efficiency in Experiment 4

	With-Hands	Without-Hands	Total
Gesturing	Male = .01 (1.00)	Male = .02 (1.02)	Male = .02 (.99)
	Female = -.12 (1.35)	Female = -.47 (1.21)	Female = -.29 (1.27)
	Total = -.05 (1.17)	Total = -.23 (1.13)	Total = -.14 (1.14)
No-Gesturing	Male = -.01 (.84)	Male = .22 (.83)	Male = .11 (.83)
	Female = -.02 (1.43)	Female = .36 (1.70)	Female = .17 (1.55)
	Total = -.02 (1.15)	Total = .29 (1.31)	Total = .14 (1.24)
Total	Male = -.0006 (.91)	Male = .12 (.92)	Male = .06 (.91)
	Female = -.07 (1.37)	Female = -.06 (1.51)	Female = -.06 (1.43)
	Total = -.03 (1.15)	Total = .03 (1.24)	Total = .00 (1.19)

Table 5.28.
Mean (SD) of Extraneous Efficiency in Experiment 4

	With-Hands	Without-Hands	Total
Gesturing	Male = .19 (.91)	Male = -.28 (.92)	Male = -.04 (.93)
	Female = .14 (1.08)	Female = -.31 (1.31)	Female = -.09 (1.20)
	Total = .17 (.98)	Total = -.29 (1.11)	Total = -.06 (1.07)
No-Gesturing	Male = -.16 (.96)	Male = -.18 (.84)	Male = -.17 (.89)
	Female = .08 (.80)	Female = .52 (1.12)	Female = .30 (.98)
	Total = -.04 (.87)	Total = .17 (1.04)	Total = .06 (.96)
Total	Male = .01 (.94)	Male = -.23 (.87)	Male = -.11 (.90)
	Female = .11 (.93)	Female = .10 (1.27)	Female = .11 (1.11)
	Total = .06 (.93)	Total = -.06 (1.09)	Total = .00 (1.01)

For intrinsic efficiency, ANCOVA results showed that there were no significant main effects for gender, hand (both $F < 1$, ns) or gesture, $F(1, 111) = 1.66$, $p = .201$, $\eta_{p2} = .15$, $MSE = 1.42$. Similarly, there were no significant interactions between gender and hand, gender and gesture, or hand and gesture (all $F < 1$, ns). Likewise, there was no three-way interaction between gender, hand and gesture ($F < 1$, ns).

For extraneous efficiency, ANCOVA results showed that there was a close-to-significant ($p < .10$) gender effect, $F(1, 111) = 3.83$, $p = .053$, $\eta_{p2} = .03$, $MSE = .96$, in which females reported a higher extraneous load efficiency ($M = .18$, $SE = .13$) than males ($M = -.18$, $SE = .13$).

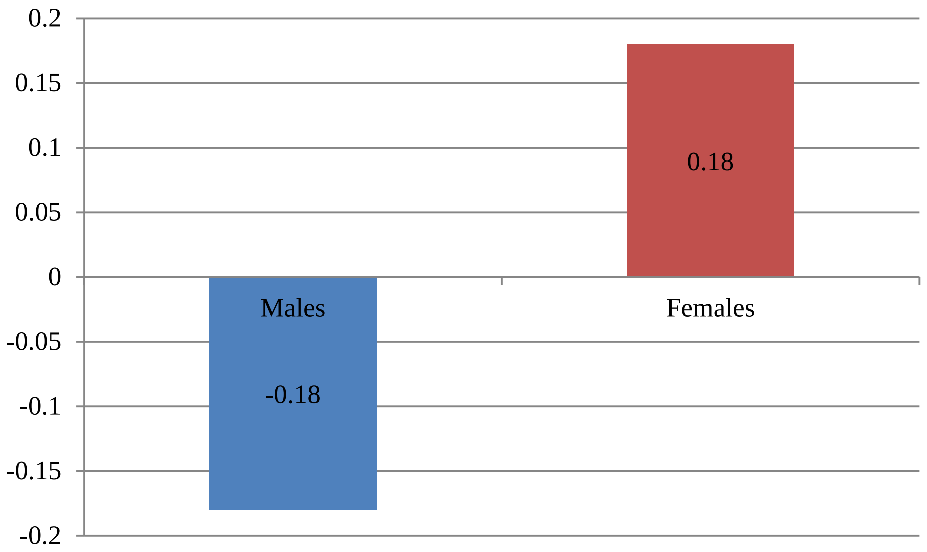


Figure 5.31. Main Gender Effect for Extraneous Efficiency in Experiment 4.

Furthermore, there was a close to significant effect ($p < .1$) in the interaction effect between hand and gesture, $F(1, 111) = 2.83$, $p = .096$, $\eta_{p2} = .03$. Simple effect test revealed that in the with-hand condition there was no difference between the gesture group ($M = .17$, $SD = .98$) and the non-gesture group ($M = -.04$, $SD = .87$), $F(1, 57) = .47$, $p = .469$, $\eta_{p2} = .01$, $MSE = .77$. In the without-hand condition, there was also no difference between the gesture group ($M = .17$, $SD = 1.04$) and the non-gesture group ($M = -.29$, $SD = 1.11$), $F(1, 57) = 2.64$, $p = .110$, $\eta_{p2} = .04$, $MSE = 1.17$.

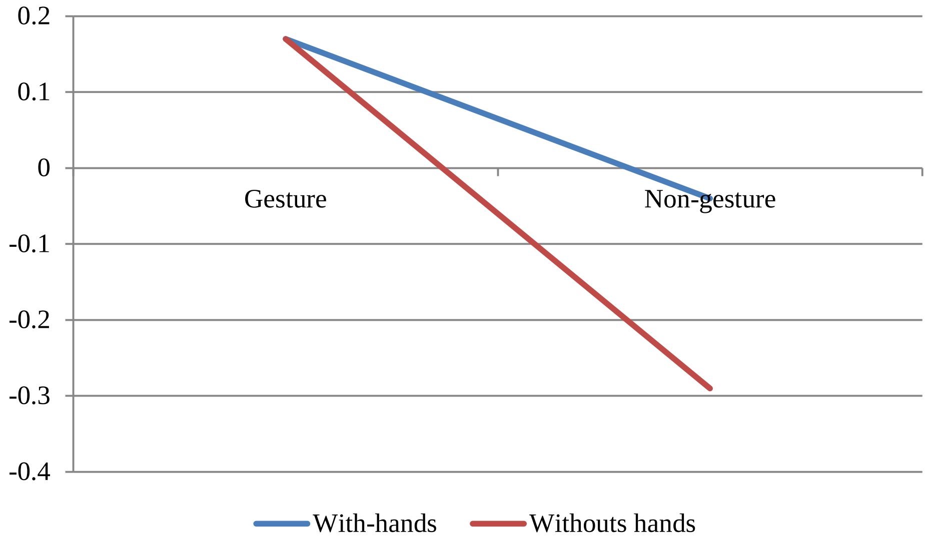


Figure 5.32. Interaction Effect for Extraneous Efficiency in Experiment 4.

There were no significant main effects for observing hand or gesture (all $F < 1$, ns). Similarly, there were no significant interaction between gender and hand, or gender and gesture (both $F < 1$, ns). Likewise, there was no three-way interaction between gender, hand and gesture ($F < 1$, ns).

5.4.4 Summary of main analysis

Hypothesis 3 predicted that observing hands would lead to superior learning than not observing hands. No direct evidence was found supporting this prediction. Moreover, there was a significant interaction between gesture and hand for the 1st attempt and the 2nd attempt. Simple effect tests showed that the non-gesture group outperformed the gesture group in the without-hand condition, but the difference disappeared in the with-hand condition. The result suggested that the benefit of hand-observation, to a certain extent, depends on the degree of hand involvement.

Hypothesis 4 predicted that the use of gestures would bring advantage to learning. However, the results from performance scores and extraneous learning efficiency found that no-gestures was more beneficial than using gestures. Hence, Hypothesis 4 was rejected.

In previous experiments females were found to have an advantage in animation conditions. In this experiment there was no significant difference in performance scores with males, however, there was a significant gender effect for extraneous load and a very close-to-significant gender effect for extraneous efficiency. Females reported lower extraneous load with greater extraneous efficiency than males in this animated environment.

5.4.5 Predictors of gender performance

To explore the potential relationship between gender, spatial ability and animations, additional regression analysis was conducted.

5.4.5.1 Test of subjective measures

Independent T-tests were conducted to compare the mean difference of self-rated mental rotational ability, self-rated overall spatial ability, frequency learning with animation and frequency learning with static picture between males and females. Result showed males (*M* = 3.20, *SD* = .80) reported a significantly higher score than females (*M* = 2.80, *SD* = .90) in self-rated mental rotation ability, $t(118) = 2.58$, $p = .011$, $d = .47$. Similarly, males (*M* = 3.40, *SD* = .72) also reported a significantly higher score than females (*M* = 2.87, *SD* = .75) in self-rated overall spatial ability, $t(118) = 3.99$, $p <.01$, $d = .72$. However, for the frequency learning with animation, there was no difference found between males (*M* = 2.47, *SD* = .87) and females (*M* = 2.55, *SD* = .91), $t(118) = -.51$, $p = .610$, $d = -.09$. Similarly, there was no difference found between males (*M* = 2.73, *SD* = .84) and females (*M* = 2.60, *SD* = 1.0), $t(118) = .79$, $p = .429$, $d = .14$, for the frequency learning with static pictures.

5.4.5.2 Regression analysis

Table 5.29 and Table 5.30 indicate the correlation coefficients between the three test scores, the MRT score, Corsi score and the two self-rating measures for both genders individually. As can be seen from the tables, only the MRT was significantly correlated with task performances for males, whereas there were a number of possible predictors of females' performance. Hence the regression analyses were conducted with different factors for each gender. MRT scores, hand-observation condition and gesturing-condition (both independent variables) were entered into the regression model using the enter method for males; whilst the Corsi score, self- rated experience playing with Lego, frequency learning with animation, frequency learning with static picture, hand-observation condition and gesturing-condition were entered into the regression model for females.

Table 5.29.
Pearson Correlation Matrix for Males in Experiment 4 (N = 60)

	Correlation							
	MRT	Corsi	Self-rated mental rotation	Overall spatial ability	Freq. learning with animation	Freq. learning with picture	Task 1 score	Task 2 score
MRT	1							
Corsi	.407**							
Self-rated MR	.286*	-.074						
Overall spatial ability	.162	-.165	.775**					
Freq. learning with animation	.307*	.262*	-.063	-.032				
Freq. learning with picture	.329*	.131	-.096	-.045	.126			
Task 1 score	.300*	.107	.096	.134	.170	.158		
Task 2 score	.453**	.197	.182	.182	.128	.173	.580**	
Transfer task score	.363**	.145	.235	.230	.076	.080	.458**	.658**

Table 5.30.
Pearson Correlation Matrix for Females in Experiment 4 (N = 60)

	Correlation							
	MRT	Corsi	Self-rated mental rotation	Overall spatial ability	Freq learning with animation	Freq learning with picture	Task 1 score	Task 2 score
MRT	1							
Corsi	.274*							
Self-rated MR	.368**	.362**						
Overall spatial ability	.425**	.186	.667**					
Freq learning with animation	.093	.347**	.220	.135				
Freq learning with picture	.110	.204	.269*	.246	.266*			
Task 1 score	.116	.393**	.245	.272*	.453**	.284*		
Task 2 score	.140	.430**	.217	.233	.463**	.370**	.757**	
Transfer task score	.143	.312*	.191	.269*	.302*	.437**	.476**	.723**

Regression results for males. For Task 1 there was a significant one-factor model (p = .020) that included the MRT measure (B = .05, SE = .021, β = .30, R^2 = .09). For Task 2 there was a significant one-factor model (p = .000) that included the MRT measure (B = .12, SE = .030, β = .45, R^2 = .21). For the transfer task there was a significant one-factor model (p = .004) that included the MRT measure (B = .07, SE = .02, β = .36, R^2 = .13).

Regression results for females. For Task 1 there was a significant three-factor model (p < .001) that included the frequency learning with animation (B = .92, SE = .33, β = .32, R^2 = .21, p = .008), gesturing-condition (B = 1.33, SE = .57, β = .26, ΔR^2 = .07, p = .023) and the Corsi measure (B = .05, SE = .02, β = .25, ΔR^2 = .06, p = .034). For Task 2 there was a significant three-factor model (p < .001) that included the frequency learning with animation (B = 1.26, SE = .49, β = .31, R^2 = .21, p = .012), the Corsi measure (B = .07, SE = .03, β = .28, ΔR^2 = .08, p = .021), and frequency learning with static picture (B = .88, SE = .43, β = .23, ΔR^2 = .05, p = .045). For the transfer task there was a significant one-factor model (p < .001) that included the frequency learning with static picture (B = 1.09, SE = .30, β = .44, R^2 = .19).

In summary, the MRT successfully predicted male performance in all 3 tasks (i.e. Task 1, Task 2 and the transfer task). For females there was no one model or factor that could predict all different task performances. However, several factors occurred twice as the frequency learning with animation successfully predicted Task 1 and Task 2 performance, the Corsi score successfully predicted Task 1 and Task 2 performance, and frequency learning with static picture successfully predicted Task 2 and transfer task performance. Interestingly, gesturing was found to be a predictor for females on the first task. Inspection of the mean scores (see Table 5.24) shows that the non-gesturing group (M = 6.63) scored higher than the gesturing group (M = 4.88) suggesting that females benefitted from no gesturing.

Chapter 6: General Summary and Discussion

6.1 Support for Hypotheses

Hypothesis 1: Animations will be superior instructional materials to static pictures

This hypothesis was tested in Experiments 1, 2 & 3. Experiments 1 & 2 had a real testing environment whilst Experiment 3 had a virtual testing environment. In addition, Experiment 1 presented learning materials (animated and statics version of Lego constructions) without showing hands whereas the learning materials in Experiment 2 & 3 showed hands manipulating the Lego bricks. In all three experiments, both animated and statics conditions were compared. Hypothesis 1 predicted that animations would be superior to static pictures. The results for this comparison are summarised in Table 6.1.

Table 6.1.
Summary of Results Directly Comparing Animation and Statics

	Performance scores			**Cognitive load scores**			**Efficiency scores**		
	1st	2nd	Trans	1st	2nd	Trans	1st	2nd	Trans
Experiment 1	ns	ns	ns	A > S^	A > S*	A > S**	ns	S > A*	S > A**
Experiment 2	ns	ns	ns	A > S*	A > S^	ns	ns	ns	ns
Experiment 3	ns	ns	ns	ns	ns	ns	ns	ns	ns

Notes:
A =Animation, S = Statics
^ .05 < p < .10, * p < .05, ** p < .01

The empirical evidence showed no difference in direct testing performance in any of the 3 experiments. However, some evidence emerged that animations induced higher cognitive load (in both Experiment 1 and Experiment 2) and were lower in learning efficiency (in Experiment 1), suggesting that static pictures were superior to animations for some measures. Thus Hypothesis 1 is rejected.

Hypothesis 2: There will be a gender-presentation format interaction

This hypothesis was tested in Experiments 1, 2 & 3. Hypothesis 2 predicted that there would be an interaction effect between gender (males vs. females) and presentation-format (animations vs. static pictures). Results from Experiment 1, Experiment 2 and Experiment 3 are summarised in Table 6.2.

Table 6.2.
Summary of Results Directly Comparing Gender and Presentation Format Interactions

	Performance scores			**Cognitive load scores**			**Efficiency scores**		
	1st	2nd	Trans	1st	2nd	Trans	1st	2nd	Trans
Experiment 1	Interact^ A: M=F S: M>F*	Interact* A: M=F S: M>F*	ns	ns	ns	ns	ns	Interact* A: M=F S: M>F*	ns
Experiment 2	ns	Interact* A: F>M* S: M=F	ns	ns	ns	ns	ns	ns	ns
Experiment 3	ns	Interact* A: M=F S: M=F	Interact* A: M=F S: M=F	ns	ns	ns	Interact^ A: M=F S: M=F	Interact^ A: M=F S: M=F	Interact^ A: M=F S: M>F^

Notes:
A =Animation, S = Statics
M = Males, F = Females
^ .05 < p < .10, * p < .05, ** p < .01

It can be seen from this table that an interaction effect was found for performance scores in all 3 experiments. Simple effect tests found that for statics presentations males were superior to females (Experiment 1, first and second attempts); and for animations females were superior to males (Experiment 2, second attempt; and Experiment 3, second attempt & transfer). Efficiency scores followed a similar pattern. Hence, Hypothesis 2 was supported. Furthermore, it can be concluded that females benefited from animations; whereas males had an advantage when observing static presentations.

Hypothesis 3: Observing hands will lead to superior learning than not observing hands

This hypothesis was tested only in Experiments 4, which had a different design to Experiments 1, 2 & 3, as only an animation condition was used. Moreover, both with-hands and without-hands versions were observed, but also the use of gestures (gestures-encouraged vs. gesture-prohibited) was added in a 2 x 2 design. Results are summarised in Table 6.3.

Table 6.3.
Summary of Results Directly Comparing Observing and Not-observing Hands and Interactions with Gesturing

	Performance scores			**Cognitive load scores**			**Efficiency scores**		
	1st	2nd	Trans	1st	2nd	Trans	1st	2nd	Trans
Experiment 4									
	Interact** H: G=NG NH: NG>G**	Interact** H: G=NG NH: NG>G**	ns	ns	ns	ns	ns	Interact^ H: G=NG NH: G=NG	ns

Notes:
M = Males, F = Females
G = gestures, NG = no gestures
H = With-hands observed, NH = No-hands observed
^ .05 < p < .10, * p < .05, ** p < .01

Hypothesis 3 predicted that observing hands would lead to superior learning than not observing hands. However, results from Experiment 4 did not show this main effect, hence, Hypothesis 3 was not supported. However, a significant interaction was found on the first and second attempts at the task. Simple effects found that non-gesturing was superior to gesturing in the no-hands condition only.

Hypothesis 4: Gesturing will lead to superior learning than no gesturing

This hypothesis was tested only in Experiment 4, which predicted that the use of gestures would bring an advantage to learning. However, the results from the first and second attempts performance scores indicated that not using gestures was more beneficial than using gestures. Hence, Hypothesis 4 is rejected. Furthermore, as shown in the interaction described above, the advantage for no-gestures was strongest in the no-hands condition. Results are summarised in Table 6.4.

Table 6.4.
Summary of Results Directly Comparing Gesture-encouraged and Gesture-prohibited

	Performance scores			**Cognitive load scores**			**Efficiency scores**		
	1st	2nd	Trans	1st	2nd	Trans	1st	2nd	Trans
Experiment 4									
	NG > G**	NG > G^	ns	ns	ns	ns	ns	ns	ns

Notes:
M = Males, F = Females
G = gestures, NG = no gestures
H = With-hands observed, NH = No-hands observed
^ .05 < p < .10, * p < .05, ** p < .01

6.2 Summary of Results for Predictive Analyses

One of the major unanswered questions in the literature is the extent of the influence of spatial ability on gender differences in performance on tasks involving visual representations. Hence, in addition to the main analysis, regression analyses were performed to identify predictors of gender performance. Subjective spatial measures (*self-rated mental rotation ability* and *self-rated overall spatial ability*), objective spatial measures (*Card Rotations Test, CRT,* in Experiments 1-3; and *Mental Rotations Test, MRT* and *the Corsi Block Tapping Test* in Experiment 4) and static-animation conditions were entered into a regression model.

Table 6.5.
Summary of Predictive Analyses Results Directly Comparing Predictors for Males and Females

	Males			**Females**		
	1st	2nd	Trans	1st	2nd	Trans
Experiment 1						
	ns	ns	ns	Self-rated overall spatial ^	Self-rated overall spatial **	CRT**
Experiment 2						
	CRT**	CRT**	CRT + self-rated overall spatial + animation *	ns	ns	Animation-static^
Experiment 3						
	ns	CRT + self-rated mental rotation**	ns	Animation-static*	ns	Animation-static*
Experiment 4						
	MRT*	MRT**	MRT**	Freq with Anim + gesturing + Corsi**	Freq with Anim + Freq with Pic + Corsi*	Freq with Pic**

Notes:
^ .05 < p < .10, * p < .05, ** p < .01

Results from Experiment 1 showed there was no significant predictor found for male performance. However for females, self-perception of overall spatial ability successfully predicted the performance in retention tasks (first and second attempts);

whilst the CRT (the objective assessment of mental rotation ability) successfully predicted the transfer task performance.

Results from Experiment 2 showed that the CRT successfully predicted males' performance in the retention tasks (first and second attempts), and the transfer task performance could be predicted from a 3-factor model with the CRT score, self-perception of spatial ability and animation-static condition. For females there was no model found for the retention tasks (first and second attempts), but the animation-static condition was found to be a significant predictor for the transfer task performance.

Results from Experiment 3 revealed that the self-rated mental rotation ability and CRT score successfully predicted males second attempt performance, but there was no significant model for the first attempt and the transfer task. In contrast, for females, there was no model found for the second attempt performance, but animation-static condition was found to be a good predictor for the first attempt and the transfer task.

Results from Experiment 4 revealed the MRT successfully predicted male performance in all three tasks (i.e. first and second attempts, and the transfer task). For females, several significant factors occurred for two scores, as the frequency of learning with animation successfully predicted performance on the first and second attempts, while the Corsi score successfully predicted performance on the first and second attempts, and the frequency of learning with static pictures successfully predicted performance on the second attempt and the transfer task. Interestingly, gesturing was found to be a predictor for females on the first attempt where inspection of mean scores demonstrated that females benefitted from no gesturing.

6.3 Theoretical Conclusions

This thesis contributes to existing instructional animation research literature by directly investigating some important factors (i.e. gender, spatial ability, observation of hands manipulating tasks, and learner gesturing) that potentially moderate the outcomes of learning from instructional animations. Overall there was strong evidence for a consistent gender difference when learning from Lego tasks using instructional animations and static pictures. The theoretical conclusions from the thesis are discussed next in more detail.

6.3.1 Finding 1: Animations were not superior to static pictures

Based on the empirical evidence from Castro-Alonso et al. (2015a) and other studies where animations were shown to be superior to static pictures with human-movement tasks (e.g. Akinlofa et al., 2013; Ayres et al., 2009; Castro-Alonso et al., 2014a, 2015a; Höffler & Leutner, 2007; Paas & Sweller, 2012; Wong et al., 2009), it was predicted (Hypothesis 1) that animations would be superior to static pictures. However, results showed an opposite trend revealing that static pictures were sometimes superior requiring less mental effort to animations on these Lego tasks. Such results are in direct contradiction with those from Castro-Alonso et al. (2015a) who showed that animations were superior to static pictures on a number of measures for the same tasks.

The results also contradict the main assumption of the human movement effect that argues that animations are particular helpful for learning human movement tasks because transient information is not such a problem as humans have evolved to learn such tasks from observational learning (Ayres & Paas, 2007a; Leahy & Sweller, 2011; Paas & Sweller, 2012; Wong et al., 2012). However, it is notable that although the Lego tasks do contain a human movement aspect, they also rely heavily on object memory placement (Choi & L'Hirondelle, 2005; Postma et al., 1998), and not a fine-tuned human motor task like knot tying or origami (Marcus et al., 2013; Wong et al., 2012; Wong et al., 2009). This reliance on this aspect of memory may have diluted the advantage of animations. Nevertheless, perhaps more importantly, in this study there was an equal percentage of males and females, unlike the Castro-Alonso et al. (2015a) study, which contained a much higher percentage of females. Consequently gender may have played a significant role as discussed next.

6.3.2 Finding 2: Gender-presentation interaction

As discussed in Chapter 3 and Chapter 4, the literature results in relation to gender and instructional animations suggest a trend that animations are more helpful for females (e.g. Cowards et al., 2012) whereas static pictures were more beneficial to males (e.g. Sánchez & Wiley, 2010). This interaction was predicted in Hypothesis 2 and the empirical evidence consistently showed this interaction effect, where females benefited from animations but males did not. Furthermore males tended to benefit from static presentations. This interaction effect may provide some insights into the unexpected lack of an animation advantage predicted in Hypothesis 1. If males perform better with static pictures, and females with animations, then an equal number of males and females will not produce an animation effect, only an interaction. It is notable that Castro-Alonso et al. (2015a) found an animation effect with the same Lego materials, but has a high percentage of females in the overall sample. Consequently, this high proportion of females may have skewed the results, rather than the overall impact of the animation. Clearly, when conducting research into the effectiveness of animations and/or static pictures then it is imperative to use an equal number of males and females.

The literature on instructional animations suggests that they are helpful for human movement tasks, because transient information is less a problem. However, in this study animations were not an advantage for males, who benefitted more from statics. This reverse effect for males compared to females suggests that animations were not necessary effective and may indicate a redundancy effect (see the expertise reversal effect in Sweller et al., 2011, pp. 155-170). Viewing animations may have increased extraneous cognitive load and reduced learning (Kalyuga & Sweller, 2014). Whereas, for statics the mental animation required by static pictures was not redundant and matched the spatial abilities of the males.

6.3.3 Finding 3: Observing hands manipulating the materials was not an advantage

With insights from the literature on embodied cognition it was expected that observing hands would be an advantage (Castro-Alonso et al., 2015a; Marcus et al., 2013; Paas & Sweller, 2012; van Gog et al., 2009). However, no direct evidences were found. Instead, there were interactions between observing-hands and using gestures. In particular the non-gesture group outperformed the gesture group in the without-hands condition, but the difference disappeared in the with-hand condition. This result suggests that the benefit of hand-observations may depend on the degree of hand involvement in the learning materials. For the human movement tasks found in previous studies (e.g. assembling key rings, paper folding or knot tying), hands are an essential part in making the moves in order to complete the tasks. However, hands play a less important role in this Lego task as they are primarily involved in placing the bricks, without the more sophisticated movements associated with knot tying or paper folding. Hence showing hands may have been redundant as it was fairly straightforward to understand the motion of the bricks. Hence any potential embodied effects may have been reduced due to the creation of extraneous cognitive load caused by redundancy.

6.3.4 Finding 4: No gesturing was superior to gesturing

Evidence from gesturing studies has shown that the use of gesture can positively influence learning (e.g. Barsalou, 1999, 2008; Chu & Kita, 2011; Glenberg, 1997; Goldin-Meadow et al., 2001; Matlen et al., 2012; Roth, 2000). Hence it was expected that gesturing would lead to superior learning than no gesturing. However, the results indicated that non-gesturing was superior. In their study Post et al. (2013) gave two explanations on why gesturing may not produce positive results: 1) simultaneously observing and making gesturing while learning may induce a redundancy effect (caused by moving hands in the video and the actual hand movement), and 2) irrelevant gestures may increase extraneous load. However in this study it was in the no-hands condition where non-gesturing was found to best, so it was impossible to attribute redundancy from observing hands unnecessarily. However, if making irrelevant gestures increased extraneous cognitive load then Post et al.'s second explanation may have merit.

Furthermore, Wagner et al. (2004) argued that the representations underlying gesture-production are in visuospatial format. Consequently gesturing could interfere with a visuospatial memory task performance, as both of them would have made use of the same working memory system (i.e. the visuospatial sketchpad). In this case, overall cognitive load could be high, leading to a lack of available working memory resources. However, this hypothesis from Wagner et al. was rejected in their experiment where no interaction effects were found between gesture and the type of tasks. Furthermore, the results from this thesis also did not follow the suggestion of cognitive overload entirely. If there was overload, the gesturing group in the with-hands condition (i.e.

simultaneously observing and making gesture) should have scored the lowest. Yet, in this study (Experiment 4) gesturing and observing hands simultaneously was slightly (but not significantly) better than gesturing without observing hands. Whereas it is not clear exactly what caused no gesturing to be superior to gesturing it is possible that a combination of different effects could have occurred including redundancy caused by a number of interactions between gestures, hand-observations and presentation-formats, as well as potential working memory overload.

6.3.5 Finding 5: Different impact of spatial ability on gender

It is claimed that females have a greater spatial location memory than males (Loring-Meier & Halpern, 1999; Sánchez & Wiley, 2010). While spatial location memory is important for the Lego task (memorising Lego brick position), the Corsi test failed to capture this female advantage showing there were no significant difference between males and females. Consequently, there is no evidence supporting or rejecting this claim.

The results in this thesis revealed that even spatial ability was measured and controlled for in each analysis, gender-presentation format interactions were still found in all experiments. Such results suggest that animations were beneficial to female performance regardless of their spatial ability. Moreover, the regression analysis results found that predictors for males' performance and females' performances were different. Successful predictors of male performance were more likely to be objective assessments (e.g. CRT in Experiment 2; and MRT in Experiment 4); however, predictors of female performances were more likely to be subjective assessments (self-rated spatial ability) and their experience using animations/static pictures, although the Corsi (which was used in only Experiment 4) was a weak objective predictor for female performance.

These results suggest that a) spatial ability is good predictor of male performance but not necessarily of female performance, and b) different measures of spatial ability may be required according to gender.

6.4 Instructional Implications

6.4.1 Tailor the use of instructional animations and static pictures to gender differences

The gender-presentation format interactions found in this study suggest that males and females may benefit from different instructional presentations. Although some caution needs to be shown in generalisations based on just the Lego tasks used, other research (e.g. Cowards et al., 2012; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006) suggests that subject areas where females are weaker in (e.g. science, maths), may benefit more from animations more than statics. As the evidence also suggests in this study that males benefitted more from statics, it is feasible that males may gain an advantage from statics in learning areas where they are weaker (e.g. languages). Regardless of whether it is only weak subject areas or not that generate these interactions, instructors should give careful consideration to choosing either animations or static pictures, to match the different genders.

6.4.2 In observational learning, showing hands or not should depend on the nature of the learning materials

As discussed in Finding 3, the effectiveness of hand-observations may depend on the degree of hand involvement in the learning materials. Hence, for learning contents where hand movements are part of the learning objective, it is recommended to show the hands. Whereas for learning contents where hands are not essential to the task completion, it is recommended not to show the hands in the instructional materials as they may distract learners' attention from the main content, inducing extraneous cognitive load and hindering learning.

6.4.3 In observational learning, gesturing should be carefully considered before including it as an additional strategy

Although gesturing has been found to be effective in some learning areas and problem solving (Chu & Kita, 2011; Matlen et al., 2012; Ouwehand et al., 2015; Post et al., 2013), the present study suggests that it should not be an automatic inclusion. Based on the results from Experiment 4, gesturing is not necessarily effective in observational learning in particular when repeating simple hand movements. Hence, it is recommended not to overuse gestures in learning, in particular when learning simple movements, or when hand movements are not part of the learning objective.

6.4.4 Caution is needed when adding supplementary learning support strategies

As found in this study adding other learning strategies such as gesturing and hand observations to observational learning tasks may be counterproductive due to redundancy. If the supporting strategies are not required (redundant) additional extraneous cognitive load may be created leading to a loss of learning.

6.5 Limitations of the Study

6.5.1 The Lego task

One of the main aims of the study was to examine the human movement effect with hand-manipulative materials. Although the Lego task in this study involved human manipulations, it is debateable whether the movements were necessary for memorising the brick locations. Blazhenkova and Kozhevnikov (2010) argued that visual-object processing and visual-spatial processing are two distinct processes. Visual-object processing operates information about visual pictorial appearance of objects and scenes in terms of shapes, colours, brightness texture and size; while visual-spatial processing operates information about spatial relations, movement and complex spatial transformation. Furthermore, Kozhevnikov, Kosslyn, and Shephard (2005) and Kozhevnikov, Hegarty, and Mayer (2002) both found individual differences between these two processes. Also a number of neuron researchers (e.g. Farah, Hammond, Levine, & Calvanio,1988; Kosslyn, 1994; Kosslyn & Koenig, 1992; Levine, Warach, & Farah, 1985; Mazard, Tzourio-Mazoyer, Crivello, Mazoyer, & Mellet, 2004; cited in Blazhenkova & Kozhevnikov, 2010) provide evidence for dichotomising the visual-spatial processing and the visual-object processing. From this perspective, the instructional materials (Lego building) used in this thesis may not have been the optimum environment for investigating the human movement effect. Nevertheless, a number of significant effects were found to provide confidence that the task was a legitimate source of investigation, as well as comparing with the previous research of Castro-Alonso, Ayres, and Paas (2015).

6.5.2 Measures of cognitive load

In this thesis, two different cognitive load measures were employed. Experiments 1-3 used the Paas 9-point mental effort rating scale while Experiment 4 used a modified version of Leppink’s 14-item cognitive load questionnaire. It could have been an advantage to use the Leppink scale throughout because of its ability to measure different types of cognitive load. Also both of the questionnaires focused only on the content and design of the Lego presentations. More insights might have been gained by including cognitive load measures, which focused on other factors as well (i.e. observing hands, gesturing, and presentation-formats).

Concerning the validity of the Leppink scale the factor analysis completed in the thesis produced 2 distinct factors consistent with previous research by Leppink (see Leppink et al., 2014). These factors were labelled intrinsic and extraneous cognitive by Leppink and are assumed to represent these individual components of cognitive load. The finding that males had significantly higher extraneous cognitive load than females, but no significant differences were found between the genders on test performance is an interesting finding. According to cognitive load theory (Sweller et al., 2011) differences in extraneous cognitive load should generate differences in performance, which were not found. However, it should be noted that there are examples in the literature of a mismatch between load and performance (Sweller, Ayres, & Kalyuga, 2011). Nevertheless, this results raises some interesting questions: a) is this factor a measure of extraneous cognitive load, or some other closely related construct, b) did males respond to the instructional format (animations) differently than females even though there were no differences in performance, c) it is just an artefact of this experiment? To answer these questions was outside the scope of the study, but future research should certainly investigate this issue further.

6.5.3 Gesturing and hand observation were only included in Experiment 4

Although investigating the effects of gesturing and hands-observations was not the main aim of the study, they were included in the last experiment. Some interesting but inconclusive findings were identified. Clearly more experiments that included these factors would have been insightful, but beyond the scope of this thesis.

6.5.4 Common issues associated with a single laboratory-based study

There are always a number of issues associated with a small sequence of laboratory-based experiments meeting the requirements of a single thesis. It is difficult to generalise using only one learning content (Lego tasks) research. Studies completed in the laboratory lack environmental validity. There was a lack of delayed testing, as well as the exclusion of other multimedia factors such as narrations. All of which could add further insights to this study.

6.6 Future Directions

Following the findings and limitations of this thesis, a number of aspects can be examined in future experiments.

6.6.1 Flowing from the limitations

As identified in the section above, the study had a number of limitations that can generate a number of future research directions. These include different types of tasks and measures of cognitive load, more extensive testing, the further investigation of gesturing and hand observations.

6.6.2 Spatial measurement

One of the major arguments for the gender difference in instructional animations is the difference in spatial ability, and hence it is important to measure spatial ability appropriately. However, the CRT test that was used in most of the experiments (i.e. Experiment 1-3) did not generate consistent results, although it has been used extensively in other studies (e.g. Castro-Alonso et al., 2015a; Geary, Gilger, & Elliott-Miller, 1992; Kozhevnikov & Hegarty, 2001; Mayer & Massa, 2003; Plass, Chun, Mayer, & Leutner, 2003). Also, the correlation tables in Chapter 5 showed that the CRT test did not have a very high correlation with performance scores. Instead, the MRT and the Corsi Test used in Experiment 4 seem to be more relevant to the performance score of males and females respectively. Moreover, taking the results of the regression analysis into consideration, none of these spatial ability assessments consistently succeeded in predicting the performance of both females and males. A strength of this study was that the use of different spatial ability measures did pinpoint some important issues associated with using spatial ability indicators. However, further research is required into spatial ability measures and their suitability for identifying gender differences as well as the impact of spatial ability on learning from visual presentations.

6.6.3 Gender differences in learning from visualisations

As discussed above, the animation advantage found by Castro-Alonso et al. (2015a) might have been caused by the high percentage of females in the study rather the overall impact of animations. This potential bias might have also occurred in the broader literature. Hence, future investigations could examine the extent that past studies have included gender bias in their samples. As well some studies that have found animations or static picture advantages with biased samples could be replicated, but with equal gender participants to confirm whether the various effects remain.

6.6.4 Embodied cognition effects

With the insights gained from Experiment 4 more investigations on gesturing and hands-observations are needed with human movement tasks. As discussed, the level of relevance between gesture and task may influence the effectiveness of both hands-observations and gesturing. One explanation may lay in the nature of the mirror neuron system. As discussed in Chapter 2 (Section 2.4.4.1), the mirror neuron system is goal-directed (di Pellegrino et al., 1992; Rizzolatti, 2005; Rizzolatti & Sinigaglia, 2010). That means the action goal needs to be understood in order to enhance successful simulation. Some evidence can be seen in result from Mavilidi, Okely, Chandler, Cliff, and Paas (2015) where integrated gesture (physical exercise that is relevant to the meaning of the learning task) produced better results than non-integrated gesture. More detailed investigations are needed to confirm such claims.

6.6.5 Movement redundancy

The results of this study suggest that showing hands in certain hand-movement tasks may be redundant (see also Post et al., 2013). The redundancy effect may have been caused by viewing and observing hands at the same time or even the learners' prior knowledge (adult learners are familiar with moving brick with hands and thus showing hands may be redundant). Hence, replicating the experiments of this study with younger children or testing with different gestures would have value, as well as broader investigations into the causes of redundancy in this domain.

6.6.6 Replication of current findings

For all experimental studies it is important to replicate the finding to increase confidence that they are robust effects. Hence future research could try to replicate these results with the present materials as well as different materials in different domains.

6.7 Final Conclusions

This thesis contributes to the research field through a direct examination of some critical factors (genders, hand-observations and gesturing) that impact on instructional visualisations (dynamics and statics), providing a number of theoretical insights. Although instructional animations were believed to be most effective in learning about hand-manipulative tasks, the interactions found in this study question this belief. By considering gender, evidence emerged that for the given Lego tasks, females benefitted most from animations, whereas males benefitted from static pictures. This interaction may explain some of the inconsistent findings with research into instructional animations. Furthermore, this study also demonstrated that potential embodied effects generated through hand-observations and gesturing are not automatic advantages if they are poorly aligned with the main content and instructional materials.

In addition to the theoretical contribution, this thesis also provided some practical contributions to practitioners. Specifically when using instructional visualisations, care must be taken to match the presentation format with gender. Furthermore, instructors should not assume that gestures and hand-observations enhance the learning environment, as under some conditions they may hinder learning.

Overall it can be further concluded that increased attention is needed on gender effects, both when conducting research into instructional animations, as well as using instructional visualisations in the classroom.

Reference

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction, 14*(3), 241–255. doi: 10.1016/j.learninstruc.2004.06.002

Akinlofa, O. R., Holt, P. O. B., & Elyan, E. (2013). Effect of interface dynamism on learning procedural motor skills. *Interacting with Computers, 25*(3), 259–269. doi: 10.1093/iwc/iws015

Akinlofa, O. R., Holt, P. O. B., & Elyan, E. (2014). The cognitive benefits of dynamic representations in the acquisition of spatial navigation skills. *Computers in Human Behavior, 30*, 238–248. doi: 10.1016/j.chb.2013.09.009

Althaus, M., Mulder, L. J. M., Mulder, G., van Roon, A. M., & Minderaa, R. B. (1998). Influence of respiratory activity on the cardiac response pattern to mental effort. *Psychophysiology, 35*, 420–430.

Antonenko, P., Paas, F., Grabner, R., & van Gog, T. (2010). Using electroencephalography to measure cognitive load. *Educational Psychology Review, 22*, 425–438. doi: 10.1007/s10648-010-9130-y

Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents. *Computers in Human Behavior, 25*, 354–350. doi: 10.1016/j.chb.2008.12.014

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. Volume 2, pp. 89–195): Academic Press

Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked example research. *Review of Educational Research, 70*(2), 181–214.

Ayres, P. (2001). Systematic mathematical errors and cognitive load. *Contemporary Educational Psychology, 26*, 227–248.

Ayres, P. (2006). Using subjective measures to detect variations of intrinsic cognitive load within problems. *Learning and Instruction, 16*, 389–400.

Ayres, P. (2015). State-of-the-art research into multimedia learning: A commentary on Mayer's handbook of multimedia learning. *Applied Cognitive Psychology, 29*(4), 631–636. doi: 10.1002/acp.3142

Ayres, P., Castro-Alonso, J. C., Singh, A.-M., Wong, M., Marcus, N., & Paas, F. (2015). *Examining the strategies and learner characteristics that make instructional animations effective.* Paper presented at the 2015 Irish International Conference on Education, Dublin, Ireland.

Ayres, P., Marcus, N., Chan, C., & Qian, N. (2009). Learning hand manipulative tasks: When instructional animations are superior to equivalent static representations. *Computers in Human Behavior, 25*, 348–353.

Ayres, P., & Paas, F. (2007a). Can the cognitive load approach make instructional animations more effective. *Applied Cognitive Psychology, 21*, 811–820.

Ayres, P., & Paas, F. (2007b). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*, 695–700.

Ayres, P., & Sweller, J. (1990). Locus of difficulty in multistage mathematics problems. *The American Journal of Psychology, 103*(2), 167–193. doi: 10.2307/1423141

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 135–146). New York: Cambridge University Press

Ayres, P., & Sweller, J. (2014). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 206–226). New York, N.Y.: Cambridge University Press

Ayres, P., & Youssef, A. (2008). *Investigating the influence of transitory information and motivation during instructional animations*. Paper presented at the 8th International Conference for the Learning Sciences, Utrecht, Netherlands

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

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423. doi: 10.1016/S1364-6613(00)01538-2

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *Psychology of learning and motivation* (Vol. Volume 8, pp. 47–89): Academic Press

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences, 7*(4), 363–379. doi: 10.1016/1041-6080(95)90007-1

Barnea, N., & Dori, Y. J. (1999). High-school chemistry students' performance and gender differences in a computerized molecular modeling learning environment. *Journal of Science Education and Technology, 8*(4), 257–271.

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences, 22*, 577–660.

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

Bartlett, F. C. (1932). *Remembering : a study in experimental and social psychology*. Cambridge: Cambridge University Press.

Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. T. Cacioppo, L. G. Tassinary & G. G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York, NY, US: Cambridge University Press

Beichner, R. J. (1996). The impact of video motion analysis on kinematics graph interpretation skills. *American Journal of Physics, 64*(10), 1272–1277. doi: 10.1119/1.18390

Benedict, R. H. B., Schretlen, D., Groninger, L., Dobraski, M., & Shpritz, B. (1996). Revision of the brief visuospatial memory test: Studies of normal performance, reliability, and validity. *Psychological Assessment, 8*(2), 145–153. doi: 10.1037/1040-3590.8.2.145

Bétrancourt, M. (2005). The animation and interactivity principles in multimedia learning. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 287–296). Cambridge, UK: Cambridge University Press

Bétrancourt, M., Dillenbourg, P., & Clavien, L. (2008). Display of key pictures from animation: Effects on learning. In J.-F. Rouet, R. Lowe & W. Schnotz (Eds.), *Understanding multimedia documents*: Springer

Bétrancourt, M., & Tversky, B. (2000). Effect of computer animation on users' performance: A review. *Le Travail Humain, 63*(4), 311–329.

Blake, T. (1977). Motion in instructional media: Some subject-display mode interactions. *Perceptual and Motor Skills, 44*(3), 975–985. doi: 10.2466/pms.1977.44.3.975

Blazhenkova, O., & Kozhevnikov, M. (2010). Visual-object ability: A new dimension of non-verbal intelligence. *Cognition, 117*(3), 276–301. doi: 10.1016/j.cognition.2010.08.021

Blomberg, G., Sherin, M., Renkl, A., Glogger, I., & Seidel, T. (2014). Understanding video as a tool for teacher education: investigating instructional strategies to promote reflection. *Instructional Science, 42*(3), 443–463. doi: 10.1007/s11251-013-9281-6

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Education, 2*(4), 1–17.

Bonanno, P., & Kommers, P. A. M. (2005). Gender differences and styles in the use of digital games. *Educational Psychology: An International Journal of Experimental Educational Psychology, 25*(1), 13–41. doi: 10.1080/0144341042000294877

Boswell, S. L. (1985). The influence of sex-role stereotyping on women's attitudes and achievement in mathematics. In S. F. Chipman, L. R. Brush & D. M. Wilson (Eds.), *Women and mathematics: Balancing the equation*. NY: Erlbaum: Lawrence Erlbaum Associates, Inc

Boucheix, J.-M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction, 19*, 112–127. doi: 10.1016/j.learninstruc.2008.03.004

Broadbent, D. E. (1958). The selective nature of learning *Perception and communication* (pp. 244–267). Elmsford, NY, US: Pergamon Press

Busch, R. M., Farrell, K., Lisdahl-Medina, K., & Krikorian, R. (2005). Corsi block-tapping task performance as a function of path configuration. *Journal of Clinical and Experimental Neuropsychology, 27*(1), 127–134. doi: 10.1080/138033990513681

Byrne, M. D., Catrambone, R., & Stasko, J. T. (1999). Evaluating animations as student aids in learning computer algorithms. *Computers and Eudcation, 33*, 253–278. doi: 10.1016/S0360-1315(99)00023-8

Campos, A. (2012). Measure of the ability to rotate mental images. *Psicothema, 24*(3), 431–434.

Carroll, J. B. (1974). Psychometric tests as cognitive tasks: A new "structure of intellect" *Technical Report 4*. Princeton, N.J.: Educational Testing Service.

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

Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*(4), 697–705.

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014a). Dynamic visualisations and motor skills. In W. Huang (Ed.), *Handbook of human centric visualization* (pp. 551–575). NY: Springer Science + Business Media New York

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

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015a). Animations showing Lego manipulative tasks: Three potential moderators of effectiveness. *Computers & Education, 85*, 1–13. doi: 10.1016/j.compedu.2014.12.022

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015b). The potential of embodied cognition to improve STEAM instructional dynamic visualizations. In X. Ge, D. Ifenthaler & J. M. Spector (Eds.), *Emerging technologies for STEAM education* (pp. 113–136): Springer International Publishing

Chandler, P. (2004). The crucial role of cognitive processes in the design of dynamic visualizations. *Learning and Instruction, 14*, 353–357.

Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction, 8*(4), 293–332.

Chandler, P., & Sweller, J. (1996). Cognitive load while learning to use a computer program. *Applied Cognitive Psychology, 10*(2), 151–170. doi: 10.1002/(SICI)1099-0720(199604)10:2<151::AID-ACP380>3.0.CO;2-U

Chang, C. C., & Yang, F. Y. (2010). Exploring the cognitive loads of high-school students as they learn concepts in web-based environments. *Computers & Education, 55*(2), 673–680. doi: 10.1016/j.compedu.2010.03.001

ChanLin, L. J. (1998). Animation to teach students of different knowledge levels. *Journal of Instructional Psychology, 25*, 166–175.

ChanLin, L. J. (2001). The effects of gender and presentation format in computer-based learning. *Educational Media International, 38*(1), 61–65.

Chase, W. G., & Simon, H. A. (1973). Perception in Chess. *Cognitive Psychology, 4*, 55–81.

Cherrier, M. M., Asthana, S., Plymate, S., Baker, L., Matsumoto, A. M., Peskind, E., . . . Craft, S. (2001). Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology, 57*(1), 80–88.

Choi, J., & L'Hirondelle, N. (2005). Object location memory: A direct test of the verbal memory hypothesis. *Learning and Individual Differences, 15*(3), 237–245. doi: 10.1016/j.lindif.2005.02.001

Chu, M., & Kita, S. (2008). Spontaneous gestures during mental rotation tasks: Insights into the microdevelopment of the motor strategy. *Journal of Experimental Psychology: General, 137*(4), 706–723. doi: 10.1037/a0013157

Chu, M., & Kita, S. (2011). The nature of gestures' beneficial role in spatial problem solving. *Journal of Experimental Psychology: General, 140*(1), 102–116. doi: 10.1037/a0021790

Çığrık, E., & Ergül, R. (2009). The investigation of the effect of simulation based teaching on the student achievement and attitude in electrostatic induction. *Procedia - Social and Behavioral Sciences, 1*(1), 2470–2474. doi: 10.1016/j.sbspro.2009.01.434

Clark, A. (2008). *Supersizing the mind: Embodiment, action, and cognitive extension*. Oxford: Oxford University Press.

Clark, J. M., & Paivio, A. (1991). Dual coding theory and education. *Educational Psychology Review, 3*(3), 149–210.

Clarke, T., Ayres, P., & Sweller, J. (2005). The impact of sequencing and prior knowledge on learning mathematics through spreadsheet applications. *Educational Technology Research and Development, 53*(3), 15–24. doi: 10.1007/bf02504794

Clements-Stephens, A. M., Rimrodt, S. L., & Cutting, L. E. (2009). Developmental sex differences in basic visuospatial processing: Differences in strategy use? *Neuroscience Letters, 449*, 155–160. doi: 10.1016/j.neulet.2008.10.094

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology, 24*(3), 329–340. doi: 10.1016/j.jenvp.2004.08.006

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition, 106*, 1047–1058. doi: 10.1016/j.cognition.2007.04.010

Cooper, G. A., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology, 79*(4), 347–362.

Cornwell, B. R., Johnson, L. L., Holroyd, T., Carver, F. W., & Grillon, C. (2008). Human hippocampal and parahippocampal theta during goal-directed spatial navigation predicts performance on a virtual Morris water maze. *The Journal of Neuroscience, 28*(23), 5983–5990. doi: 10.1523/jneurosci.5001-07.2008

Corsi, P. M. (1972). *Human memory and the medial temporal region of the brain.* (Doctor of Philosophy), McGill University, Montréal, Canada.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory? In J.-C. L. V. F. C. Wayne S. Sossin & B. Sylvie (Eds.), *Progress in brain research* (Vol. Volume 169, pp. 323–338): Elsevier

Cowan, N. (2010). The magical mystery four: How is working memory capacity limited, and why? *Current directions in psychological science, 19*(1), 51–57. doi: 10.1177/0963721409359277

Cowards, F. L., Crooks, S. M., Flores, R., & Dao, D. (2012). Examining the effect of gender and presentation mode on learning from a multimedia presentation. *GÉNEROS - Multidisciplinary Journal of Gender Studies, 1*(1), 48–69. doi: 10.4471/generos.2012.03

Dalacosta, K., Kamariotaki-Paparrigopoulou, M., Palyvos, J. A., & Spyrellis, N. (2009). Multimedia application with animated cartoons for teaching science in elementary education. *Computers & Education, 52*, 741–748. doi: 10.1016/j.compedu.2008.11.018

Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo-spatial working memory. *Journal of Neurology, 253*(2), 176–180. doi: 10.1007/s00415-005-0944-3

Darwin, C. (1859). *On the origin of species by means of natural selection, or the preservation of favored races in the struggle for life*. Champaign, IL: Project Gutenberg.

Darwin, C. (1871). *The descent of man, and selection in relation to sex*. Princeton, N.J.: Princeton, N.J. : Princeton University Press.

de Beni, R., Pazzaglia, F., Gyselinck, V., & Meneghetti, C. (2005). Visuospatial working memory and mental representation of spatial descriptions. *European Journal of Cognitive Psychology, 17*(1), 77–95. doi: 10.1080/09541440340000529

de Groot, A. D. (1965). *Thought and choice in chess*. The Hague: The Hague: Mouton.

Della Sala, S., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: a tool for unwelding visuo-spatial memory. *Neuropsychologia, 37*(10), 1189–1199.

Della Sala, S., Gray, C., Baddeley, A. D., & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall.*: Thames Valley Test Company.

Derry, S. J., Sherin, M. G., & Sherin, B. L. (2014). Multimedia Learning with video. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 785–812). New York, N.Y.: Cambridge University Press

Deyzac, E., Logie, R. H., & Denis, M. (2006). Visuospatial working memory and the processing of spatial descriptions. *British Journal of Psychology, 97*(2), 217–243. doi: 10.1348/000712605X67484

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research, 91*, 176–180.

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology, 120*(2), 237–261.

Duff, S. J., & Hampson, E. (2001). A sex difference on a novel spatial working memory task in humans. *Brain and Cognition, 47*, 470–493.

Eals, M., & Silverman, I. (1994). The Hunter-Gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology, 15*(2), 95–105. doi: 10.1016/0162-3095(94)90020-5

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*.

Falvo, D. A., & Suits, J. P. (2009). Gender and spatial ability and the use of specific labels and diagrammatic arrows in a micro-level Chemistry animation. *Journal of Educational Computing Research, 41*(1), 83–102. doi: 10.2190/EC.41.1.d

Flores, R., Cowards, F. L., & Crooks, S. M. (2010). Examining the influence of gender on the modality effect. *Journal of Educational Technology Systems, 39*(1), 87–103. doi: doi.10.2190/ET.39.1.g

French, J. W. (1963). The relationship of problem-solving styles to the factor composition of tests. *ETS Research Bulletin Series, 1963*(1), i–42. doi: 10.1002/j.2333-8504.1963.tb02349.x

Frings, L., Wagner, K., Unterrainer, J., Spreer, J., Halsband, U., & Schulze-Bonhage, A. (2006). Gender-related differences in lateralization of hippocampal activation and cognitive strategy. *NeuroReport, 17*(4), 417–421. doi: 10.1097/01.wnr.0000203623.02082.e3

Galea, L. A. M., & Kimura, D. (1993). Sex difference in route-learning. *Personality and Individual Differences, 14*(1), 53–65.

Garg, A. X., Norman, G., & Sperotable, L. (2001). How medical students learn spatial anatomy. *The Lancet, 357*(9253), 363–364. doi: 10.1016/S0140-6736(00)03649-7

Garland, T. B., & Sánchez, C. A. (2013). Rotational perspective and learning procedural tasks from dynamic media. *Computers & Education, 69*, 31–37. doi: 10.1016/j.compedu.2013.06.014

Gaulin, S. J. C., Silverman, I., Phillips, K., & Reiber, C. (1997). Activational hormonal influences on abilities and attitudes: Implications for evolutionary theory. *Evolution and Cognition, 3*(2), 191–199.

Geary, D. C. (1995). Sexual selection and sex differences in spatial cognition. *Learning and Individual Differences, 7*(4), 289–301. doi: 10.1016/1041-6080(95)90003-9

Geary, D. C. (1996). Sexual selection and sex differences in mathematical abilities. *Behavioral and Brain Sciences, 19,* 229–284. doi: 10.1017/S0140525X00042400

Geary, D. C. (2002). Principles of evolutionary educational psychology. *Learning and Individual Differences, 12,* 317–345.

Geary, D. C. (2004). Mathematics and learning disabilities. *Journal of Learning Disabilities, 37*(1), 4–15. doi: 10.1177/00222194040370010201

Geary, D. C. (2005). *The origin of mind: Evolution of brain, cognition, and general intelligence*. Washington, DC: American Psychological Association.

Geary, D. C. (2007). Educating the evolved mind: Conceptual foundations for an evolutionary educational psychology. In J. S. Carlson & J. R. Levin (Eds.), *Psychological perspectives on contemporary educational issues* (pp. 1–99). Greenwich, CT: Information Age Publishing.

Geary, D. C. (2008). An evolutionarily informed education science. *Educational Psychologist, 43*(4), 179–195. doi: 10.1080/00461520802392133

Geary, D. C., Gilger, J. W., & Elliott-Miller, B. (1992). Gender differences in three-dimensional mental rotation: A replication. *The Journal of Genetic Psychology, 153*(1), 115–117. doi: 10.1080/00221325.1992.10753706

Geiringer, E. R., & Hyde, J. S. (1976). Sex differences on Piaget's water-level task: spatial ability incognito. *Perceptual and Motor Skills, 42,* 1323–1328.

Gerjets, P., Scheiter, K., Opfermann, M., Hesse, F. W., & Eysink, T. H. S. (2009). Learning with hypermedia: The influence of representational formats and different levels of learner control on performance and learning behavior. *Computers in Human Behavior, 25*(2), 360–370. doi: 10.1016/j.chb.2008.12.015

Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction, 16*(6), 511–525. doi: 10.1016/j.learninstruc.2006.10.001

Glenberg, A. M. (1997). What memory is for: Creating meaning in the service of action. *Behavioral and Brain Sciences, 20*(1), 41–50. doi: doi:10.1017/S0140525X97470012

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining Math: Gesturing lightens the load. *Psychological Science, 12*(6), 516–522.

Griffin, A. L., MacEachren, A. M., Hardisty, F., Steiner, E., & Li, B. (2006). A comparison of animated maps with static small-multiple maps for visually identifying space-time clusters. *Annals of the Association of American Geographers, 96*(4), 740–753. doi: 10.1111/j.1467-8306.2006.00514.x

Grimley, M. (2007). Learning from multimedia materials: The relative impact of individual differences. *Educational Psychology: An International Journal of Experimental Educational Psychology Review, 27*(4), 465–485. doi: 10.1080/01443410601159795

Guillem, F., & Mograss, M. (2005). Gender differences in memory processing: Evidence from event-related potentials to faces. *Brain and Cognition, 57*(1), 84–92. doi: 10.1016/j.bandc.2004.08.026

Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed. ed.). New York: New York : Psychology Press.

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

Hasler, B. S., Kersten, B., & Sweller, J. (2007). Learner control, cognitive load and instructional animation. *Applied Cognitive Psychology, 21*, 713–729. doi: 10.1002/acp.1345

Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1084–1102. doi: 10.1037/0278-7393.18.5.1084

Hegarty, M. (2014). Multimedia learning and the development of metal models. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 673–701). New York, N.Y.: Cambridge University Press

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical for problem solving. *Journal of Educational Psychology, 91*(4), 684–689.

Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognitive and Instruction, 21*(4), 325–360.

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

Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition, 22*(4), 411–430. doi: 10.3758/BF03200867

Hegarty, M., & Waller, D. (2005). Individual difference in spatial abilities. In P. Shah & M. Akira (Eds.), *The Cambridge handbook of visuospatial thinking*. New York, NY: Cambridge University Press

Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—a meta-analytic review. *Educational Psychology Review, 22*(3), 245–269. doi: 10.1007/s10648-010-9126-7

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: A meta-analysis. *Learning and Instruction 17*, 722–738.

Höffler, T. N., & Leutner, D. (2011). The role of spatial ability in learning from instructional animations – Evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior, 27*, 209–216. doi: 10.1016/j.chb.2010.07.042

Holland, M. K., & Tarlow, G. (1972). Blinking and mental load. *Psychological Reports, 31*(1), 119–127. doi: 10.2466/pr0.1972.31.1.119

Hoogerheide, V., Loyens, S. M. M., & van Gog, T. (2016). Learning from video modeling examples: does gender matter? *Instructional Science*(44), 69–86. doi: 10.1007/s11251-015-9360-y

Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*(1), 18–29. doi: 10.1016/j.learninstruc.2008.12.001

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w and d. *American Psychologist*, *36*(8), 892–901.

Hyde, J. S., Fennema, E., & Lamon, S. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139–155.

Hyde, J. S., & Linn, M. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*(1), 53–69. doi: 10.1037/0033-2909.104.1.53

Hyde, J. S., & Mertz, J. E. (2009). Gender, culture, and mathematics performance. *Proceedings of the National Academy of Sciences, 106*(22), 8801–8807. doi: 10.1073/pnas.0901265106

Hyönä, J. (2010). The use of eye movements in the study of multimedia learning. *Learning and Instruction, 20*(2), 172–176. doi: 10.1016/j.learninstruc.2009.02.013

Jacek, L. L. (1997). *Gender differences in learning physical science concepts: Does computer animation help equalise them?* (Doctor of Philosophy), Oregon State University.

Jacobs, L. F. (1996). Sexual selection and the brain. *Trends in Ecology & Evolution, 11*(2), 82–86.

Jee, B. D., Uttal, D. H., Gentner, D., Manduca, C., Shipley, T. F., & Sageman, B. (2013). Finding faults: analogical comparison supports spatial concept learning in geoscience. *Cognitive Processing, 14*(2), 175–187. doi: 10.1007/s10339-013-0551-7

Jeung, H.-J., Chandler, P., & Sweller, J. (1997). The role of visual indicators in dual sensory mode instruction. *Educational Psychology: An International Journal of Experimental Education Psychology*, *17*(3), 329–345.

Jones, S., & Scaife, M. (2000). Animated diagrams: An investigation into the cognitive effects of using animation to illustrate dynamic processes. In M. Anderson & P. Cheng (Eds.), *Theory & applications of diagrams: Lecture notes in artificial intelligence* (pp. 231–244). Berlin: Springer-Verlag

Jordan, K., Wüstenberg, T., Heinze, H.-J., Peters, M., & Jäncke, L. (2002). Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia, 40*(13), 2397–2408. doi: 10.1016/S0028-3932(02)00076-3

Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. *Computers in Human Behavior, 24*, 852–861.

Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 576–597). New York, N.Y.: Cambridge University Press

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. *Applied Cognitive Psychology, 13*(351–371).

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology, 92*(1), 126–136. doi: 10.1037/0022-0663.92.1.126

Kalyuga, S., Chandler, P., & Sweller, J. (2001). Learner experience and efficiency of instructional guidance. *Educational Psychology, 21*(1), 5–23. doi: 10.1080/01443410124681

Kalyuga, S., Chandler, P., & Sweller, J. (2004). When redundant on-screen text in multimedia technical instruction can interfere with learning. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 46*(567–581). doi: 10.1518/hfes.46.3.567.50405

Kalyuga, S., & Sweller, J. (2014). The redundancy principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 247–262). New York, N.Y.: Cambridge University Press

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*, 211–223.

Kessels, R. P. C., van Zandvoort, M. J. E., Postma, A., Kappelle, L. J., & de Haan, E. H. F. (2000). The Corsi block-tapping task: Standardization and normative Data. *Applied Neuropsychology, 7*(4), 252–258. doi: 10.1207/S15324826AN0704_8

Kilic, E., & Yildirim, Z. (2010). Evaluating working memory capacity and cognitive load in learning from goal based scenario centered 3D multimedia. *Procedia - Social and Behavioral Sciences, 2*(2), 4480–4486. doi: 10.1016/j.sbspro.2010.03.715

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). Why minimal guidance during instruction does not work: An analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist, 41*(2), 75–86. doi: 10.1207/s15326985ep4102_1

Klingner, J., Kumar, R., & Hanrahan, P. (2008). *Measuring the task-evoked pupillary response with a remote eye tracker*. Paper presented at the Proceedings of the 2008 Symposium on Eye-Tracking Research & Applications, Savannah, Georgia.

Korakakis, G., Pavlatou, E. A., Palyvos, J. A., & Spyrellis, N. (2009). 3D visualization types in multimedia applications for science learning: A case study for 8th grade students in Greece. *Computers & Education, 52*, 390–401. doi: 10.1016/j.compedu.2008.09.011

Koroghlanian, C., & Klein, J. D. (2004). The effect of audio and animation in multimedia instruction. *Journal of Educational Multimedia and Hypermedia, 13*(1), 23–46.

Kotovsky, K., Hayes, J. R., & Simon, H. A. (1985). Why are some problems hard? Evidence from Tower of Hanoi. *Cognitive Psychology, 17*(2), 248–294. doi: 10.1016/0010-0285(85)90009-X

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745–756. doi: 10.3758/BF03200477

Kozhevnikov, M., Hegarty, M., & Mayer, R. E. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction, 20*(1), 47–77.

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

Kramer, J. H., Delis, D. C., & Daniel, M. (1988). Sex differences in verbal learning. *Journal of Clinical Psychology, 44*(6), 907–915. doi: 10.1002/1097-4679(198811)44:6<907::AID-JCLP2270440610>3.0.CO;2-8

Kühl, T., Eitel, A., Damnik, G., & Körndle, H. (2014). The impact of disfluency, pacing, and students' need for cognition on learning with multimedia. *Computers in Human Behavior, 35,* 189–198. doi: 10.1016/j.chb.2014.03.004

Kühl, T., Scheiter, K., & Gerjets, P. (2012). Enhancing learning from dynamic and static visualizations by means of cueing. *Journal of Educational Multimedia and Hypermedia, 21*(1), 71–88.

Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior, 27*(1), 29–35. doi: 10.1016/j.chb.2010.05.008

Kyun, S., Kalyuga, S., & Sweller, J. (2013). The effect of worked examples when learning to write essays in English literature. *The Journal of Experimental Education, 81*(3), 385–408. doi: 10.1080/00220973.2012.727884

Large, A. (1996). Computer animation in an instructional environment. *Library & Information Science Research, 18*(1), 3–23. doi: 10.1016/S0740-8188(96)90028-6

Large, A., Beheshti, J., Breuleux, A., & Renaud, A. (1996). Effect of animation in enhancing descriptive and procedural texts in a multimedia learning environment. *Journal of the American Society for Information Science, 47*(6), 437–448. doi: 10.1002/(SICI)1097-4571(199606)47:6<437::AID-ASI4>3.0.CO;2-W

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science, 11*(1), 65–100. doi: 10.1111/j.1551-6708.1987.tb00863.x

Law, D. J., Pellegrho, J. W., & Hunt, E. B. (1993). Comparing the tortoise and the hare: Gender differences and experience in dynamic spatial reasoning tasks. *Psychological Science, 4*(1), 35–40. doi: 10.1111/j.1467-9280.1993.tb00553.x

Lawton, C. A., Charleston, S. I., & Zieles, A. S. (1996). Individual- and gender-related differences in indoor wayfinding. *Environment and Behavior, 28*(2), 204–219. doi: 10.1177/0013916596282003

Leahy, W., & Sweller, J. (2011). Cognitive load theory, modality of presentation and the transient information effect. *Applied Cognitive Psychology, 25*(6), 943–951. doi: 10.1002/acp.1787

Lee, D. Y., & Shin, D.-H. (2012). An empirical evaluation of multi-media based learning of a procedural task. *Computers in Human Behavior, 28*(3), 1072–1081. doi: 10.1016/j.chb.2012.01.014

Lee, V. E., & Burkam, D. T. (1996). Gender differences in middle-grade science achievement: Subject domain, ability level, and course emphasis. *Science Education, 80*(6), 613–650.

Legewie, J., & DiPrete, T. A. (2014). The high school environment and the gender gap in Science and Engineering. *Sociology of Education, 87*(4), 259–280. doi: 10.1177/0038040714547770

Leppink, J., Paas, F., van der Vleuten, C. P. M., van Gog, T., & van Merriënboer, J. J. G. (2013). Development of an instrument for measuring different types of cognitive load. *Behavior Research Methods, 45,* 1058–1072. doi: 10.3758/s13428-013-0334-1

Leppink, J., Paas, F., van Gog, T., van der Vleuten, C. P. M., & van Merriënboer, J. J. G. (2014). Effects of pairs of problems and examples on task performance and different types of cognitive load. *Learning and Instruction, 30,* 32–42. doi: dx.doi.org/10.1016/j.learninstruc.2013.12.001

Leppink, J., & van den Heuvel, A. (2015). The evolution of cognitive load theory and its application to medical education. *Perspectives on Medical Education, 4*, 119–127. doi: 10.1007/s40037-015-0192-x

Lewi, D., & Barron, A. (2009). *Animated demonstrations: Evidence of improved performance efficiency and the worked example effect.* Paper presented at the First International Conference, HCD 2009, Held as Part of HCI International 2009, July 19-24, San Diego, CA, USA.

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

Logie, R. H. (1995). *Visuo-spatial working memory*. Hove: Lawrence Erlbaum.

Logie, R. H., Zucco, G. M., & Baddeley, A. D. (1990). Interference with visual short-term memory. *Acta Psychologica, 75*(1), 55–74. doi: 10.1016/0001-6918(90)90066-O

Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Hillsdale, NJ: Lawrence Erlbaum Associaties, Inc.

Lohman, D. F., & Nichols, P. D. (1990). Training spatial abilities: Effects of practice on rotation and synthesis tasks. *Learning and Individual Differences, 2*(1), 67–93. doi: 10.1016/1041-6080(90)90017-B

Loring-Meier, S., & Halpern, D. F. (1999). Sex differences in visuospatial working memory: Components of cognitive processing. *Psychonomic Bulletin & Review, 6*(3), 464–471. doi: 10.3758/BF03210836

Low, R., & Sweller, J. (2014). The modality principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 227–246). New York, N.Y.: Cambridge University Press

Lowe, R. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*, 157–176. doi: 10.1016/S0959-4752(02)00018-X

Lowe, R. (2004). *User-controllable animated diagrams: The solution for learning dynamic content?* Paper presented at the Third International Conference, Diagrams 2004, Cambridge, UK.

Lowe, R., & Schnotz, W. (2014). Animation principles in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 513–546). New York, N.Y.: Cambridge University Press

Lusk, M. M., & Atkinson, R. K. (2007). Animated pedagogical agents: Does their degree of embodiment impact learning from static or animated worked examples? *Applied Cognitive Psychology, 21*, 747–764. doi: 10.1002/acp.1347

Macken, L., & Ginns, P. (2014). Pointing and tracing gestures may enhance anatomy and physiology learning. *Medical Teacher, 36*(7), 596–601. doi: 10.3109/0142159X.2014.899684

Marcus, N., Cleary, B., Wong, A., & Ayres, P. (2013). Should hand actions be observed when learning hand motor skills from instructional animations? *Computers in Human Behavior, 29*, 2172–2178.

Marcus, N., Cooper, M., & Sweller, J. (1996). Understanding instructions. *Journal of Educational Psychology, 88*(1), 49–63. doi: 10.1037/0022-0663.88.1.49

Martens, J., & Antonenko, P. D. (2012). Narrowing gender-based performance gaps in virtual environment navigation. *Computers in Human Behavior, 28*, 809–819. doi: 10.1016/j.chb.2012.01.008

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing. *Behavior Genetics, 23*(4), 337–341. doi: 10.1007/BF01067434

Matlen, B. J., Atit, K., Göksun, T., Rau, M. A., & Ptouchkina, M. (2012). Representing space: Exploring the relationship between gesturing and geoscience understanding in children. In C. Stachniss, K. Schill & D. Uttal (Eds.), *Spatial cognition 2012* (pp. 405–415). Berlin Heidelberg: Springer-Verlag

Mavilidi, M.-F., Okely, A. D., Chandler, P., Cliff, D. P., & Paas, F. (2015). Effects of integrated physical exercises and gestures on preschool children's foreign language vocabulary learning. *Educational Psychology Review, 27*(3), 413–426. doi: 10.1007/s10648-015-9337-z

Mayer, R. E. (2001). *Multimedia learning*. Cambridge, UK: Cambridge University Press.

Mayer, R. E. (2008). *Learning and Instruction, 2nd edition.* Upper Saddle River, N.J.: Pearson Merrill Prentice Hall

Mayer, R. E. (2010). Unique contributions of eye-tracking research to the study of learning with graphics. *Learning and Instruction, 20*(2), 167–171. doi: 10.1016/j.learninstruc.2009.02.012

Mayer, R. E. (2014a). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, N.Y.: Cambridge University Press

Mayer, R. E. (2014b). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 1–24). New York, N.Y.: Cambridge University Press

Mayer, R. E., & Anderson, R. B. (1991). Animation need narrations: An experimental test of a dual-coding hypothesis. *Journal of Education Psychology, 83*(4), 484–490.

Mayer, R. E., & Chandler, P. (2001). When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages. *Journal of Educational Psychology*, *93*(2), 390–397. doi: 10.1037//0022-0663.93.2.390

Mayer, R. E., Deleeuw, K. E., & Ayres, P. (2007). Creating retroactive and proactive interference in multimedia learning. *Applied Cognitive Psychology, 21*, 795–809. doi: 10.1002/acp.1350

Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied, 11*(4), 256–265. doi: 10.1037/1076-898X.11.4.256

Mayer, R. E., & Massa, L. J. (2003). Three facets of visual and verbal learners: Cognitive ability, cognitive style, and learning preference. *Journal of Educational Psychology, 95*(4), 833–846. doi: 10.1037/0022-0663.95.4.833

Mayer, R. E., Mathias, A., & Wetzell, K. (2002). Fostering understanding of multimedia messages through pre-training: evidence for a two-stage theory of mental model construction. *Journal of Experimental Psychology: Applied, 8*(3), 147–154.

Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology, 86*(3), 389–401. doi: 10.1037/0022-0663.86.3.389

McBurney, D. H., Gaulin, S. J. C., Devineni, T., & Adams, C. (1997). Superior spatial memory of women: Stronger evidence for the gathering hypothesis. *Evolution and Human Behavior, 18*(3), 165–174. doi: 10.1016/S1090-5138(97)00001-9

McConnell, J., & Quinn, J. G. (2000). Interference in visual working memory. *The Quarterly Journal of Experimental Psychology Section A, 53*(1), 53–67. doi: 10.1080/713755873

McConnell, J., & Quinn, J. G. (2004). Complexity factors in visuo-spatial working memory. *Memory, 12*(3), 338–350. doi: 10.1080/09658210344000035

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*(5), 889–918.

McGivern, R. F., Huston, J. P., Byrd, D., King, T., Siegle, G. J., & Reilly, J. (1997). Sex differences in visual recognition memory: Support for a sex-related difference in attention in adults and children. *Brain and Cognition, 34*, 323–336. doi: 10.1006/brcg.1997.0872

McGivern, R. F., Mutter, K. L., Anderson, J., Wideman, G., Bodnar, M., & Huston, P. J. (1998). Gender differences in incidental learning and visual recognition memory: Support for a sex difference in unconscious environmental awareness. *Personality and Individual Differences, 25*, 223–232.

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

Michas, I. C., & Berry, D. C. (2000). Learning a procedural task: Effectiveness of multimedia presentations. *Applied Cognitive Psychology, 14*(6), 555–575. doi: 10.1002/1099-0720(200011/12)14:6<555::AID-ACP677>3.0.CO;2-4

Miller, G. A. (1956). The magical number seven, plus or minus two some limits on our capacity for processing information. *The Psychological Review, 63*, 81–97.

Moreno, R., & Mayer, R. E. (2007). Interactive multimodal learning environments: Special issue on interactive learning environments: Contemporary issues and trends. *Educational Psychology Review, 19*(3), 309–326. doi: 10.1007/s10648-007-9047-2

Morrison, J. B., & Tversky, B. (2001). *The (in)effectiveness of animation in instruction.* Paper presented at the ACM Conference on Human Factors in Computing Systems, Danvers, MA.

Morrison, J. B., Tversky, B., & Bétrancourt, M. (2000). *Animation: Does it facilitate learning?* Paper presented at the Smart Graphics AAAI Spring Symposium, Stanford, CA, USA.

Morsella, E., & Krauss, R. M. (2004). The role of gestures in spatial working memory and speech. *The American Journal of Psychology, 117*(3), 411–424.

Mousavi, S. Y., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334.

Mukamel, R., Ekstrom, A. D., Kaplan, J., Iacoboni, M., & Fried, I. (2010). Single-neuron responses in humans during execution and observation of actions. *Current Biology, 20*, 750–756. doi: 10.1016/j.cub.2010.02.045

Murphy, R. J. L. (1982). Sex differences in objective test performance. *British Journal of Educational Psychology, 52*(2), 213–219. doi: 10.1111/j.2044-8279.1982.tb00828.x

Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*, 279–315.

Neuschmidt, O., Barth, J., & Hastedt, D. (2008). Trends in gender differences in mathematics and science (TIMSS 1995–2003). *Studies in Educational Evaluation, 34*(2), 56–72. doi: 10.1016/j.stueduc.2008.04.002

Nievelstein, F., van Gog, T., van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less structured tasks: Learning to reason about legal cases. *Contemporary Educational Psychology, 38*(2), 118–125. doi: 10.1016/j.cedpsych.2012.12.004

Oksa, A., Kalyuga, S., & Chandler, P. (2009). Expertise reversal effect in using explanatory notes for readers of Shakespearean text. *Instructional Science, 38*(3), 217–236. doi: 10.1007/s11251-009-9109-6

Orion, N., & Kali, Y. (2005). The effect of an Earth-Science learning program on students' scientific thinking skills. *Journal of Geoscience Education 53*(4), 388–393.

Ouwehand, K., van Gog, T., & Paas, F. (2015). Effects of gestures on older adults' learning from video-based models. *Applied Cognitive Psychology, 29*, 115–128.

Owens, P., & Sweller, J. (2008). Cognitive load theory and music instruction. *Educational Psychology*, *28*(1), 29–45. doi: 10.1080/01443410701369146

Ozcelik, E., Arslan-Ari, I., & Cagiltay, K. (2010). Why does signaling enhance multimedia learning? Evidence from eye movements. *Computers in Human Behavior, 26*(1), 110–117. doi: 10.1016/j.chb.2009.09.001

Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive load approach. *Journal of Educational Psychology, 84*(4), 429–434.

Paas, F., & Sweller, J. (2012). An evolutionary upgrade of cognitive load theory: Using the human motor system and collaboration to support the learning of complex cognitive tasks. *Educational Psychology Review, 24*, 27–45.

Paas, F., Tuovinen, J. E., Tabbers, H., & van-Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, *38*(1), 63–71.

Paas, F., van-Gerven, P. W. M., & Wouters, P. (2007). Instructional efficiency of animation: Effects of interactivity through mental reconstruction of static key frames. *Applied Cognitive Psychology, 21*, 783–793. doi: 10.1002/acp.1349

Paas, F., & van Merriënboer, J. J. G. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 35*(4), 737–743. doi: 10.1177/001872089303500412

Paas, F., & van Merriënboer, J. J. G. (1994a). Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review, 6*(4), 351–371.

Paas, F., & van Merriënboer, J. J. G. (1994b). Variability of Worked Examples and Transfer of Geometrical Problem-Solving Skills: A Cognitive-Load Approach. *Journal of Educational Psychology, 86*(1), 122–133.

Paas, F., van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*, 419–430.

Pagulayan, K. F., Busch, R. M., Medina, K. L., Bartok, J. A., & Krikorian, R. (2006). Developmental normative data for the corsi block-tapping task. *Journal of Clinical and Experimental Neuropsychology, 28*(6), 1043–1052. doi: 10.1080/13803390500350977

Paivio, A. (1971). *Imagery and verbal processes*. New York: New York, Holt, Rinehart and Winston.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45*(3), 255–287.
Passig, D., & Levin, H. (2000). Gender preferences for multimedia interfaces. *Journal of Computer Assisted Learning*, *16*, 64–71.
Pazzaglia, F., & Cornoldi, C. (1999). The role of distinct components of visuo-spatial working memory in the processing of texts. *Memory, 7*(1), 19–41. doi: 10.1080/741943715
Pearson, J. L., & Ferguson, L. R. (1989). Gender differences in patterns of spatial ability, environmental cognition, and Math and English achievement in late adolescence. *Adolescence, 24*(94), 421–431.
Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition, 66*, 260–264. doi: 10.1016/j.bandc.2007.09.003
Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28*, 39–58.
Peterson, L. R., & Peterson, M. J. (1959). Short-term retention of individual verbal items. *Journal of Experimental Psychology, 58*, 193–198.
Pfeiffer, V. D. I., Scheiter, K., Kühl, T., & Gemballa, S. (2011). Learning how to identify species in a situated learning scenario: Using dynamic-static visualizations to prepare students for their visit to the aquarium. *EURASIA Journal of Mathematics, Science & Technology Education*, *7*(2), 135–147.
Piaget, J. (1926). *The language and thought of the child*. New York: Harcourt, Brace & Company.
Pillay, H. K. (1994). Cognitive load and mental rotation: structuring orthographic projection for learning and problem solving. *Instructional Science*, *22*, 91–113.
Plass, J. L., Chun, D. M., Mayer, R. E., & Leutner, D. (2003). Cognitive load in reading a foreign language text with multimedia aids and the influence of verbal and spatial abilities. *Computers in Human Behavior*, *19*, 221–243. doi: 10.1016/S0747-5632(02)00015-8
Ploetzner, R., & Lowe, R. (2012). A Systematic characterisation of expository animations. *Computers in Human Behavior, 28*, 781–794. doi: 10.1016/j.chb.2011.12.001
Pöhnl, S., & Bogner, F. X. (2012). Learning with computer-based multimedia gender effects on efficiency. *Journal of Educational Computing Research, 47*(4), 387–407. doi: 10.2190/EC.47.4.c
Pöhnl, S., & Bogner, F. X. (2013). Cognitive load and alternative conceptions in learning genetics: Effects from provoking confusion. *The Journal of Educational Research*, *106*(3), 183–196. doi: 10.1080/00220671.2012.687790
Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction, 12*, 61–86. doi: 10.1016/S0959-4752(01)00016-0
Post, L. S., van Gog, T., Paas, F., & Zwaan, R. A. (2013). Effects of simultaneously observing and making gestures while studying grammar animations on cognitive load and learning. *Computers in Human Behavior, 29*, 1450–1455. doi: 10.1016/j.chb.2013.01.005
Postma, A., Izendoorn, R., & de Haan, E. H. F. (1998). Sex differences in object location memory. *Brain and Cognition, 36*, 334–345. doi: 10.1006/brcg.1997.0974
Postma, A., Meyer, G., Tuiten, A., van Honk, J., Kessels, R. P. C., & Thijssen, J. (2000). Effects of testosterone administration on selective aspects of object-

location memory in healthy young women. *Psychoneuroendocrinology, 25*(6), 563–575. doi: 10.1016/S0306-4530(00)00010-X

Quinn, J. G. (1994). Towards a clarification of spatial processing. *The Quarterly Journal of Experimental Psychology Section A*, *47*(2), 465–480. doi: 10.1080/14640749408401120

Quinn, J. G. (2008). Movement and visual coding: the structure of visuo-spatial working memory. *Cognitive Process, 9*, 35–43. doi: 10.1007/s10339-007-0184-9

Quinn, J. G., & McConnell, J. (1996). Irrelevant pictures in visual working memory. *The Quarterly Journal of Experimental Psychology Section A*, *49*(1), 200–215. doi: 10.1080/713755613

Quinn, J. G., & Ralston, G. E. (1986). Movement and attention in visual working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 689–703. doi: 10.1080/14640748608401621

Raymond, C. L., & Benbow, C. P. (1986). Gender differences in mathematics: A function of parental support and student sex typing? *Developmental Psychology, 22*(6), 808-819. doi: 10.1037/0012-1649.22.6.808

Rebetez, C., Bétrancourt, M., Sangin, M., & Dillenbourg, P. (2009). Learning from animation enabled by collaboration. *Instructional Science, 38*(5), 471–485. doi: 10.1007/s11251-009-9117-6

Redick, T. S., Broadway, J. M., Meier, M. E., Kuriakose, P. S., Unsworth, N., Kane, M. J., & Engle, R. W. (2012). Measuring working memory capacity with automated complex span tasks. *European Journal of Psychological Assessment*, *28*(3), 164–171. doi: doi:10.1027/1015-5759/a000123

Renkl, A. (2002). Worked-out examples: instructional explanations support learning by self-explanations. *Learning and Instruction*, *12*(5), 529–556. doi: 10.1016/S0959-4752(01)00030-5

Renkl, A. (2014). The worked examples principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 391–412). New York, N.Y.: Cambridge University Press

Renkl, A., & Atkinson, R. K. (2003). Structuring the transition from example study to problem solving in cognitive skill acquisition: A cognitive load perspective. *Educational Psychologist*, *38*(1), 15–22. doi: 10.1207/S15326985EP3801_3

Riding, R., & Grimley, M. (1999). Cognitive style, gender and learning from mult-media material in 11-year-old children. *British Journal of Educational Technology*, *30*(1), 43–56.

Rieber, L. P. (1989). The effects of computer animated elaboration strategies and practice on factual and application learning in an elementary science lesson. *Journal of Educational Computing Research, 5*(4), 431–444. doi: 10.2190/961X-QQEQ-7PVK-UK9E

Rieber, L. P. (1990a). Animation in computer-based instruction. *Educational Technology Research and Development, 38*(1), 77–86.

Rieber, L. P. (1990b). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology, 82*(1), 135–140.

Rieber, L. P. (1991a). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology*, *83*(3), 318–328.

Rieber, L. P. (1991b). Effects of visual grouping strategies of computer-animated presentations on selective attention in science. *Educational Technology Research and Development*, *39*(4), 5–15.

Rieber, L. P. (1996). Animation as feedback in a computer-based simulation: Representation matters. *Educational Technology Research and Development, 44*(1), 5–22.

Rieber, L. P., Boyce, M. J., & Assad, C. (1990). The effects of computer animation on adult learning and retrieval tasks. *Journal of Computer-Based Instruction, 17*(2), 46–52.

Rizzolatti, G. (2005). The mirror neuron system and its function in humans. *Anatomy and Embryology, 210*, 419–421. doi: 10.1007/s00429-005-0039-z

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*, 169–192. doi: 10.1146/annurev.neuro.27.070203.144230

Rizzolatti, G., & Sinigaglia, C. (2010). The functional role of the parieto-frontal mirror circuit: Interpretations and misinterpretations. *Nature Reviews Neuroscience 11*, 264–274. doi: 10.1038/nrn2805

Robin, J., Hirshhorn, M., Rosenbaum, R. S., Winocur, G., Moscovitch, M., & Grady, C. L. (2015). Functional connectivity of hippocampal and prefrontal networks during episodic and spatial memory based on real-world environments. *Hippocampus, 25*, 81–93. doi: 10.1002/hipo.22352

Roth, W.-M. (2000). From gesture to scientific language. *Journal of Pragmatics, 32*(11), 1683–1714. doi: 10.1016/S0378-2166(99)00115-0

Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognise designers' styles. *Learning and Instruction, 19*(2), 185–199. doi: 10.1016/j.learninstruc.2008.03.006

Sánchez, C. A., & Wiley, J. (2010). Sex differences in science learning: Closing the gap through animations. *Learning and Individual Differences, 20*, 271–275.

Sandstrom, N. J., Kaufman, J., & Huettel, S. A. (1998). Males and females use different distal cues in a virtual environment navigation task. *Cognitive Brain Research, 6*(4), 351–360. doi: 10.1016/S0926-6410(98)00002-0

Sanes, J. N., & Donoghue, J. P. (2000). Plasticity and primary motor cortex. *Annual Review of Neuroscience, 23*(1), 393–415. doi: doi:10.1146/annurev.neuro.23.1.393

Sanger, M. J., Brecheisen, D. M., & Hynek, B. M. (2001). Can computer animations affect college biology students' conceptions about diffusion & osmosis? *The American Biology Teacher, 63*(2), 104–109. doi: 10.2307/4451051

Sanger, M. J., & Greenbowe, T. J. (1997). Students’ misconceptions in electrochemistry: Current flow in electrolyte solutions and the salt bridge. *Journal of Chemical Education, 74*(7), 819–823.

Sanger, M. J., Phelps, A. J., & Fienhold, J. (2000). Using a computer animation to improve students’ conceptual understanding of a can-crushing demonstration. *Journal of Chemical Education, 77*(11), 1517–1520.

Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: Visualizing mathematical solution procedures. *Computers in Human Behavior, 22*, 9–25. doi: 10.1016/j.chb.2005.01.009

Scheiter, K., Gerjets, P., Huk, T., Imhof, B., & Kammerer, Y. (2009). The effects of realism in learning with dynamic visualizations. *Learning and Instruction, 19*, 481–494. doi: 10.1016/j.learninstruc.2008.08.001

Schöning, S., Engelien, A., Kugel, H., Schäfer, S., Schiffbauer, H., Zwitserlood, P., . . . Konrad, C. (2007). Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia, 45*, 3203–3214.

Schram, C. M. (1996). A meta-analysis of gender differences in applied statistics achievement. *Journal of Educational and Behavioral Statistics, 21*(1), 55–70. doi: 10.3102/10769986021001055

Schwan, S., & Riempp, R. (2004). The cognitive benefits of interactive videos: learning to tie nautical knots. *Learning and Instruction, 14,* 293–305. doi: 10.1016/j.learninstruc.2004.06.005

Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology, 23*(9), 1227–1243. doi: 10.1002/acp.1526

Schworm, S., & Renkl, A. (2006). Computer-supported example-based learning: When instructional explanations reduce self-explanations. *Computers & Education, 46*(4), 426–445. doi: 10.1016/j.compedu.2004.08.011

Seidel, T., Blomberg, G., & Renkl, A. (2013). Instructional strategies for using video in teacher education. *Teaching and Teacher Education, 34,* 56–65. doi: 10.1016/j.tate.2013.03.004

Shepard, R. N., & Judd, S. A. (1975). Perceptual illusion of rotation of three-dimensional objects. *Science, 191*, 952–952.

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

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review, 76*(2), 179–193. doi: 10.1037/h0027277

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84*(2), 127–190.

Silverman, I., Choi, J., & Peters, M. (2007). The Hunter-Gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior, 36*(2), 261–268. doi: 10.1007/s10508-006-9168-6

Silverman, I., & Eals, M. (1992). Sex differences in spatial abilities: Evolutionary theory and data. In J. H. Barkow, L. Cosmides & J. Tooby (Eds.), *The adapted mind: Evolutionary psychology and the generation of culture* (pp. 533–549). New York, NY: Oxford Univerisity Press

Silverman, I., & Phillips, K. (1993). Effects of estrogen changes during the menstrual cycle on spatial performance. *Ethology and Sociobiology, 14*(4), 257–269. doi: 10.1016/0162-3095(93)90021-9

Simon, H. A., & Newell, A. (1971). Human problem solving: The state of the theory in 1970. *American Psychologist, 26*(2), 145–159. doi: 10.1037/h0030806

Singh, A.-M., Marcus, N., & Ayres, P. (2012). The transient information effect: Investigating the impact of segmentation on spoken and written text. *Applied Cognitive Psychology, 26*, 848–853. doi: 10.1002/acp.2885

Smirni, P., Villardita, C., & Zappalá, G. (1983). Influence of different paths on spatial memory performance in the block-tapping test. *Journal of Clinical Neuropsychology, 5*(4), 355–359. doi: 10.1080/01688638308401184

Smith, S. N., & Miller, R. J. (2005). Learning approaches: Examination type, discipline of study and gender. *Educaional Psychology, 25*(1), 43–53. doi: 10.1080/0144341042000294886

So, W. C., Shum, P. L. C., & Wong, M. K. Y. (2015). Gesture is more effective than spatial language in encoding spatial information. *The Quarterly Journal of Experimental Psychology, 68*(12), 2384–2401. doi: 10.1080/17470218.2015.1015431

Spanjers, I. A. E., Wouters, P., van Gog, T., & van Merriënboer, J. J. G. (2011). An Expertise reversal effect of segmentation in learning from animated worked-out examples. *Computers in Human Behavior, 27*(1), 46–52. doi: 10.1016/j.chb.2010.05.011

Speck, O., Ernst, T., Braun, J., Koch, C., Miller, E., & Chang, L. (2000). Gender differences in the functional organization of the brain for working memory. *NeuroReport, 11*(11), 2581–2585.

Sperling, R. A., Seyedmonir, M., Aleksic, M., & Meadows, G. (2003). Animations as learning tools in authentic science materials. *International Journal of Instructional Media, 30*(2), 213–221.

Speth, C., & Brown, R. (1990). Effects of college students' learning styles and gender on their test preparation strategies. *Applied Cognitive Psychology, 4*(3), 189–202. doi: 10.1002/acp.2350040304

Stern, L., Barnea, N., & Shauli, S. (2008). The Effect of a computerized simulation on middle school students' understanding of the kinetic molecular theory. *Journal of Science Education and Technology, 17*, 305–315. doi: 10.1007/s10956-008-9100-z

Stevanoni, E., & Salmon, K. (2005). Giving memory a hand: Instructing children to gesture enhances their event recall. *Journal of Nonverbal Behavior, 29*(4), 217–233. doi: 10.1007/s10919-005-7721-y

Sweller, J. (1988). Cognitive load during problem solving effects on learning. *Cognitive Science, 12*(2), 257–285. doi: 10.1207/s15516709cog1202_4

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. doi: 10.1016/0959-4752(94)90003-5

Sweller, J. (2003). Evolution of human cognitive architecture. In B. Ross (Ed.), *Psychology of learning and motivation* (Vol. 43, pp. 215–266). San Diego: Academic Press

Sweller, J. (2004). Instructional design consequences of an analogy between evolution by natural selection and human cognitive architecture. *Instructional Science, 32*(1), 9–31. doi: 10.1023/B:TRUC.0000021808.72598.4d

Sweller, J. (2008). Instructional implications of David C. Geary's evolutionary educational psychology. *Educational Psychologist, 43*(4), 214–216. doi: 10.1080/00461520802392208

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review, 22*, 123–138. doi: 10.1007/s10648-010-9128-5

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New Yok, NY: Springer.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction, 12*(3), 185–233.

Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction, 2*(1), 59–89.

Sweller, J., & Sweller, S. (2006). Natural information processing systems. *Evolutionary Psychology, 4*(1), 434–458. doi: 10.1177/147470490600400135

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296.

Tarmizi, R. A., & Sweller, J. (1988). Guidance during mathematical problem solving. *Journal of Educational Psychology, 80*(4), 424–436. doi: 10.1037/0022-0663.80.4.424

Terzis, V., & Economides, A. A. (2011). Computer based assessment: Gender differences in perceptions and acceptance. *Computers in Human Behavior, 27*, 2108–2122. doi: 10.1016/j.chb.2011.06.005

Thomas, H., & Jamison, W. (1975). On the acquisition of understanding that still water is horizontal. *Merrill-Palmer Quarterly of Behavior and Development, 21*(1), 31–44.

Thompson, S. V., & Riding, R. J. (1990). The effect of animated diagrams on the understanding of a mathematical demonstration in 11- to 14-year-old pupils. *British Journal of Educational Psychology, 60*(1), 93–98. doi: 10.1111/j.2044-8279.1990.tb00925.x

Togo, D. F., & Hood, J. N. (1992). Quantitative information presentation and gender: An interaction effect. *The Journal of General Psychology, 119*(2), 161–167. doi: 10.1080/00221309.1992.9921169

Treisman, A. M. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*(4), 242–248. doi: 10.1080/17470216008416732

Trejo, L. J., Kramer, A. F., & Arnold, J. A. (1994). Event-related potentials as indices of displaymonitoring performance. *Biological Psychology, 40*, 33–71.

Tversky, B., Morrison, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262.

Ungerleider, L. G., Doyon, J., & Karni, A. (2002). Imaging brain plasticity during motor skill learning. *Neurobiology of Learning and Memory, 78*, 553–564. doi: doi:10.1006/nlme.2002.4091

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*(3), 498–505. doi: 10.3758/BF03192720

van Gerven, P. W. M., Paas, F., van Merriënboer, J. J. G., Hendriks, M., & Schmidt, H. G. (2003). The efficiency of multimedia learning into old age. *British Journal of Educational Psychology, 73*(4), 489–505. doi: 10.1348/000709903322591208

van Gog, T., & Paas, F. (2008). Instructional efficiency: Revisiting the original construct in educational research. *Educational Psychologist, 43*(1), 16–26. doi: 10.1080/00461520701756248

van Gog, T., Paas, F., Marcus, N., Ayres, P., & Sweller, J. (2009). The mirror neuron system and observational learning: Implications for the effectiveness of dynamic visualizations. *Educational Psychology Review, 21*, 21–30. doi: 10.1007/s10648-008-9094-3

van Gog, T., & Scheiter, K. (2010). Eye tracking as a tool to study and enhance multimedia learning. *Learning and Instruction, 20*(2), 95–99. doi: 10.1016/j.learninstruc.2009.02.009

van Goozen, S. H. M., Slabbekoorn, D., Gooren, L. J. G., Sanders, G., & Cohen-Kettenis, P. T. (2002). Organizing and activating effects of sex hormones in homosexual transsexuals. *Behavioral neuroscience, 116*(6), 982–988. doi: 10.1037/0735-7044.116.6.982

van Merriënboer, J. J. G., Schuurman, J. G., de Croock, M. B. M., & Paas, F. (2002). Redirecting learners' attention during training: Effects on cognitive load, transfer test performance and training efficiency. *Learning and Instruction, 12*(1), 11–37. doi: 10.1016/S0959-4752(01)00020-2

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.

Vander Wyk, B. C., Hudac, C. M., Carter, E. J., Sobel, D. M., & Pelphrey, K. A. (2009). Action understanding in the superior temporal sulcus region. *Psychological Science, 20*(6), 771–777.

Vandierendonck, A., Kemps, E., Fastame, M. C., & Szmalec, A. (2004). Working memory components of the Corsi blocks task. *British Journal of Psychology, 95*, 57–79.

Vecchi, T., & Richardson, J. T. E. (2001). Measures of Visuospatial Short-Term Memory: The Knox Cube Imitation Test and the Corsi Blocks Test Compared. *Brain and Cognition, 46*, 291–295.

Vogel, S. A. (1990). Gender differences in intelligence, language, visual-motor abilities, and academic achievement in students with learning disabilities: A review of the literature. *Journal of Learning Disabilities, 23*(1), 44–52. doi: 10.1177/002221949002300111

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin 117*(2), 250–270.

Wagner, S. M., Nusbaum, H., & Goldin-Meadow, S. (2004). Probing the mental representation of gesture: Is handwaing spatial. *Journal of Memory and Language, 50*, 395–407. doi: 10.1016/j.jml.2004.01.002

Watson, G., Butterfield, J., Curran, R., & Craig, C. (2010). Do dynamic work instructions provide an advantage over static instructions in a small scale assembly task? *Learning and Instruction, 20*(1), 84–93. doi: 10.1016/j.learninstruc.2009.05.001

Weiss, E., Siedentopf, C. M., Hofer, A., Deisenhammer, E. A., Hoptman, M. J., Kremser, C., . . . Delazer, M. (2003). Sex differences in brain activation pattern during a visuospatial cognitive task: a functional magnetic resonance imaging study in healthy volunteers. *Neuroscience Letters, 344*(3), 169–172. doi: 10.1016/S0304-3940(03)00406-3

Wender, K. F., & Muehlboeck, J. S. (2003). Animated diagrams in teaching statistics. *Behaviour Research Methods, Instruments, & Computers, 35*(2), 255–258.

Whelan, R. R. (2007). Neuroimaging of cognitive load in instructional multimedia. *Educational Research Review, 2*(1), 1-12. doi: 10.1016/j.edurev.2006.11.001

Willemsen, E., & Reynolds, B. (1973). Sex differences in adults' judgements of the horizontal. *Developmental Psychology, 8*(2), 309.

Williamson, V. M., & Abraham, M. R. (1995). The Effects of computer animation on the particulate mental models of college chemistry students. *Journal of Research in Science Teaching, 32*(5), 521–534.

Willis, S. L., & Schaie, K. W. (1986). Training the elderly on the ability factors of spatial orientation and inductive reasoning. *Psychology and Aging, 1*(3), 239–247.

Wittig, M. A., & Allen, M. J. (1984). Measurement of adult performance on Piaget's water horizontality task. *Intelligence, 8*(4), 305–313. doi: 10.1016/0160-2896(84)90014-X

Woehlike, P. L., & Leitner, D. W. (1980). Gender differences in performance on variables related to achievement in graduate-level educational statistics. *Psychological Reports, 47*, 1119–1125.

Wong, A., Leahy, W., Marcus, N., & Sweller, J. (2012). Cognitive load theory, the transient information effect and e-learning. *Learning and Instruction 22*(6), 449–457. doi: 10.1016/j.learninstruc.2012.05.004

Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G. A., Paas, F., & Sweller, J. (2009). Instructional animations can be superior to statics when learning human motor skills. *Computers in Human Behavior, 25,* 339–347.

Wouters, P., Paas, F., & van Merriënboer, J. J. G. (2008). How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research, 78*(3), 645–675. doi: 10.3102/0034654308320320

Yarden, H., & Yarden, A. (2011). Studying biotechnological methods using animations: The teacher's role. *Journal of Science Education and Technology, 20*(6), 689–702. doi: 10.1007/s10956-010-9262-3

Yezierski, E. J., & Birk, J. P. (2006). Misconceptions about the particulate nature of matter: Using animations to close the gender gap. *Journal of Chemical Education, 83*(6). doi: 10.1021/ed083p954

Appendices

Appendix 1: Participants Questionnaire (survey of background knowledge and cognitive load measure) used in Experiments 1, 2 and 3.

Participant No.: _____

Students' self-perceptual questionnaire
School of Education, UNSW

Please fill in the blank or put a "✓" on the appropriate box to indicate your choice.

Please complete this part BEFORE the experiment
Personal Information
Gender: M ☐ F ☐ Programme: ______________________________ Year of Study: ____________
Age:________________ Level of Study: Undergraduate / Research postgraduate / Coursework postgraduate
Handedness: Left-handed ☐ Right-handed ☐

3. How would you rate your *mental rotation* ability (ie. to rotate or flip 2-D or 3-D shape mentally)?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

4. In general, how would you rate your *overall spatial ability*?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

Please complete this part DURING the experiment
5. How much mental effort did you spend in completing…

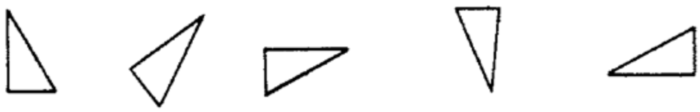
	Little			Fair					Heavy
a. Practice Task	☐	☐	☐	☐	☐	☐	☐	☐	☐
b. Task 1	☐	☐	☐	☐	☐	☐	☐	☐	☐
c. Task 2	☐	☐	☐	☐	☐	☐	☐	☐	☐
d. Task 3	☐	☐	☐	☐	☐	☐	☐	☐	☐

Please complete this part AFTER the experiment
Will you be interested in participating in other research studies from School of Education, UNSW in the future?
Yes ☐ I haven't decided yet ☐ No ☐

Appendix 2: Card Rotational Test used in Experiments 1, 2 and 3

CARD ROTATIONS TEST — S-1 (Rev.)

This is a test of your ability to see differences in figures. Look at the 5 triangle-shaped cards drawn below.



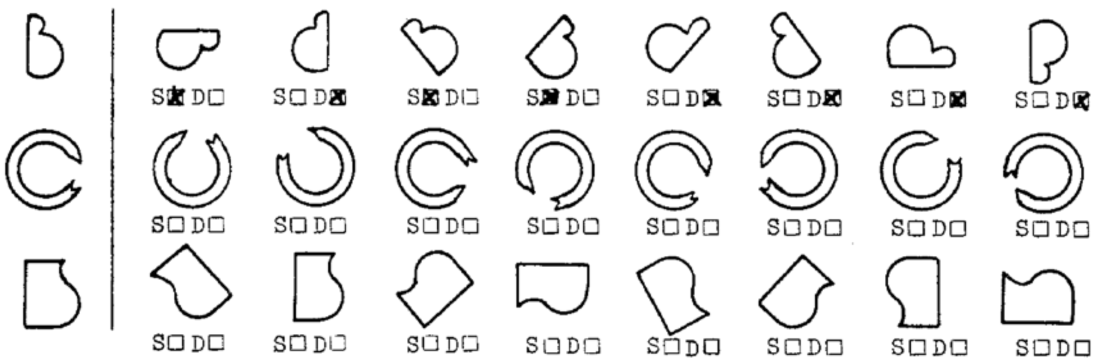
All of these drawings are of the same card, which has been slid around into different positions on the page.

Now look at the 2 cards below:

These two cards are not alike. The first cannot be made to look like the second by sliding it around on the page. It would have to be flipped over or made differently.

Each problem in this test consists of one card on the left of a vertical line and eight cards on the right. You are to decide whether each of the eight cards on the right is the same as or different from the card at the left. Mark the box beside the S if it is the same as the one at the beginning of the row. Mark the box beside the D if it is different from the one at the beginning of the row.

Practice on the following rows. The first row has been correctly marked for you.

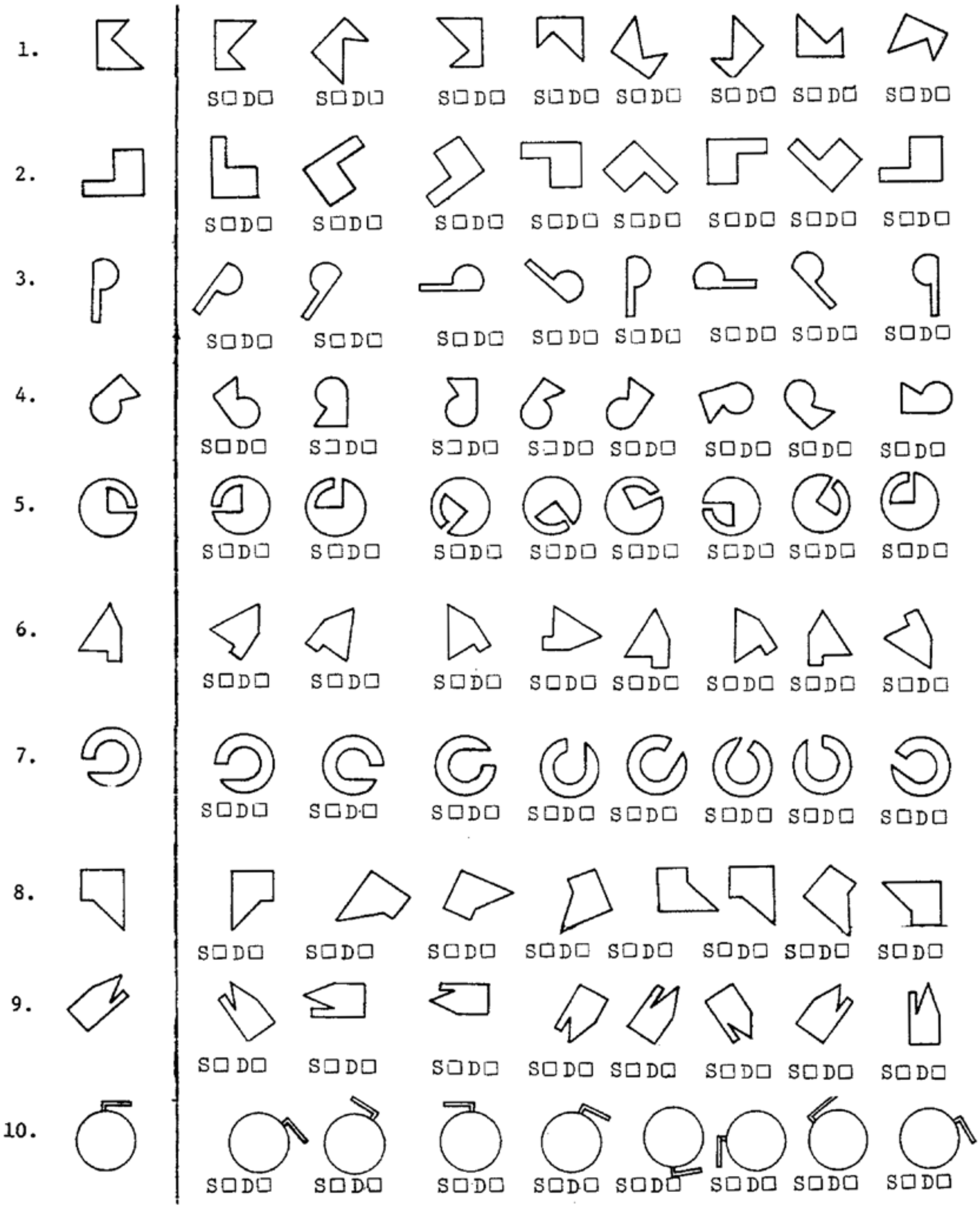


Your score on this test will be the number of items answered correctly minus the number answered incorrectly. Therefore, it will not be to your advantage to guess, unless you have some idea whether the card is the same or different. Work as quickly as you can without sacrificing accuracy.

You will have 3 minutes for each of the two parts of this test. Each part has 1 page. When you have finished Part 1, STOP. Please do not go on to Part 2 until you are asked to do so.

DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

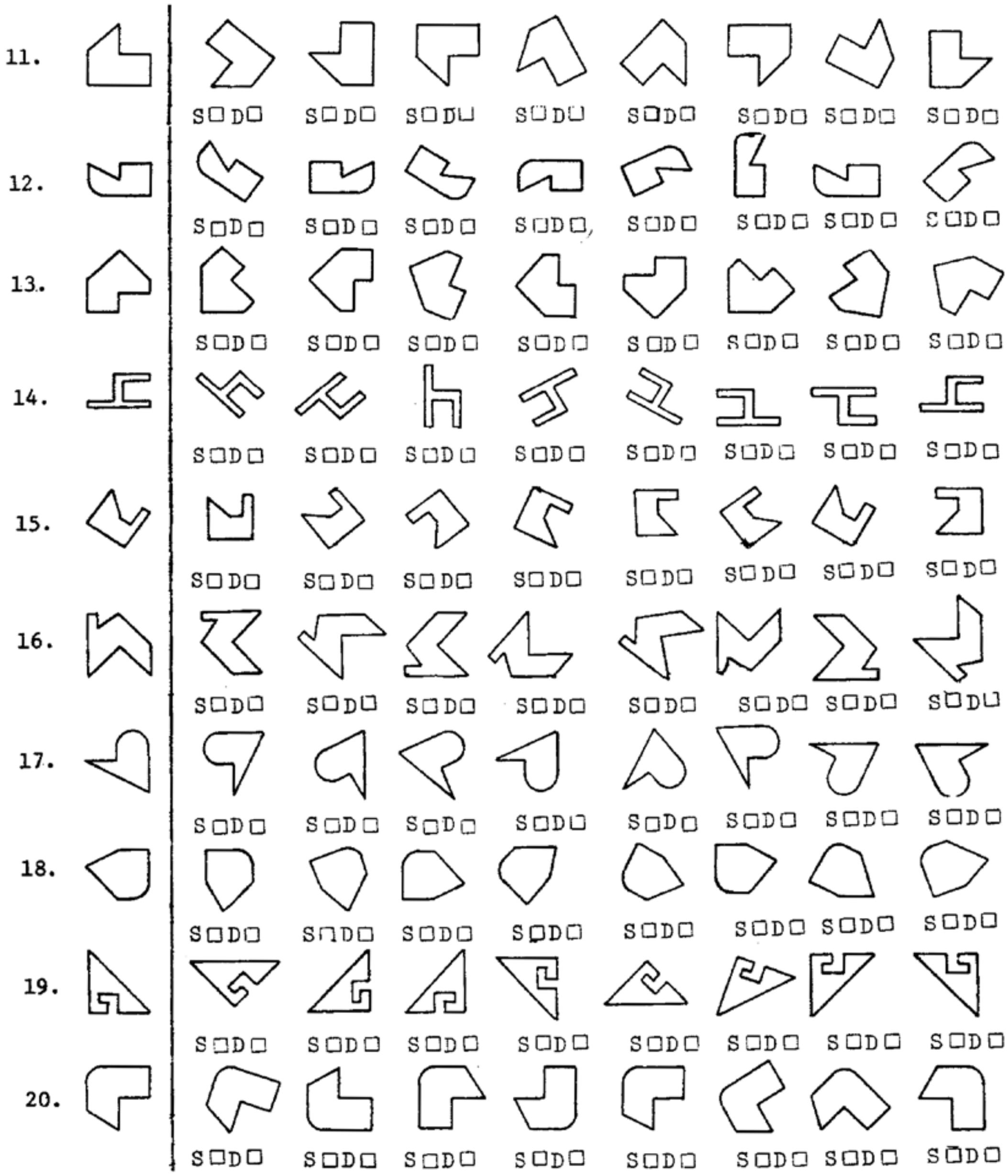
Part 1 (3 minutes)



DO NOT TURN THIS PAGE UNTIL ASKED TO DO SO.

STOP.

Part 2 (3 minutes)



STOP.

Appendix 3: Participants' Experiment Consent Form (used in Experiments 1 and 2)

UNSW

School of Education

School of Education, Faculty of Arts and Social Sciences

Approval No: 12 101

THE UNIVERSITY OF NEW SOUTH WALES

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM

The use of instructional animation: A cognitive load approach

Participant selection and purpose of study

You are invited to participate in a research study related to how instructional animation design can improve learners' effectiveness. I, Miss Mona Wong Pui Shan (research student from the School of Education, UNSW) hope to examine how animations can be improved to foster the learning of fine motor-tasks. You were invited as a possible participant in this study as you are a current student at the university.

Description of study and risks

If you decide to participate, you will need to fill in a self-perceptual questionnaire and compete two sets of spatial-ability tests. The questionnaire and test should take approximately 15 minutes to complete. You will then be presented with a video-recording showing how to construct a LEGO-shape, which you will be required to build. The task should take approximately 30-35 minutes to complete. The accuracy of the construction will be evaluated and thus please pay attention when watching the video. We cannot and do not guarantee or promise that you will receive any benefits from the study.

Confidentiality and disclosure of information

Participation in the study is anonymous and no information identifying you will be collected. The construction procedure (no more than the bricks and part of your hand) will be video-recorded for analysis purposes only with your consent. If you give us your permission by signing this document, the overall results will be used in a research postgraduate dissertation and, perhaps, in academic journal articles or conferences. In any publication, information will be provided in such a way that you cannot be identified.

Recompense to participants

Upon finishing the experiment, you will be given $20 Westfield voucher to compensate the time participating in the experiment.

Complaints may be directed to the Ethics Secretariat, The University of New South Wales, SYDNEY 2052 AUSTRALIA (phone 9385 4234, fax 9385 6648, email ethics.sec@unsw.edu.au). Any complaint you make will be investigated promptly and you will be informed out the outcome.

Feedback to participants

If you are interested in the project, please feel free to contact Mona Wong (p.wong@student.unsw.edu.au) for a summary of research findings at the completion of the study.

Your consent

Your decision whether or not to participate will not prejudice your future relations with the University of New South Wales. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have addition questions, please feel free to contact Prof Ayres (p.ayres@unsw.edu.au) or Prof Kalyuga (s.kalyuga@unsw.edu.au)

You will be given a copy of this form to keep.

THE UNIVERSITY OF NEW SOUTH WALES

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM (continued)

The use of instructional animation: A cognitive load approach

You are making a decision whether or not to participate. Your signature indicates that, having read the information provided above, you have decided to participate.

....	....
Signature of Research Participant	Signature of Witness
....	....
(Please PRINT name)	(Please PRINT name)
....	....
Date	Nature of Witness

REVOCATION OF CONSENT

The use of instructional animation: A cognitive load approach

I hereby wish to **WITHDRAW** my consent to participate in the research proposal described above and understand that such withdrawal **WILL NOT** jeopardise any treatment or my relationship with The University of New South Wales.

....	....
Signature	Date

....
Please PRINT Name

The section for Revocation of Consent should be forwarded to:
The University of New South Wales,
Sydney, NSW 2052, Australia.

c/o:

Prof. Paul Ayres Room 107, Goodsell Building, UNSW	Prof. Slava Kalyuga Room 105, Goodsell Building, UNSW

Appendix 4: Participants' Experiment Consent Form (used in Experiments 3 and 4)

UNSW

School of Education

School of Education, Faculty of Arts and Social Sciences

Approval No: 14 041

THE UNIVERSITY OF NEW SOUTH WALES

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM

The use of instructional animation: A cognitive load approach

Participant selection and purpose of study

You are invited to participate in a research study related to how instructional animation design can improve learners' effectiveness. I, Miss Mona Wong Pui Shan (a PhD candidate from the School of Education, UNSW) hope to examine how animations can be improved to foster the learning of fine motor-tasks. You were invited as a possible participant in this study as you are a current student at the university.

Description of study and risks

If you decide to participate, you will need to fill in a short questionnaire and compete a set of spatial-ability tests, which take approximately 10 minutes. You will then be presented with a short clip showing a construction procedure, which you will be required to memorise, and then to build. The task should take approximately 30 minutes to complete. The accuracy of the construction will be evaluated. We cannot and do not guarantee or promise that you will receive any benefits from the study.

Confidentiality and disclosure of information

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission, except as required by law. If you give us your permission by signing this document, the overall results will be used in a research postgraduate dissertation and, perhaps, in academic journal articles or conferences. In any publication, information will be provided in such a way that you cannot be identified.

Recompense to participants

Upon finishing the experiment, you will be given $20 Westfield voucher to compensate the time participating in the experiment.

Complaints may be directed to the Ethics Secretariat, The University of New South Wales, SYDNEY 2052 AUSTRALIA (phone 9385 4234, fax 9385 6648, email ethics.sec@unsw.edu.au). Any complaint you make will be investigated promptly and you will be informed out the outcome.

Feedback to participants

If you are interested in the project, please feel free to contact Mona Wong (p.wong@student.unsw.edu.au) for a summary of research findings at the completion of the study.

Your consent

Your decision whether or not to participate will not prejudice your future relations with the University of New South Wales. If you decide to participate, you are free to withdraw your consent and to discontinue participation at any time without prejudice.

If you have addition questions, please feel free to contact Prof Ayres (p.ayres@unsw.edu.au) or Prof Kalyuga (s.kalyuga@unsw.edu.au)

You will be given a copy of this form to keep.

THE UNIVERSITY OF NEW SOUTH WALES

PARTICIPANT INFORMATION STATEMENT AND CONSENT FORM (continued)

The use of instructional animation: A cognitive load approach

You are making a decision whether or not to participate. Your signature indicates that, having read the information provided above, you have decided to participate.

....	....
Signature of Research Participant	Signature of Witness
....	....
(Please PRINT name)	(Please PRINT name)
....	....
Date	Nature of Witness

REVOCATION OF CONSENT

The use of instructional animation: A cognitive load approach

I hereby wish to **WITHDRAW** my consent to participate in the research proposal described above and understand that such withdrawal **WILL NOT** jeopardise any treatment or my relationship with The University of New South Wales.

....	....
Signature	Date

....
Please PRINT Name

The section for Revocation of Consent should be forwarded to:
The University of New South Wales,
Sydney, NSW 2052, Australia.

c/o:

Prof. Paul Ayres Room 107, Goodsell Building, UNSW	Prof. Slava Kalyuga Room 105, Goodsell Building, UNSW

Appendix 5: Participants Questionnaire used in Experiment 4

Participant questionnaire
School of Education, UNSW

Please fill in the blank or put a "✓" in the appropriate box to indicate your choice.

Personal Information
Gender: M ☐ F ☐ Programme: ____________________________ Year of Study: _____________
Age:________________ Level of Study: Undergraduate / Research postgraduate / Coursework postgraduate
Handedness: Left-handed ☐ Right-handed ☐

1. How would you rate your mental rotation ability (ie. to rotate or flip shapes mentally)?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

2. How would you rate your *overall* spatial ability?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

3. How much experience have you had playing with Lego (or similar) bricks?
None ☐ Some ☐ Quite a bit ☐ Much ☐ Very Much ☐

4. How would you consider your ability to build shapes with Lego (or similar) bricks?
Very Weak ☐ Weak ☐ Fair ☐ Good ☐ Very Good ☐

5. When learning motor-related task (e.g. tying knots, folding paper…etc), how often did you learn it from animation/ video?
Never ☐ Rarely ☐ Occasionally ☐ Frequently ☐ Very Frequently ☐

6. When learning motor-related task (e.g. tying knots, folding paper…etc), how often did you learn it from pictures/ books?
Never ☐ Rarely ☐ Occasionally ☐ Frequently ☐ Very Frequently ☐

7. All of the following 8 questions refer to the activity that just finished. Please take your time to read each of the questions carefully and respond to each of the questions by ticking "✓" the most appropriate box.

	Not at all the case									*Completely the case*
a) The content of the video (building a Lego shape) was very complex	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
b) The problem (this particular Lego structure) covered in the video was very complex	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
c) In the video, very complex concepts were involved.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
d) I invested a very high mental effort in the complexity of this activity.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
e) The instructions in the video were very unclear.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
f) The instructions in the video were full of unnecessary components.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
g) The instructions in the video were, in terms of learning, very ineffective.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
h) I invested a very high mental effort in unclear and ineffective instructions in this activity.	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐