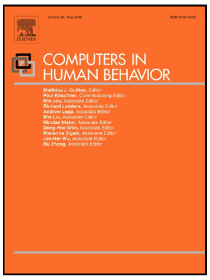
Accepted Manuscript



Investigating gender and spatial measurements in instructional animation research

Mona Wong, Juan C. Castro-Alonso, Paul Ayres, Fred Paas

PII: S0747-5632(18)30075-X

DOI: 10.1016/j.chb.2018.02.017

Reference: CHB 5380

To appear in: *Computers in Human Behavior*

Received Date: 10 March 2017

Revised Date: 31 January 2018

Accepted Date: 15 February 2018

Please cite this article as: Mona Wong, Juan C. Castro-Alonso, Paul Ayres, Fred Paas, Investigating gender and spatial measurements in instructional animation research, *Computers in Human Behavior* (2018), doi: 10.1016/j.chb.2018.02.017

Investigating gender and spatial measurements in instructional animation research

Mona Wong[a], Juan C. Castro-Alonso[b], Paul Ayres[c], and Fred Paas[d, e]

[a]Facutly of Education, The University of Hong Kong, Hong Kong
[b]Center for Advanced Research in Education (CIAE), Universidad de Chile, Chile
[c]School of Education, The University of New South Wales, Australia
[d]Department of Psychology, Education and Child Studies, Erasmus University Rotterdam, The Netherlands
[e]Early Start Research Institute, University of Wollongong, Australia

Corresponding author

Correspondence concerning this manuscript should be addressed to Mona Wong (wpsmona@hku.hk)

Funding: This work was supported by an Australian Research Council Discovery Grant [DP140103307] granted to Paul Ayres and Fred Paas

Abstract

Instructional animation research has been extensive but the results are inconsistent. Amongst a number of possible factors to explain these inconclusive results (e.g., the negative influence of transient information), the influence of spatial ability and gender are less explored. This paper reports three experiments that compared the effectiveness of learning a hand-manipulative task (Lego construction) under various conditions with direct examination of the relationship between gender, spatial ability and instructional visualisation. Regression analyses revealed that only one objective measure related to spatial ability (Corsi test) predicted overall test performance, whereas the Card Rotations Test and the Mental Rotations Test did not. However, there was a number of significant gender–spatial ability interactions showing that the spatial ability predictors of male performance were different from those of females. Furthermore a number of subjective measures of spatial ability and experience with instructional animations and static pictures were found to be significant predictors. The results suggest that gender and the type of spatial ability measures used both have a significant impact on gauging the effectiveness of instructional animations. Spatial ability measures should be tailored to gender and the specific nature of the learning domains to yield more consistent research results.

Key words: instructional animations, cognitive load theory, gender, spatial ability, visuospatial measures

1. Introduction

Animations, including all forms of dynamic visualisations such as videos and cartoons, can be more entertaining and engaging than equivalent static formats (Rieber, 1991a) and hence are often used for instructional purposes. Researchers believe that animations have a greater flexibility in depicting physical and temporal changes than static pictures (Bétrancourt & Tversky, 2000; Castro-Alonso, Ayres, & Paas, 2014a, p. 552; Lowe, 2003), and are often assumed to be superior to static pictures (Chandler, 2009). There is supporting evidence showing that animations are more effective than static pictures when teaching statistics (e.g., Wender & Muehlboeck, 2003), physics (e.g., Rebetez, Bétrancourt, Sangin, & Dillenbourg, 2010; Rieber, 1990, 1991b), geometry (e.g., Thompson & Riding, 1990), zoology (e.g., Pfeiffer, Scheiter, Kühl, & Gemballa, 2011), computer algorithm (e.g., Byrne, Catrambone, & Stasko, 1999), and motor-related tasks (e.g., Akinlofa, Holt, & Elyan, 2013; Ayres, Marcus, Chan, & Qian, 2009; Castro-Alonso, Ayres, & Paas, 2015a; Garland & Sánchez, 2013; Michas & Berry, 2000; A. Wong et al., 2009).

However, some contradictory results have emerged that challenge the effectiveness of animations (e.g., Bétrancourt, Dillenbourg, & Clavien, 2008; Höffler & Leutner, 2007). In some studies, animations have been found to be equivalent to or even inferior to static pictures. For example, 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. Castro-Alonso, Ayres, and Paas (2014b) conducted two experiments comparing animation and static pictures in memorising abstract symbol patterns. The first experiment found 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. Similarly, Mayer, Hegarty, Mayer, and Campbell (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 mechanisms, 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.

Although many reasons have been forwarded to explain these mixed results (for example, see Lowe, 2003), the main aim of the current paper is to examine the impact played by gender and the measurement of spatial ability. To set the groundwork for this focus, we initially explore the transient information effect and how animation-static outcomes are influenced by different types of knowledge.

1.1. The Transient Information Effect

It has been argued that the dynamic nature of animations produces *transient information* that can hinder their learning effectiveness (Ayres & Paas, 2007; Castro-Alonso, Ayres, & Paas, 2015b; Castro-Alonso, Ayres, Wong, & Paas, 2018; van Gog, Paas, Marcus, Ayres, & Sweller, 2009). The transient information effect is ‘a loss of learning due to information disappearing before the learner has time to adequately process it or link it with new information’ (Sweller, Ayres, & Kalyuga, 2011, p. 220). In the case of instructional animations, information on one frame will soon roll over to another frame and can be quickly lost. Learners, when learning with animations without learner-control, are required to interpret a larger amount of information within a designated time limit to attend to and identify the changes between frames, and to integrate new information (currently visible) with the old information (no longer visible) (Ayres & Paas, 2007; van Gog et al., 2009). According to cognitive load theory (see Sweller et al., 2011) such processing creates extraneous cognitive load (see Sweller, van Merriënboer, & Paas, 1998), which puts a heavy burden on our highly limited working memory. In other words, it takes precious working memory resources away from learning (schema acquisition) to deal with difficulties generated by the instructional design. Hence, learning is restricted. In comparison, a static picture may facilitate learning, because it is more permanent. This creates the possibility for re-inspection of the learning material, which reduces the need to keep everything in working memory, and consequently leaves more resources for learning.

1.2. Embodied Cognition and Observational Learning

Whereas transiency can create an unfavourable environment for learning many tasks, research also indicates that animations can be helpful in the learning of tasks that involve procedural-motor skills (Höffler & Leutner, 2007), arguably because of our innate ability for observational learning (Castro-Alonso et al., 2014a; Paas & Sweller, 2012). Humans are born with a mirror neuron system (Rizzolatti, 2005; Rizzolatti & Sinigaglia, 2010), which is a neurological mechanism assisting humans to effectively imitate actions through observations. Geary (2005, 2007, 2008) argued that, as a result of evolution, all organisms (including humans) have evolved to attend to and process movement patterns of prey and predators. It is essential to be able to learn and use movement quickly so as not to be eliminated as a species. Geary (2005) considers learning human movements as a form of *biologically primary knowledge*. Humans have evolved to learn such knowledge rather effortlessly and unconsciously without any explicit instructions. Consequently, when observing human movement either directly, or from an animation, the cognitive load is kept to a minimum when trying to reproduce the actions despite transitory information (Paas & Sweller, 2012; van Gog et al., 2009). Paas and Sweller (2012) refer to the phenomenon to learn human motor-skills from animations as the *human movement effect*. In contrast to primary knowledge, Geary (2005) refers to less evolved knowledge as *biologically secondary knowledge*. Virtually all knowledge we learn at educational institutions, such as science, is secondary knowledge, and requires considerable more conscious effort to learn. From the perspective of animations, the mirror neuron system is not helpful as most science topics are not based on human movement, and therefore animated transitory information is more problematical.

We have described how some of the mixed results of animation research can be explained by the transient information effect. If the task is learning a human motor skill then animations are likely to be helpful; however, if the task is not human movement, then animation may not provide an advantage. The transient information effect cannot fully explain all the discrepancies found in the literature, and it has been suggested that inconsistencies in the experimental methodologies might also be a significant factor (Castro-Alonso, Ayres, & Paas, 2016; Tversky, Morrison, & Bétrancourt, 2002). The following section examines some of these methodological issues.

1.3. Incomparable Methodology

During the meta-analysis of Höffler and Leutner (2007) 25 out of 57 animation studies were initially excluded, because: 1) animations were not compared with equivalent static pictures, 2) both types of visualisation were mixed together, or 3) interactivity was present with animations but not the static pictures. This cull suggested that nearly half of the animation studies identified had methodological issues. Tversky et al. (2002) also made the point that there was often an unequal amount of information shown between animations and static pictures. Recently, Castro-Alonso et al. (2016) identified seven different types of methodological issues in animation research, including appeal bias (e.g., comparing a colour animation to a black and white static image), variety bias (i.e., more visual elements, such as arrows, in the static and not the dynamic visualization), media bias (e.g., comparing static visualisations on paper to animated ones on-screen), realism bias (e.g., comparing realistic movies to abstract illustrations), number bias (i.e., number of images depicted is different in static and animated format), size bias (i.e. one of the visualisations is larger than the other) and interaction bias (i.e. comparing different types of user-interactivity). With such a number of biases present, there are considerable uncertainties surrounding the results from animation-static comparisons. We argue that this situation can be further complicated if gender and spatial ability are also not considered.

2. Gender, Spatial Ability and Animations

In regards to animation research, there is evidence that instructional animations can support females more than males (e.g., Falvo & Suits, 2009; Jacek, 1997; Sánchez & Wiley, 2010; M. Wong, Castro-Alonso, Ayres, & Paas, 2015; Yezierski & Birk, 2006). This potential gender effect has implications for research into animations because many studies have not included an equal number of males and females. Hence, studies completed with more females (those usually involving education and psychology students) may have a bias towards animations, and vice-versa if more males are included.

In learning from animations, spatial ability is an important skill. It is also important in understanding scientific and abstract concepts, such as orthographic projection (e.g., Pillay, 1994), mathematical problems (e.g., Hegarty & Kozhevnikov, 1999), and mechanical systems (e.g., Boucheix & Schneider, 2009; Hegarty, Kriz, & Cate, 2003; Hegarty & Waller, 2005). Studies have found that spatial ability is highly correlated with extracting conceptual knowledge and constructing mental animation, and consequently understanding from visual presentations (see, Hegarty et al., 2003; Hegarty, Montello, Richardson, Ishikawa, & Lovelace, 2006; Hegarty & Sims, 1994; Hegarty & Waller, 2005; Narayanan & Hegarty, 2002).

Research generally supports the findings that females have a lower spatial ability (in particular, mental rotation ability) than males (see Guillem & Mograss, 2005; Halpern, 2012, pp. 128–145; Maeda & Yoon, 2013; Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995), with some studies providing evidence in terms of differing underlying neural substrates (e.g., Koscik, O’Leary, Moser, Andreasen, & Nopoulos, 2009). It is often argued that males outperform females in mathematics and science because they have greater spatial ability that enables them to mentally visualise and understand mathematical and scientific concepts better, and also relate them to the real world more easily (e.g., Baenninger & Newcombe, 1995; Casey, Nuttall, Pezaris, & Benbow, 1995; Geary, 1996, 2004; Pearson & Ferguson, 1989; Zander, Wetzel, & Bertel, 2016). Because of differences in spatial ability, it has been argued that any advantage for females from instructional animations is due to their lower spatial ability (Falvo & Suits, 2009; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006), as animations generally benefit learners with low spatial ability (Höffler, 2010). Although, such an argument is not completely supported in animation research, as gender has not always been included directly as a variable in some studies, or directly linked with spatial ability or performance (e.g., Falvo & Suits, 2009; Sánchez & Wiley, 2010). Consequently, there is a lack of rigorous investigation linking the effectiveness of animations with gender and spatial ability measures.

2.1. Differences in Spatial Abilities According to Gender

The Theory of Evolution by Darwin (1871) provides some insights into gender differences in spatial ability. Silverman and Eals (1992) have proposed the *Hunter-Gatherer theory* of spatial sex differences, with the belief that sexual division of labour between hunting and gathering was a critical *natural selectio*n process. The theory suggested that males and females have evolved to have different capabilities for their respective roles. Male-biased spatial ability (e.g., mental rotations, maze learning…etc.) involved abilities to orient themselves in relation to objects across distances, and enables males to fulfil the hunter role. Female-biased spatial ability (e.g. memorising object locations and content) comprised of the ability to recognise the spatial configuration and location, and enables them to fulfil the foraging role.

Geary (1995) applied the evolution theory into education research and argued that the gender difference in favour of males in 3-D spatial ability was a result of *sexual selection*, ‘directly related to intra-male competition and courtship of females’ (Geary, 1995, p. 291). Males are innately better at locomotion in order to ‘find’ the females. Moreover, since humans' neurocognitive systems that support habitat representations have evolved in 3-D but not 2-D worlds, the gender difference is expected to be smaller or non-existent when processing 2-D information (Geary, 1995, p. 291).

Besides the evolutionary explanation, other researchers (e.g. A. Martens, Johns, Greenberg, & Schimel, 2006; McGlone & Aronson, 2006; Wraga, Duncan, Jacobs, Helt, & Church, 2006) have used a social psychological approach to explain gender differences in spatial ability, arguing that a stereotype threat affects females' spatial performance. In general, females are stigmatised as having lower spatial ability and this conception may influence their performance. McGlone and Aronson (2006) examined the stereotype threat effect on visuospatial ability using different social identity primes. The results showed that both males and females were sensitive to the gender prime creating a gender gap in spatial performance. Wraga et al. (2006) obtained a similar result from three experiments wherein the same stereotype and learning materials were examined on different group of participants. The authors concluded that the mental rotation ability, to a certain extent, “is not attributable solely to biological factors, but is also susceptible to environmental influences” (p. 817).

2.2. Differences in Spatial Ability Definition

While spatial ability plays an important role in learning with visualisations and gender differences, definitions of spatial ability tend to 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* (to determine the spatial relationships with respect to the orientation of oneself), *mental rotation* (mentally rotate a 2- or 3-D figure), and *spatial visualisation* (to manipulate spatial information which requires both spatial perception and mental rotation). However Coluccia and Louse (2004) and J. Martens and Antonenko (2012) argued that there was a distinctive concept namely *spatial orientation ability*, which is often confused with spatial ability. Unlike spatial ability, which focuses on determining spatial information, spatial orientation ability involves movement, spatial environment and the acquisition of knowledge about the environment.

In addition to understanding spatial performance as a result of a general spatial ability, some researchers (e.g., Darling, Sala, Logie, & Cantagallo, 2006; de Beni, Pazzaglia, Gyselinck, & Meneghetti, 2005; Logie, 1995) have argued that spatial performance is highly related to the visuospatial working memory capacity. Based on Baddeley and Hitch's working memory model (for details, see Baddeley, 2000), it is believed that the visuospatial sketchpad within working memory is responsible for processing spatial information (both visual and spatial coding, see Farah, 1988). A number of studies (Choi & L'Hirondelle, 2005; de Beni et al., 2005; Duff & Hampson, 2001; McConnell & Quinn, 2000, 2004; Pazzaglia & Cornoldi, 1999; Quinn, 1994, 2008; Quinn & McConnell, 1996; Quinn & Ralston, 1986) have identified the actual visuospatial working memory mechanism by examining the association or dissociation between visuospatial working memory and other working memory components (e.g., verbal or movement).

2.3. Different Ways to Measure Different Types of Spatial Ability

In considering the important role played by spatial ability in learning with visualisations, it is important to have a reliable measure. Based on different definitions and the various cognitive processes involved, a number of different psychometric measures have been developed. Table 1 briefly summarises some commonly used spatial measurements, taken from many of the articles discussed above. It is clear that different researchers have used a variety of categorisations and terms for spatial ability in their original papers.

Table 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)
	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	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	Sala, Gray, Baddeley, and Wilson (1997), 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)

3. Main Aims of the Present Study

As argued above, there are many explanations for the mixed results of comparing instructional animations with static pictures. For example, results can be influenced by the learning topic (human movement tasks compared with non-human movement tasks), or methodological issues such as a lack of equivalence between the two formats. There is also a real issue associated with the gender of participants, if equal gender samples are not included. Furthermore, because of the impact played by spatial ability, to draw meaningful conclusions from animation-static research, it must be accurately measured.

To help provide some clarity to this complex area of research, the findings from three experiments were re-analysed. Two of the experiments (described as Experiment 1 and 2 below) were previously published outlining animation-static interactions with gender (M. Wong et al., 2015). In all three experiments, a number of different spatial ability measures were collected, but used as covariates in the analyses. Direct relations between gender, spatial ability and the effectiveness of animations were not explored. The main aim of this article is to directly examine these relations by showing the role and relation of spatial ability measures in male and female performance when learning from animations, as well as identify spatial ability differences between males and females.

4. Method

4.1. Participants

Participants were students from an Australian university distributed according to the following samples: Experiment 1 (*N* = 59: 30 males, 29 females, mean age = 22.5), Experiment 2 (*N* = 86: M = 42 males, 44 females, mean age = 21.85), and Experiment 3 (*N* = 120: 60 males, 60 females, mean age = 24.43).

4.2. Key Conditions of the Experiments

Experiments 1 and 2 compared the learning outcomes when completing an object manipulative task (i.e. 3-D Lego construction) after being modelled either in a static or animated condition. Both experiments had a 2 (gender: males vs. females) x 2 (presentation: animation vs. static) between-subject design. Participants in Experiment 1 were required to reconstruct the 3-D shapes using actual Lego blocks on a fixed platform, whereas participants in Experiment 2 re-constructed the 3-D shapes using virtual bricks on a computer. Experiment 3 included only an animated presentation and investigated the impact of gender on learning. Table 2 summarises the allocation of participants in each condition

Table 2.

Groupings of Participants in the Three Experiments

	Experiment 1		Experiment 2		Experiment 3	
	Males	Females	Males	Females	Males	Females
Animations	16	14	22	22	60	60
Static Pictures	14	15	20	22	-	-

4.3. Materials

A number of spatial ability measures (subjective and objective), as well as indicators of motor-related learning experience, were collected over the three experiments as described below.

Subjective measures of spatial ability. Two questions assessed subjective spatial ability as shown in Table 3 (questions 1 and 2) and were collected in all three experiments. Part of the aim of including these measures was to investigate a possible stereotype threat, as discussed above.

Source of learning motor-related tasks. In Experiment 3 only, two questions assessed prior experience in learning motor-related tasks from animations and static pictures (questions 3 and 4 in Table 3).

Table 3.

Subjective Measurements Collected in the Study

1. How would you rate your mental rotation ability (i.e. 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. 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 ☐

4. When learning motor-related task (e.g. tying knots, folding paper…etc.), how often did you learn it from pictures/ books?

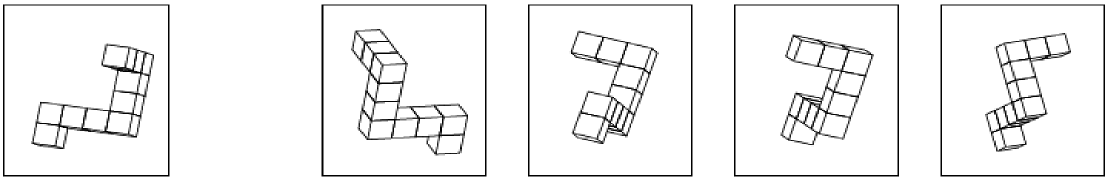
Never ☐ Rarely ☐ Occasionally ☐ Frequently ☐ Very Frequently ☐

The Card Rotations Test. The *Card Rotations Test* (CRT, Ekstrom et al., 1976) was an objective spatial ability test used in Experiments 1 and 2. This test measures mental rotation ability of 2-D figures. Each problem on the test consisted of an irregular shaped card on the left, and eight other drawings of the same card on the right (see Figure 1). Participants were required to decide whether each card on the right was only rotated or also turned over compared to the card on the left. The original test has two parts of ten questions each (three minutes each), and the whole set was used in Experiment 1. However, due to the possibility of mentally exhausting the participants before the main intervention began, the test was shortened to include only the first part (three minutes) in Experiment 2.

Figure 1. Sample Question of CRT

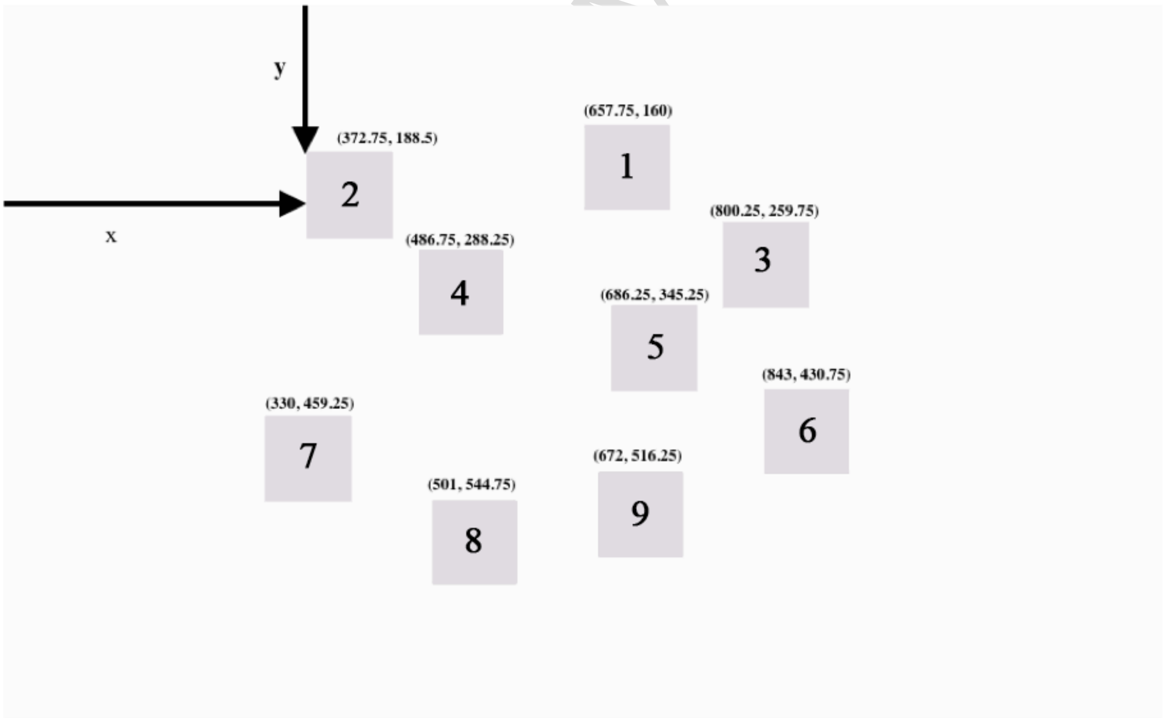
The Mental Rotations Test. In Experiment 3, a redrawn version of the *Mental Rotations Test* (MRT, Peters et al., 1995) was used to measure the mental rotation ability using 3-D figures. The MRT was originally designed by Vandenberg and Kuse (1978) based on the figures provided by Shepard and Metzler (1971). The test was previously found to be more sensitive to the gender difference in spatial ability across cultural boundaries and age cohorts when comparing with other widely available tests (cf. Peters et al., 1995). There are four different sets of MRT and in our experiments MRT-A was used. The test consists of four pages (6 problem sets each page). Each problem had a target figure shown on the left and four stimulus figures shown on the right (see Figure 2). Two of the four stimulus figures were rotated versions of the target figure and two others were different figures. The goal was to identify and mark the two rotated figures. Participants were given 3 minutes to complete 12 problem sets (half of the total questions) and a short break before starting the second half.

Figure 2. Sample Question of MRT



The Corsi Block Tapping Test. The *Corsi Block Tapping* Test was used in Experiment 3. The original version of the Corsi (1972) was built with 9 wooden blocks glued onto a large wooden board. It was originally invented to measure the spatial span in brain lesion patients. Nowadays it is widely used to measure visuospatial working memory capacity in both clinical and research practice (see, Darling et al., 2006; Kessels, van Zandvoort, Postma, Kappelle, & de Haan, 2000; Sala et al., 1999). In this experiment, a computerised version was developed (see Figure 3) and modified based on the block allocations from Kessels et al. (2000) and the sequence from Smirni, Villardita, and Zappalá (1983). The blocks would flash in yellow for one second in designated order with half a second break in between. Participants were required to memorise the sequence and then repeat back by clicking on the corresponding block in the correct order (see Figure 3 bottom for the sequence). The sequence started with 5 spans and gradually increased to 9 spans with no repetitions of any sequence. One mark would be given for each correct block in a correct order and the maximum score was 105.

Figure 3. The Corsi Test used in Experiment 3



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

Learning and testing materials. In all three experiments, an animation (video recording) was made that showed a 3-D Lego shape, consisting of three layers, being constructed block-by-block. In the first two experiments, an equivalent static presentation was made (see Figure 4; for more details see Castro-Alonso, Ayres, & Paas, 2015; Wong et al., 2015). After viewing their assigned presentation, participants were required to rebuild the shape from memory. Afterwards, a transfer test that was used in all experiments required students to rebuild a rotated first layer of the Lego bricks from memory, based on the observed presentation.

Figure 4. Learning Materials (Right to Left: Animation without Hands, Animation with Hands, Static Pictures with Hands)

4.4. Procedure

The procedure of all three experiments was approximately the same. The study sessions were conducted in a quiet room. Each session lasted 40-50 minutes with only one student. All participants in the study first completed the subjective measurements of spatial ability (see Table 3, Questions 1-2). For participants in Experiment 3, they also needed to complete Question 3-4 (see Table 3). Participants in Experiment 1 were then given six minutes to complete the whole Card Rotations Test. Participants in Experiment 2 were given three minutes to complete the first half of the test. In contrast, instead of the Card Rotations Test (CRT) participants in Experiment 3 then attempted the Mental Rotations Test (MRT) for six minutes followed by the Corsi at their own pace. After a Lego practice task where all the

Lego rules were explained and practiced, participants watched the assigned learning task materials for the first time. Immediately afterwards they were required to build the shape (Task 1) onto a designated platform. During this attempt participants had no access to the learning materials. Immediately after completion of their construction, it was then followed by a repetition of the procedure by watching the learning task for a second time and then attempting the same construction (Task 2). Immediately after the second attempt of the main task, participants were given the transfer task where they were required to rebuild a rotation of the construction without further access to the learning materials.

5. Results

5.1. Differences in Spatial Ability

Male and female group means for each spatial ability measurement collected in the three experiments were calculated (see Table 4), and independent t-tests conducted to investigate gender differences. For the first two experiments, there were no significant differences on the self-rating tests (mental rotation and overall spatial ability) or the CRT test. In Experiment 3 males reported significantly higher ratings than females on the self-rated mental rotation scale, $t(119) = 2.58$, $p < .05$, $d = .47$, and the self-rated spatial ability scale, $t(119) = 3.99$, $p < .01$, $d = .72$. For the MRT measure, there was also a significant difference where males had higher scores than females, $t(119) = 3.26$, $p < .01$, $d = .60$. However, for the Corsi test no gender difference was found.

Table 4

Mean (SD) of Spatial Ability Measurements for Males and Females across the Three Experiments

	CRT		MRT		Corsi		Self-rated mental rotation		Self-rated spatial ability	
	Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)		Mean (SD)	
	Male	Female	Male	Female	Male	Female	Male	Female	Male	Female
Experiment 1	107.87 (31.95)	101.62 (28.06)	.	.	.	.	3.43 (.63)	3.28 (.70)	3.30 (.53)	3.14 (.64)
Experiment 2	59.33 (17.46)	54.57 (14.72)	.	.	.	.	3.50 (.77)	3.45 (.79)	3.43 (.74)	3.23 (.80)
Experiment 3	.	.	20.97 (12.34)	14.48 (9.23)	73.25 (12.01)	69.32 (14.43)	3.20 (.80)	2.80 (.90)	3.40 (.72)	2.87 (.75)

Note: Experiment 2 used only the first half of CRT (full score = 80)

5.2. Regression Analysis of Experiment 1

Correlation coefficients between the three test scores, the CRT score, and the two self-rating measures were calculated and shown in Table 5. Only the CRT was significantly correlated with test scores, suggesting that this measure would be the best predictor of performance.

Table 5.

Pearson Correlation Matrix for Spatial Ability Measures in Experiment 1

	S-R mental rotation	S-R overall spatial ability	Task 1 score	Task 2 score	Transfer score
CRT	.281	.189	.363*	.421**	.352**
S-R mental rotation		.502**	.149	.058	.068
S-R overall spatial ability			.189	.112	.147

Notes: * p < .05, ** p < .01, S-R = self-rating

The regression analysis completed was *hierarchical*. At the first level (Model 1) all potential main effects were entered in a block, which included all 3 spatial ability measures and the 2 experimental factors (CRT, gender, presentation format, mental rotation, and overall spatial ability). This was followed by Model 2, which entered the CRT x Gender interaction first, followed by Model 3 which entered the CRT x Presentation format interaction next, and so on until all meaningful interactions were entered. Significant differences in changes between R^2 for each model were calculated and the final model chosen was located when no more significant differences were found between models.

Regression results. For Task 1, Model 2 [$F(6, 52) = 2.74$, $p < .05$] provided the last significant change in R^2 ($R^2 = .16$, $\Delta R^2 = .078$) from model-to-model, and identified the following significant factors: Gender ($B = -.75$, $SE = .33$, $\beta = -.38$, $p < .05$) and the CRT x Gender interaction ($B = 1.03$, $SE = .45$, $\beta = .52$, , $p < .05$). For Task 2, Model 3 [$F(7, 51) = 5.64$, $p < .001$] provided the last significant change in R^2 ($R^2 = .23$, $\Delta R^2 = .12$, $\Delta R^2 = .09$), and identified the following significant factors: the CRT x Gender interaction ($B = 1.34$, $SE = .39$, $\beta = .60$, $p < .01$), and the Gender x Presentation Format ($B = -1.19$, $SE = .42$, $\beta = -.53$, , $p < .01$), For the transfer task, Model 2 [$F(6, 52) = 3.35$, $p < .01$] provided the last significant change in R^2 ($R^2 = .18$, $\Delta R^2 = .10$), and identified the following significant factors: the CRT x Gender interaction ($B = 1.16$, $SE = .44$, $\beta = .52$, $p < .05$) and gender ($B = -.84$, $SE = .32$, $\beta = -.42$, , $p = .01$).

Interestingly for Model 1, the CRT was a significant predictor for each task, only to lose its significance when its interaction with gender was considered in the next model. Further analyses of the interaction involving the CRT measure and gender were conducted by examining the means for each cell on all 3 tasks. Male mean scores over the three tasks were 3.6, 5.1, and 2.2 for low CRT participants and 7.7, 11.4, and 5.5 for high CRT participants. Female mean scores over the three tasks were 5.8, 9.1, and 4.3 for low CRT participants and 6.6, 10.5, and 4.5 for high CRT participants. Hence, the CRT x Gender interaction can be explained by differences in the CRT having a greater impact for males than females. For the

Gender x Presentation Format interaction scores for the 2nd task were examined. Male mean scores were 9.8 for static presentations and 7.2 for animations. Female mean scores were 8.8 for static presentations and 10.5 for animations. It is therefore concluded that the Gender x Presentation Format interaction can be explained by females benefitting more from animations than males. The significant gender factor indicated that when other factors were controlled for, females scored higher than males.

In summary, the objective measurement of CRT was not found to be a significant predictor of learning performance but was moderated by gender. The Gender x CRT interaction was significant on all three tasks. The Gender x Presentation Format interaction indicated that gender again moderated the impact of presentation format. None of the subjective measures were significant predictors.

5.3. Regression Analysis of Experiment 2

Table 6 shows the correlation coefficients between the three test scores, the CRT score, and the two self-rating measures. As can be seen from the table none of these measures was significantly correlated with test measures. Consistent with the previous experiment the same hierarchical modelling was completed. For all three performance-scores, no significant regression models were identified and therefore no significant predictors of performance were found.

Table 6.

Pearson Correlation Matrix for Spatial Ability Measures Collected in Experiment 2

	S-R mental rotation	S-R overall spatial ability	Task 1 score	Task 2 score	Transfer score
CRT	.22*	.14	.10	.19	.20
S-R mental rotation		.64**	.05	.15	.13
S-R overall spatial ability			-.34	.07	.15

Notes: * p < .05, ** p < .01, S-R = self-rating

5.4. Regression Analysis of Experiment 3

Table 7 shows the correlation coefficients between the three test scores, the MRT score, Corsi score and the four self-rating measures. As can be seen from the table, all these measures were significantly correlated with the test scores suggesting a number of significant predictors. Using the same hierarchical procedure described in Experiment 1 all potential main effects were entered into the first model. This was followed by each individual meaningful interaction such as Gender x MRT for Model 2, and Gender x Corsi for Model 3, and so on. Results indicated that only Model 1 was significant for all 3 tasks as no other model led to a significant change in R^2.

Pearson Correlation Matrix for Spatial Ability Measures Collected in Experiment 3

	Corsi	S-R mental rotation	S-R overall spatial ability	F-L with animations	F-L with pictures	Task 1 score	Task 2 score	Transfer score
MRT	.36**	.36**	.35**	.19*	.23*	.24*	.23*	.29**
Corsi		.20*	.08	.30*	.18*	.29**	.34**	.25**
S-R mental rotation			.73**	.08	.13	.21*	.21*	.22*
S-R overall spatial ability				.04	.13	.24*	.22*	.26**
F-L with animations					.03	.32**	.31**	.19*
F-L with pictures						.24*	.29**	.29**

Notes: * p < .05, ** p < .01, S-R = self-rating, F-L = frequency in learning

For Task 1, Model 1 [$F(7, 112) = 4.24, p < .001, R^2 = .209$] was significant but no other model led to a significant change in R^2. The one significant predictor was frequency of learning with animations ($B = .26, SE = .10, \beta = .23, p < .05$). For Task 2 only Model 1 was significant [$F(7, 112) = 5.16, p < .001, R^2 = .244$] with 3 predictors: the Corsi ($B = .02, SE = .01, \beta = .21, p < .05$), frequency of learning with animations ($B = .21, SE = .10, \beta = .12, p < .05$), and frequency of learning with static pictures ($B = .19, SE = .10, \beta = .17, p < .05$). For the transfer task, Model 1 [$F(7, 112) = 3.61, p < .01, R^2 = .184$] again was the only significant model with 3 predictors: frequency of learning with static pictures ($B = .22$, $SE = .10$, $\beta = .20$, $p < .05$) and overall spatial ability ($B = .26$, $SE = .12$, $\beta = .21$, $p < .05$).

In summary, the subjective measures used in this experiment were the most consistent predictors of performance. Frequency of learning with animations occurred twice as did the frequency of learning with static pictures, while overall spatial ability appeared once. The Corsi measure emerged (once on Task 2) as the only objective predictor of performance. No significant interactions were found.

6. Summary

The study investigated the extent of the influence of spatial ability on gender differences in performance on tasks involving visual representations. Subjective spatial measures (*self-rated mental rotation ability* and *self-rated overall spatial ability*), objective spatial measures (*Card Rotations Test, CRT,* in Experiments 1-2; and *Mental Rotations Test, MRT* and *the Corsi Block Tapping Test* in Experiment 3) and participant learning experiences of motor-related tasks were collected (in Experiment 3). In Experiments 1 and 2 animations were compared with static pictures and in Experiment 3 only animations were used. Table 8 summarises the regression results, in order of largest β-values, of all three experiments.

Table 8.

Summary of Significant Predictors in the Regression Models of the Three Experiments

	Test 1	Test 2	Transfer
Experiment 1	1.CRT x Gender 2. Gender	1.CRT x Gender 2. Presentation format x Gender	1.CRT x Gender 2. Gender
Experiment 2	ns.	ns.	ns.
Experiment 3	1. Freq. of learning with Animation	1. Corsi 2. Freq. of learning with Animation 3. Freq. of learning with static pictures	1. Overall spatial ability 2. Freq. of learning with static pictures

The regression analysis results found that of the three objective measures of spatial ability used in the study only the Corsi test (Experiment 3) was a significant direct predictor of performance, but only on one task. However, there were a number of significant interactions between gender and the CRT measure in Experiment 1. Specifically, the CRT was found to have a greater impact on males performance than females.

Although only one objective measure of spatial ability (Corsi) was found to be a significant predictor of performance with no interaction with gender, a number of significant subjective measures were identified in Experiment 3. Ratings of frequency of learning with animations (twice), overall spatial ability (twice), and frequency of learning with static pictures (once) predicted performance independent of gender.

The regression analyses in Experiment 1 also confirmed the results reported by M. Wong et al. (2015) that females benefitted more from animations than static pictures; whereas male performance was not affected by the two presentation formats. In this experiment females were superior to males when differences in the CRT were controlled for.

Discussion

The Lego tasks required a certain amount of spatial rotation ability, especially the transfer task, and therefore the objective CRT and MRT scores, which measure spatial rotation ability, were expected to be predictors of performance for these tasks. However, the only predictor was the CRT measure, which interacted with gender, predicting male performance but not female.

The Lego tasks also required spatial location memory, which was measured by the Corsi Test in Experiment 3. It has been claimed that females have superior object location memory (see e.g. Choi & L'Hirondelle, 2005; Eals & Silverman, 1994; McBurney, Gaulin, Devineni, & Adams, 1997; Silverman, Choi, & Peters, 2007; Silverman & Eals, 1992). However, no gender differences were found on the Corsi test and no interactions were present suggesting under the present conditions males and females did not use different cognitive processes on these tasks.

In Experiment 1, a significant gender-animation interaction was found indicating that females benefitted more from animations than males. It is often argued (Falvo & Suits, 2009; Jacek, 1997; Sánchez & Wiley, 2010; Yezierski & Birk, 2006) that animation benefits females more because they have lower spatial ability. However, the CRT indicated no difference between males and females. Therefore, the argument about lower spatial ability for females and instructional animations cannot be supported using this spatial ability measure. Instead, the results indicated that males were more impacted by levels of spatial ability than females.

The findings of this study indicated that the Corsi was the only significant predictor of overall performance independent of gender. However, Experiment 3 indicated a number of significant subjective predictors (self-rating of overall spatial ability, experience of learning from animations and static pictures). Subjective measures are often considered less desirable than objective measures due to their lack of objectivity, but in this case, they produced a number of significant results for both males and females, as no interactions were identified. Hence, greater use of subjective measures may be helpful in animation research.

The subjective measure of overall spatial ability was found to be a successful predictor of performance in Experiment 3, but with no interactions with gender. Any self-affirmation demonstrated would be in both sexes. Hence, if a stereotype threat exists then based on this evidence it affects both males and females and not just females. Only once in the study did an objective scale (MRT used in Experiment 3) show males to have a higher spatial ability than females. And for both subjective measures in the same experiment males rated their spatial ability greater than females. However, this difference provided males no advantage on test results.

In Experiment 2, no significant predictors were found. Again, this may have reflected the overall general research comparing animations with statics, which can be highly variable. On the other hand, if the Corsi measure had been used in this experiment perhaps some predictors would have been identified. Although, no significant results may indicate that there may be many of other factors, not measured in this study that can influence outcomes in this domain, suggesting that the various boundary conditions associated with animation research are far from identified.

It should also be noted that Experiment 3, unlike the previous two experiments, did not include a static picture condition, and therefore the results in this experiment may not be directly compared with the other two. Nevertheless, the predictors found seem especially pertinent for animation research, if not necessarily when compared to static pictures. Research could investigate such subjective measures further in future.

In conclusion, the overall results from these three experiments showed that different spatial ability predictors varied in their capacity to strongly predict male and female performance when learning Lego construction tasks using animations and static pictures. Consequently, it is recommended that animation researchers should employ a variety of different spatial ability measures that can tap into these gender differences. If, as suggested, males and females favour different cognitive processes on such tasks but not shown in this study, then tests must be used to measure these processes. This situation may be managed more effectively by analysing the learning topics to see what types of spatial ability tests are required. Instead of using one well-known objective test such as the CRT, more appropriate matching tests should be used. It is also notable that the subjective measures had much success in predicting performance for both males and females, and therefore should not be discounted in future research.

This research contributes to the broader field of animation research in general. Although there is no direct evidence showing why animation benefits females more than males, the results showed that the different measures of spatial ability and gender have a significant impact on instructional animations. Overall, the results suggest that a failure to consider different measures of spatial ability and gender may explain some of the inconsistencies in this field of research.

References

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

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. doi: j.chb.2008.12.013

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: A cognitive load approach. *Applied Cognitive Psychology, 21*, 695–700. doi: 10.1002/acp.1343

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

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

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., 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.

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator, 2*(4), 1–17. doi: 10.1007/s00897970138a

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

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 Education, 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. (1993). *Human cognitive abilities: A survey of factor-analytical studies.* New York, NY: 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. doi: 10.1037/0012-1649.31.4.697

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. doi: 10.1007/978-1-46147485-2_22,

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. doi: 10.1007/978-3-319-02573-5_7

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2016). Comparing apples and oranges? A critical look at research on learning from statics versus animations. *Computers & Education, 102*, 234–243. doi: 10.1016/j.compedu.2016.09.004

Castro-Alonso, J. C., Ayres, P., Wong, M., & Paas, F. (2018). Learning symbols from permanent and transient visual presentations: Don't overplay the hand. *Computers & Education, 116*, 1–13. doi: 10.1016/j.compedu.2017.08.011

Chandler, P. (2009). Dynamic visualisations and hypermedia: Beyond the "Wow" factor. *Computers in Human Behavior, 25*(2), 389–392. doi: 10.1016/j.chb.2008.12.018

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

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

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

Darling, S., Sala, S. D., 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. (1871). *The descent of man, and selection in relation to sex*. Princeton, NJ: 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

Duff, S. J., & Hampson, E. (2001). A sex difference on a novel spatial working memory task in humans. *Brain and Cognition, 47*, 470–493. doi: 10.1006/brcg.2001.1326

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

Farah, M. J. (1988). Is visual imagery really visual? Overlooked evidence from neuropsychology. *Psychological Review, 95*(3), 307–317. doi: 10.1037/0033-295X.95.3.307

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

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

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. (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

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. doi: pms.1976.42.3c.1323

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, NY: Psychology Press.

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical for problem solving. *Journal of Educational Psychology, 91*(4), 684–689. doi: 10.1037/0022-0663.91.4.684

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. doi: 10.1207/s1532690xci2104_1

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. doi: 10.1016/j.learninstruc.2007.09.013

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

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

Koscik, T., O'Leary, D., Moser, D. J., Andreasen, N. C., & Nopoulos, P. (2009). Sex differences in parietal lobe morphology: Relationship to mental rotation performance. *Brain and Cognition, 69*(3), 451–459. doi: 10.1016/j.bandc.2008.09.004

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. doi: 10.2307/1130467

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

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

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

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue spatial visualization tests: Visualization of rotations (PSVT:R). *Educational Psychology Review, 25*(1), 69–94. doi: 10.1007/s10648-012-9215-x

Martens, A., Johns, M., Greenberg, J., & Schimel, J. (2006). Combating stereotype threat: The effect of self-affrmation on women's intellectual performance. *Journal of Experimental Social Pscyhology, 42*, 236–243. doi: 10.1016/j.jesp.2005.04.010

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

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

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. doi: 10.1037/0033-2909.86.5.889

McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology, 27*, 486–493. doi: 10.1016/j.appdev.2006.06.003

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

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

Narayanan, N. H., & Hegarty, M. (2002). Multimedia design for communication of dynamic information. *International Journal of Human-Computer Studies, 57*, 279–315. doi: 10.1006/ijhc.2002.1019

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. doi: 10.1007/s10648-011-9179-2

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.

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.

Pillay, H. K. (1994). Cognitive load and mental rotation: Structuring orthographic projection for learning and problem solving. *Instructional Science, 22*, 91–113. doi: 10.1007/BF00892159

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

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

Rebetez, C., Bétrancourt, M., Sangin, M., & Dillenbourg, P. (2010). 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: 10.1027/1015-5759/a000123

Rieber, L. P. (1990). Using computer animated graphics in science instruction with children. *Journal of Educational Psychology, 82*(1), 135–140. doi: 10.1037/0022-0663.82.1.135

Rieber, L. P. (1991a). Animation, incidental learning, and continuing motivation. *Journal of Educational Psychology, 83*(3), 318–328. doi: 10.1037/0022-0663.83.3.318

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. doi: 10.1007/BF02296567

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., & 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

Sala, S. D., Gray, C., Baddeley, A. D., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia, 37*(10), 1189–1199. doi: 10.1016/S0028-3932(98)00159-6

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

Sánchez, C. A., & Wiley, J. (2010). Sex differences in science learning: Closing the gap through animations. *Learning and Individual Differences, 20*, 271–275. doi: 10.1016/j.lindif.2010.01.003

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.

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.

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

Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New Yok, NY: Springer.

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. doi: 10.1023/A:1022193728205

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

Tversky, B., Morrison, J. B., & Bétrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*, 247–262. doi: 10.1006/ijhc.2002.1017

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 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

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604. doi: 10.2466/pms.1978.47.2.599

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. doi: 10.1348/000712604322779460

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. doi: 10.1016/S0278-2626(01)80086-5

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.

Wender, K. F., & Muehlboeck, J. S. (2003). Animated diagrams in teaching statistics. *Behaviour Research Methods, Instruments, & Computers, 35*(2), 255–258. doi: 10.3758/BF03202549

Willemsen, E., & Reynolds, B. (1973). Sex differences in adults' judgements of the horizontal. *Developmental Psychology, 8*(2), 309.

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

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. doi: j.chb.2008.12.012

Wong, M., Castro-Alonso, J. C., Ayres, P., & Paas, F. (2015). Gender effects when learning manipulative tasks from instructional animations and static presentations. *Educational Technology & Society, 18*(4), 37–52.

Wraga, M., Duncan, L., Jacobs, E. C., Helt, M., & Church, J. (2006). Stereotype susceptibility narrows the gender gap in imagined self-rotation performance. *Psychonomic Bulletin & Review, 13*(5), 813–819. doi: 10.3758/bf03194002

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

Zander, S., Wetzel, S., & Bertel, S. (2016). Rotate it! – Effects of touch-based gestures on elementary school students' solving of mental rotation tasks. *Computers & Education, 103*, 158–169. doi: 10.1016/j.compedu.2016.10.007

Acknowledgement

The second author acknowledges funding from CONICYT PAI, national funding research program for returning researchers from abroad, 2014, No. 82140021; and PIA–CONICYT Basal Funds for Centers of Excellence, Project FB0003.

Highlights

- Predictors of scores learning from animations are different for males and females
- Corsi test and subjective spatial measures predicts performance regardless gender
- No evidence that animations benefit females because of having lower spatial ability
- It suggests including gender and various spatial tests in future animation research