Educational Psychology Review (2019) 31:361–387
https://doi.org/10.1007/s10648-019-09469-1

META-ANALYSIS

Gender Imbalance in Instructional Dynamic Versus Static Visualizations: a Meta-analysis

Juan C. Castro-Alonso[1] · **Mona Wong**[2] · **Olusola O. Adesope**[3] · **Paul Ayres**[4] · **Fred Paas**[5,6]

Published online: 21 January 2019

Abstract

Studies comparing the instructional effectiveness of dynamic versus static visualizations have produced mixed results. In this work, we investigated whether gender imbalance in the participant samples of these studies may have contributed to the mixed results. We conducted a meta-analysis of randomized experiments in which groups of students learning through dynamic visualizations were compared to groups receiving static visualizations. Our sample focused on tasks that could be categorized as either biologically secondary tasks (science, technology, engineering, and mathematics: STEM) or biologically primary tasks (manipulative–procedural). The meta-analysis of 46 studies (82 effect sizes and 5474 participants) revealed an overall small-sized effect ($g+ = 0.23$) showing that dynamic visualizations were more effective than static visualizations. Regarding potential moderators, we observed that *gender* was influential: the dynamic visualizations were more effective on samples with less females and more males ($g+ = 0.36$). We also observed that *educational level*, *learning domain*, *media compared*, and *reporting reliability measures* moderated the results. We concluded that because many visualization studies have used samples with a gender imbalance, this may be a significant factor in explaining why instructional dynamic and static visualizations seem to vary in their effectiveness. Our findings also support considering the gender variable in research about cognitive load theory and instructional visualizations.

Keywords Dynamic and static visualization · Gender and spatial ability · STEM and manipulative–procedural tasks · Cognitive load theory · Meta-analysis

The research literature that compares the instructional effectiveness of dynamic visualizations (e.g., animations, simulations, and videos) versus static visualizations (e.g., still illustrations, slides, and photographs) is inconclusive. Although the overall findings of two relevant meta-analyses (Berney and Bétrancourt 2016; Höffler and Leutner 2007) suggest

✉ Juan C. Castro-Alonso
jccastro@ciae.uchile.cl

Extended author information available on the last page of the article

that dynamic visualizations are better instructional materials, there is an important caveat to consider before forming a definitive conclusion: many of the studies comparing dynamic versus static visualizations (including those cited in both meta-analyses) have included some methodological flaws. The issue of methodological flaws has been documented previously by several researchers. For example, Tversky et al. (2002) suggested that these comparative studies sometimes made unfair matches favoring animated depictions. In a more recent review, Castro-Alonso et al. (2016) identified seven biases (appeal, variety, media, realism, number, size, and interaction) that are not always controlled for in these types of studies. Despite such warnings, much dynamic–static research continues failing to control for moderating variables.

In this article, we extend these methodological concerns and argue that a lack of control for characteristics of the participants (e.g., see McCrudden and Rapp 2017), the intervention (e.g., Castro-Alonso et al. 2016), and the methodology (e.g., Mayer 2017) may also hinder dynamic–static visualizations research. In particular, we argue that *gender* deserves more attention as a participant characteristic. For example, the lack of attention for gender is shown in experimental studies (e.g., Garland and Sanchez 2013; Schnotz et al. 1999; Wang et al. 2011) and reviews (e.g., Castro-Alonso et al. 2016; Tversky et al. 2002) that have not mentioned the gender variable when comparing instructional visualizations. Many of the empirical studies even fail to provide the gender ratios for the whole sample or the individual conditions being compared. This call for consideration of gender in visualization research matches recent views on cognitive load theory (see Bevilacqua 2017) that argue for greater investigation of the differences between females and males in cognitive processes.

In this study, we investigated the evidence for a gender imbalance in research studying learning from dynamic versus static instructional visualizations. The gender imbalance is representative of the participant samples where many instructional visualization studies are conducted, namely with education and psychology undergraduate students (cf. Isacco et al. 2016), in which males are notably underrepresented. This imbalance is typically not considered as an issue. Therefore, to investigate whether different gender ratios produce different effects, we conducted a meta-analysis of dynamic versus static visualizations and used the percentage of females as a possible moderator.

A secondary aim of this study was to investigate other potential moderators of dynamic versus statics instructional effectiveness, especially those identified by cognitive load theory research (e.g., Castro-Alonso et al. 2016; Höffler and Leutner 2007; Paas and Sweller 2012; Wong et al. 2018). Therefore, in addition to gender, we explored those variables that can be categorized as characteristics of the participants (spatial ability and educational level), the intervention (type of task, learning domain and topic, and media compared), and the methodology (gender ratio per condition and whether pretests and reliability measures were reported). These moderators are described next.

Participant Characteristics

Gender

In addition to exploring if gender was imbalanced or neglected in the dynamic–static studies, we also investigated the percentage of females in the samples as a potential moderator for the meta-analysis. Although females and males tend to be similar in many academic aspects, they sometimes differ in variables related to learning from visualizations. For example, Zell et al. (2015) reported meta-syntheses of different meta-analyses about gender differences. When analyzing 30 meta-analyses (3611 effects) about cognitive variables (e.g., attention, memory, and problem solving), they observed that the gender differences presented an overall effect size of $d = 0.22$. The authors concluded that, although this is a small effect, it was calculated with averages, and usually larger gender differences appear between top scorers (e.g., Hedges and Nowell 1995).

Thus, gender can be influential in the effectiveness of instructional interventions (see also Bevilacqua 2017), particularly among high achievers. From the potential gender cognitive variables that can influence the learning from visualizations, we focus on spatial ability (see Höffler 2010; see also Wong et al. 2018), due to its large documented impact. This second participant characteristic is described next.

Spatial Ability

Although the spatial ability construct includes many visual and spatial subabilities (e.g., Hegarty et al. 2006; Höffler 2010; Linn and Petersen 1985; Uttal et al. 2013; Voyer et al. 1995), almost exclusively *mental rotation* and *mental folding* are used in studies of instructional visualizations and gender differences. As defined by Linn and Petersen (1985), mental rotation is the ability to mentally rotate or flip shapes quickly and accurately, and mental folding (also termed *spatial visualization*) is the ability to perform mental transformations of spatial information.

The findings generally indicate a male advantage for spatial ability tasks, which tends to be larger for mental rotation than for mental folding tasks (e.g., Linn and Petersen 1985; Stephenson and Halpern 2013; Voyer et al. 1995). For example, the study of meta-analyses by Zell et al. (2015) showed that mental rotation was among the cognitive abilities with the largest gender differences ($d = 0.57$), in favor of males. For the analyses of this study, we explored if spatial ability impacted differently on learning from dynamic or static visualizations.

As training can enhance spatial ability (see Uttal et al. 2013), it is often argued that the gender differences in spatial ability can be explained by females having less practice than males in early spatial experiences (e.g., Jirout and Newcombe 2015; Newcombe et al. 1983; see also Voyer and Jansen 2017). In other words, spatial ability is dependent on development. As spatial ability may moderate the dynamic–static studies, and as it may depend on the development (exposure) of students, we also considered developmental age (educational level) as another possible moderator for the visualization studies.

Educational Level

The literature shows that dynamic visualizations and animations are often enjoyed and have a positive impact on learning for school children (e.g., Bétrancourt and Chassot 2008), university students (e.g., Jaffar 2012), and adults (e.g., Türkay 2016). However, sometimes (e.g., Mahmud et al. 2011) enjoyment does not translate into learning. We, therefore, explored if the dynamic versus static comparisons presented different effect sizes in school children of different ages and university students.

Intervention Characteristics

Type of Task

Based on the work of David C. Geary, cognitive load theory researchers have highlighted the importance of differentiating between two type of tasks that the human species has evolved to manage differently: that is, the more primitive *biologically primary* tasks, which have evolved in humans to help in their ancient survival as a species; and the more current *biologically secondary* tasks, which have been culturally necessary to function in contemporary society (Geary 1995, 2007). Primary tasks (e.g., to manipulate things or to gesture) are learned quickly, as we have evolved a mind to acquire this information easily. In contrast, secondary tasks (e.g., to read or to understand graphs) tend to be learned slowly, as we have not evolved the mechanisms to acquire them effortlessly. Because of these differences, the easier primary tasks require less cognitive effort than the harder secondary tasks (Paas and Sweller 2012; see also Sweller et al. 2011).

In this meta-analysis, we considered instructional visualizations depicting these two types of tasks. We focused on areas where considerable research into dynamic versus statics comparisons has been completed. As such, the secondary tasks selected focused on the educational fields of science, technology, engineering, and mathematics (STEM). In comparison, the primary tasks chosen regarded object manipulations and similar manipulative–procedural tasks.

We also explored possible interactions between the dynamic versus static format, type of task, and spatial ability (see Table 1). For STEM tasks, there are conflicting arguments in making a prediction if spatial ability is more helpful for processing dynamic or static visualizations (see Mayer et al. 2005). On the one hand, the *mental animation* theoretical perspective (Hegarty 1992; see also Höffler 2010) suggests that dynamic are easier visualizations to process. Thus, spatial ability is more helpful when studying static materials, as it aids inferring the movements of the depicted STEM contents. Because a dynamic format already shows the movements, and seeing is easier than inferring, the mental animation rationale suggests that spatial ability is less necessary with dynamic depictions.

On the other hand, the perspective based on the *overwhelming processing* (Lowe 2003; see also Lowe 1999) and the *transient information effect* proposed by the cognitive load theory (see Ayres and Paas 2007; Castro-Alonso et al. 2018b) predicts that dynamic are more difficult visualizations to process, particularly those containing transient information. The transient information perspective suggests that spatial ability is a key to coping with the challenging

Table 1 Different perspectives to predict which visualization is easier and in which spatial ability is more helpful

Theoretical perspective	Rationale	Easier visualization	Spatial ability more helpful for
STEM tasks			
Mental animation	It is difficult to infer movements from statics	Dynamic	Statics
Overwhelming processing; transient information effect	It is difficult to cope with the pace of dynamic	Statics	Dynamic
Manipulative–procedural tasks			
Unnaturalness	It is difficult to cope with static (paused) or irregular primary motion	Dynamic	Statics

cognitive demands of images that leave the screen before being processed. As a static format does not contain information that disappears before being processed, it allows more time for restudying information, and therefore, learning under these static conditions is easier and requires less spatial ability. A summary of these opposite theoretical perspectives is provided in Table 1.

For manipulative–procedural tasks, we also explored interactions between the dynamic versus static format and spatial ability. The human species has evolved to learn these primary tasks more easily because they have been fundamental to survive and thrive. Arguably, since these tasks have been learned by our ancestors, it is likely that today, the best way to learn these tasks is in similar learning conditions to those of our forefathers. In support, diverse evidence (e.g., Press et al. 2005; Shimada and Oki 2012; VanArsdall et al. 2015) has shown that the natural scenarios of prehistoric ages are better learning conditions for these tasks, rather than more modern and artificial scenarios. For example, for modern humans to learn imitative hand actions and manipulations, other humans should be better teaching agents than robots (e.g., Press et al. 2005; see also Cracco et al. 2018). The type of movement shown is also critical, as the fluent movement of manipulations activates to a greater extent our evolved imitative systems, as compared to paused or unnatural motions (e.g., Shimada and Oki 2012; see also Matthews et al. 2007). This effect also has links to the literature showing that autonomously moving objects are better memorized than nonmoving elements (e.g., Bonin et al. 2014; VanArsdall et al. 2015). For the current analysis of dynamic versus static visualizations, this *unnaturalness* perspective (see Table 1) suggests that dynamic are easier visualizations to learn from, and thus, spatial ability is more helpful to deal with static images that do not present the natural movement for manipulations that we evolved to learn more easily (see also Paas and Sweller 2012).

In this meta-analysis, we explored which of the opposite theoretical perspectives outlined above would most apply to STEM tasks. In other words, we investigated if dynamic or static visualizations would be more effective for learning STEM tasks. Also, for manipulative–procedural tasks, we explored whether dynamic visualizations would be more effective than the static depictions.

Learning Domain

In addition to the general distinction between STEM and manipulative–procedural tasks, these categories contain subgroups. Among the STEM topics, the meta-analysis by Berney and Bétrancourt (2016) revealed trends (nonsignificant differences) in which more technological domains (e.g., aeronautics, informatics, mathematics, and mechanics) presented smaller effects favoring animation over statics, as compared to other fields (e.g., biology, chemistry, natural sciences, and physics). In this study, we also expected differences within STEM disciplines. Among manipulative–procedural tasks, we explored if manipulations or procedures regarding the syllabi would be different to manipulative–procedural tasks not related to school or university syllabi.

Media Compared

As reviewed in Castro-Alonso et al. (2016), the instructional media used to present the visualizations could affect the dynamic versus static comparisons. The literature has provided mixed evidence for the best educational medium. One example, supporting paper over digital media (computer and mobile devices), is the meta-analysis by Delgado et al. (2018). This analysis of over 170,00 school and university participants revealed an overall small effect size of the paper material being more effective for reading comprehension. In contrast, and on a much smaller scale, Nikou and Economides (2016) provided an example supporting computer over paper media. The authors investigated 66 high school students (49% females) learning physics (electromagnetism) through three different media conditions: (a) pen-and-paper, (b) computer, and (c) mobile device. Results showed that only the computer and mobile device produced higher knowledge gains from pre- to posttests. In our moderator analyses, we expected different outcomes when using the same medium to present the visualizations (e.g., computer dynamic vs. computer statics), as compared to when employing different media (e.g., computer dynamic vs. paper statics).

Methodological Characteristics

Following the research agenda proposed by Mayer (2017), which called for the need to improve the methodological rigor of educational multimedia research, we investigated three variables that sometimes lack in dynamic versus static comparisons. First, we contrasted studies reporting the gender distribution per compared groups, against studies that did not report it and could only be assumed they had a distribution representative of the whole sample (as the participants were randomly assigned to the groups). Second, we compared studies including or not including a pretest as a measure of prior knowledge of the participants. Last, we also explored if reporting a reliability measure of the learning tests, as compared to not reporting these data, affected the dynamic versus static comparisons.

Research Questions and Hypotheses

In the present meta-analysis, we examined the following research questions: (a) How does gender moderate the effects of dynamic versus static visualizations? (b) Are the effects of dynamic versus static visualizations moderated by other variables, including participants, intervention, and methodological characteristics?

To answer these research questions, we tested the following hypotheses:

- Gender is a participant characteristic that moderates the effects of dynamic versus static visualizations (Hypothesis 1).
- Spatial ability and educational level are participant characteristics that moderate the effects of dynamic versus static visualizations (Hypothesis 2).
- Type of task and learning domain are intervention characteristics that moderate the effects of dynamic versus static visualizations (Hypothesis 3).
- Media compared is an intervention characteristic that moderates the effects of dynamic versus static visualizations (Hypothesis 4).
- Methodological characteristics moderate the effects of dynamic versus static visualizations (Hypothesis 5).

Method

Selection Criteria

For the meta-analysis, a study was deemed eligible for inclusion if it:

1. Was published between 1990 and 2017.
2. Was written in English.
3. Was a peer-reviewed journal article.
4. Compared, in a between-subjects design, the learning effects of at least one dynamic visualization with at least one static visualization, depicting either a STEM or a manipulative–procedural task. We excluded text-only formats and mixed conditions in which both dynamic and static visualizations were included in the same group. By "dynamic," in addition to common dynamic visualizations such as videos and animations, we also considered depictions that other researchers have called "static–sequential" (e.g., Imhof et al. 2011) or "successive static" (e.g., Lowe et al. 2011).
5. Included an experimental design in which participants were randomly assigned to groups. We excluded the studies in which this random assignment was not explicitly stated.
6. Consisted of sole school or university samples.
7. Reported measurable outcomes of performance, such as retention and transfer tests.
8. Included sufficient data to allow for effect size calculations.
9. Reported the gender ratio for the total sample.

Literature Search and Selection of Studies

We used the query *animation* OR *animated* AND (*visualization* OR *picture*) as keywords to conduct a comprehensive and systematic search on the following electronic databases: ProQuest–ERIC, ProQuest–APA (PsycARTICLES and PsycINFO), and Web of Science (Social Sciences Citation Index, SSCI, Categories: *Education & Educational Research*; *Education*, *Scientific Disciplines*; *Psychology*, *Educational*; and *Psychology*, *Experimental*). The databases search procedure returned a total of 1470 articles. Following removal of duplicates, 1269 studies remained.

There were two filtering phases to determine whether these studies should be included in the meta-analysis or not. In the first filtering phase, we applied the nine inclusion criteria when screening the abstract of the articles, to determine eligibility for further examination. Two authors of the present study read 145 abstracts (approximately 10% of the total) to adjust the inclusion criteria and confirm that their rater agreement was 100%, before screening the total of abstracts. After inspecting all the abstract, 1107 results were excluded, and full-text copies were obtained for the 162 articles that passed the first filtering phase. Disagreements between both authors were discussed until consensus was reached.

In the second filtering phase, the two authors reviewed the full-text copies by applying the selection criteria stated above and excluded further 125 publications. Revealing that gender can be overlooked in these studies, many articles (72, 58% of the total) were discarded because they did not report the gender ratio for the total sample (criterion 9, see above). In fact, 48 studies met all inclusion criteria except for this information of the gender composition. This omission was observed both in earlier studies (e.g., Ardac and Akaygun 2005; Chanlin 2001;

Hays 1996; Williamson and Abraham 1995) and in more current evidence (e.g., Chen et al. 2015; Schwartz and Plass 2014; Wang et al. 2011). In total, 37 articles from the databases met all inclusion criteria and were retained in this meta-analysis.

In addition, we searched the reference sections of five classic papers (Ayres and Paas 2007; Höffler et al. 2010; Lowe 2003; Mayer et al. 2005; Tversky et al. 2002), and three meta-analyses (Berney and Bétrancourt 2016; Höffler 2010; Höffler and Leutner, 2007) which investigated the effects of animated and static pictures. These reports added eight eligible studies that met all inclusion criteria. We also included one additional study (Lusk and Atkinson 2007) that met the criteria.

In total, 46 articles were included in the meta-analysis. These articles included 82 effect sizes comparing dynamic and static visualizations. From the $k = 82$ comparisons, 19 (23% of the total) investigated school students, and 63 (77%) investigated university participants. Also, 60 (73%) corresponded to STEM and 22 (27%) to manipulative–procedural tasks. A summary of the selection of articles is provided in the flow diagram in Fig. 1.

Next, the selected articles were carefully read, to extract relevant data for the meta-analysis. First, the two authors in parallel read the same ten experiments (approximately 10% of the total) and obtained a rater agreement of 100%. Then, each author read and coded approximately half of the remaining articles. After all the data was collected, all authors agreed on the relevant information and the coding.

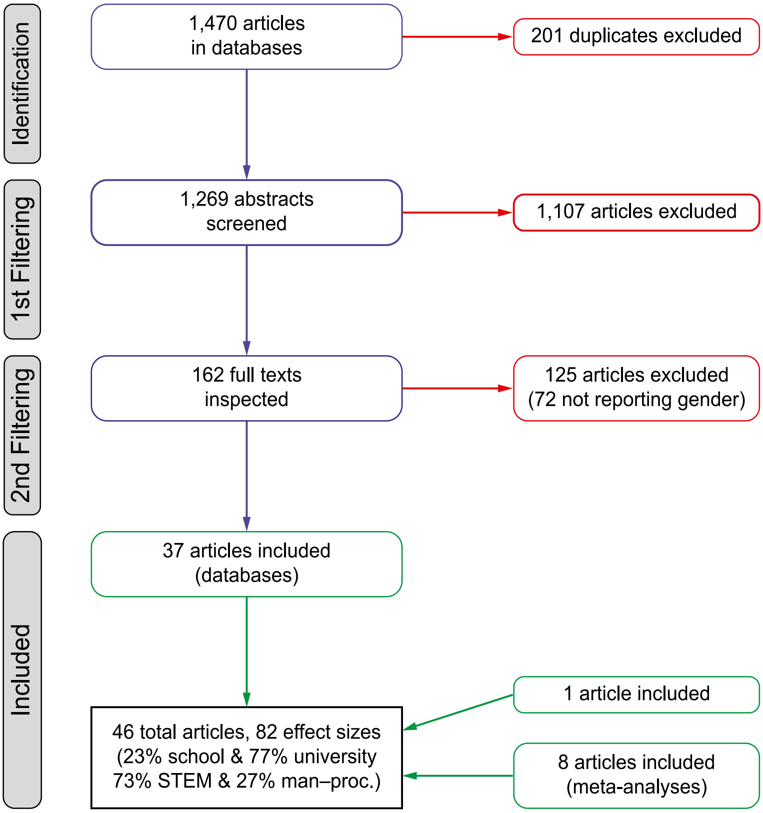

Fig. 1 Flow diagram of the selection of studies

Extraction of Effect Sizes

For each study included in this meta-analysis, we calculated Cohen's *d* effect size, a standardized estimate of the difference in achievement scores between students who studied with dynamic visualizations compared with those who studied with static-only visualizations. Cohen's *d* was computed as the difference between the mean scores of the dynamic and static groups divided by the pooled standard deviations of the two groups. Because differential sample sizes across studies may bias the effect size obtained by Cohen's *d*, Hedges' *g* (Hedges and Olkin 1985) was computed and reported throughout this meta-analysis as an unbiased estimate of the standardized mean difference effect size. Throughout this meta-analysis, a positive effect size indicates benefits of dynamic visualizations over static visualizations. Conversely, a negative effect size indicates that students who learned with static formats outperformed those who learned with dynamic visualizations.

Data Analysis

Throughout the data analyses, we followed standard guidelines for conducting a meta-analysis (Adesope et al. 2017; Bernard et al. 2009; Lipsey and Wilson 2001). We analyzed data with Comprehensive Meta-Analysis (CMA) 2.2.064 (Borenstein et al. 2008) and IBM™ SPSS™ version 24 for Windows. The weighted mean effect sizes were aggregated to form an overall weighted mean estimate of the effect of learning with dynamic presentations (i.e., $g+$). The use of weighted mean effect sizes allowed more weight to be assigned to studies with larger sample sizes. The significance of each weighted mean effect size was determined by its 95% confidence interval. When the lower limit of a confidence interval was greater than zero, a positive mean effect size was interpreted as indicating a statistically significant result in favor of the dynamic visualization. When both limits of a confidence interval were smaller than zero, the negative mean effect size was interpreted as indicating a statistically significant result in favor of the static visualization.

Homogeneity of variance was examined by the Q_B statistic to assess if the observed effect sizes that were combined into a mean all estimated the same population effect size. The CMA software reported Q_B and its concomitant p value for each subcategory to determine if the distribution of effect sizes within each subcategory was homogeneous or not. We used the I^2 statistic computed by CMA to more comprehensively interpret the result of the homogeneity test (Higgins and Thompson 2002; Huedo-Medina et al. 2006). I^2 value of 0% indicates no observed heterogeneity, and larger values show increasing heterogeneity. Researchers have suggested that percentages of around 25% ($I^2 = 0.25$), 50% ($I^2 = 0.50$), and 75% ($I^2 = 0.75$) should be interpreted to mean low, medium, and high heterogeneity, respectively (Higgins and Thompson 2002).

Results

A total of 46 research reports yielding 82 independent effect sizes ($N = 5474$) were analyzed. Three studies produced outlying effect sizes ($Z > 3.3$). Because the three studies met all inclusion criteria and were methodologically similar to other studies in our distribution, a decision was made to retain the studies in this meta-analysis, but we windsorized the effect sizes by adjusting them to values closer to the next-largest or next-lowest effect size in our distribution, as recommended by Tabachnick and Fidell (2018). Figure 2 shows the distribution of effect sizes for the meta-analysis after the three outliers were windsorized. The effect sizes ($M = 0.20$, SD = 0.57) are mainly clustered between − 0.40 and 0.80 standard deviations. These data suggest that in most studies the group that learned with dynamic learning materials outperformed the groups that learned with static visualizations.

Table 2 shows a summary of the variables coded for each study, including the study identification, the percentage of females in the whole experimental sample of the study, the spatial ability measured and its positive effect for the dynamic or the static visualization, the educational level of the sample, the learning domain and topic, whether the media compared between visualizations was the same or different, whether the gender percentage in each compared group was reported or not, whether a pretest was included in the experiment, and the associated unbiased effect size (Hedges' g). The top of the table includes the 35 articles of STEM learning tasks, and the bottom part shows the 11 articles of manipulative–procedural tasks.

Overall Effect of Dynamic Versus Static Visualizations

Table 3 shows the overall effect of the meta-analysis. The table includes the number of participants (N) in each category, the number of effect sizes (k), the weighted mean effect size ($g+$) and its standard error (SE), the 95% lower and upper confidence intervals (CI), the results of a test of homogeneity (Q_B) with its associated degrees of freedom (df) and probability (p), and the percentage of variability that could be attributed to true heterogeneity or between-studies variability (I^2). The same format was used for Tables 4 and 5.

As shown in the first row of Table 3, there is an overall ($N = 5474$; $k = 82$) statistically significant positive effect of learning with dynamic visualizations ($g+ = 0.23$), which

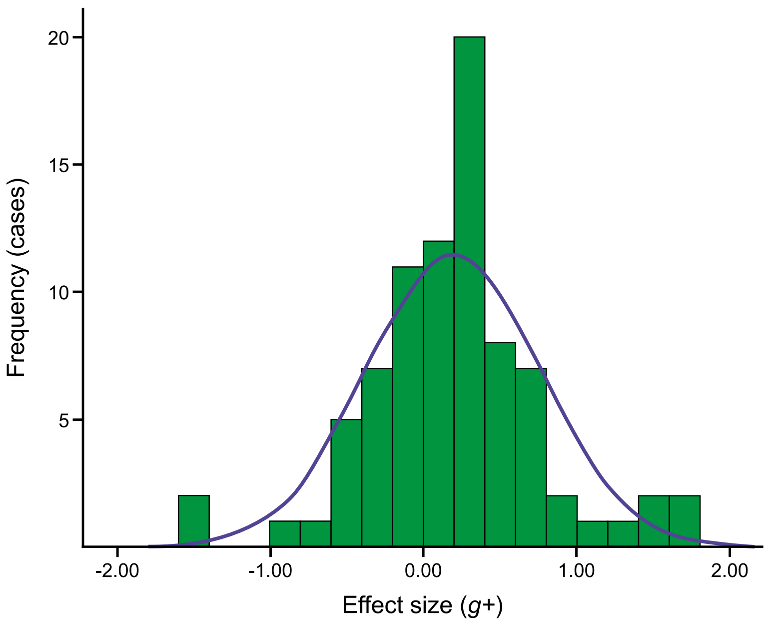

Fig. 2 Distribution of the 82 effect sizes ($M = 0.20$, SD = 0.57)

Table 2 Descriptive information and effect sizes for the coded studies (separated for STEM and manipulative–procedural tasks)

Study	% Fem	Spatial ability, effect[a]	Educational level[b]	Learning domain and topic[c]	Media compared	Gender % per group?	Pretest?	Effect size ($g+$)
STEM tasks								
Adesope and Nesbit (2013)	46	–	University	B: human nervous system	Same	Yes	Yes	0.17
Berney et al. (2015)	18	SpAb, S	University	B: scapula and shoulder flexion	Same	No	No	0.03
Boucheix and Schneider (2009), exp 1	89	SpAb, S	University	P: three-pulley system	Same	No	No	−0.12
Chien and Chang (2012)	100	–	High Sch	G: Abney Level topographic measure	Same	Yes	No	0.31
Fiorella and Mayer (2016), exp 3	73	–	University	P: Doppler effect of sound waves	Same	No	Yes	−0.02
Goff et al. (2017)	73	–	University	B: photosynthesis	Same	Yes	Yes	0.39*
Höffler and Leutner (2011), exp 1	92	MF, S	University	P: surfactants cleaning dirt	Same	No	Yes	0.70
Höffler and Leutner (2011), exp 2	41	SpAb, S	High Sch	P: surfactants cleaning dirt	Same	No	Yes	0.41
Höffler et al. (2010)	62	–	High Sch	B: reactions in photosynthesis	Same	No	Yes	−0.17
Höffler and Schwartz (2011) (1)	68	–	University	P: surfactants cleaning dirt	Same	No	Yes	−0.34
Höffler and Schwartz (2011) (2)	68	–	University	P: surfactants cleaning dirt	Same	No	Yes	0.63*
Imhof et al. (2012), exp 1	71	SpAb, A	University	B: fish movement patterns	Same	No	Yes	0.43
Imhof et al. (2011) (1)	76	MR3D, A	University	B: fish movement patterns	Same	No	Yes	−0.15
Imhof et al. (2011) (2)	76	MR3D, A	University	B: fish movement patterns	Same	No	Yes	−0.04
Kalyuga (2008)	76	–	University	T: graph/equation transformations	Same	No	Yes	0.16
Lewalter (2003)	70	–	University	P: optical gravitational lensing	Same	Yes	Yes	0.27
Lin (2011)	56	–	University	B: blood circulation	Same	No	No	0.34*
Lin and Dwyer (2010)	56	–	University	B: blood circulation	Same	No	No	0.29*
Lin and Atkinson (2011)	49	–	University	G: rock cycle	Same	No	Yes	1.70*
Lowe et al. (2011)	90	–	University	B: kangaroo hopping cycle	Same	No	No	0.11
Lusk and Atkinson (2007) (1)	75	–	University	T: worked examples of proportions	Same	No	Yes	0.30
Lusk and Atkinson (2007) (2)	75	–	University	T: worked examples of proportions	Same	No	Yes	0.09
Lusk and Atkinson (2007) (3)	75	–	University	T: worked examples of proportions	Same	No	Yes	0.38
Mayer et al. (2007), exp 1	59	–	University	T: hydraulic car brakes mechanics	Different	No	Yes	−0.17
Mayer et al. (2007), exp 2	68	–	University	T: hydraulic car brakes mechanics	Different	No	Yes	0.21
Mayer et al. (2005), exp 1	88	–	University	G: lightning development	Different	No	No	−0.31
Mayer et al. (2005), exp 2	84	–	University	T: toilet flushing system mechanics	Different	No	No	−0.50
Mayer et al. (2005), exp 3	70	–	University	G: formation of ocean waves	Different	No	No	−0.57
Mayer et al. (2005), exp 4	74	–	University	T: car brakes mechanics	Different	No	No	−0.64
Münzer et al. (2009)	77	SpAb, N	University	B: ATP enzyme synthesis	Same	No	Yes	0.27
Paik and Schraw (2013)	75	MF, NR	University	T: toilet flushing system mechanics	Same	No	No	−0.06

Table 2 (continued)

Study	% Fem	Spatial ability, effect[a]	Educational level[b]	Learning domain and topic[c]	Media compared	Gender % per group?	Pretest?	Effect size ($g+$)
Park (1998) (1)	39	–	University	P: electronic circuit	Same	No	Yes	−0.25
Park (1998) (2)	39	–	University	P: electronic circuit	Same	No	Yes	−0.56
Park and Gittelman (1992) (1)	71	–	University	P: electronic circuit	Same	No	Yes	−0.26
Park and Gittelman (1992) (2)	71	–	University	P: electronic circuit	Same	No	Yes	−0.82*
Park and Gittelman (1992) (3)	71	–	University	P: electronic circuit	Same	No	Yes	−0.26
Patwardhan and Murthy (2015)	19	–	University	T: electrical signals and systems	Same	No	No	−0.39
Rieber (1990) (1)	54	–	Elm Sch	P: Newton's laws of motion	Same	No	No	−0.23
Rieber (1990) (2)	54	–	Elm Sch	P: Newton's laws of motion	Same	No	No	0.66
Rieber (1990) (3)	54	–	Elm Sch	P: Newton's laws of motion	Same	No	No	0.45
Rieber (1991) (1)	49	–	Elm Sch	P: Newton's laws of motion	Same	No	No	0.77*
Rieber (1991) (2)	49	–	Elm Sch	P: Newton's laws of motion	Same	No	No	0.55
Sanchez and Wiley (2014)	54	SpAb, S	University	G: plates and volcanic eruptions	Same	Yes	Yes	0.28
Scheiter et al. (2006)	71	MR3D, NR	University	T: examples of probability problems	Same	No	Yes	−0.19
Schmidt-Weigand (2011), exp 1 (1)	60	–	University	G: lightning development	Same	No	Yes	0.09
Schmidt-Weigand (2011), exp 1 (2)	60	–	University	G: lightning development	Same	No	Yes	−0.15
Schmidt-Weigand (2011), exp 2 (1)	73	–	University	G: lightning development	Same	No	Yes	0.35
Schmidt-Weigand (2011), exp 2 (2)	73	–	University	G: lightning development	Same	No	Yes	0.27
Schmidt-Weigand and Scheiter (2011) (1)	68	–	University	G: lightning development	Same	No	Yes	1.53*
Schmidt-Weigand and Scheiter (2011) (2)	68	–	University	G: lightning development	Same	No	Yes	0.33
Stebner et al. (2017), exp 1 (1)	47	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.74*
Stebner et al. (2017), exp 1 (2)	47	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.03
Stebner et al. (2017), exp 1 (3)	47	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.34
Stebner et al. (2017), exp 2 (1)	54	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.54*
Stebner et al. (2017), exp 2 (2)	54	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.39
Stebner et al. (2017), exp 2 (3)	54	MF, A	Mid Sch	P: surfactants cleaning dirt	Same	No	Yes	0.34
Tekdal (2013)	44	–	University	T: programming logic operations	Same	Yes	Yes	0.81*
Thompson and Riding (1990)	50	–	Mid + High Sch	T: Pythagoras' theorem	Same	No	No	0.38
Wu and Chiang (2013) (1)	40	–	University	T: orthographic views of object	Same	Yes	No	1.03*
Wu and Chiang (2013) (2)	40	–	University	T: orthographic views of object	Same	Yes	No	0.24
Manipulative–procedural tasks								
Arguel and Jamet (2009), exp 1	82	–	University	S: first aid procedures	Same	No	Yes	0.71*
Ayres et al. (2009), exp 1	42	–	University	NS: knot tying	Same	No	No	1.60*
Ayres et al. (2009), exp 2	44	–	University	NS: keyring puzzle	Same	No	No	1.20*

Table 2 (continued)

Study	% Fem	Spatial ability, effect[a]	Educational level[b]	Learning domain and topic[c]	Media compared	Gender % per group?	Pretest?	Effect size ($g+$)
Castro-Alonso et al. (2015), exp 1	73	–	University	NS: Lego blocks assembling	Same	No	Yes	0.18
Castro-Alonso et al. (2015), exp 2 (1)	60	–	University	NS: Lego blocks assembling	Same	Yes	Yes	0.06
Castro-Alonso et al. (2015), exp 2 (2)	60	–	University	NS: Lego blocks assembling	Same	Yes	Yes	0.32
Marcus et al. (2013)	0	–	University	NS: knot tying	Same	Yes	No	0.61
Michas and Berry (2000), exp 1	66	–	University	S: first aid bandaging of a hand	Same	No	No	0.59*
Soemer and Schwan (2016), exp 1	63	–	University	S: writing Chinese pseudocharacters	Same	Yes	No	−0.57*
Soemer and Schwan (2016), exp 2	76	–	University	S: writing Chinese pseudocharacters	Same	Yes	No	−0.49*
Soemer and Schwan (2016), exp 3	72	–	University	S: writing Chinese pseudocharacters	Same	Yes	No	−0.13
Soemer and Schwan (2016), exp 4 (1)	76	–	University	S: writing Chinese pseudocharacters	Same	Yes	No	0.53*
Soemer and Schwan (2016), exp 4 (2)	76	–	University	S: writing Chinese pseudocharacters	Same	Yes	No	0.30
Swezey et al. (1991)	59	–	University	S: troubleshooting an engine model	Different	No	No	−0.05
Wong et al. (2012), exp 1 (1)	58	–	Mid Sch	NS: origami paper folding	Same	No	Yes	1.41*
Wong et al. (2012), exp 1 (2)	58	–	Mid Sch	NS: origami paper folding	Same	No	Yes	0.14
Wong et al. (2009), exp 2	62	–	Elm Sch	NS: origami paper folding	Same	No	Yes	0.99*
Wong et al. (2009), exp 3	46	–	Elm Sch	NS: origami paper folding	Same	No	Yes	0.52
Wong et al. (2015), exp 1	49	MR2D, N	University	NS: Lego blocks assembling	Same	Yes	No	0.06
Wong et al. (2015), exp 2	51	MR2D, N	University	NS: Lego blocks assembling	Same	Yes	No	0.10
Zacks and Tversky (2003), exp 2 (1)	75	–	University	NS: toy bug assembling	Same	No	Yes	−1.50*
Zacks and Tversky (2003), exp 2 (2)	75	–	University	NS: toy bug assembling	Same	No	Yes	−1.55*

The studies presenting numbers under parentheses indicate different compared groups per study

% Fem = percentage of females in the whole experimental sample. *Gender % per group?* = Was the gender percentage per compared groups reported?

[a] The spatial ability measured was reported and its positive effect for the dynamic or the static visualization. The spatial abilities were as follows: *MR3D* = mental rotation in three dimensions; *MR2D* = mental rotation in two dimensions; *MF* = mental folding (spatial visualization); SpAb = other spatial abilities, or aggregated scores. The positive effects of spatial abilities were reported for: *A* = the spatial ability favored learning *all* visualization formats; *S* = it favored learning *static* visualizations; *N* = it favored learning *none*; *NR* = the effect of spatial ability on the separate visualizations was *not reported*

[b] *Elm Sch* = elementary school; *Mid Sch* = middle school

[c] For STEM tasks, the domains were as follows: *B* = biology and medicine science; *P* = physics and chemistry science; *G* = geology and other sciences; *T* = technology, engineering, and mathematics. For manipulative–procedural tasks, the domains were as follows: *S* = syllabi; *NS* = nonsyllabi

*$p < 0.05$

Table 3 Overall effect and weighted mean effect sizes for participant characteristics

Moderator	N	k	Effect size		95% CI		Test of heterogeneity			
			$g+$	SE	LCI	UCI	Q_B	df	p	I^2 (%)
Overall effect	5474	82	0.23*	0.03	0.17	0.28	252.64	81	< 0.001	67.94
Percentage of females										
59% or less	2932	35	0.36*	0.04	0.29	0.43	114.66	34	< 0.001	70.35
60% or more	2542	47	0.07	0.04	− 0.01	0.15	109.57	46	< 0.001	58.02
Total within							224.23	80	< 0.001	
Total between							28.41	1	< 0.001	
Spatial ability favoring										
Spatial ab. not measured	4269	63	0.23*	0.03	0.17	0.29	233.31	62	< 0.001	73.43
All (static and dynamic)	552	9	0.30*	0.09	0.13	0.47	9.55	8	0.30	16.20
Static only	287	5	0.20	0.12	− 0.03	0.43	4.26	4	0.37	6.17
None/not reported	366	5	0.06	0.11	− 0.15	0.27	2.20	4	0.70	0.00
Total within							249.32	78	< 0.001	
Total between							3.32	3	0.34	
Educational level										
Elementary school	199	7	0.53*	0.14	0.25	0.80	5.96	6	0.43	0.00
Middle school	423	8	0.44*	0.10	0.25	0.64	11.26	7	0.13	37.81
Middle + high school	72	1	0.38	0.24	− 0.08	0.84	0.00	0	1.00	0.00
High school	130	3	0.12	0.18	− 0.22	0.46	2.41	2	0.30	17.00
University	4650	63	0.19*	0.03	0.13	0.25	221.54	62	< 0.001	72.01
Total within							241.17	77	< 0.001	
Total between							11.47	4	0.02	

$^*p < 0.05$

corresponds to a small size (Cohen 1988). In other words, there is an overall advantage of learning from dynamic visualizations as compared to static visualizations.

The overall distribution was highly heterogeneous, $Q_B(81) = 252.64$, $p < 0.001$, $I^2 = 68\%$. The total variability that could be attributed to true heterogeneity or between-studies variability was 68%, indicating that 68% of the variance was between-studies variance (i.e., could be explained by study-level covariates) and 32% of the variance was within-studies based on sampling error. This heterogeneity suggests that there was more variability among the independent effect sizes than would be expected for samples from a single population. Significant heterogeneity warrants robust exploration of study features that may moderate the overall effect. Hence, moderator analyses were conducted. Tables 3, 4, and 5 show the results of these moderating factors, which we describe next.

Participant Characteristics

Below the overall effect, Table 3 presents the weighted mean effect sizes for three characteristics of participants as moderator variables: percentage of females in the samples, spatial ability favoring which type of visualization, and educational level of the participants. The median for the percentage of females in the samples was 60%. We took a median split to compare studies including 59% or less females versus studies with 60% or more females (see Table 3). Dynamic visualizations were associated with statistically significant effect sizes for studies that had 59% or less females ($g+ = 0.36$), but not for studies with 60% or more females ($g+ = 0.07$). The between-levels difference was statistically significant, $Q_B(1) = 28.41$, $p < 0.001$. Post hoc analysis revealed that dynamic visualizations with studies that had 59%

or less females were associated with higher weighted mean effect size and were significantly different than studies that had 60% females or more. This is our most important finding, as it supports our claim that different gender ratios affect the comparisons of dynamic and static visualizations. The result suggests that in samples with less female representation, dynamic visualizations are advantageous, but this advantage may disappear in samples with more females. In other words, males may benefit more from dynamic visualizations than females.

A second participant characteristic considered as a moderator was which visualization was preferentially favored by spatial ability. As shown in Table 3, most of the studies in our meta-analysis ($k = 63$, 77% of the total effects) did not measure any spatial ability (see also Table 2). From the studies that did measure any spatial ability, statistically significant benefits for dynamic over static visualizations tended to be higher in the studies in which the ability favored both types of visualizations ($g+ = 0.30$). In other words, the studies that revealed that the dynamic format was more effective than the static ($k = 9$, 11%) were those in which spatial ability was helpful to learn from both formats. Table 3 also indicates that fewer studies ($k = 5$, 6%) showed that spatial ability favored learning from static visualizations. We could not find any study in which spatial ability helped to learn only from dynamic presentations. In all, these results are more supportive of the theoretical perspectives (see Table 1) termed as mental animation (STEM tasks) and unnaturalness (manipulative–procedural tasks).

We exercise caution with findings about spatial ability because the majority of the studies did not measure any type of these abilities. That is why we did not consider the different spatial abilities assessed (mental rotation in three and in two dimensions, mental folding, and other spatial abilities or aggregated scores) for moderator analyses. However, as shown in Table 2, mental folding (MF) was the most assessed spatial ability ($k = 8$, 10%) and mental rotation in two dimensions (MR2D) was the least investigated ($k = 2$, 2%).

The last participant moderator analyzed was the educational level of the students. As shown in Table 3, dynamic visualizations produced statistically significant benefits over static presentations when used by elementary school, middle school, and university samples of students. The between-levels difference of educational level was statistically significant, $Q_B(4) = 11.47, p = 0.02$. Showing an age or educational level effect, the dynamic visualizations were more effective for elementary school students ($g+ = 0.53$), than for middle school students ($g+ = 0.44$), than for university students ($g+ = 0.19$).

Intervention Characteristics

Table 4 shows the weighted mean effect sizes for three characteristics of the interventions: type of task, learning domain, and media compared. Regarding the type of task, most of the effect sizes concerned STEM tasks ($k = 60$, 73%) and the manipulative–procedural tasks were less represented ($k = 22$, 17%). For both tasks, dynamic visualizations were statistically more effective than static formats, representing small effect sizes ($g+ = 0.24$ for STEM, and $g+ = 0.18$ for manipulative–procedural), without between-levels significant differences. These results support both the mental animation and unnaturalness theoretical perspectives.

The second intervention characteristic of Table 4 concerns learning domain (see also Table 2). For STEM tasks, there were four domains: biology and medicine science (B, $k = 11$, 13%); physics and chemistry science (P, $k = 23$, 28%); geology and other sciences (G, $k = 11$, 13%); and technology, engineering, and mathematics (T, $k = 15$, 18%). For manipulative–procedural tasks, there were two domains: syllabi (S, $k = 8$, 10%) and nonsyllabi (NS, $k = 14$,

Table 4 Weighted mean effect sizes for intervention characteristics

Moderator	N	k	Effect size		95% CI		Test of heterogeneity			
			g+	SE	LCI	UCI	Q_B	df	p	I^2 (%)
Type of task										
STEM	4380	60	0.24*	0.03	0.18	0.30	164.43	59	< 0.001	64.12
Manipulative–procedural	1094	22	0.18*	0.06	0.06	0.31	87.65	21	< 0.001	76.04
Total within							252.08	80	< 0.001	
Total between							0.56	1	0.45	
Learning domain[a]										
STEM (B)	1866	11	0.27*	0.05	0.17	0.36	9.86	10	0.45	0.00
STEM (P)	1017	23	0.19*	0.06	0.06	0.31	41.68	22	0.01	47.22
STEM (G)	565	11	0.38*	0.09	0.21	0.55	68.15	10	< 0.001	85.33
STEM (T)	932	15	0.15*	0.07	0.02	0.28	39.28	14	< 0.001	64.36
Manipulative–proc. (S)	523	8	0.01	0.09	− 0.17	0.19	24.94	7	< 0.001	71.93
Manipulative–proc. (NS)	571	14	0.34*	0.09	0.17	0.51	55.79	13	< 0.001	76.70
Total within							239.69	76	< 0.001	
Total between							12.95	5	0.02	
Media compared										
Same	5050	75	0.26*	0.03	0.21	0.32	223.78	74	< 0.001	66.93
Different	424	7	− 0.20*	0.10	− 0.39	− 0.01	8.12	6	0.23	26.14
Total within							231.90	80	< 0.001	
Total between							20.74	1	< 0.001	

[a] For STEM tasks, the domains were as follows: *B* = biology and medicine science; *P* = physics and chemistry science; *G* = geology and other sciences; *T* = technology, engineering, and mathematics. For manipulative–procedural tasks, the domains were as follows: *S* = syllabi; *NS* = nonsyllabi

*$p < 0.05$

17%). Because the between-levels difference was statistically significant, $Q_B(5) = 12.95$, $p = 0.02$, the domains showed that dynamic visualizations were more effective than statics in different degrees. For STEM, geology and other sciences ($g+ = 0.38$) and biology and medicine science ($g+ = 0.27$) showed higher effects than physics and chemistry science ($g+ = 0.19$) and technology, engineering, and mathematics ($g+ = 0.15$). For manipulative–procedural, nonsyllabi ($g+ = 0.34$) was significantly higher than syllabi ($g+ = 0.01$). In short, dynamic visualizations seem to be best for biology and medicine science, geology and other sciences, and for manipulative–procedural tasks outside the syllabi.

Similarly, concerning media compared, a group of studies investigated the effects on the same medium ($k = 75$, 91%), while much fewer investigations concerned different media ($k = 7$, 9%). All studies on the same medium compared computer dynamic versus computer static visualizations. Different media research involved either television dynamic versus 35 mm slide statics (Swezey et al. 1991) or computer dynamic versus paper statics (Mayer et al. 2005, 2007). The between-levels difference was statistically significant, $Q_B(1) = 20.74$, $p < 0.001$. Post hoc analysis revealed that, when the medium was the same (computers), dynamic visualizations were associated with higher weighted mean effect size ($g+ = 0.26$) and were significantly different than when the visualization media was different. In fact, for different media, the effects were in the opposite direction, showing that statics outperformed dynamic ($g+ = -0.20$). In conclusion, dynamic visualizations outperformed static visualizations to a greater extent when they were compared in computers, than when they were shown in different media. In contrast, in different media, statics (in paper or slides media) were more effective than dynamic visualizations (in computers or television).

As with spatial ability, due to the small number of studies using different media, these findings should be interpreted cautiously.

Methodological Characteristics

Table 5 presents the effect size variations related to the methodological quality of the research. This includes whether or not the studies included three variables: the gender percentage for every experimental condition, pretests to show prior knowledge differences, and reliability data for the learning measures.

Concerning the first methodological characteristic, Table 5 shows that there were more studies not reporting the gender ratio per compared groups ($k = 64$, 78%), as compared to those that explicitly mentioned how each experimental condition was represented by females and males ($k = 18$, 22%). These two groups did not show significantly different weighted mean effect sizes for dynamic over statics. In other words, dynamic visualizations produced statistically significant benefits over static presentations regardless of whether studies reported gender distributions for every compared group ($g+ = 0.21$) or not ($g+ = 0.23$).

Table 5 also shows that there were more studies reporting pretests ($k = 50$, 61%) than those not reporting pretests ($k = 32$, 39%). As the between-levels difference was not significant, it can be concluded that dynamic presentations produced statistically significant differences over statics regardless of if pretests were used ($g+ = 0.26$) or not ($g+ = 0.19$).

Concerning the last methodological characteristic, there were more studies not reporting reliability measures ($k = 51$, 62%) than those reporting them ($k = 31$, 38%). Table 5 shows that dynamic presentations produced statistically significant benefits regardless of whether reliability measures were reported or not. However, the between-levels difference was statistically significant, $Q_B(1) = 13.77$, $p < 0.001$. Post hoc analysis revealed that, in studies that reported

Table 5 Weighted mean effect sizes for methodological characteristics

Moderator	N	k	Effect size		95% CI		Test of heterogeneity			
			$g+$	SE	LCI	UCI	Q_B	df	p	I^2 (%)
Gender % per group?[a]										
No	4177	64	0.23*	0.03	0.17	0.29	207.25	63	< 0.001	69.60
Yes	1297	18	0.21*	0.06	0.10	0.32	45.32	17	< 0.001	62.49
Total within							252.58	80	< 0.001	
Total between							0.06	1	0.80	
Pretest reported?										
No	2796	32	0.19*	0.04	0.12	0.27	102.69	31	< 0.001	69.81
Yes	2678	50	0.26*	0.04	0.18	0.34	148.31	49	< 0.001	66.96
Total within							251.00	80	< 0.001	
Total between							1.64	1	0.20	
Reliability reported?[b]										
No	2668	51	0.12*	0.04	0.04	0.20	184.97	50	< 0.001	72.97
Yes	2806	31	0.32*	0.04	0.25	0.40	53.90	30	< 0.001	44.34
Total within							238.87	80	< 0.001	
Total between							13.77	1	< 0.001	

[a] Was the gender percentage per compared groups reported?

[b] Were reliability measures for the learning tests reported?

*$p < 0.05$

reliability of their learning test, dynamic presentations were associated with higher weighted mean effect size ($g+ = 0.32$) and were significantly different than studies that did not report reliability of their outcome measures ($g+ = 0.12$).

How Valid Are the Findings? Examining Publication Bias

We examined the potential publication bias of the meta-analysis favoring published studies that report statistically significant effect sizes. We examined this threat to the validity of our findings through three approaches computed with the CMA software. First, the funnel plot (which reveals the estimates of the unbiased effect size compared with the standard error) showed a symmetrical distribution around the weighted mean effect. These symmetric funnel plots suggest the absence of publication bias (Duval and Tweedie 2000). Second, Egger's linear regression test (Egger et al. 1997) was used to more fully investigate the results of the funnel plot through an examination of the unbiased effect sizes and standard errors. Results of this test further corroborated the result of the funnel plot, clearly showing the absence of publication bias ($p = 0.42$). Third, a "classic fail-safe N" test (e.g., Rosenthal 1979) was performed to determine the number of null effect studies needed to raise the p value associated with the average effect above an arbitrary alpha level (set at $\alpha = 0.05$). Results from classic fail-safe N test revealed that 974 additional qualified studies would be required to invalidate the overall effect size found in this meta-analysis. These three different tests suggest that findings from the present meta-analysis are not threatened by publication bias to the extent that it could invalidate the findings.

Discussion

The main aim of this study was to investigate a possible gender imbalance in the research about dynamic and static visualizations. We conducted a meta-analysis to explore if different gender ratios produced different effects on these comparisons. As a secondary goal of the meta-analysis, other potential moderators, besides gender, were also investigated.

The meta-analysis of 46 studies and 82 independent comparisons ($N = 5474$) revealed an overall small effect size ($g+ = 0.23$) of dynamic visualizations being more effective learning tools than static visualizations. This finding is consistent with the two previous meta-analyses that compared dynamic to static visualizations and also found effects favoring the dynamic formats. As such, the analysis by Höffler and Leutner (2007) of 26 studies and 76 comparisons showed an overall medium effect size ($d = 0.37$), and the analysis by Berney and Bétrancourt (2016) of 61 studies and 140 comparisons showed an overall small effect size ($g+ = 0.23$) in the same directions as our current study. Nevertheless, as found in those meta-analyses, we also observed significant heterogeneity between the effect sizes, indicating that different variables were influencing these results. Moderator analyses were conducted for participant, intervention, and methodological characteristics, as discussed next.

Participant Characteristics

The main participant characteristic in this study was gender. The meta-analysis revealed that dynamic visualizations with studies that had 59% or less females showed a significantly higher mean effect size than studies that had 60% females or more. This result supports our main claim that a gender imbalance may affect the comparisons investigating the learning effectiveness of dynamic and static visualizations. Specifically, the finding suggests that in samples with less females (and more males), dynamic visualizations are advantageous, but this advantage may disappear in samples with more females (and less males). In other words, males may benefit more from dynamic visualizations than females. This is opposite to the single study included in this meta-analysis in which gender was an independent variable (Wong et al. 2015). As shown in Table 2, the study measured university students attempting a manipulative–procedural task with Lego blocks. Although the overall effect of the two experiments reported in the study showed advantages for dynamic visualizations (nonsignificant effects, see Table 2), static visualizations were more beneficial for males, and dynamic presentations were more beneficial for females (not reported here). In contrast, our present meta-analysis showed that dynamic visualizations might be more beneficial for males. The direction of effects, supporting either the dynamic or the static visualization as more effective for a certain gender, warrants further investigations.

Nevertheless, the key finding, supporting Hypothesis 1, is that learning from dynamic or static visualizations was influenced by the gender of the student. This result aligns with the comments by Bevilacqua (2017) that cognitive load theory should investigate the gender effects in cognitive processes. Regrettably, when conducting the literature search for this meta-analysis, we observed that gender was often neglected as a potential variable for instructional visualization research. For example, many studies (48), including recent ones (e.g., Chen et al. 2015; Schwartz and Plass 2014; Wang et al. 2011), were not included in our analyses solely because they failed to provide the gender ratio of the sample. Also, many included comparisons ($k = 64$, 78%) did not give details about the gender ratio for every compared group (see also "Methodological Characteristics"). In consequence, we believe that cognitive load theory, and other theoretical approaches, should include gender when researching instructional visualizations.

In addition to gender, another participant variable investigated as a potential moderator was spatial ability. Results showed that in cases where the dynamic visualizations were mostly effective, spatial ability (commonly, mental folding) was equally effective for improving learning from dynamic and static visualizations. Hence, the beneficial role of dynamic visualizations may surpass the beneficial role of spatial ability. In addition, we could not find any study showing that spatial abilities were only helpful for dynamic visualizations (without also being helpful for static visualizations). Nevertheless, as only 23% ($k = 19$) of the effects in this meta-analysis included some measure of spatial ability, any conclusions concerning this variable need further investigation.

Regarding the educational level moderator, we observed that the dynamic visualizations were more effective for elementary school students than for middle school students than for university participants. In other words, there appears to be a decline in the instructional effectiveness of dynamic visualizations as students develop. As the literature has shown positive motivational and learning effects for dynamic visualizations presented to students from all ages (e.g., Bétrancourt and Chassot 2008; Höffler and Leutner 2007; Mahmud et al. 2011), this lower effect for more adult students was not predicted. In all, there is partial support for Hypothesis 2, as the educational level was a moderator in the effectiveness of dynamic versus spatial visualizations, but for spatial ability, further investigation is needed.

Intervention Characteristics

Comparing the effects of STEM and manipulative–procedural tasks, it was observed that dynamic visualizations were more effective than static visualizations for both tasks equally.

For STEM tasks, the two meta-analyses included had considered mostly studies about STEM tasks, as the meta-analysis by Höffler and Leutner (2007) included 77% of STEM studies and the meta-analysis by Berney and Bétrancourt (2016) incorporated 90% of studies about STEM tasks. As those analyses showed an overall advantage of dynamic visualizations, the results are consistent with our current findings of an advantage of dynamic visualizations for STEM tasks. Concerning the theoretical perspectives of Table 1, this dynamic advantage aligns better with the mental animation perspective (dynamic are easier for secondary tasks) than with the perspectives presented as the overwhelming processing or the transient information effect of cognitive load theory (static are easier for secondary tasks).

For manipulative–procedural tasks, the meta-analysis by Höffler and Leutner (2007) revealed the largest effects favoring dynamic visualizations ($d = 1.06$) when the tasks involved procedural–motor knowledge. Our current study shows the same direction of effects favoring dynamic formats for manipulative–procedural tasks, although our effect size is smaller ($g+ = 0.18$). The differences in effect sizes are largely due to the differences in defining procedural–motor knowledge, as compared to our manipulative–procedural tasks. In any case, the results align with the unnaturalness theoretical perspective shown in Table 1 (dynamic are easier for primary tasks).

Regarding learning domain for the STEM tasks, dynamic visualizations may be more effective for biology and medicine science, and geology and other sciences, as compared to the more technology-oriented tasks of technology, engineering, and mathematics. A similar trend (nonsignificant) was reported in the meta-analysis by Berney and Bétrancourt (2016), in which dynamic visualizations were less effective in the technological domains (e.g., informatics, mathematics, and mechanics), as compared to other scientific fields (e.g., biology, chemistry, and natural sciences). For the manipulative–procedural tasks, dynamic visualizations were more effective for nonsyllabi tasks than for syllabi tasks. A possible explanation is that the nonsyllabi tasks that we included may have been more biologically primary (e.g., knot tying, paper folding, Lego assembling) than the more secondary syllabi tasks (e.g., writing and troubleshooting problems in an engine). Hence, it is possible that these nonsyllabi tasks activated more the evolved mechanisms to deal better with dynamic visualizations than with static visualizations (unnaturalness perspective of Table 1). In short, there is partial support for Hypothesis 3, as learning domain was a moderator in the effectiveness of dynamic versus spatial visualizations, but the type of task was not.

The last intervention characteristic provided evidence that not under all conditions dynamic visualizations are better than the static formats. Only when the visualizations were compared in the same computer medium, dynamic was advantageous. In contrast, when the comparisons were made with different media, there was a better performance of the static visualization (in paper or 35 mm slides) as compared to the dynamic format (in computers or television). As these comparisons always showed statics in a medium without screens, this suggests that screen media (computers and television) may be less effective than paper or slides. Recently, the meta-analysis by Delgado et al. (2018) also showed that paper material was more effective than digital resources (computer and mobile devices) for the task of reading comprehension. This supports Hypothesis 4, as the media being compared was a moderator in the effectiveness of dynamic versus spatial visualizations. However, the small number of comparisons employing different media that we included ($k = 7$, 9%) hinders reaching a strong conclusion.

Methodological Characteristics

Following the suggestion for improving methodological rigor by Mayer (2017), we assessed three methodological characteristics that could affect the overall advantage of dynamic visualizations: (a) reporting the gender ratio per experimental conditions, (b) including a pretest to measure prior knowledge of the participants, and (c) including reliability measures for the learning tests.

For the first variable, it was a concern that the majority of the comparisons ($k = 64$, 78%) did not report the gender ratio per experimental conditions. As this meta-analysis concludes that a gender imbalance affects dynamic versus statics research, not reporting the gender ratio in every condition that is compared should be avoided in future investigations. Despite this concern, whether the studies reported these data or not did not significantly change the advantage of dynamic over static visualizations.

In contrast to the above concern, it was encouraging that most of the comparisons ($k = 50$, 61%) included pretests to control for prior knowledge differences, although again this variable was not influential, as dynamic visualizations presented similar advantages in studies with or without the use of pretests.

Last, it was also a concern that most of the effects compared ($k = 51$, 62%) were from studies that did not report the reliability of their learning measures. In this case, this methodological variable was influential, as those studies reporting the reliability of their learning tests showed larger effects favoring dynamic over static visualizations. This is a reassuring result, as it is indicating that the positive effects of dynamic visualizations were also present (in fact, they were higher) when the studies employed more strict learning measures with reliability scores. Altogether, there is weak support for Hypothesis 5: only reporting reliability measures was a moderator in the effectiveness of dynamic versus spatial visualizations, but neither reporting gender per condition nor reporting pretests moderated the effects.

Limitations and Future Directions

One limitation of the present study concerns the stringent inclusion criteria that we had to use to investigate the effects of a gender imbalance in the samples. The criteria (e.g., criterion 9) meant that many dynamic versus static studies with different spatial ability measures were discarded. Future research may investigate how different spatial abilities (cf. Castro-Alonso et al. 2018a) affect the effectiveness of learning visualizations.

A second limitation is that we did not consider other moderating variables, such as level of realism, interaction features, and similar variables that are known to affect learning from visualizations (see Castro-Alonso et al. 2016). This was beyond the scope of the present study, but in future studies, gender differences could be considered by controlling for these other moderating variables.

Last, the university samples included were largely drawn from many different disciplines. A future direction is to compare these gender differences and spatial ability effects in different disciplinary areas, such as those requiring more spatial ability (e.g., geometry or physics) versus those requiring less (e.g., history or literature).

Conclusion

In this meta-analysis, we have provided additional evidence of the positive effects of dynamic visualizations for learning, and have shown that many moderators are involved, including variables of participants, intervention, and methodology. From these moderators, we noted that gender is a key participant characteristic to consider when investigating the instructional effectiveness of dynamic and static visualizations. As many studies have not included gender as a variable, this may have had a major influence on visualizations studies and may be a significant factor in explaining why instructional dynamic and static visualizations seem to vary in their effectiveness. Our findings support that dynamic visualizations may be more effective for males than for females. Future studies controlling the gender variable will provide further evidence to inform visualizations and possibly cognitive load theory research. To control gender, we recommend that these future studies (a) include an equal gender proportion in every condition being compared or (b) employ the same number of females and males.

Acknowledgements We are thankful to Mariana Poblete and Monserratt Ibáñez for their assistance.

Funding Funding from PIA-CONICYT Basal Funds for Centers of Excellence Project FB0003 is gratefully acknowledged.

Compliance with Ethical Standards

Conflict of Interest The authors declare that they have no conflict of interest.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

*Adesope, O. O., & Nesbit, J. C. (2013). Animated and static concept maps enhance learning from spoken narration. Learning and Instruction, 27, 1–10. https://doi.org/10.1016/j.learninstruc.2013.02.002.

Adesope, O. O., Trevisan, D. A., & Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Review of Educational Research, 87*(3), 659–701. https://doi.org/10.3102/0034654316689306.

Ardac, D., & Akaygun, S. (2005). Using static and dynamic visuals to represent chemical change at molecular level. *International Journal of Science Education, 27*(11), 1269–1298. https://doi.org/10.1080/09500690500102284.

*Arguel, A., & Jamet, E. (2009). Using video and static pictures to improve learning of procedural contents. Computers in Human Behavior, 25(2), 354–359. https://doi.org/10.1016/j.chb.2008.12.014.

*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(2), 348–353. https://doi.org/10.1016/j.chb.2008.12.013.

Ayres, P., & Paas, F. (2007). Making instructional animations more effective: a cognitive load approach. *Applied Cognitive Psychology, 21*(6), 695–700. https://doi.org/10.1002/acp.1343.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., & Bethel, E. C. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research, 79*(3), 1243–1289. https://doi.org/10.3102/0034654309333844.

Berney, S., & Bétrancourt, M. (2016). Does animation enhance learning? A meta-analysis. *Computers & Education, 101*, 150–167. https://doi.org/10.1016/j.compedu.2016.06.005.

*Berney, S., Bétrancourt, M., Molinari, G., & Hoyek, N. (2015). How spatial abilities and dynamic visualizations interplay when learning functional anatomy with 3D anatomical models. Anatomical Sciences Education, 8(5), 452–462. https://doi.org/10.1002/ase.1524.

Bétrancourt, M., & Chassot, A. (2008). Making sense of animation: how do children explore multimedia instruction? In R. K. Lowe & W. Schnotz (Eds.), *Learning with animation: research implications for design* (pp. 141–164). New York, NY: Cambridge University Press.

Bevilacqua, A. (2017). Commentary: Should gender differences be included in the evolutionary upgrade to cognitive load theory? *Educational Psychology Review, 29*(1), 189–194. https://doi.org/10.1007/s10648-016-9362-6.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: further evidence from word and picture stimuli. *Memory & Cognition, 42*(3), 370–382. https://doi.org/10.3758/s13421-013-0368-8.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2008). *Comprehensive meta-analysis (version 2. 2.048) [computer software]*. Englewood, NJ: Biostat.

*Boucheix, J.-M., & Schneider, E. (2009). Static and animated presentations in learning dynamic mechanical systems. Learning and Instruction, 19(2), 112–127. https://doi.org/10.1016/j.learninstruc.2008.03.004.

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

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. https://doi.org/10.1016/j.compedu.2016.09.004.

Castro-Alonso, J. C., Ayres, P., & Paas, F. (2018a). Computerized and adaptable tests to measure visuospatial abilities in STEM students. In T. Andre (Ed.), *Advances in human factors in training, education, and learning sciences: proceedings of the AHFE 2017 international conference on human factors in training, education, and learning sciences* (pp. 337–349). Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-60018-5_33.

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

Chanlin, L.-J. (2001). Formats and prior knowledge on learning in a computer-based lesson. *Journal of Computer Assisted Learning, 17*(4), 409–419. https://doi.org/10.1046/j.0266-4909.2001.00197.x.

Chen, S.-C., Hsiao, M.-S., & She, H.-C. (2015). The effects of static versus dynamic 3D representations on 10th grade students' atomic orbital mental model construction: evidence from eye movement behaviors. *Computers in Human Behavior, 53*, 169–180. https://doi.org/10.1016/j.chb.2015.07.003.

*Chien, Y.-T., & Chang, C.-Y. (2012). Comparison of different instructional multimedia designs for improving student science-process skill learning. Journal of Science Education and Technology, 21(1), 106–113. https://doi.org/10.1007/s10956-011-9286-3.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Cracco, E., Bardi, L., Desmet, C., Genschow, O., Rigoni, D., De Coster, L., et al. (2018). Automatic imitation: a meta-analysis. *Psychological Bulletin, 144*(5), 453–500. https://doi.org/10.1037/bul0000143.

Delgado, P., Vargas, C., Ackerman, R., & Salmerón, L. (2018). Don't throw away your printed books: a meta-analysis on the effects of reading media on reading comprehension. *Educational Research Review, 25*, 23–38. https://doi.org/10.1016/j.edurev.2018.09.003.

Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x.

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*(7109), 629–634. https://doi.org/10.1136/bmj.315.7109.629.

*Fiorella, L., & Mayer, R. E. (2016). Effects of observing the instructor draw diagrams on learning from multimedia messages. Journal of Educational Psychology, 108(4), 528–546. https://doi.org/10.1037/edu0000065.

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

Geary, D. C. (1995). Reflections of evolution and culture in children's cognition: implications for mathematical development and instruction. *American Psychologist, 50*(1), 24–37. https://doi.org/10.1037/0003-066X.50.1.24.

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). Charlotte, NC: Information Age Publishing.

*Goff, E. E., Reindl, K. M., Johnson, C., McClean, P., Offerdahl, E. G., Schroeder, N. L., & White, A. R. (2017). Variation in external representations as part of the classroom lecture: an investigation of virtual cell animations in introductory photosynthesis instruction. Biochemistry and Molecular Biology Education, 45(3), 226–234. https://doi.org/10.1002/bmb.21032.

Hays, T. A. (1996). Spatial abilities and the effects of computer animation on short-term and long-term comprehension. *Journal of Educational Computing Research, 14*(2), 139–155. https://doi.org/10.2190/60y9-bqg9-80hx-ueml.

Hedges, L. V., & Nowell, A. (1995). Sex differences in mental test scores, variability, and numbers of high-scoring individuals. *Science, 269*(5220), 41–45. https://doi.org/10.1126/science.7604277.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. San Diego, CA: Academic.

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. https://doi.org/10.1037/0278-7393.18.5.1084.

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. https://doi.org/10.1016/j.intell.2005.09.005.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558. https://doi.org/10.1002/sim.1186.

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

Höffler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction, 17*(6), 722–738. https://doi.org/10.1016/j.learninstruc.2007.09.013.

*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(1), 209–216. https://doi.org/10.1016/j.chb.2010.07.042.

*Höffler, T. N., Prechtl, H., & Nerdel, C. (2010). The influence of visual cognitive style when learning from instructional animations and static pictures. Learning and Individual Differences, 20(5), 479–483. https://doi.org/10.1016/j.lindif.2010.03.001.

*Höffler, T. N., & Schwartz, R. N. (2011). Effects of pacing and cognitive style across dynamic and non-dynamic representations. Computers & Education, 57(2), 1716–1726. https://doi.org/10.1016/j.compedu.2011.03.012.

Huedo-Medina, T. B., Sánchez-Meca, J., Marín-Martínez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: *Q* statistic or I^2 index? *Psychological Methods, 11*(2), 193–206. https://doi.org/10.1037/1082-989X.11.2.193.

*Imhof, B., Scheiter, K., Edelmann, J., & Gerjets, P. (2012). How temporal and spatial aspects of presenting visualizations affect learning about locomotion patterns. Learning and Instruction, 22(3), 193–205. https://doi.org/10.1016/j.learninstruc.2011.10.006.

*Imhof, B., Scheiter, K., & Gerjets, P. (2011). Learning about locomotion patterns from visualizations: effects of presentation format and realism. Computers & Education, 57(3), 1961–1970. https://doi.org/10.1016/j.compedu.2011.05.004.

Isacco, A., Hammer, J. H., & Shen-Miller, D. S. (2016). Outnumbered, but meaningful: the experience of male doctoral students in professional psychology training programs. *Training and Education in Professional Psychology, 10*(1), 45–53. https://doi.org/10.1037/tep0000107.

Jaffar, A. A. (2012). YouTube: an emerging tool in anatomy education. *Anatomical Sciences Education, 5*(3), 158–164. https://doi.org/10.1002/ase.1268.

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: evidence from a large, representative U.S. sample. *Psychological Science, 26*(3), 302–310. https://doi.org/10.1177/0956797614563338.

*Kalyuga, S. (2008). Relative effectiveness of animated and static diagrams: an effect of learner prior knowledge. Computers in Human Behavior, 24(3), 852–861. https://doi.org/10.1016/j.chb.2007.02.018.

*Lewalter, D. (2003). Cognitive strategies for learning from static and dynamic visuals. Learning and Instruction, 13(2), 177–189. https://doi.org/10.1016/s0959-4752(02)00019-1.

*Lin, H. (2011). Facilitating learning from animated instruction: effectiveness of questions and feedback as attention-directing strategies. Educational Technology & Society, 14(2), 31–42.

*Lin, H., & Dwyer, F. M. (2010). The effect of static and animated visualization: a perspective of instructional effectiveness and efficiency. Educational Technology Research and Development, 58(2), 155–174. https://doi.org/10.1007/s11423-009-9133-x.

*Lin, L., & Atkinson, R. K. (2011). Using animations and visual cueing to support learning of scientific concepts and processes. Computers & Education, 56(3), 650–658. https://doi.org/10.1016/j.compedu.2010.10.007.

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. https://doi.org/10.2307/1130467.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Lowe, R. K. (2003). Animation and learning: selective processing of information in dynamic graphics. *Learning and Instruction, 13*(2), 157–176. https://doi.org/10.1016/S0959-4752(02)00018-X.

*Lowe, R. K., Schnotz, W., & Rasch, T. (2011). Aligning affordances of graphics with learning task requirements. Applied Cognitive Psychology, 25(3), 452–459. https://doi.org/10.1002/acp.1712.

*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(6), 747–764. https://doi.org/10.1002/acp.1347.

Mahmud, W., Hyder, O., Butt, J., & Aftab, A. (2011). Dissection videos do not improve anatomy examination scores. *Anatomical Sciences Education, 4*(1), 16–21. https://doi.org/10.1002/ase.194.

*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(6), 2172–2178. https://doi.org/10.1016/j.chb.2013.04.035.

Matthews, W. J., Benjamin, C., & Osborne, C. (2007). Memory for moving and static images. *Psychonomic Bulletin & Review, 14*(5), 989–993. https://doi.org/10.3758/bf03194133.

Mayer, R. E. (2017). Using multimedia for e-learning. *Journal of Computer Assisted Learning, 33*(5), 403–423. https://doi.org/10.1111/jcal.12197.

*Mayer, R. E., DeLeeuw, K. E., & Ayres, P. (2007). Creating retroactive and proactive interference in multimedia learning. Applied Cognitive Psychology, 21(6), 795–809. https://doi.org/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. https://doi.org/10.1037/1076-898x.11.4.256.

McCrudden, M. T., & Rapp, D. N. (2017). How visual displays affect cognitive processing. *Educational Psychology Review, 29*(3), 623–639. https://doi.org/10.1007/s10648-015-9342-2.

*Michas, I. C., & Berry, D. C. (2000). Learning a procedural task: effectiveness of multimedia presentations. Applied Cognitive Psychology, 14(6), 555–575. https://doi.org/10.1002/1099-0720(200011/12)14:6<555::aid-acp677>3.0.co;2-4.

*Münzer, S., Seufert, T., & Brünken, R. (2009). Learning from multimedia presentations: facilitation function of animations and spatial abilities. Learning and Individual Differences, 19(4), 481–485. https://doi.org/10.1016/j.lindif.2009.05.001.

Newcombe, N. S., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles, 9*(3), 377–386. https://doi.org/10.1007/bf00289672.

Nikou, S. A., & Economides, A. A. (2016). The impact of paper-based, computer-based and mobile-based self-assessment on students' science motivation and achievement. *Computers in Human Behavior, 55*, 1241–1248. https://doi.org/10.1016/j.chb.2015.09.025.

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*(1), 27–45. https://doi.org/10.1007/s10648-011-9179-2.

*Paik, E. S., & Schraw, G. (2013). Learning with animation and illusions of understanding. Journal of Educational Psychology, 105(2), 278–289. https://doi.org/10.1037/a0030281.

*Park, O.-C. (1998). Visual displays and contextual presentations in computer-based instruction. Educational Technology Research and Development, 46(3), 37–50. https://doi.org/10.1007/bf02299760.

*Park, O.-C., & Gittelman, S. S. (1992). Selective use of animation and feedback in computer-based instruction. Educational Technology Research and Development, 40(4), 27–38. https://doi.org/10.1007/BF02296897.

*Patwardhan, M., & Murthy, S. (2015). When does higher degree of interaction lead to higher learning in visualizations? Exploring the role of 'interactivity enriching features'. Computers & Education, 82, 292–305. https://doi.org/10.1016/j.compedu.2014.11.018.

Press, C., Bird, G., Flach, R., & Heyes, C. (2005). Robotic movement elicits automatic imitation. *Cognitive Brain Research, 25*(3), 632–640. https://doi.org/10.1016/j.cogbrainres.2005.08.020.

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

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

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641. https://doi.org/10.1037/0033-2909.86.3.638.

*Sanchez, C. A., & Wiley, J. (2014). The role of dynamic spatial ability in geoscience text comprehension. Learning and Instruction, 31, 33–45. https://doi.org/10.1016/j.learninstruc.2013.12.007.

*Scheiter, K., Gerjets, P., & Catrambone, R. (2006). Making the abstract concrete: visualizing mathematical solution procedures. Computers in Human Behavior, 22(1), 9–25. https://doi.org/10.1016/j.chb.2005.01.009.

*Schmidt-Weigand, F. (2011). Does animation amplify the modality effect—or is there any modality effect at all?. Zeitschrift für Pädagogische Psychologie, 25(4), 245–256. https://doi.org/10.1024/1010-0652/a000048.

*Schmidt-Weigand, F., & Scheiter, K. (2011). The role of spatial descriptions in learning from multimedia. Computers in Human Behavior, 27(1), 22–28. https://doi.org/10.1016/j.chb.2010.05.007.

Schnotz, W., Böckheler, J., & Grzondziel, H. (1999). Individual and co-operative learning with interactive animated pictures. *European Journal of Psychology of Education, 14*(2), 245–265. https://doi.org/10.1007/bf03172968.

Schwartz, R. N., & Plass, J. L. (2014). Click versus drag: user-performed tasks and the enactment effect in an interactive multimedia environment. *Computers in Human Behavior, 33*, 242–255. https://doi.org/10.1016/j.chb.2014.01.012.

Shimada, S., & Oki, K. (2012). Modulation of motor area activity during observation of unnatural body movements. *Brain and Cognition, 80*(1), 1–6. https://doi.org/10.1016/j.bandc.2012.04.006.
*Soemer, A., & Schwan, S. (2016). Task-appropriate visualizations: can the very same visualization format either promote or hinder learning depending on the task requirements? Journal of Educational Psychology, 108(7), 960–968. https://doi.org/10.1037/edu0000093.
*Stebner, F., Kühl, T., Höffler, T. N., Wirth, J., & Ayres, P. (2017). The role of process information in narrations while learning with animations and static pictures. Computers & Education, 104, 34–48. https://doi.org/10.1016/j.compedu.2016.11.001.
Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence, 41*(5), 341–357. https://doi.org/10.1016/j.intell.2013.05.006.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory*. New York, NY: Springer.
*Swezey, R. W., Perez, R. S., & Allen, J. A. (1991). Effects of instructional strategy and motion presentation conditions on the acquisition and transfer of electromechanical troubleshooting skill. Human Factors, 33(3), 309–323. https://doi.org/10.1177/001872089103300306.
Tabachnick, B. G., & Fidell, L. S. (2018). *Using multivariate statistics* (7th ed.). New York, NY: Pearson.
*Tekdal, M. (2013). The effect of an example-based dynamic program visualization environment on students' programming skills. Educational Technology & Society, 16(3), 400–410.
*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. https://doi.org/10.1111/j.2044-8279.1990.tb00925.x.
Türkay, S. (2016). The effects of whiteboard animations on retention and subjective experiences when learning advanced physics topics. *Computers & Education, 98*, 102–114. https://doi.org/10.1016/j.compedu.2016.03.004.
Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247–262. https://doi.org/10.1006/ijhc.2002.1017.
Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402. https://doi.org/10.1037/a0028446.
VanArsdall, J. E., Nairne, J. S., Pandeirada, J. N. S., & Cogdill, M. (2015). Adaptive memory: animacy effects persist in paired-associate learning. *Memory, 23*(5), 657–663. https://doi.org/10.1080/09658211.2014.916304.
Voyer, D., & Jansen, P. (2017). Motor expertise and performance in spatial tasks: a meta-analysis. *Human Movement Science, 54*, 110–124. https://doi.org/10.1016/j.humov.2017.04.004.
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. https://doi.org/10.1037/0033-2909.117.2.250.
Wang, P.-Y., Vaughn, B. K., & Liu, M. (2011). The impact of animation interactivity on novices' learning of introductory statistics. *Computers & Education, 56*(1), 300–311. https://doi.org/10.1016/j.compedu.2010.07.011.
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. https://doi.org/10.1002/tea.3660320508.
*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. https://doi.org/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(2), 339–347. https://doi.org/10.1016/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.
Wong, M., Castro-Alonso, J. C., Ayres, P., & Paas, F. (2018). Investigating gender and spatial measurements in instructional animation research. *Computers in Human Behavior, 89*, 446–456. https://doi.org/10.1016/j.chb.2018.02.017.
*Wu, C.-F., & Chiang, M.-C. (2013). Effectiveness of applying 2D static depictions and 3D animations to orthographic views learning in graphical course. Computers & Education, 63, 28–42. https://doi.org/10.1016/j.compedu.2012.11.012.
*Zacks, J. M., & Tversky, B. (2003). Structuring information interfaces for procedural learning. Journal of Experimental Psychology: Applied, 9(2), 88–100. https://doi.org/10.1037/1076-898X.9.2.88.
Zell, E., Krizan, Z., & Teeter, S. R. (2015). Evaluating gender similarities and differences using metasynthesis. *American Psychologist, 70*(1), 10–20. https://doi.org/10.1037/a0038208.

Affiliations

Juan C. Castro-Alonso[1] **· Mona Wong**[2] **· Olusola O. Adesope**[3] **· Paul Ayres**[4] **·**
Fred Paas[5,6]

Mona Wong
wpsmona@hku.hk

Olusola O. Adesope
olusola.adesope@wsu.edu

Paul Ayres
p.ayres@unsw.edu.au

Fred Paas
paas@essb.eur.nl

1 Center for Advanced Research in Education, Universidad de Chile, Santiago, Chile

2 Faculty of Education, The University of Hong Kong, Hong Kong, Hong Kong

3 College of Education, Washington State University, Pullman, WA, USA

4 School of Education, University of New South Wales, Sydney, Australia

5 Department of Psychology, Education, and Child Studies, Erasmus University Rotterdam, Rotterdam, The Netherlands

6 Early Start/School of Education, University of Wollongong, Wollongong, Australia