



Unifying the Ability-as-Compensator and Ability-as-Enhancer Hypotheses

Tim Kühl^{1,2} · Benedict C. O. F. Fehringer¹ · Stefan Münzer¹

Accepted: 28 November 2021
© The Author(s) 2021

Abstract

Spatial abilities have been found to interact with the design of visualizations in educational materials in different forms: (1) spatial abilities enhanced learning with optimized visual design (ability-as-enhancer) or (2) spatial abilities compensated for suboptimal visual design (ability-as-compensator). A brief review of pertinent studies suggests that these two forms are viewed as mutually exclusive. We propose a novel unifying conceptualization. This conceptualization suggests that the ability-as-enhancer interaction will be found in the low-medium range of a broad ability continuum whereas the ability-as-compensator interaction will be found in the medium-high range. The largest difference in learning outcomes between visual design variations is expected for medium ability. A corresponding analytical approach is suggested that includes nonlinear quadratic interactions. The unifying conceptualization was confirmed in an experiment with a consistent visual-spatial task. In addition, the conceptualization was investigated with a reanalysis of pooled data from four multimedia learning experiments. Consistent with the conceptualization, quadratic interactions were found, meaning that interactions depended on ability range. The largest difference between visual design variations was obtained for medium ability, as expected. It is concluded that the unifying conceptualization is a useful theoretical and methodological approach to analyze and interpret aptitude-treatment interactions that go beyond linear interactions.

Keywords Aptitude-treatment-interaction · Ability-as-enhancer · Ability-as-compensator · Spatial ability · Visual design · Multimedia learning

All three authors contributed equally to this work and should be considered as first authors.

✉ Tim Kühl
tim.kuehl@uni-mannheim.de

Extended author information available on the last page of the article

Introduction

There is abundant evidence that visualizations can foster learning in educational contexts (cf. Levie & Lentz, 1982; Mayer, 2021). In processing and understanding visualizations, learners' spatial abilities are involved (cf. Hegarty & Waller, 2005). Correspondingly, substantive correlations between spatial ability measures and learning outcomes for learning with visualizations have been found, suggesting that spatial abilities play an important role in learning with visualizations (for reviews see Hegarty & Kriz, 2008; Höffler, 2010).

However, spatial abilities do not only correlate with learning outcomes, but they may interact with the design of the visualization, such as animations compared to static pictures or two-dimensional compared to three-dimensional visualizations (cf. Höffler, 2010). In this regard, two aptitude (spatial ability) \times treatment (design of visualization) interaction hypotheses are stated (cf. Mayer & Sims, 1994): the ability-as-enhancer hypothesis and the ability-as-compensator hypothesis. According to the ability-as-enhancer hypothesis, a certain amount of spatial ability is necessary to profit from an optimized visualization. This means that the potential beneficial effect of an optimized visualization compared to a suboptimal visualization will not unfold when learners' spatial abilities are low. The ability-as-compensator hypothesis posits that only learners with higher spatial abilities, but not learners with lower spatial abilities can compensate for a suboptimal (e.g., static) visualization, while this compensating effect of spatial abilities is not necessary when learners receive an optimized (e.g., dynamic) visualization.

The ability-as-enhancer and ability-as-compensator hypotheses are viewed as mutually exclusive. In contrast, we assume on a conceptual level that the two forms of the aptitude-treatment interactions are not necessarily exclusive but can be united — by considering them as two ends of an ability continuum. We introduce an analytical approach that corresponds to our conceptualization. We then present an original empirical study and a reanalysis of data from four multimedia learning studies that show first empirical evidence for the unifying view.

In the following, we will first provide a brief review concerning the empirical evidence with respect to the different roles spatial abilities can play in learning with visualizations, specifically with respect to the ability-as-enhancer and the ability-as-compensator hypotheses. Thereafter, we will introduce the unifying conceptualization of the two hypotheses.

A Brief Review of Empirical Results Concerning the Role of Spatial Ability in Learning with Visualizations

Visualizations are an important source of information in instructional materials. Visualizations can either be the only source of information or they can be combined with other sources, such as verbal explanations. In multimedia instructions, learners benefit from the combination of visualizations with text compared with learning from text alone (multimedia effect, cf. Levie & Lentz, 1982; Mayer, 2021), because

visualizations convey information that cannot be easily understood or inferred from text (e.g., Kühl et al., 2018; Larkin & Simon, 1987). In order to understand and profit from a visualization, the visualization needs to be adequately processed by a learner.

Learners' spatial abilities are involved in processing visualizations. Even though a comprehensive definition of spatial ability is not available, it is widely agreed on that spatial abilities are related to mentally performing processes such as apprehending, encoding, remembering, or mentally transforming visual-spatial objects (cf. Hegarty & Waller, 2005; Uttal et al., 2013). Spatial abilities are not considered as a unitary construct but consist of a set of related but different abilities (cf. Carroll, 1993; Hegarty & Waller, 2005). On the one hand, there is a factor-analytic tradition that is rooted in intelligence testing. On the other hand, there is a cognitive research tradition, including investigations of the structure and processes of working memory (see Hegarty & Waller, 2005, for a review of both traditions). In the factor-analytic tradition, different sub-factors of spatial ability have been described. Lohman (1988) distinguished three spatial factors: (1) "speeded rotation" of simple items, (2) "spatial orientation" (involving perspective change), and (3) "spatial visualization" (referring to complex materials and sequences of transformations). Carroll (1993) identified five main factors based on extensive data sets: (1) "visualization" involving complex and multi-step spatial transformations, (2) "spatial relations" requiring mental rotation with simple figures under speeded instruction, (3) "closure speed" requiring quick recognition of incomplete figures, (4) "closure flexibility" involving identification of hidden figures in complex spatial patterns, and (5) "perceptual speed" requiring speeded comparisons of simple figures. Measures addressing the factor "visualization" (complex materials, multiple transformation steps) measure a rather broad notion of spatial ability.

In the cognitive-processing tradition, measures of working memory (WM) capacity are commonly used as predictors of individual differences. Such measures can be specific for different codes of information, such as verbal and visual-spatial information — which is for instance accounted for in Baddeley's working memory model (Baddeley, 1986) that distinguishes a phonological loop and a visuo-spatial sketchpad (i.e., visuo-spatial working memory). Measures of visuo-spatial working memory predict performance in visuo-spatial tasks (e.g., Miyake et al., 2001). Individual differences in mental spatial processing, to which we refer with the broad term of spatial ability in the following, can be addressed with measures both from the factor-analytic tradition as well as from the cognitive-processing tradition (cf. Hegarty & Waller, 2005).

Spatial abilities play a crucial role in many domains in which the understanding of visualizations is important, such as in STEM domains (science, technology, engineering, and mathematics), in architecture, or in geography. Accordingly, studies have repeatedly demonstrated that individual differences in spatial abilities play a decisive role as a predictor of learning success with visualizations, for instance, in STEM domains (Allen et al., 2019; Wai et al., 2009), in medical training (Cohen & Hegarty, 2007; Garg et al., 1999; Keehner et al., 2006), or in route and layout learning about large-scale spaces from virtual environments and videos (Münzer & Stahl, 2011; Münzer & Zadeh, 2016; Hegarty et al., 2006; Waller, 2000).

Visualizations may demand higher or lower cognitive engagement in mental spatial processing, depending on their design. For instance, static pictures often — but not necessarily always — require inferring spatial transformations mentally, whereas dynamic visualizations (animations) can show spatial transformations explicitly. In such cases, dynamic visualizations compared to static visualizations may facilitate spatial understanding because the external presentation can support otherwise effortful mental visual-spatial processing (supplantation, Salomon, 1994; facilitation function, Schnotz & Rasch, 2005). Under such circumstances, the dynamic visualization may be considered as the optimized visualization and the static visualization as the suboptimal visualization. However, it should be noted that dynamic visualizations cannot not automatically be considered as the optimized and the static visualization as the suboptimal design; rather, which visualization format can be considered as the optimized one, and which as the suboptimal one, depends on the processes that dynamic and static visualizations require and support, respectively (Ploetzner et al., 2020).

Höffler (2010) reviewed the role of spatial abilities in multimedia learning with visualizations, considering 27 experiments from 19 primary studies published between 1994 and 2009. The meta-analytic review suggested a general advantage of higher spatial ability learners over lower-ability learners. Moreover, in this review, the potentially moderating role of spatial abilities in learning with different visualization formats was investigated. It was found that learners with lower spatial ability profited from dynamic visualizations (animations) compared to static visualizations, whereas differences between dynamic and static visualizations were less pronounced for learners with higher spatial ability. These results are in line with the ability-as-compensator hypothesis (Höffler, 2010). It should be noted though that in five of the reviewed studies, no significant interaction between spatial ability and treatment condition was found (Hannafin et al., 2008; Hegarty et al., 2003; Massa & Mayer, 2006; Wender & Muehlboeck, 2003; Westerman, 1997), and that in four primary studies the ability-as-enhancer interaction was actually obtained (Diaz & Sims, 2003; Huk, 2006; Huk & Steinke, 2007; Mayer & Sims, 1994).

More recent studies not yet included in this meta-analytic review have been similarly inconclusive. On the one hand, several studies found empirical evidence that is consistent with the ability-as-compensator hypothesis. In these studies, empirical evidence was observed (on at least one learning outcome measure) that animations compared to static pictures supported especially individuals with lower spatial abilities in multimedia learning, but not with higher spatial abilities (e.g., Kühl et al., 2018b; Barrett & Hegarty, 2016; Berney et al., 2015; Bogomolova et al., 2020; de Back et al., 2020; Höffler & Leutner, 2011; Lee & Wong, 2014; Sanchez & Wiley, 2014). Similarly, in studies examining spatial transformations needed for the understanding of three-dimensional spatial structures from two-dimensional visualizations, empirical evidence was found that students' performances were less dependent on the respective spatial ability if animations of the spatial transformation were shown, compared to static pictures (Münzer, 2012, 2015). In these studies, the visualizations were not accompanied by text. The visual designs and the corresponding spatial ability tests focused on only one selected spatial transformation process.

On the other hand, there were also at least two studies where the pattern of results corresponded to the ability-as-enhancer interaction (on at least one learning outcome measure). In one study about spatial anatomy comprehension, particularly participants with higher spatial abilities profited from the sequence “first static, then dynamic visualizations” compared to the sequence “first dynamic, then static visualizations,” but not participants with lower spatial abilities (Wu et al., 2013). In another study in the domain knot tying (de Koning et al., 2019), only students with higher spatial abilities performed better when they received a dynamic visualization compared to a static visualization, while students with low spatial abilities performed equally well when receiving a dynamic compared to a static visualization. In the first study by Wu et al. (2013), the visualizations were accompanied by text, whereas the visualizations in the study by de Koning et al. (2019) were not accompanied by text.

With the meta-analytic review by Höffler (2010), there were also studies that explicitly investigated the interaction between spatial ability and static compared to dynamic visualizations, but observed no interaction (Kühl et al., 2011b; Brucker et al., 2014; Chen et al., 2015; Imhof et al., 2011; Imhof et al., 2012; Wang & Tseng, 2020). It should also be noted that in several studies, it was not reported (since it was not in the focus of the research questions) whether an interaction between spatial abilities and type of visualization was observable (e.g., Kühl et al., 2011a; Kühl et al., 2012; Kühl et al., 2018a; Kühl & Münzer, 2019; Castro-Alonso et al., 2014; Castro-Alonso et al., 2018; Garland & Sanchez, 2013; Paik & Schraw, 2013; Stebner et al., 2017).

Summing up, the empirical evidence regarding the interplay of spatial abilities in learning with different visualization formats is mixed. It seems far from obvious why one would predict either an ability-as-compensator or an ability-as-enhancer-interaction. In current research, the two hypotheses are viewed as mutually exclusive.

A Unifying Conceptualization of the Ability-as-Enhancer and Ability-as-Compensator Hypotheses

We argue that the two hypotheses are not necessarily mutually exclusive, but propose a unifying conceptualization of the ability-as-enhancer and the ability-as-compensator hypotheses as two different interaction forms (Fig. 1): These two different forms of interaction will be found if comparisons are made between different points of the ability continuum. At the point of *low ability*, the potential of an optimized visualization might not unfold, because learners lack resources to profit from an optimized visualization. If the level of ability, however, is appropriate relative to task demands such that attempts to perform the task *can* be successful (*medium ability*), then the optimized visualization might have the intended supportive effect. The ability-as-enhancer interaction would result from the comparison between these two ability points (i.e., between low and medium ability). That is, an ability-as-enhancer interaction will be obtained if low-ability learners are compared with medium-ability learners.

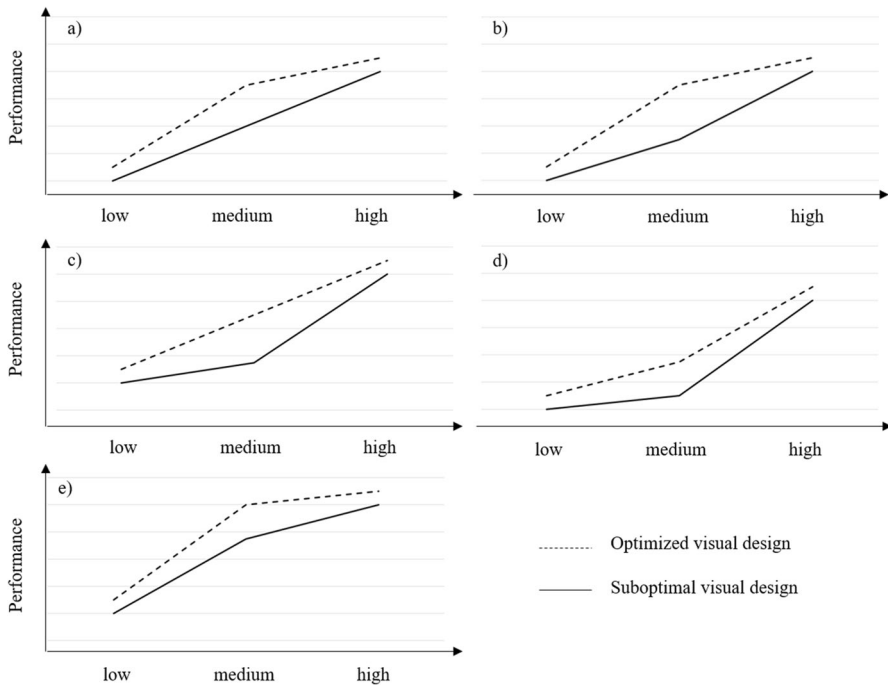


Fig. 1 Illustrations of the conceptualization of the ability-as-enhancer and the ability-as-compensator aptitude-treatment-interaction depending on ability level (low, medium, high), considering a broad range of ability relative to the task demand. The conceptualization may be reflected in different actual forms of regression lines (a–e), but all forms represent the same principle, the unifying conceptualization. The ability-as-enhancer interaction would result from comparing the low-to-medium ability range; the ability-as-compensator interaction would result from comparing the medium-to-high ability range

The medium-ability learners, in contrast to the low-ability learners, can make use of the optimized design to perform successfully. Ability is thus thought to function as an enhancer given the optimized visualization. Thereby, the regression line for the relation between ability and performance is steeper for the optimized visualization compared to the suboptimal visualization. In our illustrations of the conceptualization provided in Fig. 1, this comparison can be found between the “low” and the “medium” ability levels. For all illustrated forms of the unifying conceptualization depicted in Fig. 1a–e, it is true that the regression line for the optimized visual design is steeper than the regression line for the suboptimal visual design between the low and medium ability levels. This reflects the ability-as-enhancer interaction.

If the ability increases more (*high ability*), then a learner may have appropriate mental resources available to perform successfully, may the visual design be optimized or not. The ability-as-compensator hypothesis will be obtained if high-ability learners are compared with medium-ability learners. Here, the comparison includes learners who perform well regardless of the design (high ability) and learners for whom the optimized design is supportive (medium ability).

In particular, high-ability learners perform the task well also in the suboptimal design variant, in contrast to the medium-ability learners who experience difficulties with the suboptimal design. High ability is thus considered to be compensating for the suboptimal design. Thereby, the regression line for the relation between ability and performance is flatter for the optimized visualization compared to the suboptimal visualization. In our illustrations of the conceptualization shown in Fig. 1, this comparison can be found between the “medium” and the “high” ability levels. For all illustrated forms of the unifying conceptualization depicted in Fig. 1a–e, it is true that the regression line for the optimized visual design is flatter than the regression line for the suboptimal visual design between medium and high ability levels. This reflects the ability-as-compensator interaction.

To conclude, the two different forms of the aptitude-treatment interaction between spatial ability and visual design are not mutually exclusive but can be united. The present conceptualization integrates the at first-glance exclusive ability-as-enhancer and ability-as-compensator hypotheses into one superordinate model. The unifying conceptualization implies that a difference between an optimized design and a suboptimal design would be most pronounced for medium-ability learners, but less accentuated for low-ability or high-ability learners.

Analytical Approach for the Unifying Conceptualization

The unifying conceptualization is reflected in a corresponding analytical approach. The regression model treats the aptitude variable as a continuous predictor (our verbal descriptions and the illustrations in Figure 1 refer to low, medium, and high ability ranges for illustrative purposes only). The challenge is to define one single model that is able to cover all corresponding regression lines. Without violation of the underlying principle of the conceptualization, the regressions of the suboptimal and optimized ability design can be linear, u-shaped, or inverted u-shaped in different combinations (see Fig. 1). The regression model to cover a linear regression line for the suboptimal visual design and an inverted u-shaped optimized visual design regression line (Fig. 1a) should contain a dummy variable coding “suboptimal” as “0” and “optimized” as “1” as well as a quadratic interaction term:

$$\text{perf} = b_0 + b_1 \text{abi} + b_2 \text{dumOpt} + b_3 \text{abi} \times \text{dumOpt} + b_4 (\text{abi} \times \text{dumOpt})^2 \quad (1)$$

with “perf” indicating performance, “abi” ability, “dumOpt” the dummy variable, and b_i the regression weights. In this case, b_1 would indicate the slope of the suboptimal linear regression. Since the interaction term itself is not linear (there is a positive interaction from low to medium and a negative interaction from medium to high), the interaction has an inverted u-shaped progression, and this should be indicated by a negative b_4 coefficient.

In the case of a linear progression with the optimized visual design and a u-shaped progression with the suboptimal visual design (Fig. 1c), the dummy variable should be reverse coded with “0” indicating “optimized” and “1” indicating “suboptimal”:

$$\text{perf} = b_0 + b_1\text{abi} + b_2\text{dumSub} + b_3\text{abi} \times \text{dumSub} + b_4(\text{abi} \times \text{dumSub})^2 \quad (2)$$

In contrast to the previous case, b_4 should be positive indicating a u-shaped interaction (a negative interaction from low to medium and a positive interaction from medium to high).

However, the three other cases (Fig. 1b–e) could not be covered with this multiple regression equation as in these cases both (suboptimal and optimized) conditions have non-linear progression lines. To solve this issue, it was decided to include two dummy variables in the regression equation, one indicating the optimized visual design (as dumOpt , see Eq. 1) and one indicating the suboptimal visual design (as dumSub , see Eq. 2):

$$\text{perf} = b_0 + b_1\text{abi} + b_{21}\text{dumOpt} + b_{22}\text{dumSub} + b_{31}\text{abi} \times \text{dumOpt} + b_{32}\text{abi} \times \text{dumSub} + b_{41}(\text{abi} \times \text{dumOpt})^2 + b_{42}(\text{abi} \times \text{dumSub})^2 \quad (3)$$

Since dumOpt and dumSub are perfectly colinear, the coefficients b_{21} and b_{22} as well as b_{31} and b_{32} are redundant. Therefore, one of each pair can be removed from Eq. 3. Note that the quadratic terms express the non-linear regression specific for each visual design and they are, therefore, not redundant. After removing (arbitrarily) the b_{22} and b_{32} terms, the final multiple regression model results:

$$\text{perf} = b_0 + b_1\text{abi} + b_2\text{dumOpt} + b_3\text{abi} \times \text{dumOpt} + b_{41}(\text{abi} \times \text{dumOpt})^2 + b_{42}(\text{abi} \times \text{dumSub})^2 \quad (4)$$

The relevant regression coefficients are b_{41} and b_{42} indicating the specific progression of the regression lines presented in Fig. 1: (a) $b_{41} < 0$ and $b_{42} = 0$; (b) $b_{41} < 0$ and $b_{42} > 0$; (c) $b_{41} = 0$ and $b_{42} > 0$; (d) $b_{42} > b_{41} > 0$; (e) $b_{41} < b_{42} < 0$. However, these five different combinations of the two visual design regression lines all correspond to the same underlying principle of the unifying conceptualization and can be summarized by the simple hypothesis “ $b_{42} > b_{41}$ ” indicating a higher benefit from the optimized visual design than from the suboptimal visual design in the medium ability range compared to the low and high ability range.

The Present Study: Aims and Research Questions

The aim of the present study is to empirically substantiate the theoretical conceptualization of a unifying view of the ability-as-compensator and the ability-as-enhancer hypotheses. Therefore, an experiment (Study 1) was conducted as well as data from a larger research project with four experiments were reanalyzed (Study 2).

Study 1 focused on performing a specific spatial task with static and dynamic visual designs. The task requires to mentally integrate symbolic information in a spatial structure. This task resembles a basic spatial process in learning about complex spatial structures from visualizations with different zoom levels. The variation of visual design was realized using a within-subjects experimental design. For this task, the dynamic visual design can be considered as the optimized and the static visual design as the suboptimal visual design (for more details, see the “Study 1: Experimental Approach” section). No additional text

accompanied the visualizations. A measure of spatial ability was selected that is rooted in the cognitive-processing tradition and measures the capacity of visual-spatial working memory (VSWM). This measure requires to integrate spatial information and varies the amount of to-be-rehearsed information in VSWM as a span measure. These features match mental processes of the experimental task. This experiment was thus designed to study the interactions between spatial ability and visual design variants in a focused and controlled way.

However, the advantage of high internal validity of the experiment might come along with lower ecological validity for learning about more complex, meaningful phenomena. Therefore, we reanalyzed in Study 2 pooled data from four published experiments with the same multimedia learning materials and examined whether the proposed unifying conceptualization would also be observable in multimedia learning with meaningful learning materials. The four experiments had been conducted in the context of a larger research project on learning with static and dynamic visual designs, at which the dynamic visual design can be considered as the optimized visual design (for more details, see the “Study 2: Reanalysis of Four Multimedia Experiments” section). In all of the experiments, multimedia learning materials consisting of a visualization and written text were presented, and in all of the experiments, the same variation of visual design (static picture vs. animation) was realized. Spatial ability was measured with the same ability test in all four experiments. This ability test is rooted in the factor-analytic tradition and measures the broad spatial factor of visualization. The measure is most commonly used in multimedia learning experiments (cf. Höffler, 2010). The data were collected in laboratory setting as well as school settings. The reanalysis thus contributes to examining the ecological validity of the unifying conceptualization.

The main hypothesis for the experiment as well as the reanalysis of the existing multimedia learning data addressed the unifying conceptualization of the ability-as-enhancer and -compensator hypotheses: The regression lines reflecting the relation between ability and task performance will not only differ for the suboptimal (here: static) visual design and the optimized (here: dynamic) visual design, but also change as a function of ability. In particular, the regression line for the optimized dynamic design will be steeper than the regression line for the suboptimal static design in the low-to-medium ability range, whereas in the medium-to-high ability range, the regression will be flatter for the dynamic compared to the static design. As a consequence, the largest difference between suboptimal static design and optimized dynamic design will be found for medium ability. If learners’ spatial abilities are low, then their attempts to perform the task might not be very effective, regardless of the design of the visualization. If learners’ spatial abilities are in a medium range, then their attempts to perform the task will be supported through the optimized, dynamic visual design, compared with the static design. A consideration of learners in the low-medium ability range would thus result in an ability-as-enhancer interaction. If learners’ spatial abilities are high, then performing the task will be successful, regardless of the design of the visualization. Thus, a consideration of learners in the medium-high ability range would result in an ability-as-compensator interaction.

Study 1: Experimental Approach

The goal of this experiment is to examine the aptitude-treatment interaction between spatial ability and static versus dynamic visual designs with a visual-spatial task. This task is termed mental integration task. It requires to link symbolic detail information with spatial locations, as well as to memorize and to integrate these combinations in a simple spatial structure. The mental integration task is thought to reflect common visual-spatial processes when learning about spatial structures required in domains such as anatomy, mechanical systems, architecture, or geography. Information obtained from two-dimensional visualizations of complex spatial structures depends on the zoom level. If the zoom level is adjusted such that details can be viewed, then the overall structure is not visible at the same time. If the zoom level is adjusted such that the overall structure can be overviewed, then the details are not visible. Thus, a learner has to mentally integrate detail-location combinations in order to know about the details at particular spatial locations.

The dynamic variant of the present mental integration task supports the formation of the link between the detail information and its location through an animation. In the static variant, the link has to be formed mentally from two separate static pictures. Thus, the dynamic visual design is considered to be the optimized visualization, and the static visualization as the suboptimal visualization. In both variants of the task, learners have to maintain a varying number of such links in integrated form. The integration is particularly important because memory is tested with regard to the integrated form in the mental integration task (see the “Method” section of Study 1 for details).

Corresponding to the mental integration task, the capacity of visual-spatial working memory (VSWM) is the spatial ability measure in the experimental study. VSWM functioning and capacity is considered an important and central facet of spatial ability (Hegarty et al., 2006; Hegarty & Waller, 2005; Miyake et al., 2001). More specifically, VSWM has been appraised as a key factor in learning about spatial configurations and structures, because different parts of the configuration have to be maintained and mentally integrated (Münzer, 2012; Hegarty et al., 2006). Accordingly, the present measure of VSWM capacity requires to imagine putting subsequently presented symbols (consecutive letters of the alphabet) into adjacent cells of an imagined spatial matrix. The separately presented pieces of information form an integrated, meandering sequence through the matrix. In previous studies with samples drawn from the same student population, this measure of VSWM capacity showed a wide range of individual differences (Münzer, 2012; Münzer et al., 2018). In a previous study, this measure of VSWM capacity was related to the static variant of the mental integration task with $r = .40$ (Münzer et al., 2018). Furthermore, this measure of VSWM capacity predicted spatial layout learning from direct navigation experience in a real environment (Münzer et al., 2012).

The items of the mental integration task vary with respect to the number of to-be-integrated detail-location-combinations in order to provide some fluctuation in

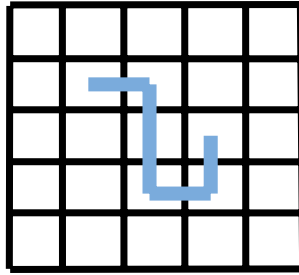


Fig. 2 Mental pathway span task with an illustration of a possible path of adjacent cells through the 5×5 matrix to be imagined and memorized. The starting cell was always the same cell (second row, second column from top-left). The shown example would represent the easiest level (span = 5). Participants were tested up to a span of 13

demands between items. However, all items require the same mental processes. Performance in the mental integration task is ultimately indicated by the mean accuracy across all presented items.

Method

Participants

All participants ($N = 79$) were students of a German University. They received course credit for their participation. Two participants were excluded because they were not able to solve the easiest level of the measure of VSWM capacity (the mental pathway span task, see below). The final sample consisted of $N = 77$ (57 female, 20 male) participants. Their mean age was $M = 21.51$ years ($SD = 2.38$ years) and ranged from 18 to 32 years.

Materials

The capacity of visual-spatial working memory was measured by means of a mental pathway span task, a variant of the Brooks task (Brooks, 1967; Quinn & Ralston, 1986) with computer-based administration. The task required the mental imagination of an empty 5 × 5 matrix. Participants listened to a sequence of verbal-auditory assignments via headphones. These assignments described a path of adjacent cells through the imagined matrix (Fig. 2). The indicated cells should mentally be filled with characters in alphabetical order, with the start-cell always located in the second column, second row (from top-left) containing an “A.” The assignments thus instructed participants to imagine putting the characters in adjacent cells of the matrix (e.g., “in the cell to the left put a ‘B’”; “in the cell above put a ‘C’”). After the description of the path, participants clicked the filled cells on a 5 × 5 matrix shown on the computer screen. (The particular letters themselves did not play a particular role in the task and had not to be memorized, because they followed the

alphabetical order.) The number of assignments to be imagined (“span level”) varied between five and 13 (maximum level). The present task does not change the size of the matrix, but only the length of the sequence (in contrast to variants used by Bosco et al., 2004, and Vecchi & Cornoldi, 1999). For each span level, there were three trials. The test started with a span level of eight, corresponding to average performance found in previous studies (Münzer, 2012). If at least two out of three trials of a level were correct, then the level was increased. If only one trial out of three trials was correct, the level was decreased. The test terminated if none of the three trials of the level was correct. Particular span levels were only tested once. That means if someone failed in a specific level or was successful, this participant would not repeat this level again. This adaptive procedure aimed at finding the participant’s highest span level that she/he could achieve, starting with the level corresponding to average performance. The mental pathway span of a particular participant corresponds to the level in which at least two out of three paths were correctly remembered.

In the mental integration task, participants were asked to mentally integrate a series of detail and location information pieces within a larger spatial structure. The spatial structure had the simple form of a 3×3 matrix, with each cell representing a room. Within a room, a symbol could be found on a wall (Fig. 3a). These symbols were not visible when looking at the 3×3 matrix structure. The combination of a certain location (a cell in the matrix) and the detail information to be found there (the symbol on the wall) was provided as two separate static pictures (static visual design condition) or as an animation (dynamic visual design condition). In the static visual design condition, the cell was indicated in the matrix by a red square and the symbol was shown separately (Fig. 3a). In the dynamic visual design condition, an animation showed a continuous zoom starting from an “aerial” view on the 3×3 matrix, then focusing on the location cell and finally “flying” into the cell to view the wall with the symbol on it. The animation thus started with the “location” picture and ended with the “detail” picture of the static version. It linked them with continuous zoom and camera movement. One item consisted of two to five such location-symbol combinations. These were shown subsequently in the study phase. Participants were asked to memorize all locations with their corresponding detail information. After the study phase of a trial, a test picture appeared which showed an integrated partial structure (part of the matrix) with visible symbols. This partial structure (Fig. 3b) always consisted of three adjacent cells (rooms) and contained two of the locations and symbols presented in the study phase. In addition, four alternative pictures showed different possible positions of the partial structure within the complete structure (matrix). Only one of the four alternatives matched simultaneously all the location-detail information presented in the study phase. Participants indicated as accurately and quickly as possible which of the four alternatives was the position of the partial structure within the overall structure. Static and dynamic conditions were blocked. In each visual design condition (static, dynamic) and for each span length (2, 3, 4, 5), there were eight items, resulting in 32 (4×8) items per visual design condition. Participants thus received a block of 32 static tasks and a block of 32 dynamic tasks. The order of these blocks varied between participants. Within a visual design condition, items were randomized. The presentation was system-controlled. Participants were not able to stop or repeat the presentation. The

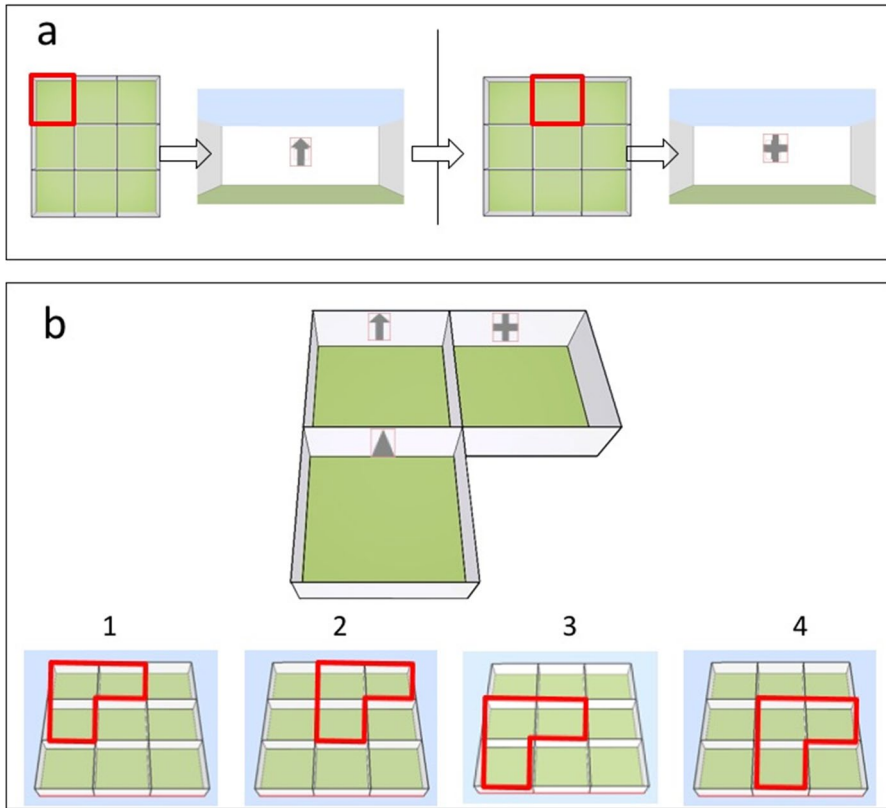


Fig. 3 Mental integration sample trial (static visual design version). The location and the detail information at that location were provided as separate, successive pictures (a). After several (two up to five) such combinations, the integrated test picture appeared (b). Participants decided which of the four indicated selections represent the position of the partial integrated structure shown above

presentation time was 6 s for the presentation of one link between a location–detail combination of the spatial structure. Correspondingly, one dynamic zoom video had a duration of 6 s. Because of the zooming-in, the complete spatial structure was visible when the video started but disappeared during the animation. In the static condition, there were two static pictures, one showing the location in the spatial structure and one showing the detail. Each picture was shown for 3 s. Thus, presentation times were the same in both conditions. The mental integration task was presented by the experimental software E-Prime Version 2.0 (Psychology Software Tools Inc., 2012).

Procedure

First, the mental pathway span task was conducted. Then, the mental integration task followed. Finally, demographic data were collected. The participants were tested in

groups of up to six persons in an experimental laboratory with separation panels between the desks.

Analytical Approach for Study 1

Spatial ability was measured with the mental pathway span task which is a measure of visual-spatial working memory capacity. The analyses included the mental pathway span as a continuous predictor variable. It was z -standardized across all participants. Visual design (static vs. dynamic) was included as a categorical within-subject factor in the analyses using two dummy variables (see the “Analytical Approach for the Unifying Conceptualization” section). The two-factor interaction was tested with a multi-level-model (MLM) with two levels (visual design was on Level 1 and subjects on Level 2).

According to the main hypothesis, the relationship between spatial ability (measured with the mental pathway span) and performance in the mental integration task (mean accuracy of the location-detail combinations integrated and remembered correctly over all trials) should correspond to the following pattern in this experiment: The benefit of the dynamic visual design compared to the static visual design should increase with mental pathway span in the low-to-medium range of mental pathway span, meaning that the slope of spatial ability and mental integration performance would be comparatively steeper for the dynamic compared to the static visual design (cf. left sides of Fig. 1a–e). However, this benefit should decrease for greater mental pathway span (i.e., in the medium-to-high range of mental pathway span), meaning that the slope of the relation of spatial ability and mental integration performance would be comparatively flatter for the dynamic compared to the static visual design (cf. right sides of Fig. 1a–e). As a result, differences between static and dynamic visual designs should be highest for participants with a medium mental pathway span compared to participants with either low or high mental pathway spans.

To account for different possible linear and non-linear forms of the regression lines that all correspond to the unifying conceptualization (illustrated in Figure 1a–e), the regression model included two dummy variables for coding the visual design. As explained above (in the “Analytical Approach for the Unifying Conceptualization” section), coding the treatment with one treatment variable (e.g., static = 0, dynamic = 1) and including a corresponding quadratic interaction term would make specific predictions about the linear and the non-linear regressions. With the coding static = 0, dynamic = 1 for the dummy variable, a linear regression would be predicted for the static visual design and a non-linear would be predicted for the dynamic visual design (with a coding static = 1, dynamic = 0, reverse predictions would be made). Therefore, a second dummy variable for visual design was included (as described above, Eq. 4): The two dummy variables for coding the visual design condition were “dumOpt” (0 = static, 1 = dynamic) and “dumSub” (0 = dynamic, 1 = static).

The linear interaction was considered by the interaction term of the first dummy variable with mental pathway span. Note that the second dummy variable was only considered for its corresponding quadratic interaction term, but not for the main effect or for the linear interaction term, as both dummy variables are perfectly

Table 1 Descriptive statistics of the mental pathway span and the mental integration task (separated for static and dynamic visual design). Acc = accuracy. RT = reaction times (in milliseconds)

	<i>M</i>	<i>SD</i>	Min	Max	Skewness	Kurtosis
Mental pathway (span)	8.05	2.01	5	13	0.46	-0.47
Mental integration task static (acc)	.75	.22	.12	1.00	-1.15	0.43
Mental integration task dynamic (acc)	.76	.22	.16	1.00	-1.12	0.10
Mental integration task static (RT)	6332	2644	2519	18274	1.63	4.41
Mental integration task dynamic (RT)	6291	3050	2478	17238	1.86	4.24

complementary (see the “Analytical Approach for the Unifying Conceptualization” section).

Follow-up illustrative analyses were conducted comparing performances in the mental integration task between the static and the dynamic visual designs for low-ability, medium-ability, and high-ability participant groups separately. For these analyses, three distinct ability groups were built: a group of low-ability participants with a span < 8, a group of medium ability participants with a span of 8 and 9, and a group of high-ability participants with a span > 9. These span values were based on the known distribution of the mental pathway span from two previous studies. One study (Münzer et al., 2018) comprised university students, as in the present study. The other study (Münzer, 2012) involved $N = 152$ younger students from secondary schools who had participated in two experiments on real-world spatial learning. These students thus belonged to a population with a presumably broader distribution of spatial ability. Data from both studies show that mental pathway spans of 8 and 9 formed the medium ability category.

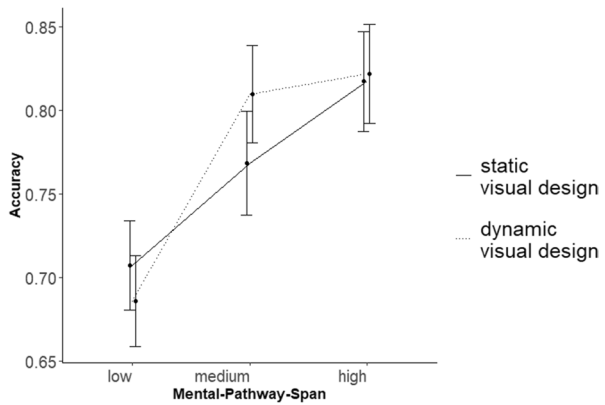
Results and Discussion

Table 1 shows the descriptive results of the mental pathway task (span) and the mental integration task in the static and dynamic visual design (accuracy). Apparently, there was no general advantage of the dynamic visual design (mean accuracy = 0.76) for the accuracy in the mental integration task, compared with the static visual design (mean accuracy = 0.75). Results show a moderate right-skewed distribution of the mental pathway span task and left-skewed distributions of the mental integration task with static and dynamic visual design. The maximum span of the mental pathway task of 13 was only reached by two participants.

Table 2 shows the results of the fixed effects of the analyzed multi-level model (MLM, see the “Method” section for Study 1). The significant quadratic terms for the interaction of mental pathway span with visual design, dynamic and static, b [(mental pathway span \times dumOpt)²] = $-.08$, $p < .001$ and b [(mental pathway span \times dumSub)²] = $-.04$, $p = .036$ (Table 2), indicate non-linear relations with negative quadratic trends for both presentation forms. The negative sign indicates an inverted U-progression (see also Fig. 1e). However, the effects were significantly different ($z = -2.60$, $p = .009$), indicating that the non-linear and inverted U-progression was

Table 2 Results for accuracy of the fixed effects of the MLM

	<i>b</i>	df	<i>t</i> -value	<i>p</i>
Intercept	.79	75	26.72	< .001
Visual design (static vs. dynamic)	.04	73	1.81	.075
Mental pathway span	.08	75	3.30	.002
Mental pathway span × visual design	.02	73	1.40	.164
(Mental pathway span × dumOpt) ²	−.08	73	−3.98	< .001
(Mental pathway span × dumSub) ²	−.04	73	−2.14	.036

Fig. 4 Mean accuracy in the mental integration task, separated for static vs. dynamic visual design. Ability (VSWM capacity as indicated by mental pathway span) is the (continuous) predictor of mean accuracy

more strongly accentuated for the dynamic visual design compared to the static visual design. Correspondingly, this reflects a higher mental integration performance gain for the group in the medium ability range with the dynamic compared to the static visual design. The main effect of mental pathway span indicates the positive relation between mental pathway span and mental integration performance, b (mental pathway span) = .08, $p < .002$.

For illustrative purposes, participants were assigned to three ability groups, as described above. There were $n = 35$ in the low ability group (span < 8); $n = 22$ in the medium ability group ($8 \leq \text{span} \leq 9$); and $n = 20$ in the high-ability group (span > 9). Fig. 4 illustrates the differences in the progressions for static vs. dynamic visual design with regard to the relation between mental pathway span and mental integration accuracy for the three ability groups. For the dynamic visual design, the relation between mental pathway span and mental integration accuracy is best described with an inverted U-shaped, nonlinear progression, while for the static visual design, this nonlinear progression is far less pronounced (and hardly discernible in Fig. 4). This pattern corresponds to the principle of the unifying conceptualization. There was a steeper accuracy gain from the low-to-medium ability group in the condition of the dynamic visual design compared to the static visual design. This pattern reflects the ability-as-enhancer effect. The ability-as-compensator effect can be found between the medium and high-ability group. There was a lower accuracy

gain in the condition of dynamic visual design compared to the static visual design (Fig. 4).

Additional analyses compared the accuracy in the mental integration task between the static and dynamic design for each ability group separately (the descriptive data are depicted in Fig. 4). For the low-span group, performance did not differ between the static and the dynamic visual design, $t(34) = 0.73$, $p = .765$, Cohen's $d = 0.09$, $CI = [-0.39, 0.56]$. The same was true for the high-span group, $t(19) = 0.17$, $p = .433$, Cohen's $d = 0.03$, $[-0.61, 0.67]$. The difference between static and dynamic visual design just failed to reach significance for the medium-span group, $t(21) = 1.60$, $p = .062$, Cohen's $d = 0.20$, $[-0.41, 0.81]$. For the medium-span group, a descriptively higher difference between visual design conditions in terms of effect size was found, compared to the low-span or the high-span groups.

The results of the experiment — and especially our analytical approach to examine the overall pattern with nonlinear regressions — revealed first empirical evidence for our proposed unifying conceptualization of the ability-as-enhancer and the ability-as-compensator hypotheses. Dividing the sample in three ability groups served primary for illustrative purposes: Also, these results speak — admittedly only descriptively — for the notion that the optimized visual design especially supported performing the mental integration task if learners' ability was in the medium range, but not if learners' ability was in the lower or higher ability range. However, the effect for learners in the medium ability range failed to reach statistical significance and was only marginal. The fact that we observed an overall significant effect with the nonlinear regression approach but failed to illustrate this effect in terms of significance by means of dividing the sample in subgroups and conducting t -tests may be attributable to a power problem (Irwin & McClelland, 2003). It is important to note that the latter results of the t -tests for each subgroup do not speak against the unifying conceptualization (but descriptively support it), since these t -tests were primarily conducted for illustrative purposes. The major analysis that drives the claim the data support the unifying conceptualization is the nonlinear regression approach. Hence, it can be concluded that the unifying conceptualization (including both the enhancer and the compensator interaction) was confirmed.

In Study 2, we wanted to substantiate our findings from Study 1, and examine whether the proposed unifying conceptualization of the aptitude-treatment-interaction would also be observable with more complex and meaningful multimedia learning materials that include visualizations and text. Meaningful multimedia learning materials that require learners to form a cognitive schema given inter-related pieces of information presented in visual and verbal codes might change the role of spatial ability. The predictive value of spatial abilities for learning with visualizations might be less pure and somewhat overshadowed when the visualizations are accompanied with text, since in this case spatial abilities also need to account for the comprehension of the text and the ability to integrate text with visualizations (Hegarty & Sims, 1994). Thus, in contrast to Study 1, which adopted an experimental approach that allowed us to focus on visual processing, the role of spatial ability might be an indirect one with complex and meaningful multimedia materials. Therefore, a higher number of participants representing a wide range of spatial ability was included to

look at general trends for the relation between learning outcomes and ability in treatment conditions that varied the visual design.

Study 2: Reanalysis of Four Multimedia Experiments

We reanalyzed pooled data obtained in a larger research project comprising four experiments (Kühl, 2021; Kühl & Münzer, 2019; Kühl et al., 2018a, 2018b) utilizing the same instructional materials. The instructional materials were typical multimedia presentations, i.e., a visualization was accompanied with explanatory verbal information (written text). Each of the four experiments included specific learning conditions. However, the learning content, the visual design variants (static picture vs. animation), and the measure of spatial ability were the same in all four experiments. In all four experiments, there were retention knowledge questions and transfer knowledge questions on the learning content with only small changes on specific questions between the experiments. Spatial ability was assessed with a shortened form of the Paper Folding Test (PFT; Ekstrom et al., 1976) in all four experiments. The PFT is a measure of the broad “visualization” factor of spatial ability rooted in the factor-analytic tradition and is the most used spatial ability measure in multimedia learning research (cf. Höffler, 2010).

The instructional material used in the four experiments was about the topic of Kepler’s second law. This law describes how the velocity of a planet changes while orbiting the sun on an ellipse depending on the changes of the distance between sun and planet. The dynamic visualization (animation) possessed a clear informational advantage over the static picture, since only the animation depicted the changes in the planet’s velocity directly. The static picture only depicted the ellipse, the sun, and the planet as well as graphic indications of different spatial regions of the ellipse corresponding to the distance between the sun and the planet. This graphical indication was provided both in the picture and in the animation. It was intended to support the understanding of the relation between distance and velocity considering differences between the indicated spatial regions. All four single experiments have shown a main effect of visualization with an advantage of the dynamic visual design compared with the static visual design, especially for transfer tasks and partly also factual knowledge scores (Kühl, 2021; Kühl & Münzer, 2019; Kühl et al., 2018a; Kühl et al., 2018b). Thus, the dynamic visualization is demonstrably the optimized visual design for this topic and the static picture is demonstrably the suboptimal visual design.

Method

Participants and Design

Of originally 853 participants, three participants with spatial ability scores lower or greater than three standard deviations from the mean within each of all four experiments were excluded from the data set. The pooled data of the remaining 850

participants (600 female; 250 male), who took part in one of the four experiments from the larger research project, were reanalyzed. Of these 850 participants, $n = 132$ took part in Experiment 1 of the larger research project¹, $n = 260$ took part in Experiment 2, $n = 259$ took part in Experiment 3, and $n = 199$ took part in Experiment 4. Participants of Experiments 1, 2, and 3 were students from a German university ($M = 21.73$ years, $SD = 2.89$ years, 499 female, 152 male), while participants in Experiment 4 were students from German secondary schools ($M = 15.27$ years, $SD = 0.81$ years, 101 female, 98 male).

In all of the experiments, spatial ability was assessed, and the instructional effectiveness of the dynamic compared to the static visual design was investigated using a between-subject design. The data of the four experiments were merged, resulting in $n = 424$ participants that received dynamic visualizations and $n = 426$ participants that received static visualizations. In addition, specific learning conditions (such as complete vs. incomplete text, prompts to stimulate the drawing of inferences, pre-activation of relevant prerequisite knowledge before the presentation) had been examined in each of the four experiments. These differences, however, were not considered in the present reanalysis of the data, which focuses on the learning condition variation that was common in all four experiments, i.e., the variation of visual design (static vs. dynamic).

Materials

The instructional material dealt with Kepler's second law, which implicates how a planet's velocity changes depending on the distance to the sun when it is orbiting the sun on an ellipse. In all four experiments, participants either received a dynamic visualization that displayed a planet moving along the ellipse, showing directly the changes in velocity, or a static visualization that was identical to the dynamic visualization with the exception that no planet was shown moving and hence changes of its velocity were not depicted. A written text that described Kepler's second law was always placed below the respective visualization. These characteristics were constant in all experiments. The four experiments differed with respect to further independent variables that addressed the amount of the verbal information, instructions for specific learning activities, time of testing, and prior knowledge. The four experiments also differed with regard to pacing, with the opportunity of self-pacing in Experiments 2 and 3, but not in Experiments 1 and 4 (system-paced). In Experiment 1 (Kühl et al., 2018b), the text information was manipulated (by explicitly describing in the text the planet's changes in velocity vs. not describing this particular information in the text). In Experiment 2 (Kühl et al., 2018a), the between-subjects factors prompting learners to think about the changes in velocity (prompting vs. not prompting) as well as time of testing (immediate vs. one week delayed) were additionally implemented. In Experiment 3 (Kühl, 2021), the between-subjects factors prerequisite knowledge (receiving a prerequisite knowledge instruction vs. not receiving

¹ In Experiment 1, there were additionally 66 participants that only received text without visualizations. These data were therefore omitted from this reanalysis.

this instruction) as well as time of testing (immediate vs. one week delayed) were additionally implemented. In Experiment 4 (Kühl & Münzer, 2019), the influence of additional information that was associated but not essential to understand Kepler's second law was manipulated (receiving vs. not receiving additional information).

Measures

Spatial ability was assessed with a shortened form of the Paper Folding Test (PFT; Ekstrom et al., 1976) in all four experiments. The shortened form consists of the first part of the PFT and includes ten items. Each item has one correct solution and four distractors. In order to prevent participants from guessing, participants were informed that one point would be subtracted for each item in case of an incorrect answer, resulting in a minimum of -10 and a maximum of 10 points. Participants were given 3 min to work on the shortened form of the PFT.

Learning outcomes were assessed by two scores. One score comprised factual knowledge², and one score assessed learning outcomes for transfer tasks. The factual knowledge is knowledge about Kepler's second law that was basically presented in the multimedia message. In contrast to factual knowledge, transfer knowledge concerns thoughts and conclusions that were not presented in the multimedia message but could be derived based on a correct cognitive schema about Kepler's second law.

For Experiments 1, 2 and 3, the factual knowledge scores were assessed based on an open-ended retention task, a drawing task, a computerized static picture task and a computerized animation task. For Experiment 4, for which the data collection took place in a school, factual retention knowledge was assessed based on the retention task and the drawing task only, while the two computerized visual tasks were not applied. In the retention task, participants had to write down everything they could remember about Kepler's second law. For the drawing task, participants were asked to draw Kepler's second law (which basically corresponded to the static visualization). For the computerized static picture task, participants had to choose one among four pictures that correctly depicted Kepler's second law, and for the computerized animation task, participants had to choose one among four animations that correctly depicted the course of a planet's velocity according to Kepler's second law.

For transfer, there were four tasks in which participants had to apply implications of Kepler's second law to new scenarios. For instance, participants had to reason about the influence it would have on the course of the velocity of a planet when the planet would be circling on a more elliptic orbit (compared to a similar but less elliptic orbit), or when the sun could be closer to the center of the ellipse. The questions of the four tasks were open-ended. These four tasks were essentially the same across all four experiments, but the wording as well as the instructions for solving the tasks and correspondingly the scoring differed slightly between the experiments.

² Factual knowledge was further divided in non-dynamic and dynamic factual knowledge scores in Experiments 2–4, but not in Experiment 1. To be better able to put together the different factual knowledge scores of all four experiments, and as this distinction is moreover not in the scope of the present reanalysis, non-dynamic and dynamic factual knowledge were summed up to one score.

Learning outcome measures of each experiment were z -standardized, so that these scores could be merged for the reanalysis without differences in scores due to slight variations in scoring between single experiments.

Analytical Approach for Study 2

The analytical approach for the unifying conceptualization was applied (see the “Analytical Approach for the Unifying Conceptualization” section). A single model was estimated for each of the two dependent variables (factual knowledge, transfer knowledge). Multiple regression analyses were performed due to the between-subject design. Scores of all considered variables were z -standardized within each experiment. The independent variables were the visual design (static vs. dynamic), the spatial ability measure (PFT scores, z -standardized), and their linear and quadratic interaction terms. As in Study 1 (see also the “Analytical Approach for the Unifying Conceptualization” section, Eq. 4), two variables were defined for coding the visual design condition (“dumOpt”: 0 = static, 1 = dynamic; “dumSub”: 0 = dynamic, 1 = static). As before, only the variable of the dynamic visual design (“dumOpt”) was used for the main effect as well as for the linear interaction term, which was the product of the visual design treatment variable and the z -standardized spatial ability scores. The quadratic interaction terms were the quadratic values of the two linear interaction terms with the dynamic as well as the static visual design.

As for Study 1, additional analyses were conducted for illustrative purposes, comparing learning outcomes between visual design conditions for three ability groups separately. Participants were assigned to three ability groups based on their z -standardized spatial ability scores. Participants with z -scores lower than -0.5 were assigned to the low-ability group ($n = 246$). Participants with z -scores from -0.5 to 0.5 formed the medium ability group ($n = 313$). The high-ability group consisted of participants with z -scores greater than 0.5 ($n = 291$).

Results and Discussion

The spatial ability score (measured by the PFT) as well as both dependent variables, factual and transfer knowledge, were group-mean z -standardized for all analyses. Therefore, means and standard deviations were not exactly 0 and 1 (Table 3).

Results will be reported and discussed for each of the two dependent learning scores separately. For additional analyses, three ability groups (low, medium, high) were considered. The descriptive results for the three ability groups for the factual as well as the transfer knowledge can be found in Table 4.

Fig. 5 depicts the descriptive data for factual knowledge. The data show the expected general pattern. The regression model (Table 5) was significant, $F(5, 844) = 19.28, p < .001$. The included predictors explained $R^2 = .10$ of the variance (adjusted) in factual knowledge. Spatial ability (PFT) and visual design were highly significant predictors. The significant quadratic interaction term of the dynamic visual design was negative, indicating a non-linear, inverted u-shaped trend. The

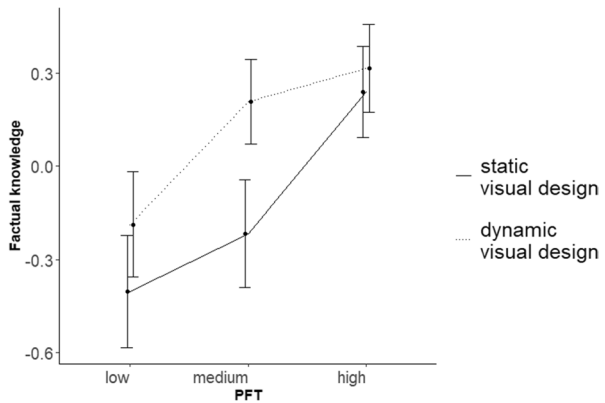
Table 3 Descriptive statistics of the spatial ability score (PFT) and both dependent variables, factual, and transfer knowledge

	<i>M</i>	<i>SD</i>	Skewness	Kurtosis
PFT	0.01	0.98	-0.42	-0.39
Factual knowledge	0.01	0.99	-0.41	-0.24
Transfer knowledge	0.03	1.00	-0.10	-1.09

All variables are *z*-standardized

Table 4 Descriptive results of the three ability groups for factual as well as transfer knowledge

	Ability Group		
	Low	Medium	high
Factual knowledge (mean)	-0.30	0.00	0.28
Factual knowledge (standard deviation)	1.00	1.01	0.89
Transfer knowledge (mean)	-0.34	-0.05	0.43
Transfer knowledge (standard deviation)	0.90	1.00	0.95

Fig. 5 Relation between visualization ability (PFT) and factual knowledge differentiated between static and dynamic presentation**Table 5** Results of the multiple regression analysis with factual knowledge as dependent variable

	<i>b</i>	<i>SE</i>	<i>t</i> -value	<i>p</i>
Intercept	-.14	.06	-2.30	.022
Visual design (static vs. dynamic)	.35	.08	4.16	< .001
PFT	.34	.05	6.98	< .001
PFT × visual design	-.13	.07	-1.92	.055
(PFT × dumOpt) ²	-.09	.04	-2.23	.026
(PFT × dumSub) ²	.02	.04	0.49	.627

Table 6 Results of the multiple regression analysis with transfer knowledge as dependent variable and with the treatment coding 0 for the static visual design and 1 for the dynamic visual design

	<i>b</i>	SE	<i>t</i> -value	<i>p</i>
Intercept	-.26	.06	-4.45	< .001
Visual design (static vs. dynamic)	.44	.08	5.31	< .001
PFT	.44	.05	9.38	< .001
PFT × visual design	-.14	.07	-1.98	.048
(PFT × dumOpt) ²	-.00	.04	0.09	.925
(PFT × dumSub) ²	.14	.04	3.54	< .001

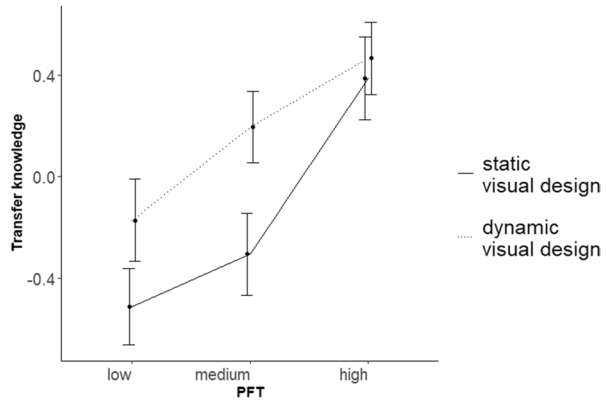
quadratic interaction term of the static visual design was not significant. Therefore, the found pattern corresponds to the sample pattern of Fig. 1a. The slope of the relation between ability and performance was steeper for the dynamic compared to the static visual design in the low-medium ability range and indicates the ability-as-enhancer effect, whereas in the medium-high ability range the slope was flatter for the dynamic compared to the static visual design, which corresponds to the ability-as-compensator effect (see Fig. 5).

Additional analyses, that served illustrative purposes, compared the static and the dynamic visual design within each ability group for the dependent variable factual knowledge using separate *t*-tests. For low-ability participants, the difference between the static and the dynamic visual design was not significant, $t(244) = 1.79$, $p = .074$, Cohen's $d = 0.23$, $CI = [-0.02, 0.48]$. For the high-ability group, there was no significant difference, $t(289) = 0.73$, $p = .468$, Cohen's $d = 0.09$, $[-0.15, 0.32]$. The difference between the static and the dynamic visual design was highly significant for the medium-ability group, and the effect size of the difference was in the medium range, $t(290.30) = 3.79$, $p < .001$, Cohen's $d = 0.43$, $[0.21, 0.66]$. The effect sizes reflect the expected pattern, meaning that the highest difference between the visual design variations was found for the group of participants with medium ability.

Table 6 shows the results of the multiple regression analysis for transfer knowledge. The regression model was significant, $F(5, 844) = 31.25$, $p < .001$, and the predictors explained $R^2 = .15$ of the variance (adjusted) in transfer knowledge. Both the static vs. dynamic visual design and spatial ability were highly significant predictors. The quadratic interaction term of the dynamic visual design was not significant, indicating a rather linear relation between ability and transfer knowledge in the dynamic visual design condition. In contrast, the quadratic interaction term of the static visual design condition was significant with a positive sign, indicating a non-linear, u-shaped relation between ability and transfer knowledge in the static visual design condition (Figure 6). The pattern reflects the conceptualization (see sample pattern depicted in Fig. 1c), because in the low-medium ability range, the slope of the relation between ability and transfer scores was steeper for the dynamic compared to the static visual design, whereas in the medium-high ability range, the slope was flatter for the dynamic compared to the static visual design.

For illustrative purposes, the transfer knowledge was compared between static and the dynamic visual design conditions for each ability group using separate

Fig. 6 Relation between visualization ability (PFT) and transfer knowledge differentiated between static and dynamic visual design



t-tests. For the low-ability group and the medium ability group, significant differences between animation and static picture condition were found (low-ability group, $t(244) = 2.96, p = .003$; medium ability group, $t(311) = 4.58, p < .001$). The effect size of the medium-ability group was greater than the effect size of the low-ability group (low-ability group, Cohen's $d = 0.38, [0.12, 0.63]$; medium-ability group, Cohen's $d = 0.52, [0.29, 0.74]$). In contrast, no difference in transfer knowledge was found between the static visual design condition and the dynamic visual design condition when participants had high spatial ability, $t(286.46) = 0.72, p = .474$, Cohen's $d = 0.08, [-0.15, 0.31]$. Similar as for the factual knowledge, the overall pattern suggests an ability-as-enhancer effect between the low and medium ability group and an ability-as-compensator effect between the medium- and high-ability group.

Taken together, the pattern of results for the learning outcome measures of factual knowledge and transfer knowledge are essentially in line with the unifying conceptualization of the enhancer- and compensator-hypotheses: For both learning outcome measures, in the low-medium ability range, the slope for the relation of ability and performance was relatively steeper for the dynamic than for the static visual design condition, whereas in the medium-high ability range, the slope for the relation of ability and performance was relatively steeper for the static than for the dynamic visual design condition. This pattern of results is corroborated by the significant quadratic interaction for the non-linear trend of one of the regressions, corresponding to the forms introduced above (the pattern found for factual knowledge corresponds to the form illustrated in Fig. 1a, the pattern found for transfer knowledge corresponds to the form illustrated in Fig. 1c). For both dependent outcome measures, the largest difference between visual design conditions was found in the medium ability range.

General Discussion

Up to now, the ability-as-enhancer interaction and the ability-as-compensator interaction have been treated as mutually exclusive. By taking nonlinear relations into account, we introduced in the present study a unifying conceptualization that considers the two aptitude-treatment interactions as two ends of one ability continuum. Next to our theoretical approach, we found first empirical evidence in two studies that support the unifying conceptualization. Thus, this unifying conceptualization has the potential to resolve — at least partly — the inconclusive results concerning the ability-as-enhancer and ability-as-compensator hypotheses and can thus advance the field and our understanding of the role of spatial abilities in learning with different visualizations.

The unifying conceptualization and its relation to the two hypotheses can be briefly summarized as follows: With (very) low ability, it is expected that learners will lack resources to benefit from the optimized design. Learners with medium ability, however, may have the resources available to perform successfully with, and hence benefit from the optimized design. Following, the ability-as-enhancer hypothesis is found in the low-to-medium ability range. Learners with high ability may compensate for a suboptimal design, so that learning will be successful regardless of optimizing the visual design. Thus, when comparing the medium with the high-ability learners, an ability-as-compensator interaction can be observed. It follows that an effect of the optimized vs. the suboptimal design is expected to be largest for participants with medium ability.

We have reasoned that non-linear relations need to be considered to detect the unifying conceptualization. We have illustrated different patterns of non-linear and linear relations for optimized and suboptimal design variants that are all compatible with our conceptualization (Figure 1). We have suggested a regression approach with two quadratic interaction terms. This approach provides flexibility to model the different patterns. We have applied the approach in two studies.

The results of the experiment in Study 1, that was characterized by a high internal validity, revealed first empirical evidence for our proposed unifying conceptualization. The purpose of the reanalysis of four published studies from one larger research project was to substantiate the unifying conceptualization by examining it with more meaningful and complex learning materials that combine visualizations with verbal information (i.e., higher ecological validity). General assumptions of the theoretical unifying conceptualization were confirmed in the reanalysis. The illustrative follow-up analyses with subsamples additionally suggested patterns corresponding to the ability-as-enhancer and ability-as-compensator interactions in the low-medium and medium-high ability range. Since both effects can occur in the same study (i.e., with the same sample and same stimulus materials), it is recommended to use the nonlinear approach of the unifying conceptualization that we propose here.

Implications

It has been an unsolved issue in previous research whether ability enhances learning with an optimized design or whether ability compensates for a suboptimal design. Both hypotheses have been treated as mutually exclusive and been tested with linear interaction terms, implying that only one of the two hypotheses can be observed. The consideration as mutually exclusive interactions might have contributed to the inconsistent results considering the aptitude-treatment interaction in the previous literature.

In contrast, the unifying conceptualizations integrates the ability-as-enhancer and the ability-as-compensator hypotheses into one superordinate model. This conceptualization can serve as a basis for (at least partly) resolving the heterogeneous results considering the at first-glance contradicting hypotheses (see the “A Brief Review of Empirical Results Concerning the Role of Spatial Ability in Learning with Visualizations” section above). Next to the theoretical conceptualization, we provided an analytical approach to test the unifying conceptualization by means of nonlinear interaction terms. Finally, we provided empirical evidence that support our conceptualization. Based on the current study, we advocate to consider the unifying conceptualization and its associated analytical approach in future research about the role of spatial abilities in learning with different visualization formats.

Concerning the ability range, the unifying conceptualization provides a more general approach for assessing the aptitude-treatment interactions. If the strongest effect between an optimized and a suboptimal visual design does not occur around medium ability, but more in the direction of lower ability, then an overall linear regression model would suggest an ability-as-compensator effect. If, however, the strongest effect between the visual design conditions would be found around the higher ability range, then an overall linear regression model would suggest an ability-as-enhancer effect. This consideration might also explain the more frequent occurrence of the compensator than of the enhancer effect in previous studies. Studies often examine university students as participants who might possess above-average mental abilities (including spatial abilities). Therefore, low-ability participants with respect to the ability measure itself are often not included because the samples are not representative of the general population. This might additionally contribute to obtaining the ability-as-compensator interaction rather than the ability-as-enhancer interaction since the largest effect between both visual designs is nearer to the low ability end than the high ability end. Correspondingly, it appears that the ability-as-compensator interaction is observed more often than the ability-as-enhancer interaction. In the meta-analytic review by Höffler (2010), there was overall more evidence for the ability-as-compensator hypothesis. This problem may be addressed with large samples and/or with samples drawn from different populations (such as university students and secondary school students, as in the present study) in order to obtain a wider range of abilities. In the present study, the ability ranges might have been wider because not only university students but also samples from secondary schools were considered.

It is an open question whether the unifying conceptualization for the role of spatial ability in learning with visualizations, and particularly the consideration of

nonlinear (quadratic) interaction terms, may also be transferable to other aptitudes (e.g. prior knowledge) and treatments that form aptitude-treatment interactions that are usually investigated with linear interaction terms. Future research may apply this approach to other learning prerequisites. This may also have profound practical implications considering the design of adaptive learning environments, that in turn are often based on aptitude-treatment interactions (cf. Kalyuga, 2014).

Limitations, Challenges and Outlook

One concern that relates to both studies is the validity of the used ability measures with regard to addressing an identifiable spatial process. Measures of spatial ability might capture different factors (and hence, different spatial processes, such as spatial orientation, visualization, or mental rotation), but they correlate with each other and with measures of general intelligence. This raises the question whether an ability measure truly addresses a specific spatial ability and, therefore, to what degree the assessed measures are actually related to the processing of the visualization. In the experiment of Study 1, a measure of VSWM capacity was chosen because the measure apparently shared important characteristics with the dependent variable, namely the mental integration task. However, alternative measures of spatial abilities, such as a measure of perspective taking ability as well as a measure of inductive reasoning, were correlated with the mental integration task in its static variant as well (Münzer et al., 2018). For Study 2, where applied multimedia learning materials were used, spatial ability was measured with the PFT, which is the most commonly used measure in this context (cf. Höffler, 2010). The PFT also correlates with other measures of spatial ability as well as general intelligence (cf. Carroll, 1993). Since we assessed for each study only one measure, but not different measures, it is unclear whether the used measures were the most specific ones to tap the demands that were posed by the respective visualizations. To be better able to derive conclusions about the specificity of the applied tests for the respective visualizations, it would be preferable in future research to control for further predictors when assessing individual differences (cf. Münzer et al., 2018).

The exact trends and interactions in specific ability ranges might depend on additional factors, such as the above discussed ability range, but also the effectiveness of the visual design. Concerning the visual design, the fact that in the two above presented studies the dynamic visualization could be considered as the optimized visualization does not mean that this is an automatism. Rather, the processing demands that are associated with the design of the visualization influences which visualization format can be considered as the optimized one – and this could also be a series of static pictures (e.g., Imhof et al., 2013). Considering a wide range of multimedia learning studies for interactions between spatial ability and visual design variations (see the “A Brief Review of Empirical Results Concerning the Role of Spatial Ability in Learning with Visualizations” section presented above), it might be noted that the studies have been heterogeneous with respect to the measures of spatial abilities and the various domains and disciplines from which the instructional materials were taken. The studies have differed correspondingly with respect to learning demands

in different visual design conditions and outcome measures. Therefore, the effectiveness of the visual design variation differs between studies (cf. Ploetzner et al., 2020), and this might influence the form of the interactions.

It seems reasonable to not only consider a learner's ability per se, but to consider the learner's ability level in relation to the objective task demands (i.e., relative difficulty). Following this reasoning, in studies with generally lower difficulty — which means that participants' ability in relation to the task demands is rather medium to high — learners are generally able to understand the learning materials, resulting in an ability-as-compensator interaction. In contrast, studies with medium to higher difficulty, which means that participants' ability in relation to the task requirements is rather low to medium, only medium ability participants are able to make use of an optimized visualization. This might provoke an ability-as-enhancer interaction. Following, it might be speculated that the four primary studies cited above in which the ability-as-enhancer interaction was found (Diaz & Sims, 2003; Huk, 2006; Huk & Steinke, 2007; Mayer & Sims, 1994) might have used stimulus materials for the learning task that was more complex and difficult than stimulus materials used in studies in which the ability-as-compensator interaction was obtained (e.g., 3D-visualization in Huk, 2006; asynchronous spoken text in Mayer & Sims, 1994). These considerations suggest that a number of factors such as the effectiveness of the optimized visualization, the ability range, and the general difficulty of the learning materials may influence the exact form of the aptitude-treatment interactions.

In general, the aptitude-treatment-interactions will be the result of the relation between measurable aptitude (e.g., spatial ability, general intelligence, prior knowledge) and the demands of processing the given learning materials (e.g., the visual-spatial mental processes required to understand the depicted changes in the learning materials, the knowledge about the meaning of basic concepts mentioned in the learning materials). Thus, both "aptitude-per-se" as well as demands and difficulties of the learning materials contribute to the resulting interactions.

We consider the introduced unifying conceptualization of the ability-as-enhancer and ability-as compensator hypotheses as a very promising and valuable model that helps in resolving heterogenous findings about the role of spatial ability in learning with different visualizations. However, this is not to misunderstand that we think this model will resolve all problems associated with the moderating role of spatial abilities. First, whether the unifying conceptualization will be observable will, as abovementioned, depend on the ability range of the examined sample (also in relation to the task demands). As a positive side effect, this may stimulate researchers to reason beforehand why and which interaction they assume. Second, as discussed above concerning the specificity of the ability measure, a moderating role will hardly be observable when the assessed ability measure does not strongly relate to the demands the optimized and suboptimal visualizations pose. Third, the ability-as-enhancer and the ability-as-compensator hypotheses — and correspondingly the unifying conceptualization — may not be the only conceivable role spatial abilities play. For example, spatial abilities may also hinder learners to deeper process the content — for instance due to metacognitive factors such as feelings of overconfidence (cf. inhibiting function; Schnotz & Rasch, 2005). Thus, we do not claim that the unifying conceptualization will resolve all problems associated with the complex

interplay of spatial abilities in learning with different visualizations. However, the unifying conceptualization provides a broader view on this subject and can be useful for analyzing and interpreting aptitude-treatment-interactions. Thereby, it may support us in gaining a deeper understanding of this complex subject.

Conclusions

To the best of our knowledge, this is the first study that resolved the issue of the form of the aptitude-treatment interaction between spatial ability and visual design with a unifying conceptualization for the ability-as-enhancer and the ability-as-compensator hypotheses — by showing that the ability-as-enhancer and ability-as-compensator hypotheses can be considered as two ends of one ability continuum. A flexible analytical approach was introduced that can account for different patterns of non-linear and linear trends for the relation between ability and performance in optimized vs. suboptimal design conditions. Both, in the experiment which served to test the conceptualization under controlled conditions with a consistent visual-spatial task and in the reanalysis of experiments with more meaningful multimedia learning materials, the conceptualization was generally confirmed. In addition, we discussed factors such as sample characteristics and effectiveness of the visual design variation. These factors may contribute to an explanation under which conditions particular interactions can be found. The conceptual and methodological approach presented here can be useful for analyzing and interpreting aptitude-treatment-interactions and for determining conditions and factors that influence the form of interactions observed. Thereby, this study informs researchers to not (only) consider linear interactions between spatial abilities and type of visualizations, but also to consider quadratic interaction terms. In addition, the opportunity to obtain more fine-grained results regarding the effectiveness of optimized designs in different ability ranges can have valuable practical implications for adaptation of instruction to ability.

Code Availability Not applicable.

Author Contribution All three authors contributed equally to this work and should be considered as first authors: All authors developed the hypotheses. SM developed the design and materials of the experiment. BCOFF was responsible for data collection of the experiment and data analysis of both the experiment and the reanalysis. TK contributed the data of the four multimedia experiments and prepared the reanalysis. All authors were responsible for writing and editing the final manuscript.

Funding This work was supported by the German Research Foundation (DFG) under grant agreement KU 3160/2–1 as well as grant agreement MU 1786/3–1. Open Access funding enabled and organized by Projekt DEAL.

Data Availability The data are available on request from the corresponding author.

Declarations

Ethics Approval and Consent to Participate The ethics approval for this project was waived by the local ethics committee, because the funding by the German Research Foundation (Deutsche Forschungsgemeinschaft) implied compliance with ethical guidelines for empirical studies with human participants and

the corresponding declarations. The study complied with the ethical guidelines for studies with human participants. In particular, all participants volunteered to participate, signed a written consent for study participation (their parents signed for the experiment that took place in school) and were allowed to end the experiment at any time.

Consent for Publication Not applicable

Conflict of Interest The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Allen, K., Higgins, S., & Adams, J. (2019). The relationship between visuospatial working memory and mathematical performance in school-aged children: A systematic review. *Educational Psychology Review*, 1-23. <https://doi.org/10.1007/s10648-019-09470-8>
- Baddeley, A. D. (1986). *Working memory*. Oxford University Press.
- Barrett, T. J., & Hegarty, M. (2016). Effects of interface and spatial ability on manipulation of virtual models in a STEM domain. *Computers in Human Behavior*, 65, 220–231. <https://doi.org/10.1016/j.chb.2016.06.026>
- 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, 452–462. <https://doi.org/10.1002/ase.1524>
- Bogomolova, K., van der Ham, I. J., Dankbaar, M. E., van den Broek, W. W., Hovius, S. E., van der Hage, J. A., & Hierck, B. P. (2020). The effect of stereoscopic augmented reality visualization on learning anatomy and the modifying effect of visual-spatial abilities: A double-center randomized controlled trial. *Anatomical Sciences Education*, 13, 558–567. <https://doi.org/10.1002/ASE.1941>
- Bosco, A., Longoni, A. M., & Vecchi, T. (2004). Gender effects in spatial orientation: Cognitive profiles and mental strategies. *Applied Cognitive Psychology*, 18(5), 519–532. <https://doi.org/10.1002/acp.1000>
- Brooks, L. R. (1967). The suppression of visualization by reading. *The Quarterly Journal of Experimental Psychology*, 19(4), 289–299. <https://doi.org/10.1080/14640746708400105>
- Brucker, B., Scheiter, K., & Gerjets, P. (2014). Learning with dynamic and static visualizations: Realistic details only benefit learners with high visuospatial abilities. *Computers in Human Behavior*, 36, 330–339. <https://doi.org/10.1016/j.chb.2014.03.077>
- Carroll, J. B. (1993). *Human cognitive abilities*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511571312>
- Castro-Alonso, J. C., Ayres, P., & Paas, F. (2014). Learning from observing hands in static and animated versions of non-manipulative tasks. *Learning and Instruction*, 34, 11–21. <https://doi.org/10.1016/j.learninstruc.2014.07.005>
- Castro-Alonso, J. C., Ayres, P., Wong, M., & Paas, F. (2018). Learning symbols from permanent and transient visual presentations: Don't overlay the hand. *Computers & Education*, 116, 1–13. <https://doi.org/10.1016/j.compedu.2017.08.011>
- 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>

- Cohen, C. A., & Hegarty, M. (2007). Individual differences in use of external visualisations to perform an internal visualization task. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 21(6), 701–711.
- de Back, T. T., Tinga, A. M., Nguyen, P., & Louwerse, M. M. (2020). Benefits of immersive collaborative learning in CAVE-based virtual reality. *International Journal of Educational Technology in Higher Education*, 17, 1–18.
- de Koning, B. B., Marcus, N., Brucker, B., & Ayres, P. (2019). Does observing hand actions in animations and static graphics differentially affect learning of hand-manipulative tasks? *Computers & Education*, 141. <https://doi.org/10.1016/j.compedu.2019.103636>
- Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). Manual for kit of factor-referenced cognitive tests. *Educational Testing Service*.
- Diaz, D. D., & Sims, V. K. (2003). Augmenting virtual environments: The influence of spatial ability on learning from integrated displays. *High Ability Studies*, 14(2), 191–212.
- Garg, A., Norman, G. R., Spero, L., & Maheshwari, P. (1999). Do virtual computer models hinder anatomy learning? *Academic Medicine*.
- 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>
- Hannafin, R. D., Truxaw, M. P., Vermillion, J. R., & Liu, Y. (2008). Effects of spatial ability and instructional program on geometry achievement. *The Journal of Educational Research*, 101(3), 148–157.
- Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4), 411–430. <https://doi.org/10.3758/BF03200867>
- Hegarty, M., & Kriz, S. (2008). Effects of knowledge and spatial ability on learning from animation. In R. Lowe & W. Schnotz (Eds.), *Learning with animation: Research implications for design* (pp. 3–29). Cambridge University Press.
- Hegarty, M., & Waller, D. A. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *The cambridge handbook of visuospatial thinking* (pp. 121–169). Cambridge University Press.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition and instruction*, 21(4), 209–249.
- 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, 151–176. <https://doi.org/10.1016/j.intell.2005.09.005>
- Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations - A meta-analytic review. *Educational Psychology Review*, 22, 245–269.
- 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.
- Huk, T. (2006). Who benefits from learning with 3D models? The case of spatial ability. *Journal of computer assisted learning*, 22(6), 392–404.
- Huk, T., & Steinke, M. (2007). Learning cell biology with close-up views or connecting lines: Evidence for the structure mapping effect. *Computers in Human Behavior*, 23(3), 1089–1104.
- 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.
- Imhof, B., Scheiter, K., Edelman, J., & Gerjets, P. (2012). How temporal and spatial aspects of presenting visualizations affect learning about locomotion patterns. *Learning and Instruction*, 22(3), 193–205.
- Imhof, B., Scheiter, K., Edelman, J., & Gerjets, P. (2013). Learning about locomotion patterns: Effective use of multiple pictures and motion-indicating arrows. *Computers & Education*, 65, 45–55.
- Irwin, J. R., & McClelland, G. H. (2003). Negative consequences of dichotomizing continuous predictor variables. *Journal of Marketing Research*, 40, 366–371.
- Kalyuga, S. (2014). The expertise reversal principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 576–597). Cambridge University Press.
- Keehner, M., Lippa, Y., Montello, D. R., Tendick, F., & Hegarty, M. (2006). Learning a spatial skill for surgery: How the contributions of abilities change with practice. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, 20(4), 487–503.
- Kühl, T. (2021). Prerequisite knowledge and time of testing in learning with animations and static pictures: Evidence for the expertise reversal effect. *Learning and Instruction*, 73, 101457. <https://doi.org/10.1016/j.learninstruc.2021.101457>

- Kühl, T., Scheiter, K., Gerjets, P., & Edelmann, J. (2011a). The influence of text modality on learning with static and dynamic visualizations. *Computers in Human Behavior*, *27*, 29–35. <https://doi.org/10.1016/j.chb.2010.05.008>.
- Kühl, T., Scheiter, K., Gerjets, P., & Gemballa, S. (2011b). Can differences in learning strategies explain the benefits of learning from static and dynamic visualizations? *Computers & Education*, *56*, 176–187. <https://doi.org/10.1016/j.compedu.2010.08.008>.
- Kühl, T., & Münzer, S. (2019). The moderating role of additional information when learning with animations compared to static pictures. *Instructional Science*, *47*, 659–677. <https://doi.org/10.1007/s11251-019-09498-x>.
- Kühl, T., Scheiter, K., & Gerjets, P. (2012). Enhancing learning from dynamic and static visualizations by means of cueing. *Journal of Educational Multimedia and Hypermedia*, *21*, 71–88.
- Kühl, T., Navratil, S. D., & Münzer, S. (2018a). Animations and static pictures: The influence of prompting and time of testing. *Learning and Instruction*, *58*, 201–209. <https://doi.org/10.1016/j.learninstruc.2018.07.006>.
- Kühl, T., Stebner, F., Navratil, S. D., Fehringer, B. C. O. F., & Münzer, S. (2018b). Text information and spatial abilities in learning with different visualization formats. *Journal of Educational Psychology*, *110*, 561–577. <https://doi.org/10.1037/edu0000226>.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, *11*(1), 65–100.
- Lee, E. A. L., & Wong, K. W. (2014). Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, *79*, 49–58.
- Levie, W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication and Technology Journal*, *30*, 195–232.
- Lohman, D. F. (1988). Spatial abilities as traits, processes, and knowledge. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 181–248). Lawrence Erlbaum Associates.
- Massa, L. J., & Mayer, R. E. (2006). Testing the ATI hypothesis: Should multimedia instruction accommodate verbalizer-visualizer cognitive style? *Learning and Individual Differences*, *16*(4), 321–335.
- Mayer, R. E. (2021). *Multimedia learning (third edition)*. New York: Cambridge University Press.
- Mayer, R. E., & Sims, V. K. (1994). For whom is a picture worth a thousand words? Extensions of a dual-coding theory of multimedia learning. *Journal of educational psychology*, *86*(3), 389.
- Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130*, 621–640. <https://doi.org/10.1037/0096-3445.130.4.621>
- Münzer, S. (2012). Facilitating spatial perspective taking through animation: Evidence from an aptitude-treatment-interaction. *Learning and Individual Differences*, *22*, 505–510. <https://doi.org/10.1016/j.lindif.2012.03.002>.
- Münzer, S. (2015). Facilitating recognition of spatial structures through animation and the role of mental rotation ability. *Learning and Individual Differences*, *38*, 76–82. <https://doi.org/10.1016/j.lindif.2014.12.007>.
- Münzer, S., & Stahl, C. (2011). Learning routes from visualisations for indoor wayfinding: Presentation modes and individual differences. *Spatial Cognition and Computation*, *11*(4), 281–312. <https://doi.org/10.1080/13875868.2011.571326>.
- Münzer, S., & Zadeh, M. V. (2016). Acquisition of spatial knowledge through self-directed interaction with a virtual model of a multi-level building: Effects of training and individual differences. *Computers in Human Behavior*, *64*, 191–205. <https://doi.org/10.1016/j.chb.2016.06.047>.
- Münzer, S., Zimmer, H. D., & Baus, J. (2012). Navigation assistance: A trade-off between wayfinding support and configural learning support. *Journal of Experimental Psychology: Applied*, *18*(1), 18–37. <https://doi.org/10.1037/a0026553>.
- Münzer, S., Fehringer, B. C. O. F., & Kühl, T. (2018). Specificity of mental transformations involved in understanding spatial structures. *Learning and Individual Differences*, *61*, 40–50. <https://doi.org/10.1016/j.lindif.2017.11.004>.
- Paik, E. S., & Schraw, G. (2013). Learning with animation and illusions of understanding. *Journal of Educational Psychology*, *105*, 278–289. <https://doi.org/10.1037/a0030281>
- Ploetzner, R., Berney, S., & Bétrancourt, M. (2020). A review of learning demands in instructional animations: The educational effectiveness of animations unfolds if the features of change need to be learned. *Journal of Computer Assisted Learning*, *36*, 838–860. <https://doi.org/10.1111/jcal.12476>

- Psychology Software Tools Inc. (2012). *E-Prime 2.0*. Retrieved from <http://www.pstnet.com>
- Quinn, J. G., & Ralston, G. E. (1986). Movement and attention in visual working memory. *The Quarterly Journal of Experimental Psychology Section A*, *38*(4), 689–703. <https://doi.org/10.1080/14640748608401621>
- Salomon, G. (1994). *Interaction of media, cognition and learning*. Lawrence Erlbaum Associates.
- 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>
- Schnotz, W., & Rasch, T. (2005). Enabling, facilitating, and inhibiting effects of animations in multimedia learning: Why reduction of cognitive load can have negative results on learning. *Educational Technology Research and Development*, *53*(3), 47.
- Stebner, F., Kühl, T., Höffler, T., 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>
- 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*, 352–402.
- Vecchi, T., & Cornoldi, C. (1999). Passive storage and active manipulation in visuo-spatial working memory: Further evidence from the study of age differences. *European Journal of Cognitive Psychology*, *11*(3), 391–406. <https://doi.org/10.1080/713752324>
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817–835. <https://doi.org/10.1037/a0016127>
- Waller, D. (2000). Individual differences in spatial learning from computer-simulated environments. *Journal of Experimental Psychology Applied*, *6*(4), 307–321.
- Wang, T. L., & Tseng, Y. K. (2020). The effects of visualization format and spatial ability on learning star motions. *Journal of Computer Assisted Learning*, *36*, 61–69.
- Wender, K. F., & Muehlboeck, J. S. (2003). Animated diagrams in teaching statistics. *Behavior Research Methods, Instruments, & Computers*, *35*(2), 255–258.
- Westerman, S. J. (1997). Individual differences in the use of command line and menu computer interfaces. *International Journal of Human-Computer Interaction*, *9*(2), 183–198.
- Wu, H.-K., Lin, Y.-F., & Hsu, Y.-S. (2013). Effects of representation sequences and spatial ability on students' scientific understandings about the mechanism of breathing. *Instructional Science*, *41*, 555–573.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Tim Kühl^{1,2}  · Benedict C. O. F. Fehringer¹  · Stefan Münzer¹

Benedict C. O. F. Fehringer
b.fehringer@uni-mannheim.de

Stefan Münzer
stefan.muenzer@uni-mannheim.de

¹ Psychology of Education, University of Mannheim, Postfach 103462, 68131 Mannheim, Germany

² Department of Education, University of Potsdam, Potsdam, Germany