The handle http://hdl.handle.net/1887/25194 holds various files of this Leiden University dissertation.

**Author:** Chabani, Elaheh (Ellahe)
**Title:** Enhancing visuospatial processing skills in children
**Issue Date:** 2014-04-15
Chapter 3

Face-to-face versus computer-based visuospatial training in children

Ellahe Chabani & Bernhard Hommel
Manuscript submitted for publication
Abstract

Growing evidence highlights the importance of visual-spatial processing skills (VSPS) but teaching and training of these skills at early age in schools remain understudied. In this study, we compared the effectiveness of an experimental computerized VSPS-enhancing approach, a conventional face-to-face training regime, and a non-training control group in improving performance in a tangram game. We also compared the effect of training on a simple transformations test tapping into visual memory and on a complex transformations test involving mental rotation. Findings suggest that both computer-based and face-to-face training can reliably improve VSPS and that the two training methods are equally effective. Most positive findings were restricted to performance on the complex transformations test.

Keywords: Visuospatial skills; visualization; mental rotation; computer based instruction; visual cues.
Introduction

Over the past two decades, various abilities and skills have become increasingly valued in schools, including visual spatial processing skills (VSPS). The continued growths of multimedia that rely heavily on VSPS have added further emphases to the issue. Broadly speaking, VSPS refer to the ability of carry out processes responsible for generating, retaining, retrieving, and transforming visual images (i.e., non-linguistic information; e.g., Linn & Petersen, 1985; Lohman, 1993). Among other things, these skills play an important role for performance in the STEM (science, technology, engineering, and mathematics) domain (e.g. Lubinski, 2010; Miller & Halpern, 2013; Sorby, Casey, Veurink & Dulaney, 2013; Uttal et al., 2012; Wai, Lubinski & Benbow, 2009) and in the acquisition of related academic skills (e.g., math, reading, or writing; e.g., Assel et al., 2003; Cheng & Mix, 2012). Yet, it has frequently been observed that VSPS at school is not just under-supported but under-valued (National Council of Teachers of Mathematics, 2010; National Research Council, 2006; Webb, Lubinski & Benbow, 2007). The Learning to Think Spatially (2006) report has emphasized the need of identifying and developing methods and procedures that support and enhance VSPS at early ages, to consider how to integrate VSPS-related performance into the curriculum in more general ways, and to think about how and when the use of new technologies with young children can lead to improvement. Recently, Krakowski, Ratliff, Gomez, and Levine (2010) from the Spatial Intelligence and Learning Center reported that important hindering factors are teachers’ lack of awareness of the VSPS concept, and lack of knowledge about how to teach VSPS and how to integrate that teaching into their curriculum, difficulties that go hand in hand with the lack of adapted training instruments or materials, in particular for school age children.

While addressing such calls will require considerable concerted efforts, the present project aimed to take a preliminary, first step in this direction and set the stage for further, more extended and differentiated investigations. To this end, we have developed a tangram based game (TangSolver) that we think can support VSPS enhancement at earlier ages and facilitate the integration of these skills into the school curriculum. In the present study, our main focus was to evaluate the efficacy of traditional training (face to face) employing multimodal instruction (verbal and visual cues) to computerized training that relies on visual-only cues (image scaffolding).

Malleability of VSPS

VSPS involve the process of perceiving, transforming, and recreating different aspects the visual and spatial world (e.g., Linn & Petersen, 1985; Lohman, 1993). The importance of VSPS has been widely acknowledged, however, a great deal of confusion regarding the underlying components, definition, and the classification of these remain (e.g., Carroll, 1993; Eliot & Smith, 1983; Lohman, 1988; Kaufman, 2007; Sutton & Williams, 2007). These skills and their development have been attributed to a number of
variables, including cognitive development, spatial experiences, aptitude, age, and gender (e.g., Hegarty & Waller, 2006).

Numerous researchers have asked whether these skills can improve as a result of training and experience. Does the training in particular spatial tasks transfers to other spatial tasks, is there a critical or sensitive period for influencing the development of VSPS, and what are the major determinants of individual differences in response to training (e.g., gender, age, mental rotation capacity)? Recently, Uttal, Meadow, Tipton, Hand, Alden, and Warren (2012) have carried out a meta-analysis of over 217 studies on VSPS to identify possible training moderators and the magnitude, durability, and generalizability of training effects. The analysis confirmed the theoretical and practical importance of VPSP at any age and indicated that they can be significantly improved by even short training procedures. The authors suggest that adding spatially-challenging activities to standard courses can further improve VSPS and can lead to transfer to other spatially-demanding tasks. They also highlight the relevance of videogames for improving spatial skills: “playing active games has the potential to enhance spatial thinking substantially, even when compared to a strong control group” (Uttal et al., 2012; p. 54). Importantly for our present purposes, the authors also emphasize the lack of studies in younger children (four out of 217 studies investigated children below 13 years), which contrasts with the large amounts of studies involving adolescent and adults in STEM education.

**Importance of teaching and training VSPS at early age**

Recent findings suggest that practicing VSPS is not just important for STEM education, as indicated by the bulk of research considered by Uttal et al., (2012) and even more in recent studies (Miller & Halpern, 2013; Sorby, Casey, Veurink & Dulaney, 2013), but it is also highly relevant for younger learners. This is because VSPS develops through lifetime and as emerging deficits in one sub-skill or ability can often be compensated by excellence in others (e.g., Van Garderen & Montague, 2003). Mastery and understanding scale, quantity, direction, interval, size, shape recognition, and sequence of number or letter ordering are the basis for comprehending arty and academic topics (math, arithmetic skills, reading, and writing) that rely greatly on VSPS (e.g., Aunio & Niemivirta, 2010; Cheng & Mix, 2012; Holmes, Adams & Hamilton, 2008; Levine, Kwon, Huttenlocher, Ratliff & Dietz, 2009; Newcombe & Frick, 2010; Passolunghi & Mammiarella 2010; Vasilyeva & Huttenlocher, 2004). Considering that developing children’s spatial thinking at young age improves symbolic and numerical representation, it seems important to aim for the earliest-possible training regimes.

Furthermore, in a world of constant change, our “cyber children” are growing up immersed in an increasingly visually oriented and technology-driven society. This suggests that mere exposure to digital multimedia and everyday living is sufficient to impact cognitive development and there is evidence that various interventions can enhance the development of VSPS in particular (Blazhenkova & Kozhevnikov,
2009; Kaufman, Steinbügl, Dünser & Glück, 2005; Wright et al., 2008; Uttal et al., 2012). Indeed, it can be argued that individual and collective success in our technological era will depend largely on VPSP that include visualization, speed in grasping the big picture, thinking graphically, visual memory, and pattern-finding (Carr, 2008, 2010; Newcombe & Frick, 2010; Pink, 2005). In short, “Literacy in the future will include the ability to read both text and image, together and separately. The creative use of images is no longer an added extra for a text but a vital link in the cognitive processing of information and essential in the creation of sound pedagogy” (Sankey, 2002, p.1). If so, students with underdeveloped VSPS might increasingly lag behind their peers.

The integration of VSPS-enhancing activities into standard courses remains limited, while there is strong evidence that VSPS-related skills do not need to be taught formally. For instance, playing videogames such as Tetris has been found to increase visual-spatial attention and a number of other spatial skills (Gee, 2007; Green & Bavelier, 2003, 2006; Spence & Feng, 2010; Rafi, Samsudin & Said, 2008). In addition, studies on games and instructional software as viable medium for learning have provided solid ground for making use of them in the classroom (e.g., Doering & Veletsianos, 2009; Klopfer, Osterweil & Salen, 2009). We argue that an optimal solution might be game-based classroom training, as possible with the tangram game. Although uncountable numbers of physical and computerized tangram games exist, these are not designed for the testing and training of particular VSPS sub-skills, which motivated us to develop the TangSolver test.

**Development of VSPS’s game based training: TangSolver**

The Chinese Tangram puzzle game requires assembling seven geometrical pieces to form a bigger shape or figure. Such a game taps into the processing of shapes and relationships between shapes (e.g., between two triangles that make a square) and has been used to evaluate various visual spatial skills or diagnostic for visuospatial problem solving abilities (Bohning & Althouse, 1997; Crawford, 2002; Ford, 2003; Foster, 2007; Gardner, 1974; Slocum, et al., 2003; Van Hiele, 1999). As shown by various authors (e.g., Lee, Lee & Collins, 2009; Siew & Abdullah, 2012; Yang & Chen, 2010), and by Van Hiele (1999) in particular, the solution of geometrical puzzles relies on and promotes skills in handling qualitative spatial relations between elements in visual scenes and categorical representations to reason about shapes or the space they encode—VSPS that is.

While the definitions and assumed components of VSPS vary among authors, one widely accepted differentiation stems from Linn and Petersen (1985), who divide VSPS into spatial perception, spatial visualization, and mental rotation. According to these authors, spatial perception is the ability “... to determine spatial relationships with respect to the orientation of their own bodies, in spite of distracting information”, spatial visualization is “the ability in which complex spatial information are manipulated when several stages are needed for solving the tasks”, and mental rotation is the ability to mentally rotate
two- or three-dimensional figures as quickly and accurately as possible. In the development of TangSolver we focused on spatial visualization and mental rotation.

a) Spatial visualization

According to Linn & Petersen (1985), spatial visualization skills involve multi-step manipulations of spatially presented information, which require analysis of the relationship between forms and different spatial representations. Typically, learner’s spatial visualization performance is related to the focus of attention in shape characteristics, such as size, contour, lines, space, or color (Gibson, 1969). To manipulate size and complexity of forms, we used the seven classical geometric forms and combinations thereof (e.g., a triangle and a square), which we will refer to as master pieces (MPs). The same shape could be constructed by assembling 4 MPs, 5 MPs, 6 MPs or 7 MPs, which resulted in four difficulty levels (L1: 4 MPs, L2: 5 MPs, L3: 6 MPs, and L4: 7 MPs; see fig. 1 and fig.2). Another important characteristic of shapes is color (Gibson, 1969). We used monochrome MPs for the test items and colored MPs (red, yellow, and blue) for the training items. Theoretically, we assumed that children who have better developed spatial visualization would demonstrate better performance at baseline. As training MPs were colored, we thought that children who have better developed analogical reasoning capacity would recognize better how different master pieces were either combined or divided, and this would lead to more improvement.

Partial screenshots of 5 MPs, 6 MPs and 7 MPs of this item.

**Monochrome visual cue (MVC)**
Master pieces are highlighted step by step in an unrelated color than master pieces in use. This visual cue was provided automatically after 1 min of unsuccessful try. After that the flowing visual cues were enabled and the child could active as much as necessary.

**Color visual cue (CVC)** - master pieces are highlighted step by step in the same color of master pieces in use.

**Back and forward arrow**
Allows backward and forward to a lower/higher level or only checking how MPs were combined/splintered.

**Out-line** - the shape is presented on the working spaces that allow the drag and drop on the top of

---

Fig 1. Screenshots of the Computer training and different visual cues.
b) Mental rotation

The second VSPS sub-skill we considered was mental rotation, which involves the ability to rapidly and accurately rotate mentally two or three-dimensional figures (Linn & Petersen, 1985). Mental rotation performance correlates with level in math, geometry, and reading (Dehaene, Cohen, Sigman & Vinckier, 2005; Dehaene, 2009; Hegarty & Kozhevnikov, 1999). In typical mental rotation tasks (e.g., Shepard & Metzler, 1971; Vandenberg & Kuse, 1978), individual are asked to match pairs of shapes that differ in orientation (mirrored vs. non-mirrored, or rotated), with speed and accuracy being the main measures. Some authors have argued that different test items might be prone to different solution strategies (e.g., Cherney & Neff, 2004; Geiser, Lehmann & Eid, 2006; Glück & Fitting, 2003): test items can often be solved by holistic matching (involving mental rotation) or by means of feature comparison (an analytic step-by-step process).
step strategy rather than one relying on mental rotation proper)—often with better performance related to the former than to the latter.

To assess both of these strategies, we developed two subtests. One did not require any mental rotation but could easily be solved by media dragging and dropping MPs—we will refer to this version as “simple transformations test” (STT; see fig 3). The other did require mental rotation: an initial flip or rotation of MPs followed by drag and drop was needed—to this version we will refer to as “complex transformations test” (CTT; see fig 3). In a nutshell, the former taps more into visual memory while the latter assesses mental rotation capacity proper. Theoretically, it can be expected that performance differences between children with lower and children with higher mental rotation capacity be more pronounced in the complex-transformations task than in the simple-transformations task.

![Fig. 3](An illustration of Simple transformation test & Complex transformation test of item with 4 MPs.)

**Instructional software vs. traditional (face-to-face) instruction**

An extensive body of research has promoted the use of instructional software in the classroom (e.g., Doering & Veletsianos, 2009; Klopfert, Osterweil & Salen, 2009) but the debate about the effectiveness of instructional software as compared to traditional instruction is still open (e.g., Cheung & Slavin, 2012; Fleischer, 2012; Tamim et al., 2011; see [www.edutopia.org](http://www.edutopia.org)). This debate is partly fueled by insufficient software-induced improvement in particular content areas (mathematics, reading, etc.) and the expectation that instructional software might eventually replace teachers altogether. However, as pointed by Ross, Morrison, and Lowther (2010), “educational technology is not a homogeneous ‘intervention’ but a broad variety of modalities, tools, and strategies for learning. Its effectiveness, therefore, depends on how well it helps teachers and students achieve the desired instructional goals” (p.19). Indeed, the intent of most instructional software is to maximize student’s growth and individual success by meeting each student’s needs, not to replace teachers. Thus, the overarching goal of successful education cannot be to conceive of instructional software and traditional instruction as mutually exclusive alternatives but to create successful learning experiences that facilitate learning and transfer of knowledge (e.g., Archer & Hughes, 2011; Fadel, 2008; Picciano, 2009).
Although our study will not be able to close the debate, we investigated whether and to what degree visuospatial processing skills in young children can be enhanced, and which method might be best suited. In particular, we compared a custom-made, computerized visuospatial training program to the same program delivered in the traditional way (face to face). Even though there are reasons to assume that children might benefit from VSPS training (e.g., Lee, Lee & Collins, 2009; Siew & Abdullah, 2012; Yang & Chen, 2010; Uttal, et al., 2012), to date there is no evidence which method might be most successful in reaching that goal.

**Multimodal instruction vs. visual cues**

Arguably, the kind of instruction is as important for successful learning as the medium being used, in particular in the context of VSPS enhancement. Unfortunately, however, individual differences between learners make it difficult to determine which kind of medium might be the most suitable. Various authors have suggested to give consideration to individual information processing differences and their particular strength or weakness regarding the given task (e.g., Gardner, 1983; Omrod, 2008; Pashler, McDaniel, Rohrer & Bjork, 2008). According to Gardner, everyone possesses all forms of intelligence or capacity (i.e., musical, interpersonal, spatial-visual and linguistic) but to different degrees. One individual might thus have high verbal/linguistic capacity, which would suggest providing verbal information rather than graphs or pictures, while another with high visual intelligence would strongly benefit from visual material. Likewise, learners with language-based disabilities are likely to prefer visual over verbal communication (Newhall, 2012; Smith & Tyler, 2009). Hence, “different modes of instruction might be optimal for different people because different modes of presentation exploit the specific perceptual and cognitive strengths of different individuals” (Pashler et al., 2008, p.109).

Even though the development of a fully individualized learning program was beyond the scope of the current study, we want to emphasize that our comparison of face-to-face instruction, which combined verbal and visual information, and computer-based instruction, which relied on visual-only cues, implied a comparison of multimodal and unimodal instruction means. This is important because some theories suggest that the efficiency of multimodal and unimodal instruction may differ in principle. For instance, dual coding theory assumes that information is processed along two separate processing routes, one dedicated to verbal information and another to nonverbal, visual information (Paivio, 1986). This account has been extended to literacy, written composition, spelling, of reading comprehension (e.g., Kintsch, 2004; Krasny, Sadoski & Paivio, 2007; Krasny & Sadoski, 2008; Sadoski & Paivio, 2001, 2004; Sadoski, Willson, Holcomb & Boulware-Gooden, 2005). It would suggest that multimodal (verbal/visual) information might lead to better learning because two rather than one system are activated. Other accounts allow for different predictions. For instance, the cognitive theory of multimedia learning (Mayer, 2001, 2009; Mayer & Moreno, 2003) suggests that learning occurs when relevant information can be selected and organized.
into a coherent representation and integrated into the existing knowledge base. Given that information needs to pass the learner’s working memory, which is considered a limited cognitive resource, less information may be more, as too much information can easily exhaust working memory capacity. If so, restricting learning cues to just one modality might be beneficial. However, while visual cues have been studied in the context of learning with text and pictures for comprehension (Anglin et al., 2004; Eitel et al., 2013; Fletcher & Tobias, 2005; Hegarty & Just, 1993; Hegarty, 2011; Schnottz, 2005), to our knowledge there is no study that compared visual-only to multimodal cues.

Typically, visual cuing refers to “the addition of design elements that direct the learner’s attention to important aspects of the learning material” (Plass, Homer & Hayward, 2009, p. 39). In tasks such as puzzle construction, visual cuing might consist in simply presenting the solution or by attracting attention to the critical part (e.g., by increasing the luminance of relevant parts in the visual display). The former technique can be considered to relate to global visual memory, while the latter refers to spatial visual memory, which has been aimed at in the development of TangSolver. Accordingly, our study does not speak to the relative efficiency of the two cuing techniques.

Given the conflicting theoretical account, it was difficult to predict the impact of training modality. According to dual-coding theory, it would make sense to assume that the typical learner would benefit more from traditional (multimodal) instruction (Fadel, 2008) as compared to the unimodal computer-based instruction. However, cognitive theory seems to make the opposite prediction, as our unimodal computer-based instruction might be less likely to overload working memory.

**Overview of main questions**

In sum, our guiding questions were:

i. Will the two experimental groups show better performance than the control group after the training? Even though there is some evidence that VSPS can improve through training (Uttal et al., 2012), more evidence is necessary, especially with regard to our new, custom-made program.

ii. Will the two training modalities (face-to-face versus computer-based) differ in efficiency, will thus one training regime produce stronger improvement than the other? As pointed out, different theoretical approaches suggest different directions into which such a difference might go.

iii. Will training effects be visible in simple and complex-transformations tests alike (and, thus, rather non-specific) or be stronger in the more visually demanding complex-transformations test?

Note that the current study does not aim to reconcile different conceptions of VSPS but, rather, advocate the integrating assessment and training of VSPS in primary schools. Obviously, TangSolver is just a means to an end in the process of promoting the assessment, and training of VSPS within standard courses. It is
hoped that implementing such application in school promotes teaching and training VSPS-related skills in young children and inspires researcher to further develop assessment and training procedures.

**Method**

**Participants**

A total of 104 typical developing children (48 boys, mean age=8.5 years +/- 9.9 months; range 7-10.5 months; 56 girls, mean age=8.5 years +/- 8.9 months; range 7-10.5 months) from a number of standard primary schools in the Netherlands was tested. Participants had no specific academic, learning or behaviour problems and came from diverse ethnic backgrounds and socio-economic classes.

**Design**

The study involved a pre-test, two training sessions, and a post-test, with two experimental training groups and a control group. Children were matched as much as possible for gender and age and were randomly assigned to the Computer training (n=41), Face To Face training (n=42), and Control group (n=21) with a ratio of 2:2:1 respectively. This unequal RCT ratio was chosen because of time constraints. It is assumed that only ratios of 3:1 or more are likely to reduce the power of a study significantly (Pocock, 1995) and according to Torgerson and Campbell, (1997) oversampling of experimental groups (below the critical ratio of 3:1) in evaluations of new learning technologies is not problematic.

**Procedure and material**

Children were tested and trained during school time in separate rooms at their own school by three psychology master students. For the computer training group, the respective software was installed on the school’s PCs. Testing and training sessions were scheduled weekly within respect to school constraints, and each session took approximately 35-40 minutes over a period of three months.

**Pretest and posttest.** During the first and last session, all children’s VSPS were assessed by means of the TangSolver test. The TangSolver\(^1\) application that served for both assessment and training of participants’ VSPS is composed of three layers (TangSolver-Try-out, TangSolver Test, TangSolver Training). The application contains features that support interaction via computer mouse; it can be run on PC or Mac and requires installation of Java’s virtual machine, which is freely available. The screen consists of two windows: a smaller “model window” that shows the required shape and a “working space” in which participants can drag and drop MPs, and rotate or flip them by means of the mouse (see fig. 1).

Before starting the pre-test all children received training on how to drag, rotate and flip with the computer mouse by means of the TangSolver-Try-out. The try-out comprises of three parts requiring

\(^{1}\)The application used can be obtained by request from the first author (echabani@gmail.com).
dragging, rotating, and flipping, respectively; it was not time limited and participants could practice until the mouse manipulation was satisfactory. To evaluate children’s VSPS before and after training in a standardized way (without any feedback or guidance) we used the TangSolver Test. TangSolver Test is composed of two subtests (described above), a simple transformations test (STT) and a complex transformations test (CTT) (see fig 3). Each subtest contained eight items. The items were similar in terms of difficulty with two items at each of the four difficulty levels (2x4MP, 2x5MP, 2x6MP, and 2x7MP). None of the STT items required mental rotation, all solutions could be achieved by dragging and dropping MPs (see fig 3). In contrast, CTT items did require mental rotation; they called for an initial flip or rotation of MPs followed by drag and drop (see fig 3).

The pre-test and the post-test were time-limited (max time of 1:30 min/item). However, children who were quick (task completion <= 1:30) could make use of the “Next” bottom press, which displayed the following item. For children who were slow (task completion > 1:30), a window asking “Do you need more time? Yes - No” appeared, which allowed them one extra minute, after which the next item appeared automatically. Each test took 10-20 min, depending on how much extra time was used.

Cronbach’s alpha for the 16 items of pre- and post-test task completed scores based in the current sample (n=104) was .73, and .77, respectively. The internal consistency for the three groups (Computer, Face to Face and Control) was .70, .82, and .70 at pre-test and .57, .81, and .79 at the post-test, respectively—indicating rather high scale reliability. The test-retest reliability in the control group (n=21), who did not receive any training and completed the post-test at least 6 weeks after the pre-test, was .84, with a correlation coefficient of .8, indicating sufficient stability over the time.

Training sessions. The aim of the training was to support the participant when s/he could not solve the problem independently. The two training modalities we considered were Computer and Face to Face, while children in the Control group were engaged in discussions or drawing tasks. The two types of training differed in respect to the material used (computer vs. manipulative material) and in the manner children were tutored. The former practiced on computers and the guidance was exclusively through different visual cues (see fig. 1). The Face to Face groups practiced with manipulative material which required the presence of one assessor per child and the guidance was through visual hints and verbal prompts (teacher based instruction i.e., are you sure that should be here? Please have a look at the previous board) (see fig. 2). There were two training sessions and during each session the child was trained on three items. Both types of training used six different items similar to those used in the pre/post-test at each of the four difficulty levels; training started with the easiest level (4 MPs) and progressed to the most difficult level (7 MPs). The child could start on a new level or item only when the task was completed successfully at the previous level. Therefore, neither the time for completing the task nor the instruction was limited.

Measures and scoring. Although the range of outcomes measures provided by the application is extensive, in the current study VSPS were scored according to three criteria for each subtests (simple and
complex transformation test). The *Tasks-completed score* counted the number of items (max 8 pts.). *Accuracy scores* considers partial task performance by referring to the number of MPs correctly placed in incomplete puzzles (max 44 pts). For instance, in a puzzle with 7 MPs, the range is between 6 MPs correct (85.7%) to 2 MPs correct (28.6%), irrespective of the fact that the entire range would be associated with a “0” score in the binary Tasks-completed score. Finally, the *Time on task* score refers to the total time spent on the test, which we took to reflect individual differences in processing speed (Jensen, 1998). As the variance in time limited tests does not accurately reflect processing speed (Karweit, 1984), we calculated the score by dividing the total time (sec) spent on successful tasks by the number of successful tasks.

**Results**

Our main questions were whether and how training would change performance from pre- to post-test in “simple” or “complex” transformations tests and the two training groups (Computer vs. Face to Face), and whether these changes would be more pronounced in the experimental groups than the control group. To identify these effects, we analysed each of the three dependent measures (Accuracy, Tasks Completed scores and Time-on-task) by means of three-way ANOVAs for repeated measures with Session (pre- and post-test) and Test (simple transformations test (STT) vs. complex transformations test (CTT)), as within-participants factors and Training Group (Computer, Face to Face and Control) as between-participants factor. The theoretically most interesting result pattern would consist of a three-way interaction involving Session. Higher-order interactions were disentangled by means of independent samples t-tests, and effect sizes were assessed by means of Cohen’s d (where effect sizes of .20 are considered small, of .50 as medium, and of .80 or more as large). The analysis of gender effects was beyond the scope of this study but to inform future studies, we will provide gender-specific descriptive statistics for each of the variables. An alpha level of .05 was used for all statistical tests.

Before assessing the effect of training, we checked whether the three training groups were initially comparable. To do so, we entered the three pre-test scores (Accuracy, Tasks Completed and Time on Task) of the two subtests (STT and CTT) into one-way ANOVAs within Training Group (Computer, Face to Face and Control) as factor. There were no statistically significant group effects for either STT (p=.9, p=.8, p=.6 for Accuracy, Tasks Completed, and Time on Task, respectively) or CTT (p=.3, p=.8, p=.2, respectively).

**Accuracy**

The three-way ANOVA yielded a main effect of Session (F(1,101)=46.78, p<.001, ηp²=.32), indicating that participants improved from the first to the last session. The interaction effect of Session by Group (F(1,101)=3.8, p=.026, ηp²=.070) showed the three training groups improved differently however. A one-way ANOVA of gain scores (posttest minus pretest) for each Test (STT, CTT) revealed no significant differences between groups (p=.37) in STT but a significant effect in CTT (F(2,101)=4.42, p=.01 ηp²=.08). The
latter was due to that performance in the Control group differed from that in both the Computer group ($t(60)=3.24$, $p=.002$, $d=.087$) and the Face to Face group ($t(61)=2.03$, $p=.046$, $d=-.56$), while performance was comparable in the Computer and the Face to Face group ($p=.4$). As can be seen in fig. 4, the control group improved on the simple transformations test that relies on memory while in the complex transformations test involving mental rotation only the two training groups improved. This highlights the importance of training in a mental rotation task, irrespective of training modality.

**Table 1.**
Descriptive Statistics of Accuracy, of Simple and Complex Transformation Tests (Pre- and Post-test), per Training Group (Computer, Face to Face and Control) and by gender.

<table>
<thead>
<tr>
<th>Training Group</th>
<th>N</th>
<th>Simple Transformation Test</th>
<th>Complex Transformation Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td><strong>Computer</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
<td>22.53</td>
<td>8.86</td>
</tr>
<tr>
<td>Female</td>
<td>24</td>
<td>22.54</td>
<td>7.10</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>22.54</td>
<td>7.77</td>
</tr>
<tr>
<td><strong>Face to Face</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>24.14</td>
<td>7.09</td>
</tr>
<tr>
<td>Female</td>
<td>21</td>
<td>20.81</td>
<td>6.22</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>22.48</td>
<td>6.80</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td>21</td>
<td>23.00</td>
<td>8.77</td>
</tr>
</tbody>
</table>

**Fig. 4**
Accuracy (max. 44 pts) of Pre and Post-test by Training Groups (Computer, Face to Face and Control) of Simple and Complex Transformation Tests.
Tasks completed

The three-way ANOVA yielded main effects of Session (F(1,101)=60.62, p<.001, ηp²=.37), indicating that participant improved from the first to the last session, and of Test (F(1,101)=48.48, p<.001, ηp²=.33), showing that performance was better on STT than on CTT. The interaction effect of Session by Training group (F(1,101)=3.22, p=.044, ηp²=.060) indicated that the three training groups improved differently. A one-way ANOVA of the gain score of each test revealed that the significant difference between groups were again on CTT (F(2,101)=4.7, p=.01, ηp²=.085) but not on STT (p=.4). Performance on CTT differed between Computer (1.9) and Control groups (9.14), (t(60)=57, p=.003, d=0.85), and between Face to Face (2.07) and Control groups (.14), (t(60)=2.29, p=.02, d=.62), while the two training groups did not differ—even though numerically the Computer group showed better performance (see fig. 5). That is, the two training modalities were about equally efficient.

Table 2.
Descriptive Statistics of Tasks Completed, of Simple and Complex Transformation Tests (Pre- and Post-test), per Training Group (Computer, Face to Face and Control) and by gender.

<table>
<thead>
<tr>
<th>Training Group</th>
<th>Simple Transformation test</th>
<th>Complex Transformation test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pretest</td>
<td>Posttest</td>
</tr>
<tr>
<td><strong>Computer</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>17</td>
<td>3.41</td>
</tr>
<tr>
<td>Female</td>
<td>24</td>
<td>2.96</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>3.15</td>
</tr>
<tr>
<td><strong>Face to Face</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>21</td>
<td>3.38</td>
</tr>
<tr>
<td>Female</td>
<td>21</td>
<td>2.52</td>
</tr>
<tr>
<td>Total</td>
<td>42</td>
<td>2.95</td>
</tr>
<tr>
<td><strong>Control</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>10</td>
<td>2.80</td>
</tr>
<tr>
<td>Female</td>
<td>11</td>
<td>3.36</td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td>3.16</td>
</tr>
</tbody>
</table>

Fig. 5
Tasks Completed (max. 8 pts) of Pre and Post-test by Training Groups (Computer, Face to Face and Control) of Simple and Complex Transformation Tests.
Time on task

The three-way ANOVA yielded a main effects of Session ($F(1,85)=24.6$, $p<.001$, $\eta^2=.22$), showing that time on task decreased from the first to the last session (420 vs. 259 sec.), and of Test ($F(1,85)=10.24$, $p=.002$, $\eta^2=.10$), due to that participants spend more time on CTT (382 sec.) than on STT (297 sec.). There were no other significant interactions (see fig. 6).

Table 3.
Descriptive Statistics of Time on task (sec), of Simple and Complex Transformation Tests (Pre- and Post-test), per Training Group (Computer, Face to Face and Control) and by gender.

<table>
<thead>
<tr>
<th></th>
<th>Simple Transformation test Pretest</th>
<th></th>
<th></th>
<th>Complex Transformation Posttest</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N</td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
<td>N</td>
</tr>
<tr>
<td>Computer</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>341.27</td>
<td>232.91</td>
<td>116.72</td>
<td>54.69</td>
<td>387.23</td>
</tr>
<tr>
<td>Female</td>
<td>20</td>
<td>457.55</td>
<td>369.96</td>
<td>247.16</td>
<td>277.28</td>
<td>463.89</td>
</tr>
<tr>
<td>Total</td>
<td>36</td>
<td>405.87</td>
<td>317.78</td>
<td>189.19</td>
<td>217.57</td>
<td>429.82</td>
</tr>
<tr>
<td>Face to Face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>18</td>
<td>268.52</td>
<td>131.14</td>
<td>175.33</td>
<td>219.05</td>
<td>404.53</td>
</tr>
<tr>
<td>Female</td>
<td>16</td>
<td>453.01</td>
<td>324.47</td>
<td>138.40</td>
<td>97.71</td>
<td>407.02</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>355.34</td>
<td>255.84</td>
<td>157.95</td>
<td>171.49</td>
<td>405.70</td>
</tr>
<tr>
<td>Control</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>385.83</td>
<td>413.42</td>
<td>136.21</td>
<td>92.53</td>
<td>520.14</td>
</tr>
<tr>
<td>Female</td>
<td>10</td>
<td>434.20</td>
<td>307.80</td>
<td>366.23</td>
<td>394.38</td>
<td>503.91</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td>412.70</td>
<td>348.06</td>
<td>264.00</td>
<td>315.75</td>
<td>511.12</td>
</tr>
</tbody>
</table>

Fig. 6
Time on Task (sec) of Pre and Post-test by Training Groups (Computer, Face to Face and Control) of Simple and Complex Transformation Tests.
In order to assess the potential of computerized VSPS training vis-à-vis traditional (face to face) training, we compared the effectiveness of these training methods against a control group that did not receive any training. Our results showed that both training methods were equally efficient in significantly improving VSPS in terms of accuracy and the number of completed tasks (see fig. 4, 5 and 6). These observations support the conclusions of Uttal et al., (2012) and provide evidence that the combination of computer-based instruction and visual-only cues can be very effective (Mayer & Moreno, 2003; Mayer, 2009). This encourages the use of visual information in training children with language-based learning disabilities (e.g., Guarnera, Commodari & Peluso, 2013; Dalton & Proctor, 2007; Newhall, 2012; Smith & Tyler, 2009).

We suggest that applications like TangSolver provide methods and procedures that enhance VSPS at earlier ages, and that can assist teachers in several ways. For example, school-based interventions have been reported to suffer from numerous limitations, such as the necessity of well-trained tutors, the difficulty to find a way to provide standard instructions, and very high time demands on the involved personnel. Computer applications like Tangsolver have the potential to overcome such limitations, as they provide very standardized training conditions (which need not exclude individualized training levels) and an environment in which the learner can practice independently, at his or her own pace. Not only does this help reducing demands on teachers and well-trained tutors, it also encourages autonomous learning and self-management. As stressed by Black et al., (2006), the principle of learner’s autonomy implies that learners need to be given opportunities for strategic thinking and reflection about their own learning, and this is what such computer-based training can offer (Doering & Veletsianos, 2009). In conclusion, greater focus on the individual’s performance deficiencies would help concentrating training resources—the aim of blended learning (e.g., Picciano, 2009), which seeks to combine technology and traditional instruction rather than pitting one against the other (e.g., Cennamo, Ross & Ertemr, 2013).

However, the probably most important result of our study was the effect of training on the simple and complex transformations tests. Interestingly, the positive findings were mainly restricted to performance on tasks that required flipping or rotating MPs as assessed through the complex transformations test (see fig. 4, 5 and 6). In contrast, training had little impact on performance in the simple transformations test, which assesses visual memory. Thus, while a simple transformations test might be appropriate for younger children or children with structural mental-rotation impairments (e.g., Guarnera, Commodari & Peluso, 2013), only performance on the complex transformations test seems diagnostic for VSPS proper. This supports previous claims that VSPS relies on the ability to grasp complex systems and to discover complex spatial relationships and possible transformations (e.g., Davis, Rimm & Siegle, 2011; Golon, 2008; Subotnik, Olszewski-Kubilius & Worrell, 2011). As noted by many authors, the failure to
identify and nurture these abilities does not only do a disservice to the children involved, but also to society as a whole (Dai, Swanson & Cheng, 2011; Subotnik, et al., 2011).

While our findings demonstrate that training in spatially demanding construction tasks is effective, further analysis is needed to identify the role of individual strategies and changes therein—which include holistic mental rotation and step-by-step feature-based comparison (Cherney & Neff, 2004; Geiser, Lehmann & Eid, 2006; Glück & Fitting, 2003). In particular, there is a need to investigate training data by using mathematically based data mining methods (see, www.ed.gov/technology and http://myweb.fsu.edu/vshute/publications.html). Indeed, the wide variety of outcome measures, such the number of click, moves, and so forth allows the construction of knowledge tracing models (e.g., Aleven et al., 2010; Baker et al., 2010; Feng & Heffernan, 2010; Shute & Zapato-Rivera, 2012). Hence, the potential of applications like TangSolver goes beyond demonstrating training outcomes by inviting process-based analyses. However, good models guiding such finer-grained analyses are rare and a main challenge will be to find the most diagnostic performance indicators to predict the trainees’ performance (Shute & Ke, 2012) and persistence (Ventura & Shute, 2013), self-regulation, control strategies, motivation, etc. (Shute & Ventura, 2013).

The current study presents several shortcomings. First, without a follow-up study we cannot know to which degree the increase of performance we observed was able to induce long-term learning. In particular, while keeping items constant from pre- to posttest was important for valid comparison, we cannot tell task mastery from learning proper (Guskey, 2007). Thus, further studies should use test items that are significantly different from training items (e.g., convex tangrams) and should look into longer post-training intervals (Uttal et al., 2012). Moreover, even though processing speed is frequently assessed in mental rotation tasks to check for gender differences in performance, future studies should consider dropping the time limit to allow for different processing styles. Finally, while our findings show that visual cues can be effective, more empirical research is needed to test for the most efficient format of cues and instructions, as well as for possible interactions with individual processing styles.

Our present study is but one step into the direction of cognitive enhancement in children and there is certainly a need for further research to substantiate our findings. It would be interesting to see whether more, or longer spatial training, or training on different spatial tasks, lead to more, or more enduring enhancement. Also interesting is whether our training effects scale up to a less selected population, especially to children with special needs, and to older children and teens. Finally, in evaluating the efficacy of training, it is important to consider that it needs to be an on-going process (an integral part of an individual’s learning) rather than a short event assessed mainly by experimental research, which calls for more collaboration between research and educational agents.