Instr Sci (2014) 42:995–1012
DOI 10.1007/s11251-014-9316-7

Effectiveness of visual and verbal prompts in training visuospatial processing skills in school age children

Ellahe Chabani · Bernhard Hommel

Received: 2 September 2012 / Accepted: 26 March 2014 / Published online: 5 April 2014

Abstract Recent decades have witnessed a growing interest in intervention-based assessment to promote and enhance children's learning. In this study, we explored the potential effect of an experimental visual–spatial intervention procedure and possible training benefits of two prompting modalities: one group received training with verbal and visual prompts, a second group training with visual prompts only, while a third, control group did not receive any training. The two training methods led to significant improvements of performance in visuospatial tasks as compared to control group, and they did so about equally well. Our findings provide evidence for the efficiency and benefits of interventions targeting visuospatial processing skills. The success of such interventions does not seem to be bounded by age or gender, and it seems that visual cues are particularly effective.

Keywords Visuospatial · Problem solving · Individual differences · Verbal prompting · Visual prompting

Introduction

The recent years have seen a trend away from a unitary concept of human intelligence and towards concepts that allow for multiple types and varieties of intelligence. Among those, visual–spatial processing skills (VSPSs), which reflect the ability to generate, retain, retrieve, and transform visual stimulus material (e.g., Gardner 1983; Linn and Petersen 1985; Lohman 1988; Sternberg 2003; Van Garderen and Montague 2003), have been

E. Chabani · B. Hommel (✉)
Cognitive Psychology, Institute of Psychology, Leiden University, Leiden, The Netherlands
e-mail: hommel@fsw.leidenuniv.nl

E. Chabani
e-mail: echabani@gmail.com

B. Hommel
Leiden Institute for Brain and Cognition, Leiden, The Netherlands

considered particularly important. Indeed, there is evidence suggesting a pivotal role of VSPSs in performance related to STEM (science, technology, engineering, and mathematics; e.g., Lubinski 2010; Uttal et al. 2012; Wai et al. 2009) and in early academic skills (e.g., math, reading, writing; Assel et al. 2003; Cheng and Mix 2012; Holmes et al. 2008; Passolunghi and Mammuarella 2010; Rasmussen and Bisanz 2005).

Furthermore, education is undergoing a profound change worldwide and coming generations will grow up in an increasingly visual multimedia environment; recent technological developments (Webs, App's applications) rely heavily on VSPS. Theoretically, success at school and at future workplaces will thus largely depend on visualization, grasping the big picture, visual memory, pattern-finding and thinking graphically (Carr 2008, 2010). Yet, it has been frequently observed that VSPS are not adequately practiced, addressed, and assessed at school (National Research Council 2006; Webb et al. 2007). For instance, recently, the National Council of Teachers of Mathematics (2010) and the US-American National Research Council (2006) have warned that visuospatial intelligence is not just under-supported but under-valued, and therefore under-instructed—which has been taken to call for a national commitment to the development of visuospatial thinking across all domains of the school curriculum.

To move one step forward into this direction, our project aimed at developing a VSPS instrument that can serve for both assessing and enhancing VSPS in school age children, which we thought might not only promote teaching and training in that domain but may also stimulate researchers to further develop related assessment and training procedures.

Visuospatial processing skills (VSPS)

Researchers and theorists in different areas have acknowledged that VSPS is not a unitary construct, but rather can be broken into a collection of sub-skills or components (e.g., Carroll 1993; Eliot and Smith 1983; Lohman 1988; Kaufman 2007; Sutton and Williams 2007). Unfortunately, meta-analyses, factor-analytical assessments (Carroll 1993), and other approaches have failed to find clear evidence for a particular number of separable factors, so there is currently no consensus on how many factors are involved. What seems to be clear, however, is that spatial perception, spatial visualization, spatial orientation, spatial sequencing, mental rotation, and working memory are among them (Allen 2003; Carroll 1993; Eliot and Smith 1983; Lohman 1988; Kaufman 2007; Linn and Petersen 1985; Hegarty and Waller 2004; Sutton and Williams 2007; Willcutt et al. 2005). Regardless of confusion regarding the definition, its underlying factors or sub-skills, and the classification (e.g., D'Oliveira 2004; Hegarty and Waller 2005), there is evidence for the malleability of VSPS (Uttal et al. 2012). Uttal et al.'s meta-analysis of over 217 studies on VSPS confirmed the theoretical and practical importance of visuospatial skills at any age and indicated that even short training procedures can significantly improve VSPS. 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. According to Uttal et al. (p. 54) "playing active games has the potential to enhance spatial thinking substantially, even when compared to a strong control group." One potential explanation for the lack of studies in younger children is the lack of child-friendly testing and assessment instrument. Even though the experimental material developed for the present study does not aim to reconcile the different theories and conceptions of VSPS, it aimed at providing means to overcome this shortcoming.

Testing and assessment of VSPS

Eliot and Smith (1983) have distinguished between VSPS recognition tasks (e.g., copying task, embedded figure and visual memory, mental rotation of shapes) and manipulation tasks (e.g., block rotation, block counting, solving mazes, and paper folding). A wide variety of tests and assessment instruments exists and many psychometric intelligence tests for children include visuospatial tasks such as Pattern Reasoning, Block Design of WISC-IV (2004), Matrix Reasoning of Kaufman Assessment Battery for Children (2004), or the Test of Visual-Perceptual Skills-Revised TVPS-R; Gardner 1996). While there is significant debate about what exactly these tests measure (e.g., Mathewson 1999), it is recognized that the subtests for children provide an objective and standardized measure of particular subskills—i.e. visualization. For instance, in the Block Design test individuals are asked to reproduce a design from colored plastic blocks. Such tasks require the ability to analyze and synthesize an abstract design, which is considered a measure of spatial visualization. A critical point is that such tests measure the broad concept of spatial visualization but do not address isolated subskills, such as mental rotation or visual discrimination.

Others have argued that the main goal of such summative and normative sub-tests is to compare a given child's scores to the age-group standards (Flanagan and Kaufman 2004; Haywood and Lidz 2007; Sternberg and Grigorenko 2002). In particular, the purpose of such tests is often to detect learning disabilities and eligibility or special education or related services. This leads to a strong focus on current performance rather than on the potential that a given child may possess. Accordingly, most available tests do not provide enough information for educators to create programs to remedy a child's learning problems (e.g., Haywood and Lidz 2007; Sternberg and Grigorenko 2002). This shortcoming has motivated the idea of a more intervention-based assessment of cognitive abilities that considers both current performance and the potential to improve.

Intervention-based assessment

Intervention-based assessment or formative evaluation, as coined by Black and Wiliam (1998), refers to "all those activities undertaken by teachers, and/or by students, which provide information to be used as feedback to modify the teaching and learning activities in which they are engaged." The last decade has witnessed the development of different types of formative evaluation—classroom-based assessment, such as dynamic assessment (DA) or response to intervention (RTI). Notwithstanding differences within and between such approaches in terms of theoretical premises, historical roots, and procedure (for an overview see, Archer and Hughes 2011; Fuchs and Fuchs 2006; Grigorenko 2009), both DA and RTI approaches are learner and process oriented. Basically, both concepts aim at systematic screening and information gathering procedures to monitor students' progress efficiently. Ideally, information provided through assessments enable the identification of instructional modalities, material, and technologies to promote active learning, as well as developing pedagogical strategies for children with special strength/weakness or educational needs. Theoretically, by implementing screening, progress monitoring, and outcome assessments in a reliable and valid way, it is possible to reduce the use of time-consuming and expensive formal diagnostic instruments (Cortiella 2011) and provide more efficient help to learners.

In view of the role played by VSPSs in learning (Assel et al. 2003; Cheng and Mix 2012; Holmes et al. 2008; Passolunghi and Mammuarella 2010; Rasmussen and Bisanz 2005), it is reasonable to assume that some of the difficulties children exhibit at school and in other learning environments can be explained by weaknesses in VSPSs rather than in some general capacity to learn. Therefore, assessment and trainability of such skills can have important implications for guiding educational interventions and/or helping teachers in their teaching approach.

Verbal and visual feedback in VSPS learning

In any intervention-based assessment or formative evaluation, instruction and feedback play a critical role. Providing guidance and feedback to learners about the performed action or task is an important factor that affects learning and skill acquiring (see Hattie and Timperley 2007; and Shute 2008, for two particularly influential studies). The main purpose of feedback is considered "to reduce discrepancies between current understandings and performance and a goal" (Hattie and Timperley 2007, p. 86) and to "signal a gap between a current level of performance and some desired level of performance or goal" (Shute 2008, p. 157). Hattie and Timperley have identified four levels of feedback with differential effect on learning: (1) feedback on the task (2) feedback about the processing of the task, (3) feedback about self-regulation, and (4) feedback about the self as a person. Their conclusion was that feedback about the self (i.e. "good girl/boy," "great try," etc.) represents the least effective form of feedback. Such feedbacks have no instruction-related content and might improve the student's investment of effort or attitude toward learning but do not affect achievement. In contrast, feedback on the task and feedback about the processing of the task are effective in enhancing and facilitating depth learning. The third level about self-regulation guide learners how to engage in future learning situations and helps students attribute their success or failure at a task to a particular and specific cause rather than to their self-efficacy. Along the same lines, Shute (2008) considers feedback effective to the degree that it focuses on the task, and he suggests that it should be presented in manageable units and not be too elaborated.

While these frameworks are helpful in guiding the design of interventions using feedback and instructions, it remains unclear of which kind and modality efficient feedback should be. Classroom-based instructions are commonly verbal (e.g., "check this part", "give concrete examples"), sometimes accompanied by visual information, such as images or graphs. Even though verbal instructions are certainly important in guiding the child's attention to the relevant information, there is a need to also consider students with different learning needs and preferences. For instance, students with hearing impairments (Dye et al. 2008) and children with language-based learning disabilities or different linguistic backgrounds may have a hard time decoding verbal instructions (e.g., Cortiella 2011; Paul 2007) and would thus profit more from purely visual feedback such as visual scaffolding, or showing complete or partial solutions.

Visual feedback or visual cues has generally been studied within multimedia learning (e.g., Butcher 2006; Hegarty and Just 1993; Kalyuga et al. 2003; Moreno and Mayer 1999; Olina et al. 2006). Two major theoretical frameworks have guided empirical research in multimedia learning. The cognitive load theory (CLT) describes learning in terms of information processing system involving working memory storage (e.g., Moreno 2010; Paas et al. 2004; Schnotz and Kurschner 2007). In short, if in the learning process mental resources (working memory) are exhausted then learning may fail to occur. The optimal solution is then modifying the instructional material to lower the level of cognitive load. CLT-motivated studies have argued that cognitive processes involved in active coordination of visual and verbal information during learning can promote students' understanding, in particular with complex materials. Another influential theory is the cognitive theory of multimedia learning (Mayer 2005). It is based on Paivio's (1986) assumption that information processing occurs in two complementary channels: a visual/pictorial channel and an auditory/verbal channel, which are sensitive for different kinds of information. Mayer (p. 47) states that "people learn more deeply from words and pictures than from words alone" and he suggests that instructional designs should avoid cognitive overload in learners by using both channels to provide instructions. However, while multimedia studies have provided evidence for efficient learning with visual–verbal prompts, to our knowledge no study has used manipulative tasks and classroom-based instructions.

The current study

The primary objectives of the present study were threefold. Firstly, we were interested to evaluate whether the VSPSs of children could be improved by training at all. We expected that children who were trained would perform better on the post-test than children in a control group in which no VSPS training was provided. Second, we compared two types of instruction and guidance in a pre-train-post-test design. As mentioned above, there is evidence for the efficacy of combining visual and verbal prompts during training (e.g., Resing and Elliott 2011), but little is known about the efficacy of purely visual prompting—which however would be more suited for children with verbal difficulties. Accordingly, we compared two training modalities by providing some children with both verbal and visual hints and other children with visual hints only. Although this manipulation was thought to inform later studies on children with verbal problems, the present study focused on typically developing, healthy children. We considered that children with no particular verbal or language difficulty might profit more from multimodal support (Mayer 2005).

Our third aim was more explorative. Studies of gender differences—in particular on VSPS—have generated considerable controversy among researchers (for an extensive review, see Halpern 2012; Newcombe and Learmonth 2005). Among other issues, one important question raised by Uttal et al. (2012) was whether the gender differences that were observed in a wide variety of spatial tasks reflect true (structural) gender differences or mainly differential degrees of practice. While Uttal et al. pointed out that the "gender gap in spatial skills did not shrink due to training" (p. 43), other researchers (e.g., Terlecki et al. 2008; Tzuriel and Egozi 2010) reported that gender differences can often be removed by training. A similar discussion in the literature refers to age. Piaget (1977) considered that early spatial understanding is topological in nature, while Euclidean representations would emerge no earlier than at the age of 9 or 10. This prediction is not consistent with findings reported by Sophian (2000), who demonstrated that 4- and 5-year-olds can compare proportions and figures and are able to correctly match a shrunken picture to the original. Even though we are not committed with regard to the existence and cause of gender and age effects, and even though we consider the available evidence as too inconclusive to justify directed predictions, we were interested to see whether gender and age might mediate possible training benefits.

Method

Participants

The sample consisted of 281 typical children (152 boys and 129 girls) with a mean age of 95 months, SD = 13.14, with no known histories of developmental, neurological, or learning problems. Children came from a number of primary schools in the Netherlands and France. The languages spoken by the children and the trainers were Dutch, English and French. In all cases parental consent for participation was obtained.

Design and procedure

The study utilized a pre-test—training—post-test control-group design, with two training groups and a control group. Before dynamic testing started, the children were quasi-randomly assigned to three groups, but matched for general inductive reasoning ability. The three groups consisted of the verbal plus visual training group, the visual training group, and the control group. Pre-test and post-tests of the dynamic test were administered to all children. Children in both training groups received two trainings between pre- and post-test, while the control group was engaged in discussion and drawing tasks (see Table 1 for the design). The prompting during the combined verbal–visual training consisted of verbal instructions in the children's native language and of visual aids. Children in the visual training group received only visual aids. Children were tested individually by three students of psychology and the first author. Testing and training sessions were scheduled weekly, in separate rooms at the children's own school and each session took approximately 35–40 min. We have tried to keep the time between pre- and post-testing as equivalent as possible; nonetheless due to school's activities (i.e. holidays, school trips, end of the school year party etc.) some post-testing was delayed by ±2–3 weeks.

Instruments

Development of VSPS instrument

The rationale The experimental VSPS developed for this study is based on the Tangram Chinese game (putting together seven geometrical forms to form shape). This choice was initially motivated by the fact that geometric knowledge has been used to evaluate various visuospatial solving problem abilities (Dehaene et al. 2006; Lee et al. 2009; Lovett and Forbus 2010; National Council of Teacher's Mathematics 2003). For instance, Block Design of WISC or Kohs' blocks assess an individual's ability to analyse, synthesize, and reproduce an abstract design. Bodies of studies on spatial visualization ability and studies on general problem-solving ability suggest that such tasks tap into the same cognitive abilities and are useful in many advanced disciplines such STEM (science technology engineering and math) (for a review see, Hegarty et al. 2007). Arranging geometrical forms displays the grouping and the fact that there is not a unique solution links the task to problem solving theories (e.g., Ford 2003; Foster 2007; Mayer and Wittrock 2006; Slocum et al. 2003).

From an educational point of view, Tangram assists in developing geometrical knowledge, reasoning, geometrical imagination, development of creative thinking, including the understanding of geometrical shapes, size, and position in space, as well as the reliance of perceived shape on position in space (Van Hiele 1983). For example, Tangram games allow the consideration of shapes and relationships between shapes (e.g., two triangles can make a square), which links performance to other domains, such as mathematics, without having to resort to formulas but rather by developing a geometric, basic understanding of concepts such as "area" and "congruence" (for an overview, see, Bohning and Althouse 1997; Gardner 1996). Another argument for a Tangram-based intervention is that it requires or at least benefits from all visuospatial abilities that so far

Table 1 Experimental design

Groups (N)	RPM	Pre-test (VS-T)	Dynamic intervention (VS-I)		Post-test (VS-T)
			Session 1	Session 2	
VeVis (88)	X	X	X	X	X
Vis (99)	X	X	X	X	X
Control (94)	X	X	–	–	X

VeVis verbal and visual dynamic intervention, *RPM* Raven progressive matrices, *VS-T* visuo-spatial test

have been related to but are not limited to the ability to understand how objects appear in different positions, which is referred as spatial visualization (Lohman 1988; Kaufman 2007; Linn and Petersen 1985). Spatial visualization includes the ability to manipulate information sequentially and spatially, the skill to conceptualize how objects relate to each other in space, the ability to visualize mental rotation of objects, 2-dimensional understanding, recall of something seen some time ago, or immediate and delayed memory related visual memory (Carroll 1993; Eliot and Smith 1983; Lohman 1988; Kaufman 2007; Linn and Petersen 1985; Hegarty and Waller 2004; Sutton and Williams 2007; Willcutt et al. 2005).

Furthermore, with Tangram material, understanding the task is straightforward. The child can evaluate the correctness of his/her actions and his/her progress relatively easily. Moreover, solving Tangram puzzles does not rely on verbal capacity or typical academic knowledge (i.e. reading, writing, calculating), the puzzles are challenging and yet manageable and they provide a motivating context in which children are more likely to experience enjoyment rather than the stress of a testing situation. In short, the foremost argument for our choice was that in geometric-puzzle construction individuals are likely to use qualitative, and/or categorical, representations to reason about shapes or space (spatial visualization and visual discrimination), and process the spatial relations between elements in a visual scene.

Composition of Items

In a typical Tangram game a large variety of shapes can be created by arranging seven geometrical forms. In our Tangram game, we made use of master pieces (MPs) that were constructed by pre-combined forms (e.g., a MP could be the combination of a triangle and square) and as well the standard forms. Thus shapes could be created by using range of MPs starting with 4 MPs, 5 MPs, 6 MPs to 7 MPs (that correspond to the seven classical geometric forms). The number of MPs needed for making each shapes was considered an index of the difficulty level of the items.

Testing items and testing procedure

To evaluate children's VSPS before and after training in standard way (without any feedback or guidance), eight items (eight different figures)—two items at each of four difficulty levels—were selected. All MPs were made of green sticky plastic that were placed on a white board (see Fig. 1). For both pre- and post-test the same material was used. During the tests a puzzle figure, printed on a card, and a white board with the necessary MPs were presented to the child. The child task was to pick up the forms and solve the puzzle as quickly as possible. However, the testing was time limited, and for

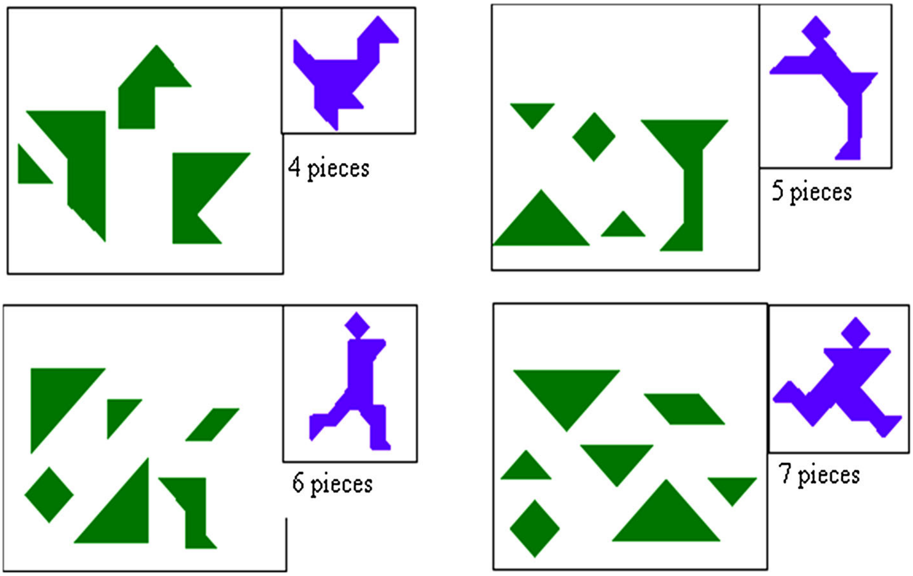

Fig. 1 Examples of items according to each difficulty level of pre- and post-test

solving each item a max of 2 min was attributed, after which independently of in/correctness of the task the next item was presented to the child. During the tests the feedback consists in encouragement i.e. "well done", or in case of unsuccessful attempt i.e. "that is a really difficult one, let's try another one". The condition for termination was three consecutive failures.

Inductive reasoning

The Raven progressive matrices (RPM) test (Raven et al. 1998) was used to match children with regard to their inductive reasoning ability. The Raven test is a broadly used non-verbal multiple-choice test of visuospatial inductive reasoning. In each test item, the child is asked to identify the missing element that completes a pattern.

Intervention items and intervention procedure

The intervention consisted in providing guidance when a child could not solve the problem alone. The two intervention were Verbal and Visual (VeVis) training and Visual (Vis) training. Six different items to those used in the testing were selected, and for each item four boards (with 4, 5 6 and 7 MPs) were made. The aim was to provide intervention at all four levels for each of the six selected items. We used the same material but with three different colors (blue, red, and yellow) (see Fig. 2). It was assumed that children who have more developed analogical capacity would recognize how different master pieces were either combined or divided, and that this would lead to more independent successful task accomplishment.

The verbal prompts as of metacognitive level were mainly based on self-exploration, and gradually moved to more specific instruction (see Table 2) that were similar to studies

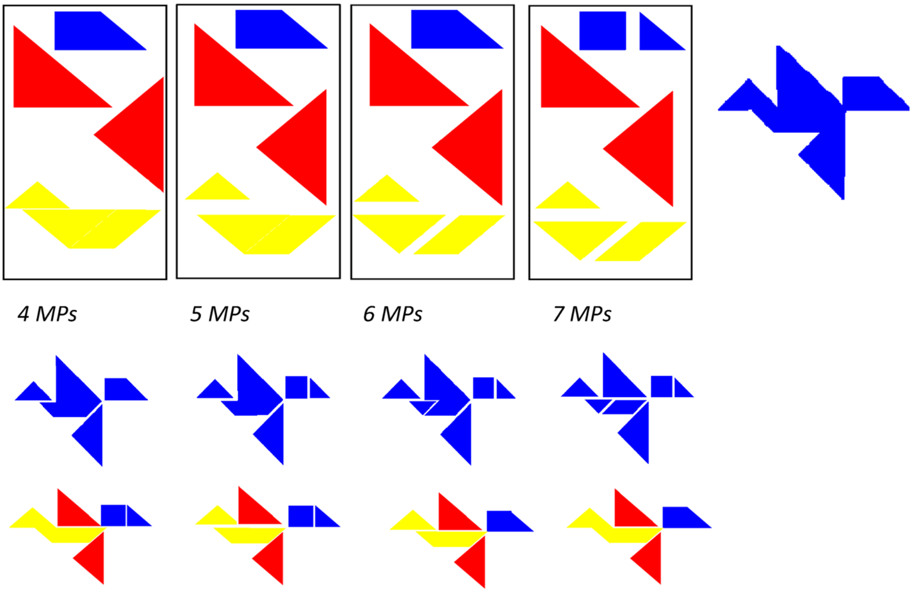

Fig. 2 An example of intervention material and visual prompts

of Resing and Elliott (2011). There were two kinds of visual prompts: Monochrome cues showing how MPs could be broken up and colored cues doing the same, only that here all pieces were in the same color as in the MP (see Fig. 2).

Both interventions started with the presentation of a board with 4 MPs (the easiest level) and then progressed to the most difficult level with 5, 6, and 7 MPs, respectively. The children were asked to make the puzzle in their own rhythm and were told that, if they would encounter difficulty in solving the problem, we would work together. After 1–2 min of unsuccessful trying, the intervention took place, in the modality of the respective group (VeVis or Vis). In the VeVis group, visual cues were provided when verbal prompts were not sufficient. With both kinds of interventions, the child would first be presented with the monochromic cards (see Table 2). If the children would be able to derive the general spatial composition principle from these cards, they should be able to solve the task. However, we considered that some children might not be able to abstract from the actual colors of the puzzle components, which is why we presented the colored cards. If the child was struggling with an item or at a particular level, the intervention was abandoned momentarily to avoid frustration. After a short break of a few minutes, the intervention was continued with the next item or level and all previously completed boards were put on the table in front of the child.

Scoring

Performance during pre- and post-test was assessed by creating three scores: the *Time On Task* score, mainly a "bonus" variable (in the sense of "going through" in the face of difficulty) that represents the time taken to complete the task with a maximum of 2 min/item; the *Accuracy* score, a variable that represents the total number of correctly pieces placed per item; and the *Tasks Completed* score, another variable that counts whether a given task was completed (1) or not (0). Of main interest were changes in these variables

Table 2 Order of verbal and visual prompts offered during the training procedure

	Prompts type and order	Frequency of use
Verbal graduated prompts		
Self-exploration	Please compare those boards Which changes do you see? Can you recognize which pieces were combined or were divided? Are you sure that is how they were combined? Are you sure that this form should be here? Please check again?	The amount was adapted to child responsiveness and assessor's judgment
Specific	This form should be here This is the head and not the body or foot	
Visual cues		
Solution card *AVC*	Puzzle picture with one color with breaking lines	VeVis: 10 s and 2 times Vis: 10 s and up to 3 times
CVC	Puzzle picture with color match to master pieces with breaking lines	VeVis: 10 s and 2 times Vis: 10 s and up to 3 times

AVC abstract visual cue, *CVC* concrete visual cue, *VeVis* visual&verbal training, *Vis* visual training

from pre- to post-test, and in particular changes that were restricted to, or more pronounced in the two actual training groups than in the control group.

Results

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 (Time On Task, Accuracy, and Tasks Completed scores) into separate three-way ANOVAs with Training Group (verbal–visual, visual, and control group), Gender (male vs. female), and Age Group (three age intervals: AgeG1 = 6–7.5 years, AgeG2 = 7.6–8.5 years, and AgeG3 = >8.5 years) as between-participants factors. Neither Training Group [Time On Task, $F(1,281) = 2.40$, $p = .09$; Accuracy, $F(1,281) = 2.40$, $p = .06$; and Tasks Completed, $F(1,28) < 1$] nor Gender [Time On Task, $F(1,281) < 1$; Accuracy, $F(1,281) < 1$; Tasks Completed, $F(1,281) < 1$] showed any significant pre-intervention difference between groups. Age Group also showed no effect for two of the three scores [Accuracy, $F(2,281) < 1$; Tasks Completed, $F(2,281) < 1$]. The only significant difference was found for Age Group regarding Time On Task, $F(2,281) = 3.44$, $p = .033$, $\eta p^2 = .027$. LSD post hoc tests indicated that AgeG1 ($M = 681.7$, $SD = 82.08$) spent less time on completing puzzles than AgeG2 ($M = 709.63$, $SD = 82.08$, $p = .028$) and AgeG3 ($M = 712.65$, $SD = 106.67$, $p = .026$), while there were no significant differences between AgeG2 and AgeG3.

As pointed out, our main interest was whether and where changes from pre- to post-test were more pronounced in the two training groups than in the control group. To identify these effects, we analyzed each of the three dependent measures (Time On Task, Accuracy and Tasks Completed score) by means of a four-way ANOVA for repeated measures with session (pre- and post-test) as the within-participant factor, and Training Group (verbal–

visual, visual, and control), Gender (male vs. female), and Age Group (see above) as between-participants factors. Gender and Age Group were included to identify possible individual differences in the effectiveness of the training. The theoretically most interesting result pattern would consist of a two-way interaction involving Session and Training Group and higher-order interactions including these two factors. See Table 3 for descriptive statistics. An alpha level of .05 was used for all statistical tests.

Time on task

The four-way ANOVA yielded main effects of Session, $F(1,263) = 83.39$, $p < .001$, $\eta p^2 = .24$, Training Group, $F(2,263) = 3.18$, $p = .043$, $\eta p^2 = .02$, and Age Group, $F(2,263) = 4.09$, $p = .018$, $\eta p^2 = .03$, indicating that participants spent more time on task after the intervention (763 vs. 700 s), that less time was spent in the control group (717 s) than in the two Training Groups (742 and 734 s for Vis and VeVis, respectively), and that the youngest group spent less time on task (715 s) than the two older groups (739 and 740 s). More importantly, however, Session interacted with Training Group, $F(2,263) = 11.07$, $p < .001$, $\eta p^2 = .08$, was involved in a reliable three-way interaction including Session, Training Group, and Age Group, $F(4,263) = 2.59$, $p = .037$, $\eta p^2 = .04$. The latter was due to that the interaction of Session and Training Group was significant in the two older age groups, $F(2,93) = 9.44$, $p < .001$, $\eta p^2 = .17$, and $F(2,66) = 5.16$, $p < .01$, $\eta p^2 = .14$, respectively, but not in the youngest group, $p > .5$.

As Fig. 3 indicates, the time on task increased from the pre- to the post-test in the youngest group, but it did so for all three types of training alike, that is, independent of the presence and the type of training. In other words, training had no specific impact on the youngest age group. In contrast, in the two oldest groups the time on task did not significantly increase over session in the control group (p's $> .37$), while purely visual training led to an increase of time on task in both the medium, $t(32) = 7.17$, $p < .001$, and the oldest age group, $t(26) = 5.49$, $p < .001$. The effect of combined verbal and visual training was less clear, producing a significant effect in the medium age group, $t(31) = 2.07$, $p = .047$, and no significant effect in the oldest age group, $t(20) < 1$. Hence, intervention-specific effects on the Time on Task "bonus" score in the two older groups were strongest with purely visual interventions.

Accuracy

The four-way ANOVA yielded main effects of Session, $F(1,263) = 67.40$, $p < .001$, $\eta p^2 = .20$, and Training Group, $F(2,263) = 6.75$, $p < .001$, $\eta p^2 = .05$, indicating that participants were more accurate after the intervention and that they were most accurate in the combined (verbal–visual) Training Group (18.1), least accurate in the control group (15.4) and intermediate in the Vis group (16.9). Furthermore, Session interacted with gender, $F(1,263) = 7.17$, $p < .01$, $\eta p^2 = .08$, and was involved in a significant three-way interaction including Session, Gender, and Age Group, $F(2,263) = 4.04$, $p = .02$, $\eta p^2 = .03$. The latter was due to that the interaction of Session and Gender was reliable in the two older age groups, $F(1,97) = 8{,}64$, $p = 004$, $\eta p^2 = .08$, and $F(1,70) = 6{,}41$, $p = .014$, $\eta p^2 = .08$, respectively, but not in the youngest group, $p > .57$. As Fig. 4 indicates, all groups benefitted from training but girls from the older age groups did so in particular. The interaction of Session and Training Group just missed the significance criterion ($p = .066$) as did the four-way interaction ($p = .062$). However, numerically

Table 3 Descriptive statistics of pre-test and post-test for Time On Task, Accuracy and Tasks Completed scores per condition, gender and age groups

	N	Time on Task(s)		Accuracy		Task completed	
		M	SD	M	SD	M	SD
Pre-test							
Groups							
Control	94	691.44	96.19	14.71	4.80	.85	2.82
Vis	99	691.79	90.80	16.11	5.85	.69	.86
VeVis	88	716.69	87.91	16.58	5.92	.81	1.04
Total	281	699.47	92.17	15.79	5.58	.78	1.80
Gender							
Male	152	700.49	97.11	15.93	5.56	.84	2.28
Female	129	698.27	86.35	15.62	5.61	.71	.99
Total	281	699.47	92.17	15.79	5.58	.78	1.80
Age							
AgeG1 (6–7.5)	110	681.70	88.47	15.93	5.43	.87	2.67
AgeG2 (7.6–8.5)	99	709.63	82.08	15.72	5.58	.69	.80
AgeG3 ≥8.5	72	712.65	106.67	15.68	5.86	.76	1.00
Total	281	699.47	92.17	15.79	5.58	.78	1.80
Post-test							
Groups							
Control	94	738.38	110.81	15.97	4.35	1.11	.99
Vis	99	792.78	74.33	17.89	5.44	1.85	1.42
VeVis	88	753.36	72.33	19.48	5.11	1.98	1.55
Total	281	762.24	90.46	17.74	5.17	1.64	1.38
Gender							
Male	152	759.22	98.35	17.33	5.33	1.59	1.33
Female	129	765.79	80.41	18.23	4.96	1.70	1.46
Total	281	762.24	90.46	17.74	5.17	1.98	1.54
Age							
AgeG1 (6–7.5)	110	750.17	96.05	17.73	5.44	1.60	1.42
AgeG2 (7.6–8.5)	99	767.37	87.60	17.67	4.94	1.55	1.34
AgeG3 ≥8.5	72	773.61	84.39	17.88	5.13	1.83	1.39
Total	281	762.24	90.46	17.74	5.17	1.64	1.38

speaking the strongest improvement across sessions was observed with VeVis training (15.4, 16.9, and 18.1 for Control, Vis, and VeVis, respectively).

Tasks Completed

The four-way ANOVA yielded main effects of Session, $F(1,263) = 201.03$, $p < .001$, $\eta p^2 = .43$, and Training Group, $F(2,263) = 7.01$, $p < .001$, $\eta p^2 = .05$, indicating that participants completed more tasks after the intervention (.7 vs. 1.7) and that children in the verbal–visual training group completed the most tasks (1.4), the control group the fewest (.9), while the Vis group fell in between (1.3). More importantly, however, Session

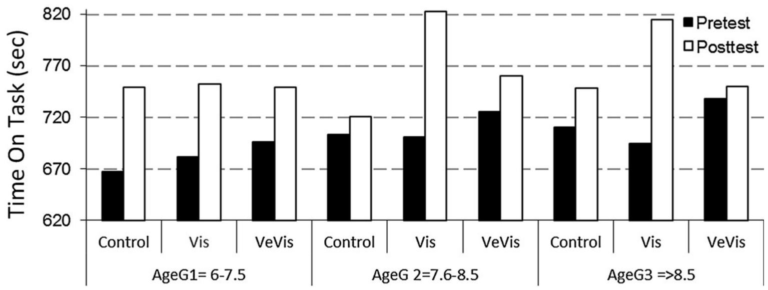

Fig. 3 Changes from pre-test to post-test of Time on Task, for groups (verbal–visual, visual and control), and age group

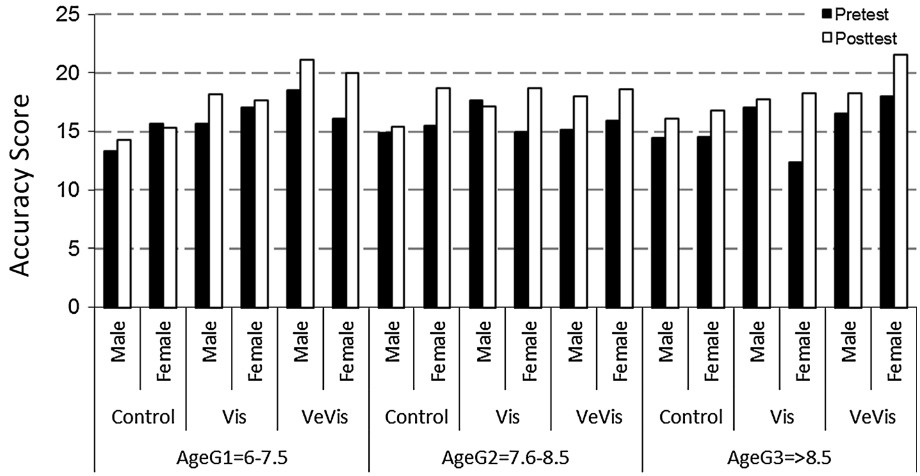

Fig. 4 Changes from pre-test to post-test of accuracy, for groups (verbal–visual, visual and control), age and gender

interacted with Training Group, $F(2,263) = 10.91, p < .001, \eta p^2 = .08$, and was involved in a significant three-way interaction including Session, Training Group, and Age Group, $F(4,263) = 3.05, p = .018, \eta p^2 = .04$. The latter was due to that the interaction of Session and Training Group was significant only in the youngest group, $F(2,104) = 12.84$, $p < .001$, $\eta p^2 = .20$, but did not reach significance in the two older age groups, $F(2,93) = 2.24$, $p = .11$, and $F(2,66) = 2.33$, $p = .10$, respectively. The effect in the youngest group showed that Session had no effect in the control group, $p = .4$, but improved performance in both intervention groups, ps $< .001$ (Fig. 5).

Discussion

The aim of the present study was to evaluate the potential effect of VSPS interventions with an experimental intervention. To this end, we compared two training modalities: verbal and visual hints versus visual hints only. We expected that children who received

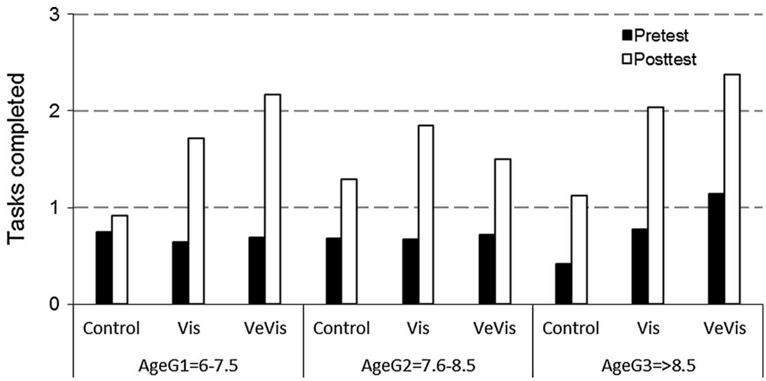

Fig. 5 Changes from pre-test to post-test of Tasks Completed scores for Training Groups (verbal–visual, visual and control) and age

training in one of these two groups would improve more than children in the control group, considered that VeVis training might be more effective than Vis training, and we explored whether training effects might be mediated by age and/or gender. With regard to the intervention effect on VSPS, our results are consistent with recent studies on dynamic assessment (e.g., Resing and Elliott 2011). Both VeVis and Vis training boosted performance in a VSPS-sensitive task from pre- to post-test as compared to the control group. This provides strong evidence that VSPSs can be improved by relative simple forms of training, and supports arguments in favor of more dynamic testing methods to reveal the true potential of cognitive skills, at least in children.

Surprisingly, the outcome pattern suggests that, if anything, Vis training provides more benefits than VeVis training. From a dual-coding perspective (Mayer 2005), this could be taken to suggest that our visual cues were particularly effective, and more effective than the verbal cues—perhaps because spatial information is particularly suited for visual processing, and vice versa (Paivio 1986). This is of particular importance for the training of students with hearing impairments or langue difficulties (Cortiella 2011; Dye et al. 2008; Paul 2007). However, it is also possible that the type of verbal prompts used in the present study (which were inspired by Resing and Elliott 2011) were not appropriate for this task and/or the subjects being tested. For example, in Resing and Elliott's analogy task, self-evaluation of one's response is not as straightforward as in solving puzzles. In puzzle solving, children do not necessarily rely on extra cues to evaluate whether the task is completed, and it may be easier to see how to modify one's actions. Thus, prompts such "Are you sure that pieces should be here, please check, or which changes do you see" might be too simple or inappropriate and might have distracted and interfered with learning more than they helped. Furthermore, is it well known that outcome can be highly dependent on the variability of expertise and experience of the assessors (Haywood and Lidz 2007; Jeltova et al. 2007). Moreover, judging the necessity of the most suitable degree of prompting is difficult and subjective, which might explain part of the problem with verbal cues. Besides, two training sessions may not have been enough to demonstrate reliable intervention effects (Uttal et al. 2012). In future studies, a direct contrast of verbal-only and visual-only cues, together with extended training, might increase our insight into these possibilities.

A limitation of our study method is that it does not allow for more detailed analyses is of the components underlying VSPSs (Mathewson 1999). A potential solution might be computerized task versions, which might allow for the independent assessment of sub-skills, such as imagery or mental rotation. Such an approach would also address another limitation that relates to the experimental material we used. Note that solving a puzzle depends not only on cognitive skills, which were the target of our study, but also on motor skills (e.g., Grissmer et al. 2010), which may or may not be sensitive to training. Another concern is data recording; reaction times were recorded by having the experimenter click on a key as soon as the child started and completed the task. Obviously, the accuracy of this measure depends on the attention the experimenter devoted to the task. Again, computerized versions would help addressing this problem and improve accuracy and reliability (e.g.; Resing et al. 2011). Moreover, they would have the advantage of increasing the game-like character of the test and allow for a swifter and more systematic presentation of feedback.

Another concern is the distinction between performance and learning. Soderstrom and Bjork (in press) have argued that it is only performance that is measurable (in terms of improvement from pre- to post-test) while learning must be inferred from performance. They point out "there are many instances where learning occurs but performance in the short term doesn't improve, and there are instances where performance improves, but little learning seems to happen in the long term." From that perspective, all we can say is that our interventions improved performance, but we cannot or should be sure whether this was due to actual learning. Assessing learning proper would require a follow-up test, which should be included in further studies.

With regard to age as a potentially mediating factor, our results support the findings of Peter et al. (2010) that children younger than 7.5 years old are capable of creating symbolic representations and use spatial relations. The intervention did not affect all outcome measures alike and they were mediated by age and to some degree by gender. Both motivational and cognitive measures suggest that specific training effects are restricted to, or at least drastically stronger in older children, that is, children at an age of 7.6 years or older. The only exception to this trend is the Tasks Completed score, which showed the only specific training effect in the youngest group. The fact that the three dependent variables we were considering were not equally affected by age makes it difficult to rule out possible effects in even younger children. However, a closer look at the outcome pattern suggests that this is not so much due to a lack of training effects in the older groups but a reflection of the fact that in these older groups even the control condition shows a strong improvement. This issue was addressed in the review of Grigorenko and Sternberg (1997), who found that approximately 30 % of the investigated children improved to a statistically significant extent simply because of retesting. In terms of gender differences, our findings do not fit with those of Tzuriel and Egozi (2010), who found a gender difference at baseline while we did not. A specific result was that in the oldest group girls were particularly benefiting from training. This might be because girls are or became more interested in making such puzzles than boys.

Taken altogether, our study provides evidence for the efficiency and benefits of interventions targeting VSPCs. The success of such interventions does not seem to be bounded by age or gender, and it seems that visual cues are particularly effective. At the same time, we consider our findings preliminary and note that more research on the functional implications of different outcome measures, on suitable verbal cues, ideally with computerized versions, is necessary.

References

Allen, G. L. (2003). Functional families of spatial abilities: Poor relations and rich prospects. *International Journal of Testing, 3*(3), 251–262.

Archer, A., & Hughes, C. A. (2011). *Explicit instruction: Efficient and effective teaching*. New York, NY: Guilford Publications.

Assel, M. A., Landry, S. H., Swank, P., Smith, K. E., & Steelman, L. M. (2003). Precursors to mathematical skills: Examining the roles of visual–spatial skills, executive processes, and parenting factors. *Applied Developmental Science, 7*, 27–38.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education, 5*(1), 7–74.

Bohning, G., & Althouse, J. K. (1997). Using tangrams to teach geometry to young children. *Early Childhood Education Journal, 24*(4), 239–242.

Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology, 98*(1), 182–197.

Carr, N. (2008). *The big switch: Rewiring the world, from Edison to Google*. New York: Norton.

Carr, N. (2010). *The shallows: What the internet is doing to our brains*. New York: Norton.

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

Cheng, Y. L., & Mix, K. S. (2012). Spatial training improves children's mathematics ability. *Journal of Cognition and Development*,. doi:10.1080/15248372.2012.725186.

Cortiella, C. (2011). *The state of learning disabilities*. New York: National Center for Learning Disabilities. Retrieved from http://www.ncld.org/images/stories/OnCapitolHill/PolicyRelatedPublications/stateofld/2011_state_of_ld_final.pdf.

D'Oliveira, T. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *The International Journal of Aviation Psychology, 14*, 19–38.

Dehaene, S., Izard, V., Pica, P., & Spelke, E. (2006). Core knowledge of geometry in an Amazonian indigene group. *Science, 311*, 381–384.

Dye, M. W. G., Hauser, P. C., & Bavelier, D. (2008). Visual skills and cross-modal plasticity in deaf readers: Possible implications for acquiring meaning from print. *Annals of the New York Academy of Sciences, 1145*, 71–82.

Eliot, J., & Smith, I. M. (1983). *An international directory of spatial tests*. Windsor Berkshire: NFER-Nelson.

Flanagan, D. P., & Kaufman, A. S. (2004). *Essentials of WISC-IV assessment*. Hoboken, NJ: Wiley.

Ford, B. E. (2003). *Tangrams: The magnificent seven piece puzzle*. Vallejo, CA: Tandora's Box Press.

Foster, T. E. (2007). *The legend of the tangram prince*. Charleston, SC: BookSurge.

Fuchs, D., & Fuchs, L. (2006). Introduction to response to intervention: What, why, and how valid is it? *Reading Research Quarterly, 41*, 93–99.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Gardner, M. F. (1996). *Test of visual perceptual skills (n-m) revised*. Hydesville, CA: Psychological and Educational.

Grigorenko, E. L. (2009). Dynamic assessment and response to intervention: Two sides of one coin. *Journal of Learning Disabilities, 42*, 111–132.

Grigorenko, E. L., & Sternberg, R. J. (1997). Styles of thinking, abilities, and academic performance. *Exceptional Children, 63*(3), 295–312.

Grissmer, D., Grimm, K., Aiyer, S., Murrah, W., & Steele, J. (2010). Fine motor skills and early comprehension of the world: Two new school readiness indicators. *Developmental Psychology, 46*(5), 1008–1017.

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

Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research, 77*(1), 81–112.

Haywood, H., & Lidz, C. (2007). *Dynamic assessment in practice: Clinical and educational applications*. New York: Cambridge University Press.

Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language, 32*, 717–742.

Hegarty, M., Keehner, M., Cohen, C., Montello, D. R., & Lippa, Y. (2007). The role of spatial cognition in medicine: Applications for selecting and training professionals. In G. Allen (Ed.), *Applied spatial cognition* (pp. 285–315). Mahwah, NJ: Lawrence Erlbaum.

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence, 32*, 175–191.

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah & A. Miyake (Eds.), *Handbook of visuospatial thinking* (pp. 121–169). New York: Cambridge University Press.

Holmes, J., Adams, J. W., & Hamilton, C. (2008). The relationship between visuo-spatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology, 20*, 272–289.

Jeltova, I., Birney, D., Fredine, N., Jarvin, L., Sternberg, R. J., & Grigorenko, E. L. (2007). Dynamic assessment as a process-oriented assessment in educational settings. *Advances in Speech-Language Pathology, 9*, 1–13.

Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 38*(1), 23–31.

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35*, 211–223.

Lee, J., Lee, J. O., & Collins, C. (2009). Enhancing children's spatial sense using tangrams. *Childhood Education, 86*(2), 92–94.

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

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

Lovett, A., & Forbus, K. (2010). *Shape is like space: Modeling shape representation as a set of qualitative spatial relations*. Guilford Publications. Proceedings of the AAAI spring symposium on cognitive shape processing. Palo Alto. Retrieved from www.cs.northwestern.edu/~aml758/Papers/ShapeRepresentationComparison.pdf.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49*, 344–351.

Mathewson, J. H. (1999). Visual–spatial thinking: An aspect of science overlooked by educators. *Science Education, 83*, 33–54.

Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 31–48). New York, NY: Cambridge University Press.

Mayer, R. E., & Wittrock, R. C. (2006). Problem solving. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (2nd ed., pp. 287–304). Mahwah, NJ: Erlbaum.

Moreno, R. (2010). Cognitive load theory: More food for thought. *Instructional Science, 38*(2), 135–141.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358–368.

National Council of Teacher's Mathematics. (2003). *Developing geometry understandings and spatial skills through puzzle like problems with tangrams: Tangram challenges*. Retrieved from www.nctm.org.

National Council of Teachers of Mathematics. (2010). *Curriculum focal points for prekindergarten through grade 8: A quest for coherence*. Reston, VA: NCTM. Retrieved from www.nctm.org.

National Research Council. (2006). *Learning to think spatially: GIS as support system in K-12 curriculum.* Washington, DC: National Academies Press.

Newcombe, N. S., & Learmonth, A. E. (2005). Development of spatial competence. In P. Shah & A. Miyake (Eds.), *Handbook of visuospatial thinking* (pp. 213–256). Cambridge: Cambridge University Press.

Olina, Z., Reiser, R., Huang, X., Lim, J., & Park, S. (2006). Problem format and presentation sequence: Effects on learning and mental effort among U.S. high school students. *Applied Cognitive Psychology, 20*, 299–309.

Paas, F. G. W. C., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science, 32*, 1–8.

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford: Oxford University Press.

Passolunghi, M. C., & Mammuarella, I. (2010). Spatial and visual working memory ability in children with difficulties in arithmetic word problem solving. *European Journal of Cognitive Psychology, 22*, 944–963.

Paul, R. (2007). *Language disorders from infancy through adolescence* (3rd ed.). St. Louis: Mosby.

Peter, M., Gluck, J., & Beiglbock, W. (2010). Map understanding as a developmental marker in childhood. *Journal of Individual Differences, 31*, 64–67.

Piaget, J. (1977). In H. E. Gruber & J.J. Vonèche (Eds.), *The essential Piaget*. New York: Basic Books.

Rasmussen, C., & Bisanz, J. (2005). Representation and working memory in early arithmetic. *Journal of Experimental Child Psychology, 91*, 137–157.

Raven, J., Raven, J. C., & Court, J. H. (1998, updated 2003). *Manual for Raven's Progressive Matrices and Vocabulary Scales*. San Antonio, TX: Harcourt Assessment.

Resing, W. C. M., & Elliott, J. G. (2011). Dynamic testing with tangible electronics: Measuring children's change in strategy use with a series completion task. *British Journal of Educational Psychology, 81*, 579–605.

Resing, W. C. M., Steijn, W. M. P., Xenidou-Dervou, I., Stevenson, C. E., & Elliott, J. G. (2011). Dynamic testing with an electronic console: Do children listen to the computer? *Journal of Cognitive Education and Psychology, 10*(2), 178–194.

Schnotz, W., & Kurschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review, 19*(4), 469–508.

Shute, V. J. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153–189.

Slocum, J., Botermans, J., Gebhardt, D., Ma, M., Ma, X., Raizer, H., et al. (2003). *The tangram book*. New York: Sterling Publishing Co.

Soderstrom, N. C., & Bjork, R. A. (in press). Learning versus performance. In D. S. Dunn (Ed.), *Oxford bibliographies online: Psychology*. New York: Oxford University Press. Retrieved from http://bjorklab.psych.ucla.edu/pubs/Soderstrom_Bjork_Learning_versus_Performance.pdf.

Sophian, C. (2000). Perceptions of proportionality in young children: Matching spatial ratios. *Cognition, 75*, 145–170.

Sternberg, R. J., & Grigorenko, E. L. (2002). *Dynamic testing: The nature and measurement of learning potential*. New York: Cambridge University Press.

Sternberg, R. J. (2003). Contemporary theories of intelligence. In W. M. Reynolds & G. E. Miller (Eds.), *Comprehensive handbook of psychology: Educational psychology* (Vol. 7, pp. 23–46). New York: Wiley.

Sutton, K. J., & Williams, A. P. (2007). *Spatial cognition and its implications for design*. Hong Kong: International Association of Societies of Design Research.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*, 996–1013.

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children. *Child Development, 81*, 1417–1430.

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2012). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin,*. doi:10.1037/a0028446.

Van Garderen, D., & Montague, M. (2003). Visual–spatial representation, mathematical problem solving, and students of varying abilities. *Learning Disabilities Research & Practice, 18*, 246–254.

Van Hiele, P. M. (1983). *Structure and insight*. Dordrecht: Kluwer Academic.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*, 817–835.

Webb, R. M., Lubinski, D., & Benbow, C. P. (2007). Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology, 99*, 397–420.

Willcutt, E. G., Doyle, A. E., Nigg, J. T., Faraone, S. V., & Pennington, B. F. (2005). Validity of the executive function theory of attention-deficit/hyperactivity disorder: A meta-analytic review. *Biological Psychiatry, 57*, 1336–1346.