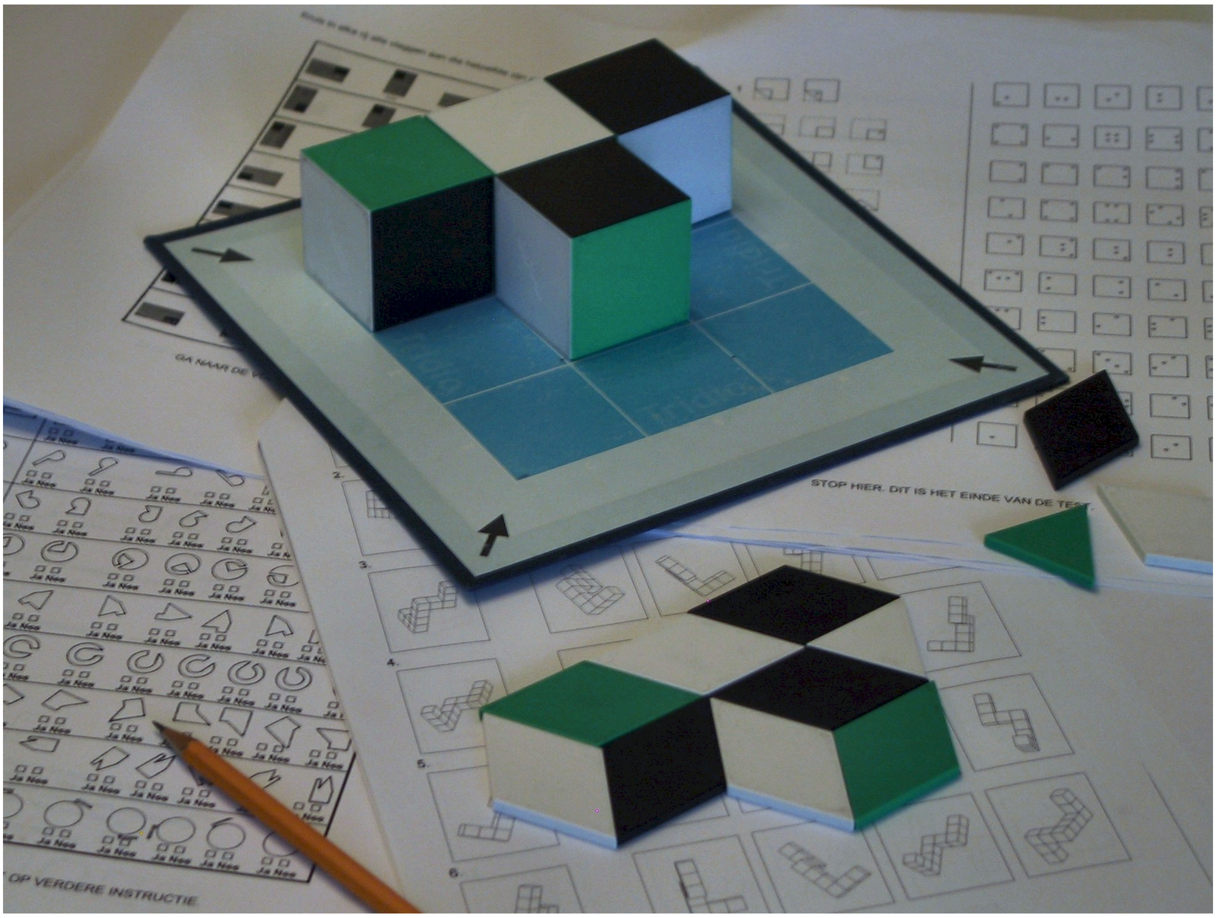
University of Twente
Enschede - The Netherlands

Spatial Ability in Primary School: Effects of the Tridio® Learning Material.

Marjoke Bakker



Master Thesis of Psychology

University of Twente, Enschede

November 2008

Spatial Ability in Primary School: Effects of the Tridio® Learning Material.

Marjoke Bakker

November 2008

Instructional Technology

Faculty of Behavioral Sciences

University of Twente

Enschede

The Netherlands

Committee:

dr. Casper D. Hulshof	Instructional Technology, University of Twente
Jan-Maarten Luursema	Cognitive Psychology and Ergonomics, University of Twente

External advisor:

mw. drs. Fenna T. van Nes	Freudenthal Institute for Science and Mathematics Education, Utrecht University

Table of Contents

Abstract

The Tridio® learning material, consisting of cubes and mosaic pieces and accompanying exercises involving isometric and orthogonal views, has been developed with the aim of enhancing children's spatial ability. This study investigated the effects of Tridio on 5th-graders' (11-year-olds') spatial ability, categorized in J. B. Carroll's (1993) factors spatial relations (SR) and visualization (VZ). A matched-pairs pretest-posttest design (25 pairs) was used. Experimental group children received a training with Tridio, consisting of five individual 30 min sessions. SR was measured using the Card Rotations test and the Flags test; the Paper Folding test and the Mental Rotations test were used as tests of VZ. The appropriateness of these tests for 11-year-olds was first examined in a pilot study. Furthermore, a content-specific test of Tridio performance was administered. Partial correlations between the content-specific and the spatial ability test scores, controlling for school performance, indicated that Tridio adds to the general school curriculum in focusing on spatial ability, but that not all types of Tridio exercises contribute to this. Transfer effects of the Tridio training on spatial ability were, however, not found. With a higher power, effects on SR may be found, but probably not on VZ. The found lack of effect on VZ may be due to the young age of the participants, or to the fact that many children did not get to the more complex Tridio exercises. Content-specific effects were present, indicating that children possibly learned something other than spatial ability.

Samenvatting

Het leermateriaal Tridio, dat bestaat uit kubussen en mozaiekstukjes en bijbehorende opdrachten met isometrische en orthogonale projecties, is ontwikkeld om het ruimtelijk inzicht van kinderen te verhogen. In deze studie is het effect van Tridio op het ruimtelijk inzicht van leerlingen van groep 7 (11-jarigen) van de basisschool onderzocht. Ruimtelijk inzicht werd hierbij onderverdeeld in J. B. Carroll's (1993) factoren *spatial relations* (SR) en *visualization* (VZ). Er is gebruik gemaakt van een matched-pairs pretest-posttest onderzoek (25 paren). De experimentele groep werd getraind met Tridio, gedurende vijf individuele sessies van 30 minuten. SR werd gemeten met de Card Rotations test en de Flags test; de Paper Folding test en de Mental Rotations test werden gebruikt als VZ tests. Deze tests werden eerst in een pilot studie onderzocht op hun geschiktheid voor 11-jarigen. Verder werd er een content-specifieke Tridio prestatietest afgenomen. Partiële correlaties tussen de content-specifieke en de ruimtelijk inzicht test scores, gecorrigeerd voor schoolprestaties, lieten zien dat Tridio iets toevoegt aan het standaard curriculum in het behandelen van ruimtelijk inzicht, maar dat niet alle soorten Tridio opgaven hieraan bijdragen. Transfer effecten van de Tridio training op ruimtelijk inzicht werden echter niet gevonden. Met een hogere power zouden er effecten op SR gevonden kunnen worden, maar waarschijnlijk niet op VZ. Het ontbreken van een effect op VZ kan worden verklaard door de jonge leeftijd van de proefpersonen, of het feit dat veel kinderen niet toekwamen aan de complexere Tridio opgaven. Er waren wel content-specifieke effecten, wat aangeeft dat de kinderen mogelijk iets anders dan ruimtelijk inzicht hebben geleerd.

Introduction

Spatial ability is the ability to construct, retain, retrieve, and manipulate visual images of two- and three-dimensional objects (Lohman, 1993). This ability has been found to be related to mathematics and science achievement (e.g., Guay & McDaniel, 1977; Hegarty & Waller, 2005; Tracy, 1987, 1990), and is even argued to be essential to scientific and mathematical thinking (Clements & Battista, 1992). Furthermore, spatial ability has been linked to success in several occupations, such as piloting, mechanics, engineering drawing and surgery (Hegarty & Waller). Therefore, it seems important to teach spatial skills in schools. The National Council of Teachers of Mathematics (NCTM) acknowledged the importance of spatial ability by including spatial skills in the US curriculum standards for primary and secondary school geometry education (NCTM, n.d.). Also, the TAL-team, which established learning trajectories and achievement targets for Dutch primary school by order of the government, identified several spatial activities as being important in primary school education (Gravemeijer et al., 2007; Van den Heuvel-Panhuizen & Buys, 2005).

In the Netherlands, the learning material Tridio® was developed with the aim of enhancing primary school children's spatial ability (Productief B.V., 2006b, 2006c). The material consists of cubes and mosaic pieces (see Appendix A), which can be used in different types of exercises. Currently, around 40% of primary schools in the Netherlands have purchased the Tridio material (as estimated by M. van Herel at Productief B.V., personal communication, January 10, 2008), in part because of its assumed benefits for advancing children's spatial skills. Interestingly, however, the effect of Tridio on children's spatial ability has never been tested experimentally. Studying this effect was the primary goal of the reported research.

This introduction starts with a description of spatial ability. A subsequent section is dedicated to the trainability of spatial ability. The Tridio learning material and its potential in improving spatial ability are discussed. Furthermore, because Tridio has not been developed for use with a specific age group, different sources are employed in deciding on the age group best to include in this study. A further section deals with the issue of measuring spatial ability. Also, we discuss the frequently found gender differences in spatial ability, which is an additional focus of the current research. Finally, we lay out the plan for the current study and list our research questions.

Categorization of spatial abilities

To be able to measure the effect of Tridio on spatial ability, we needed a clear sense of what spatial ability is. Ability in general can be defined as a person's current capabilities on a certain class of tasks. When discussing ability, we have to distinguish between achievement and aptitude. Achievement deals with the outcomes of prior learning, and can thus be improved by training. Aptitude, on the other hand, is the potential of future learning: If an ability helps in predicting future learning in addition to a prediction from a measure of current achievement, it can be considered as an aptitude (Carroll, 1993). Researchers do not agree on whether a person's aptitude can be improved: Carroll, for example, considers aptitude as a relatively constant attribute, whereas Snow and Swanson (1992) assume that it can be developed through education. Since Tridio is aimed at improving children's spatial capacities, and not necessarily their future learning potential, this study primarily focuses on spatial ability in the sense of achievement, thus on the current level of spatial performance as determined by prior learning. We aim to discover whether this level is increased by working with Tridio. Of course, children's spatial aptitudes prior to working with Tridio play a role in whether and how much their spatial ability can be improved.

When defining spatial ability, most researchers do not consider it as a unitary trait. Rather, they think of it as consisting of different abilities or categories (e.g., Carroll, 1993; Hegarty & Waller, 2005, Linn & Petersen, 1985). As noted by Linn and Petersen, such categorizations can be based on a psychometric or a cognitive perspective. In the psychometric tradition, factor analysis is employed to divide spatial ability into categories on the basis of correlations between scores on different tests. In the cognitive perspective, categories are based on similarities in the processes used to solve several spatial tasks.

Of the many psychometric studies carried out to categorize spatial ability (for an overview, see Hegarty & Waller, 2005), the most comprehensive is probably the one by Carroll (1993). He used over 90 datasets for a factor analysis of visual perception, a broader category of which he argues spatial abilities are a subset. The five factors he found were visualization, spatial relations, closure speed, closure flexibility and perceptual speed. The factor visualization consists of “test variables that appeared to reflect processes of apprehending, encoding, and mentally manipulating spatial forms” (p. 309). Here, the emphasis is on power rather than on speed. Tests loading on this factor have complex items and liberal time limits. Examples of these tests are the Cube Comparisons test, the Paper Folding test and the Surface Development test (Ekstrom, French, Harman, & Dermen, 1976) and the three-dimensional Mental Rotations test (MRT; Vandenberg & Kuse, 1978). The factor spatial relations loads on simple speeded tests, such as Cards, Figures, and Flags (Thurstone & Thurstone, 1941), in which two-dimensional figures have to be mentally rotated. In tests of closure speed, the task mainly is to identify a not previously specified spatial form that is “in some way disguised or obscured by a ‘noisy’ or distracting context” (p. 310). In closure flexibility tests, the subjects are required to search a visual field with a distracting context to find a previously specified spatial form, as is the case in the Hidden Figures test (Ekstrom et al.). Finally, the perceptual speed factor comprises simple searching and comparison tasks: searching for a prespecified spatial form in the absence of a distracting context or deciding if two or more visual presentations are identical. The factors visualization and spatial relations can be considered as the most “spatial” in nature: the tasks require mental manipulation of spatial forms. The other three factors load on tasks with less spatial characteristics, mainly involving visual searching.

Linn and Petersen (1985) used a primarily cognitive perspective in their frequently cited categorization. They identified three categories of spatial ability: spatial perception, mental rotation, and spatial visualization. Spatial perception is defined as the ability to “determine spatial relationships with respect to the orientation of their own body” (p. 1482). An example of a spatial perception test is the Water Level task, in which the subject has to indicate a horizontal water line in a picture of a tilted glass (Piaget & Inhelder, 1956). The mental rotation category includes both two-dimensional and three-dimensional mental rotation tasks, such as Cards, Figures, and Flags, appearing in Carroll's SR factor, and Vandenberg & Kuse's (1978) Mental Rotations test, included in Carroll's VZ factor. Spatial visualization comprises spatial tasks that involve multistep, analytic procedures, and require flexibility in strategy selection. This category consists of tasks from both Carroll's VZ factor (excluding three-dimensional mental rotation) and his CF factor.

Linn and Petersen's categorization is somewhat arbitrary, as was pointed out by the researchers themselves as well as by others. Voyer, Voyer, and Bryden (1995), for example, argue that Linn and Petersen's definition of the spatial visualization category is unclear, making it serve as a sort of rest category. Furthermore, Linn and Petersen's own and other studies (e.g., Vederhus & Krekling, 1996; Voyer et al., 1995) failed to fully support this categorization. Carroll's factors, on the other hand, are more heavily grounded on real data. Therefore, for this study we decided to use Carroll's factors as the basis of measuring spatial abilities. Because we are interested in the effect of Tridio on spatial ability, not on visual perception more generally, we restrict ourselves to the most spatial factors: spatial relations (SR) and visualization (VZ).

Training spatial ability

An important question for a study on the effects of a learning material on spatial ability is, of course, whether spatial ability can be trained. If spatial ability were an innate capacity that is unlikely to change much, an intervention of relatively short duration would not be able to affect it.

In a meta-analysis, Baenninger and Newcombe (1989) found that spatial test scores typically improve by both practice (test-retest) and training. Furthermore, for training to be more effective than mere practice, the training has to be of at least medium duration, consisting of more than one training session during more than three weeks. Test-specific training, focusing on a specific spatial measure, was found to be most effective. However, since such training merely provides practice with one specific spatial task (e.g., mental rotation), effects of test-specific training may not generalize to other spatial tasks. More general spatial ability training was also found to be helpful in improving spatial ability: General training groups improved more than control groups. Later studies have largely supported this finding (e.g., Alias, Black, & Gray, 2002; Clements, Battista, Sarama, & Swaminathan, 1997; Kwon, 2003; Lord, 1985; Sanz de Acedo Lizarraga & García Ganuza, 2003), but the lengthy training by Shavalier (2004) did not improve performance on spatial ability tests.

In the case of test-specific training, one can speak of a content-specific training effect when it improves the intended test performance: The skill that is taught is improved. With more general types of training, however, effects on spatial ability test scores would be a case of transfer: improvement on skills related to, but not exactly similar to, the training content. The Tridio activities can be seen as a more general type of training, because the exercises are of various types and do not focus on specific spatial tests. A question to be answered with the current study, then, is whether training with Tridio produces transfer effects, in addition to content-specific effects on children's performance on Tridio exercises.

Types of training. Researchers have tried various types of training to enhance students' spatial ability. Some used virtual reality environments in which one can for example “walk around” and see virtual buildings form different angles (e.g., Kwon, 2003; Shavalier, 2004). Others employed paper-based exercises, sometimes accompanied with real objects that could be manipulated, for example sketching block buildings and imagining and drawing cuts through solids (e.g., Alias et al., 2002; Brinkmann, 1966; Lord, 1985; Sanz de Acedo Lizarraga & García Ganuza, 2003). Most related to the Tridio activities is probably the intervention employed by Ben-Chaim, Lappan, and Houang (1988). In their study, 5th- to 8th-grade students received a 3-week training consisting of activities like constructing cube buildings corresponding to isometric or orthogonal views, and drawing isometric and orthogonal views of cube buildings. This training resulted in significant improvements on an author-constructed spatial visualization test, but since no control group was included in the study, it is unclear whether these improvements were actually caused by the training. Furthermore, because the test scores were composed of performance on both training-related items (i.e., content-specific or “near” transfer items; Mayer, as cited in Ben-Chaim et al.) and transfer items, one can not be sure whether the score gains indicate a content-specific or a more general training effect. To improve on the design of the Ben-Chaim et al. study, the current study used a control group and separate tests of content-specific performance and general spatial ability.

The Tridio learning material

As mentioned before, the Tridio learning material was developed with the aim of enhancing children's spatial ability. The material consists of cubes with white, black, and green sides, mosaic pieces (rhombuses and triangles) in the same colors and a board to place the cubes on (see Figure 1). Black, white, and green have been chosen because the corresponding tone-values to these colors can be easily distinguished by colorblind individuals (M. van Herel at Productief B.V., personal communication, September 10, 2007). Of the cubes, opposite sides have the same color, while adjacent sides have different colors (see Appendix A for a more detailed description of the material). Accompanying this material, several exercise sets have been developed (Productief B.V., 2003, 2005, 2006a, 2006b, 2006c, 2006d, 2006e). Some exercises require the student to construct a cube building displayed in a picture. In others, an orthogonal top view and two side views of a cube building are provided, and the student is asked to make a cube building that fits all three views. Another core activity with Tridio is to lay out the isometric view of a cube building using the mosaic pieces (Figure 1). This requires the student to see a three- dimensional cube building as a two-dimensional pattern. Additionally, the exercise sets include a lot of variants of these tasks. One of them is completing a two-dimensional picture of a cube building by placing mosaic pieces in the picture. Also, students are asked to construct with mosaic pieces the isometric view one would see when viewing a cube building from another viewpoint. Because we want to examine the general effect of Tridio on children's spatial ability, instead of the effect of only a limited set of exercise types, the current study focused on all types of Tridio exercises.

The exercises can be seen as a kind of puzzle or game: Children usually think they are fun (M. van Herel, personal communication, September 10, 2007). Originally, Tridio had been developed for use in the higher grades of primary school (grades 3-6). Later on, the exercises were modified and extended to allow for a broader target group, including highly gifted children.

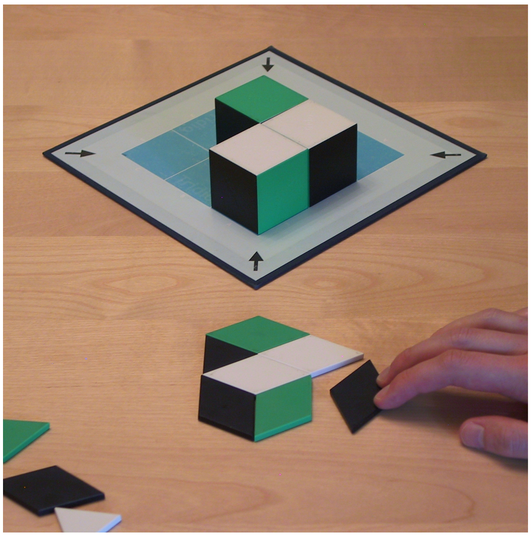


Figure 1. The Tridio® learning material, consisting of cubes, mosaic pieces and a board. Here, an isometric view of the cube building on the board is being constructed using the mosaic pieces.

Tridio and spatial ability. Spatial ability can be expected to play a role in completing Tridio exercises, since these exercises require children to form and manipulate mental images. In many exercise types, the student has to construct three-dimensional cube buildings corresponding to two-dimensional representations of them (either isometric or orthogonal views), or translate two-dimensional views to three-dimensional cube buildings. This requires the student to mentally convert the given two- or three-dimensional object to the three- or two-dimensional structure to be produced. In addition, the task of constructing the isometric view of a cube building when viewed from another viewpoint requires mental rotation of the cube building. Furthermore, for some exercise types multistep mental manipulations are required, for example when a student has to directly create an isometric view of the cube building corresponding to three orthogonal views, without first constructing the cube building. Since the Tridio exercises, thus, provide practice in forming and manipulating mental images, they may be helpful in improving spatial ability.

When looking at Carroll's factors, the Tridio activities most closely match the visualization (VZ) factor. The activities are complex and involve “apprehending, encoding, and mentally manipulating spatial forms" (Carroll, 1993, p. 309). Therefore, one can expect Tridio to most likely affect the corresponding visualization ability.

An interesting question for this study is whether Tridio adds to the standard school curriculum in offering practice in spatial skills. As mentioned earlier, spatial ability plays a role in mathematics and science, which are, of course, also part of the school curriculum. If Tridio exercises are not more related to spatial ability than is the standard school curriculum, including math and science, they may not have an additional value in improving spatial ability. Before investigating the effects of Tridio on spatial ability, then, it is useful to study the relationship between children's Tridio performance and their spatial ability, controlling for school performance.

In which grade to use Tridio?

Tridio is not aimed at a specific age group or school grade (M. van Herel at Productief B.V., personal communication, September 10, 2007). Therefore, it was unclear on what age group the current study had to focus. To get a better idea about the grade for which Tridio (including all types of exercise) is most appropriate, we looked at the current use of Tridio in schools, Tridio-related learning goals appearing in curriculum standards, and Tridio-like activities proposed in literature.

Current use of Tridio. To examine in what grades Tridio is currently being used, telephone interviews were held with 15 (remedial) teachers who use the Tridio material in school. These teachers were contacted using customer data obtained from Productief B.V.

From the interviews, it became clear that Tridio is mainly being used in primary schools, but also some secondary schools make use of it. In secondary schools, Tridio is mostly employed in the earlier grades (grades 7 and 8; ages 13-14 years). It is used as a physical aid in understanding specific mathematical topics, like views, and not directly to enhance students' spatial abilities. In contrast, many primary school teachers mentioned some broader aim of using Tridio, like enhancing general math abilities or improving spatial knowledge.

In primary schools, there is a lot of variation with respect to the grades in which Tridio is treated: some schools use it with kindergartners, whereas others present the material to 6th-graders, or to any grade in between. As can be expected, lower grade teachers generally use the more easy Tridio exercise sets in their lessons, but also when looking at a single exercise set, large variations exist with respect to the grades in which the exercises are used. The most difficult exercise types are primarily used in grades 3-5 (ages 9-11).

Thus, there is no uniform policy in what grade to use Tridio. However, it seems that Tridio as a means to enhance spatial abilities fits better in the primary than in the secondary school curriculum, since in primary school there is more room and interest for "extra" subject matter, not specifically aimed at supporting the teaching of the school subjects. Because in this study we want to include all types of Tridio exercises (i.e., easy and more difficult ones), and since the most difficult exercise types are mainly employed in the higher grades of primary school, it appears that we have to focus on the upper grades of primary school (i.e., grades 3-6).

Curriculum standards. As a second source for deciding on the optimal grade for the Tridio training, we consulted curriculum standards for mathematics education, searching for learning goals related to the Tridio exercises. When looking at the Dutch attainment targets for primary school mathematics education, put forward by the TAL-team (Gravemeijer et al., 2007; Van den Heuvel-Panhuizen & Buys, 2005), Tridio seems most appropriate for grades 5 and 6. For these grades, the attainment targets include being able to relate three-dimensional objects to two-dimensional ones, knowledge about views and projection methods with which three-dimensional objects can be represented in two dimensions, and being able to operate on two- and three-dimensional objects and to predict and analyze the consequences of these operations. In the mathematics standards by the US National Council of Teachers of Mathematics (NCTM), the standards that are most related to Tridio can be found for grades 3-5. In these grades, students should, among others: "create … mental images of objects … ; identify and build a three-dimensional object from two-dimensional representations of that object; [and] identify and draw a two-dimensional representation of a three-dimensional object" (NCTM, n.d.).

Tridio-like activities. As a third way of examining in what grade Tridio can best be treated, we looked at activities similar to the Tridio exercises, that have been proposed by other researchers to be used in schools with the aim of enhancing spatial abilities. Lappan and Winter (1982), for example, designed activities such as drawing orthogonal and isometric views of cube buildings for grades 5-9, which were later used in the previously mentioned experiment by Ben-Chaim et al. (1988). Ben-Chaim et al. suggested 7th grade to be best for teaching spatial visualization tasks, because students in this grade improved most on the spatial test that was administered. Barzel, Haug, Häger, and Rabstein (2007) recommended spatial activities for 5th grade, including constructing cube buildings corresponding to isometric or orthogonal views of them, drawing orthogonal views, and constructing isometric views using rhombuses. Furthermore, Spiegel and Spiegel (2003) suggested several activities for 4th-graders, in which students have to reason about cube buildings displayed isometrically.

Taking into account the arguments from all three sources, i.e., the current use of Tridio in schools (grade 3-6), the curriculum standards (Netherlands: grades 5-6; US: grades 3-5), and the recommendations from literature (grade 4-7), we concluded that, in the Dutch situation, 5th grade would be optimal for a Tridio training consisting of all exercise types. Thus, we chose to work with 5th-grade, i.e., 11-year-old, students in our experiment.

Measuring spatial abilities

To examine the effect of Tridio on 5th-graders spatial ability, we needed an appropriate measure of spatial ability. As mentioned before, the Tridio activities can be expected to have greatest effects on the visualization (VZ) ability. The other "spatial" factor from Carroll's factor analysis, spatial relations (SR), focuses more on speed and less on power than does the VZ factor. It is interesting to study effects of Tridio on this factor, too, as a measure of "farther" transfer. Thus, we decided to include in the experiment tests of both VZ and SR.

Testing children. A problem with testing children is that most spatial ability tests have been developed for use with adults (Johnson & Meade, 1987; Kerns and Berenbaum,1991; and Casey et al., 2008). This is also the case for the tests proposed by Carroll (1993) to measure SR and VZ. Several researchers have tried to solve this problem by adjusting adult tests for use with children. In a comprehensive study by Johnson and Meade, several VZ and SR tests, including variants of the MRT, the Flags test (Thurstone & Thurstone, 1941) and Ekstrom et al.'s Cube Comparisons test, were adapted for use in various grades. For 4th- to 6th-graders, test instructions were read aloud and model items - physical objects representing example items - were used in demonstrating the tasks to be performed. All tests administered in the study were found to be appropriate for this age group, except the Cube Comparisons variant. Model items were also used by Shavalier (2004) for facilitating 4th- to 6th-graders understanding of some tests including the Paper Folding test and the MRT, and by Vederhus and Krekling (1996) for illustrating to 9- and 10-year-olds adapted versions of the PMA Spatial Relations test (two-dimensional mental rotation) and the MRT, among others. In adapting some of their tests, Vederhus and Krekling also reduced the number of response alternatives.

Modified versions of adult spatial tests can, thus, be used with children. However, some modifications may alter the nature of the task. For example, Kerns and Berenbaum (1991) adapted the MRT by using real, three-dimensional wooden cube objects as test items. In this way, the task probably requires less mental manipulation, since participants do not have to mentally translate a two-dimensional picture to a three-dimensional object. Furthermore, a test's nature may change when the number of response alternatives is reduced, as in this case the role of guessing increases.

To be able to accurately measure Carroll's factors, it is best to use test variants that resemble as closely as possible the original tests on which the factor structure was based. Thus, adult SR and VZ tests may be adjusted to children by clarifying the test instructions – for example by reading aloud instructions and showing model items – but preferably not by removing response alternatives or changing the task's nature. Little research has been done on the appropriateness of adult spatial tests adapted for children by instructional changes alone (Shavalier [2004] did administer tests in this way, but did not report on the success of this method). Therefore, a pilot study was conducted to investigate whether several adult SR and VZ tests, with the use of read-aloud instructions and model items, are appropriate for 5th grade students. This pilot study was also used to examine whether the original tests' time limits are suitable for this age group.

Speed and power. As all SR and VZ tests have time limits, scores on tests of both factors are partly determined by speed of performance, although this is less the case for VZ than for SR, because of the more liberal time limits of the former. Because in the Tridio exercises the emphasis is on power rather than on speed, it can be expected that a training with Tridio primarily affects children's power (i.e., accuracy) in solving spatial tasks. The speed component of the spatial tests, then, may be a complicating factor in measuring the effects of Tridio on spatial ability, since this may obscure effects on power. To be able to separate the effects of Tridio on power, we wanted to compute separate speed and power scores. This can be done by administering the tests using the red pencil method proposed by Johnson and Meade (1985, 1987). According to this method, participants start working with a regular lead pencil and are asked to switch to a red pencil when reaching the test's time limit. With the red pencil, they are allowed to complete the test. The speed score can then be computed by the proportion correct of the total of items on the test, counting only the items marked with the regular pencil (Johnson & Meade, 1985). Alternatively, the test's original scoring method can be used on the regular-pencil items. The power score is determined as a proportion-correct score over all attempted items, both the ones marked with regular pencil and with red pencil (Johnson & Meade, 1985).

Gender differences in spatial ability

The topic of gender differences in spatial ability is highly recurring in literature. Males typically outperform females on several spatial tasks (e.g., Halpern & Collaer, 2005). In accordance with this research tradition, an additional aim of the current study was to examine boy-girl differences in spatial ability.

Although male advantages have been found on various spatial tests, between different spatial tasks large variation exists in the size of these gender differences (Linn & Petersen, 1985; Voyer et al., 1995). For some VZ tests, like the Paper Folding test, usually no significant gender differences are found (Halpern & Collaer, 2005). On SR tests, moderate effect sizes of gender are present, of approximately .4 standard deviations (Voyer et al.). The largest and most robust gender differences exist on Vanderberg & Kuse's MRT. On this test, effect sizes between .7 and 1.0 standard deviations have been found, depending on the scoring method used (Voyer et al.).

Researchers have come up with various possible causes for the gender differences in spatial ability (e.g., Halpern & Collaer, 2005). From an evolutionary perspective, for example, the differences between males and females can be explained by the different roles they occupied prehistorically. Other biological explanations relate gender differences in spatial ability to levels of sex hormones or differences in brain lateralization (Halpern & Collaer). Other research on the causes gender differences has focused on an experiential explanation, hypothesizing that males outperform females on spatial tests because they have had more experience with spatial activities during their life. For example, boys usually have more “spatial” toys like blocks and Lego, more often participate in mathematical or scientific activities, including geometry, and more often engage in contact sports or scouting activities, which may cause them to develop higher spatial abilities (Baenninger & Newcombe, 1989). In their meta-analysis, Baenninger and Newcombe found a positive relationship between spatial activity participation and spatial test performance. However, as this analysis involved correlational studies, no causal inferences can be made.

Gender differences in children. Although gender differences in spatial ability have frequently been found for adults, they are not always observed in children. Johnson and Meade (1987) conducted a large study on the onset of gender differences on a composite measure of spatial skill. On this measure, they found gender differences to be present from the age of 10.

Others noticed that the age at which gender differences appear varies with the tests administered. In a meta-analysis, Voyer et al. (1995) found gender differences to be significant from 10 years on the generic mental rotation task (Shepard & Metzler, 1971) and the PMA Spatial Relations test (two-dimensional mental rotation) and from age 13 on the DAT Spatial Relations (deciding what a shape would look like when folded, a VZ task). On the Vandenberg and Kuse MRT, gender differences were found in 5th- and 6th-graders (11- and 12-year-olds) by Geiser, Lehmann, Corth, and Eid (2007). In Richmond's (1980) study with 10-year-olds, boys outperformed girls on Thurstone's (1941) Cards, Figures, and Flags (SR tests).

The absence of gender differences at younger ages may be related to the hormonal explanation of these differences, because hormonal differences between genders greatly increase at puberty. Another explanation for the lack of gender differences in younger children may be the inappropriateness of several of the tests for children (e.g., Voyer et al., 1995). Some researchers have, therefore, used adapted adult spatial tests in examining gender differences in children, like Johnson and Meade (1987) did in their previously cited study. Kerns and Berenbaum (1991) and Vederhus and Krekling (1996) used adapted versions of the MRT and of two-dimensional mental rotation tests with 9- and 10-year-olds and 9- to 13-year-olds respectively, and found significant gender differences in favor of boys on these tests. Shavalier (2004), in her study with 4th- to 6th-grade students, found gender differences on the Eliot-Price Test (a relatively unknown perspective taking test, probably measuring the VZ factor), but not on the MRT and the Paper Folding test.

Concluding, in our study with 5th-grade (11-year-old) students, we can expect to find boys outscoring girls on SR tests (two-dimensional mental rotation). For VZ tests, the picture is less clear, depending on the actual tests used. The MRT may display gender differences, as well as the Eliot-Price Test, but on other VZ tests, such as the DAT Spatial Relations and the Paper Folding test, probably no gender differences will be found.

Gender differences in training effects. From the experiential explanation of the gender differences in spatial ability, one can expect spatial training to have a greater effect on females' spatial test performance than on males'. According to this theory, males already have received a lot of training by experience and thus already perform close to their maximum potential, whereas females have more room to improve (Baenninger & Newcombe, 1989). To a less extent, this may also apply to 11-year-olds, since at this age boys and girls may already differ in their toy-playing, sports, and scouting experiences. A meta-analysis by Baenninger and Newcombe, however, did not reveal significant gender differences in training effects. More recent studies also failed to find differences in training effects between males and females (Ben-Chaim et al., 1988; Casey et al., 2008; Clements et al., 1997; Sanz de Acedo Lizarraga & García Ganuza, 2003; Shavalier, 2004), thus not supporting the experiential hypothesis.

The current study

First, a pilot study was conducted to examine the appropriateness of some SR and VZ tests for 5th-grade students, and to determine suitable time limits for these tests.

Of the main study, the primary goal was to investigate the effects of the Tridio learning material on children's spatial ability. This was done by a matched pairs control group design with pre- and posttests of the SR and VZ factors of spatial ability. Matched pairs were used to increase the power of the study to detect effects on spatial ability (Cohen, 1988). Children's spatial ability pretest scores were used as the primary matching criterion, since these were expected to be related to their gain scores: Children with higher spatial ability may have less room for improvement than children having lower levels of spatial ability, whereas another hypothesis is that the spatial ability pretests may, additionally to being a measure of spatial achievement, be seen as measures of spatial aptitude, thus being predictive of future learning. From each obtained pair, one child was randomly assigned to the experimental condition and the other to the control condition. The experimental group was trained with Tridio during five sessions that spanned approximately three weeks, as recommended by Baenninger and Newcombe (1989) from the results of their meta-analysis of effects of spatial training. The control group received no intervention. The effects of the Tridio training could then be studied by comparing the gains from pretest to posttest of the experimental group children with those of their control group pairmates. In addition, the spatial ability pretest scores were used to examine gender differences in spatial ability.

To be able to measure content-specific effects of the Tridio training, in addition to its transfer effects on the spatial ability tests, a Tridio test was developed to measure children's performance on different types of Tridio exercises. This test was administered both before and after the intervention. Scores on the Tridio pretest were also used to study the relationship of children's Tridio performance with their spatial ability, as measured by the spatial ability pretests. In this way, insight could be developed into the role spatial ability plays in the Tridio exercises, which may give a clearer picture of the potential of Tridio in enhancing spatial ability. To be able to examine how much focus on spatial ability Tridio adds to the standard curriculum, we controlled for math and general school performance in investigating these relations.

The main study was aimed at answering the following research questions, for 5th-grade (11-year-old) children:

a) Is Tridio performance related to spatial ability, when controlling for school performance?
b) Does spatial ability increase by working with Tridio (i.e., is there a transfer effect)?
c) Does Tridio performance improve by working with Tridio (i.e., is there a content-specific learning effect)?
d) Is there a gender difference in spatial ability?
e) Is there a gender difference in the effect of the Tridio training on spatial ability?

Concerning research question a), a positive relationship between Tridio performance and spatial ability, controlling for school performance, was expected, especially for the factor VZ, since the Tridio activities explicitly focus on forming and manipulating complex mental images, which is presumably not the case for other school activities. Such a positive relationship would indicate that Tridio adds to the standard school curriculum in providing practice in spatial skills. Regarding question b), we assumed the Tridio training to have the potential of increasing children's spatial ability, especially the VZ factor, since the training offers practice in understanding and mentally manipulating spatial forms, and since spatial ability can be improved by general training (Baenninger & Newcombe, 1989). However, because of the shortcomings of the Ben-Chaim et al. (1988) study on the effects of a Tridio-related training on spatial ability, we did not have clear expectations on this transfer effect of Tridio. A content-specific effect of the Tridio training (question c) was expected to be found, equivalent to the high effects of test-specific training found by Baenninger and Newcombe. Next, gender differences on pretest scores (question d) were expected to be found on SR tests and three-dimensional mental rotation tests, but not on other VZ tests (except possibly the Eliot-Price Test). Finally, regarding question e), a gender difference on the transfer effect of Tridio would support the experiential explanation of gender differences (e.g., Baenninger &

Newcombe, 1989; Kass et al., 1998). However, in accordance with the findings of Baenninger and Newcombe and others (Ben-Chaim et al., 1988; Casey et al., 2008; Clements et al., 1997; Sanz de Acedo Lizarraga & García Ganuza, 2003; Shavalier, 2004), this gender difference was not expected to be found.

In addition to these research questions, children's attitudes to the Tridio training were examined, by use of a questionnaire.

PILOT STUDY

The objective of the pilot study was to examine what spatial ability tests would be appropriate for 5th-grade (11-year-old) students and what time limits on these tests would be most suitable for these children. The results were used to decide what tests to administer in the main study.

Method

Participants

The pilot test was administered in a “groep 7”(grade 5) class at a Roman Catholic primary school in Enschede, a city in the east of the Netherlands. The participants were 22 children, mean age 11.1 (range 10.4 to 12.1), of which 7 were boys and 15 were girls.

Materials

The pilot test session consisted of six spatial ability tests (a seventh test was planned to be included, but due to time limits this test was not administered). As tests of Carroll's (1993) SR factor, the Card Rotations test (Ekstrom et al., 1976) and the Flags test (Thurstone & Thurstone, 1941) were administered. To measure VZ, we selected the Paper Folding test (Ekstrom et al., 1976), the Cubes test (Johnson & Meade, 1987), and two types of Mental Rotations tests both based on Shepard and Metzler's (1971) three-dimensional objects: the Vandenberg and Kuse (1978) version and the mental rotations test employed by Johnson and Meade. These are all timed paper-and-pencil tests. The test instructions were translated into Dutch and to make the instructions comprehensible for 11-year-olds, some complex sentences were simplified and difficult words were avoided.

Card Rotations test. In the Card Rotations test (CRT), the items are grouped in rows of eight (see Figure 2). Each item shows a two-dimensional object, called a card. This card is either a rotated version of the target card displayed in front of the row or its (rotated) mirror image. The task is to decide whether the card is the same as the target card or not. In the original test, participants have to mark the “S” if the card is the *same* as the target card or the “D” if the card is *different* (i.e., its mirror image). In the Dutch translation we decided to use a box for each item, which has to be marked only if the card is the same as the target card, as shown in Figure 2.

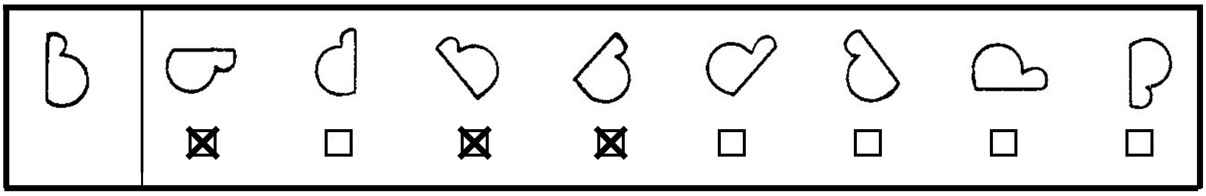


Figure 2. Sample row of items from the Dutch translation of Ekstrom et al.'s (1976) Card Rotations test. The cards that can be rotated to look the same as the leftmost card have been marked.

Flags test. In the Flags test, the task is to decide whether a pictured American flag is the same as a target flag. Each of the items shows a row of five flags, as can be seen in Figure 3. The leftmost flag is the target flag; the others are either rotations or (rotated) mirror images of the target. The participant is required to mark every flag that can be rotated to look the same as the target flag.

Figure 3. Sample item from the Flags test (Thurstone & Thurstone, 1941). The correct answers, the flags that can be rotated to look the same as the leftmost flag, have been marked.

Paper Folding test. The task of the Paper Folding test (PFT) is to predict how a piece of paper will look after it has been folded, a hole has been punched in it, and it has been unfolded again. For each item, a number of pictures on the left show how a square piece of paper is folded and where the hole is punched in (see Figure 4). On the right, five pictures are given, and the participant has to choose the picture that correctly shows what the piece of paper will look like when it is unfolded. Because of the poor quality of the pictures in the original test, we redrew the test before administering it (Figure 4 displays a redrawn item). To make sure all children understood what they had to do, two practice items were added, which were not part of the original test (see Appendix B).

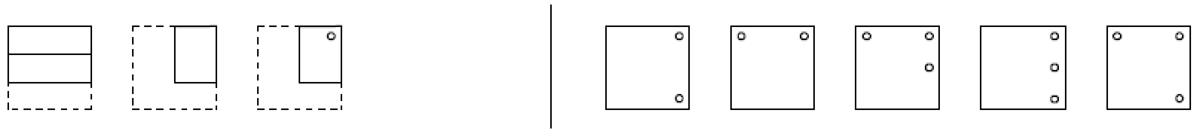


Figure 4. Sample item from the redrawn version of Ekstrom et al.'s (1976) Paper Folding test. The second picture correctly shows how the paper on the right would look after it is unfolded.

Cubes test. In the Cubes test, a test adapted for children by Johnson and Meade (1987), each item contains two pictures of a cube. The task is to decide whether these can both be pictures of the same cube or not, by marking an “S” (*same*) or a “D” (*different*) respectively. When translating this test into Dutch, these letters were replaced by the full Dutch words for *same* (*hetzelfde*) and *different* (*anders*), since using only the first letter of these words would probably be confusing for the current age group (see Figure 5).

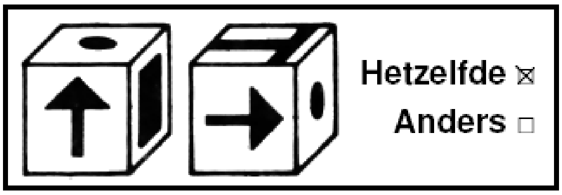


Figure 5. Sample item from the Dutch translation of the Cubes test used by Johnson and Meade (1987). “Hetzelfde” (*same*) has been marked instead of “Anders” (*different*), because the two pictures can be of the same cube.

Mental Rotations tests. The Mental Rotations test (MRT), based on the cube objects by Shepard and Metzler (1971), is probably the most studied test of spatial ability. Different versions of this test have been developed, the most well-known of which is the one by Vandenberg and Kuse (1978), which has been redrawn by Peters et al. (1995). Each item on this test consists of a target object on the left and four objects on the right (see Figure 6). Two of the objects on the right are rotated versions of the target object, whereas the other two are different objects. The participant's task is to mark those two objects that are the same as the target object.

Several researchers have argued that this version of the MRT might be too difficult for children and have adapted the test for use with children (e.g., Johnson & Meade, 1987; Kerns & Berenbaum, 1991; Vederhus & Krekling, 1996). To be able to compare the use of different versions, both Vandenberg and Kuse's MRT (the MRT-A, Peters et al.; further referred to as MRT-VK) and the adapted version used by Johnson and Meade (1987; further called MRT-JM) were included in the pilot study. In the MRT-JM, each item displays two objects, and the child has to indicate whether these objects are the same or different, by marking an “S” or a

“D” respectively. These letters were translated into Dutch in the same way as was done for the Cubes test (see Figure 7).

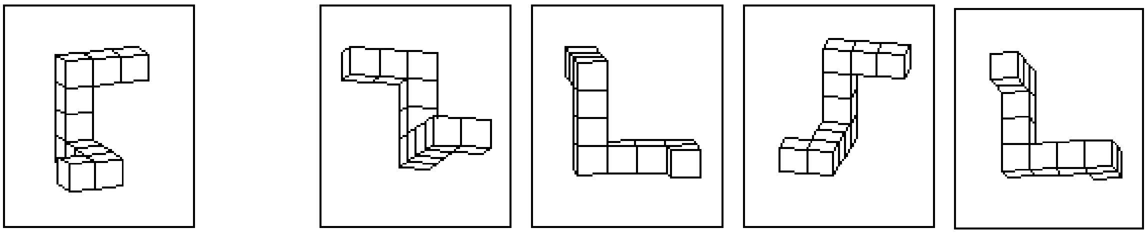


Figure 6. Sample item from the redrawn Vandenberg and Kuse Mental Rotations test (Peters et al., 1995). The second and third picture in the row of four are the correct answers, because they show a rotation of the leftmost object.

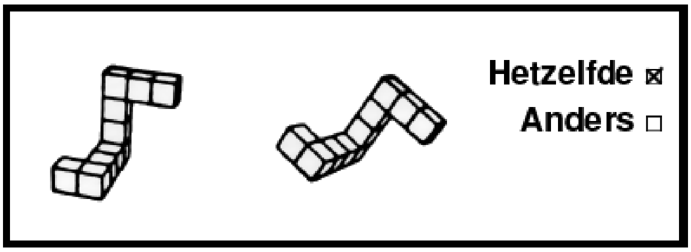


Figure 7. Sample item from the Dutch translation of the Mental Rotations test used by Johnson and Meade (1987). “Hetzelfde” (*same*) has been marked instead of “Anders” (*different*), because the two pictures display the same object.

Test Booklet. The order of the six tests was randomly determined to be as follows: 1) Flags test; 2) MRT-JM; 3) MRT-VK; 4) CRT; 5) PFT; 6) Cubes test. The tests were put together in a test booklet, starting with a page containing general instructions about the test session (additionally, the booklet contained the seventh, not administered test and a questionnaire, which was also skipped due to time limits). Because of the limited time available, of most tests only part of the items was administered. The CRT, PFT and MRT-VK originally consist of two parts. Only the first parts of these tests were used. For the Cubes test and the MRT-JM, no such natural subdivision exists. For the Cubes test, then, the first half of the items was included; of the MRT-JM, the first two of three pages of items were selected. Of the Flags test, all items were used. The characteristics of the (partial) tests included in the pilot study are displayed in Table 1.

Table 1. *Pilot Test Characteristics.*

Test	Number of items	Score formula[a]	Score range	Time limit[b] (min:sec)
Flags test	32	number correctly marked – number incorrectly marked[c]	0-60	5:00
JM Mental Rotations test	28	number correct / total number of items	0-1	2:40
VK Mental Rotations test	12	number correct	0-12	3:00
Card Rotations test	80	number correct – number incorrect	0-80	3:00
Paper Folding test	10	number correct – ¼ * number incorrect	0-10	3:00
Cubes Test	16	number correct / total number of items	0-1	1:30

Note. JM = Johnson & Meade; VK = Vandenberg & Kuse.
[a]Drawn from the tests' manuals. [b]Derived from the tests' manuals. For tests of which only part of the items was administered, the time limit was altered accordingly. [c]Unmarked items are not considered.

Procedure

The children were tested in their regular class setting, in one session in the morning. This session lasted approximately 1.5 hours. The class teacher was present while the experimenter (the author) administered the tests.

Red pencil method. To try the red pencil method proposed by Johnson and Meade (1987, 2008) for use in the main study, this method was also employed in the pilot study. In this case colored pencils (not necessarily red) were used. According to the method, on each test the children were asked to start working with a regular pencil, and to switch to a colored pencil when instructed to (i.e., after a certain time limit). While working with the regular pencil, the students were allowed to correct answers using an eraser; when working with the colored pencil, correcting was no longer allowed. To avoid confusion, the participants were asked to put on the floor the writing material that was not in use (Johnson & Meade, 2008).

Time limits. An additional benefit of the red pencil method for the pilot study was that two different time limits could be employed on each test, making it possible to compare them. Most tests had originally been constructed for adults, and, thus, their time limits could be too strict for the current age group. For all tests, then, in addition to the original time limit (Table 1) employed for the regular-pencil period, a second, less strict time limit was used for the total (regular and colored pencil) test administration. However, for most tests, many children completed all items before the second time limit was reached. In this way, the second time limit disadvantaged fast children, since they could have obtained a higher score if more items were present. Since this is much less the case for the first time limit, the two time limits could, unfortunately, not be compared. However, the appropriateness of the first (original) time limits could be assessed.

Test instructions and administration. The children were asked to place their desks apart from each other. The experimenter told them that they would be doing a test on spatial ability, which would have no influence on their school grades. It was explained that spatial ability is the ability to imagine how things look, for example when objects are rotated in the mind. All children were given a regular lead pencil with an eraser on the back, and a colored pencil. Then, the experimenter read aloud the general test instructions. These instructions explained that the children always had to look carefully at the bottom of each page whether they were allowed to turn the page or had to wait until the experimenter told them to. Also, the children were instructed not to rotate the test booklet. The use of the two pencils was explained. The students were told to switch to the colored pencil when instructed to; the use of time limits was not mentioned. Finally, the children were asked to work as quickly as possible, but to keep working accurately. They were told guessing would be of no use, because wrong answers would be subtracted from their score.

Before each test, the test instructions were read aloud. To make sure the children understood the test, model test items were shown during the instructions. The models had been constructed and were employed using Johnson and Meade's (1985, 2008) guidelines. For the Flags test, two cardboard flags were used: an ordinary, black-and-white American flag and its mirror image. By demonstrating that rotating one of the flags could not make it look like the other, it was shown that these two were different. The MRT-JM was illustrated by showing three-dimensional versions of the example objects, constructed by gluing together wooden cubes. It was demonstrated how such an object could be rotated such that it looked similar to different pictures displaying the object from different angles. Also, two three-dimensional objects that were each other's mirror images were shown. They were rotated to demonstrate that one could not be made identical to the other. For the MRT-VK, no further models were shown, as the children by then were already familiar with the objects occurring in the test. For the CRT, a cardboard model of the target card of the example item row was shown (Figure 2). For each item in this row, it was demonstrated that it was, or was not, the same as the target card. To demonstrate the folding and hole-punching process in the PFT, a square piece of paper was used. This was folded and a hole was cut in it, in the same way as was displayed in the example item. Subsequently, it was shown how this piece of paper looked when unfolded again. The example items of the Cubes test, finally, were illustrated using cardboard versions of the cubes appearing in the example items. These cubes were rotated to demonstrate that they could or could not be the same.

After the instructions to each test had been read and the model items had been shown, the participants completed the test's practice items, and their answers were checked. Students' questions were answered. The solution to the second practice item of the PFT was demonstrated with a piece of paper, because many children found this item difficult. The students were told that, when they would have finished the test, they had to wait quietly until everyone had finished. They were allowed to read a book in the meantime.

When there were no further questions, the children were instructed to start with the test items. A stopwatch was used for timing. When the test's time limit was reached, the participants were asked to put their regular pencil on the floor and to continue working with the colored pencil.

Because of the limited time available, for the Cubes test only a regular-pencil period was employed.

Results and Discussion

As mentioned before, the second time limits could not be compared to the first ones. Therefore, we only report the scores corresponding to the first (original) time limit of each test, i.e., including only the items marked with the regular pencil. The means and standard deviations of these scores are displayed in Table 2. To study the appropriateness of the tests and their time limits, boxplots were examined for floor and ceiling effects. A floor effect indicates that a test is too difficult or its time limit too strict, whereas in the unlikely case of a ceiling effect, the test can be considered as too easy or the time limit as too loose.

The Flags test and the CRT were easily understood by most children. No floor or ceiling effects were found on these tests, which indicates that the tests with their original time limits are appropriate for the current age group. On the PFT, again no floor or ceiling effects were found, showing that also this test, with its original time limit, is suitable for 5th-graders. Also for the Cubes test, a boxplot revealed no floor or ceiling effect. This test, however, seemed to be rather difficult for the current age group: Many children appeared to have problems in understanding the instructions. When looking at the ratio of correctly answered items to attempted items, the mean of this ratio was only .65, which does not exceed very much the chance proportion of .5. The apparent difficulty of this test agrees with Johnson and Meade's (1987) advise not to use the Cubes test below grade 6.

Table 2. *Descriptive Statistics of the First Score of the Pilot Tests (n = 22).*

Test	Score range[a]	*M*	*SD*
Flags test[b]	0-60	14.9	7.5
JM Mental Rotations test[c]	0-1	.32	.09
VK Mental Rotations test[c]	0-12	4.2	1.7
Card Rotations test[c]	0-80	53.0	11.7
Paper Folding test	0-10	3.2	2.2
Cubes test	0-1	.51	.18

Note. JM = Johnson & Meade; VK = Vandenberg & Kuse.
[a]Range of possible scores. [b]Two children seemed to have misunderstood the test. After eliminating these cases, the following values were obtained: first score: $M = 16.2$, $SD = 6.5$; second score: $M = 27.3$, $SD = 10.9$. [c]Before computing these values, one outlier had been eliminated.

On both the MRT-JM and the MRT-VK, we found no floor or ceiling effects. Thus, both versions of the MRT may be appropriate for this age group. The correlation between the scores on the two MRT versions was remarkably low and non-significant ($r = .20$, $p > .05$). Apparently, the two tests do not measure quite the same thing, and the MRT-JM, thus, is not a good substitute for the more generally used MRT-VK. While both tests were found appropriate for the current age group, the MRT-VK has some advantages over the MRT-JM. Firstly, the MRT-VK is more commonly used, which makes results more comparable to those of other studies. A second advantage of the MRT-VK is its higher reliability: In the MRT-JM, there are only two options to choose from for each item, whereas in the MRT-VK the participant has to select two out of four pictures, which makes performance less dependent on chance. As in this study the MRT-VK had been administered just after the MRT-JM, involving the same cube objects, students' scores on the MRT-VK were probably somewhat higher than would be the case if only the MRT-VK were administered. When administering the MRT-VK alone, a higher time limit might prevent possible floor effects. According to the test manual (Peters, 1995), a time limit of 4 min for each part may be used instead of the standard 3 min. Although Geiser et al. (2007) successfully used the MRT-VK with the 3 min time limit with 5th- and 6th-grade children, the less strict time limit of 4 min might be more appropriate for children of this age, since the test had originally been constructed for use with adults.

Conclusions

Using read-aloud instructions and model test items in illustrating these instructions, all tests in the current study were found to be appropriate for 5th-graders, except the Cubes test, which was too difficult.

Both the MRT-JM and the MRT-VK can be used with this age group. The MRT-VK is preferred over the MRT-JM, since the former is more generally used and more reliable.

For the MRT-VK, a time limit of 4 min for each part seems appropriate for children of this age. On the PFT, the CRT, and the Flags tests, the original time limits can be used with 5th-graders.

Thus, in the main study the factor SR can be measured with the CRT and the Flags test, both using the original time limits. As tests of the factor VZ, we can administer the MRT-VK, with a time limit of 4 min, and the PFT, with its original time limit.

MAIN STUDY

Method

Participants

Power analysis

To determine the number of participants required for the study to have sufficient power to detect possible transfer effects of the Tridio training on spatial ability, a power analysis was performed using the software program G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007). As an estimate of the effect size, the medium value of .5 (Cohen, 1988) was chosen. For some earlier studies on transfer effects of training on spatial ability, effect sizes were reported or could be computed following Cohen. These effect sizes are all very high: Brinkmann (1966) found an effect size of .85; for Kwon's (2003) study, effect sizes of .81 and .97 were computed; and Sanz de Acedo Lizarraga and García Ganuza (2003) reported an effect size of .92. This makes an estimate of .5 for the effect size reasonably conservative.

When entering an α level of .05, a power of .80 (the convention proposed by Cohen, 1988), and an effect size of .5, a required sample size of 27 pairs was obtained for a one-tailed test.

Participants in the study

The experiment was conducted at a Roman Catholic primary school in Oldenzaal, a small town in the east of the Netherlands. All three classes of “groep 7” (grade 5) participated in the study. In total this amounted to 62 children. After a permission letter had been returned by the parents, two children were excluded from the study due to parent refusal. Another child was excluded, because she was absent during the spatial ability pretest week. This resulted in a total of 59 participants, mean age 11.1 (range 10.4 to 12.3), of which 27 were boys and 32 were girls.

The matching procedure, which will be described in the *Procedure* subsection, created 25 matched pairs. This is a little less than the 27 pairs suggested by the power analysis, but with an effect size of .51 instead of .5, still rather conservative, this sample size would be large enough to detect training effects on spatial ability. After randomly assigning one child of each pair to the experimental group and the other to the control group, the experimental group contained 13 boys and 12 girls, mean age 11.1, and the control group comprised 11 boys and 14 girls, mean age 11.1. The remaining, non-matched children received the same treatment as the matched control group. For some of the analyses, the total group of 59 participants was used. This will be indicated where relevant.

Materials

Spatial ability tests

To measure children's spatial ability both before and after the intervention, four spatial ability tests were administered: two measuring SR and two measuring VZ. The tests were selected based on the conclusions of the pilot study: For measuring SR, the Card Rotations test (CRT) and the Flags test were employed, and as tests of VZ we used the Paper Folding test (PFT) and the Vanderberg and Kuse Mental Rotations test (MRT-VK).

Dutch translations. We used the same Dutch translations of the tests as in the pilot study, with a few adaptations. For the CRT, the translation of the “S” and “D” boxes to one box that had to be checked when the item was the same as the first card of the row, proved to be inappropriate, because in this way no distinction could be made between items that were decided to be different from the target card and items that were not attempted. This caused ambiguities in scoring the test, since the number of wrong answers had to be subtracted from the number of correct answers. For the main study, then, two boxes for each item were again used. These boxes were labeled with the Dutch words for *yes* (*ja*) and *no* (*nee*), because using the first letters of the Dutch words for *same* and *different* would probably be confusing for 11-year-olds, and displaying the full (rather long) words would take too much space. In the adapted version, then, the participant had to mark “Ja” when the card was the same as the target one, and “Nee” when it was not the same (see Figure 8). For the Flags test, the translation of the instructions was slightly adapted to prevent a misunderstanding that arose in the pilot study.

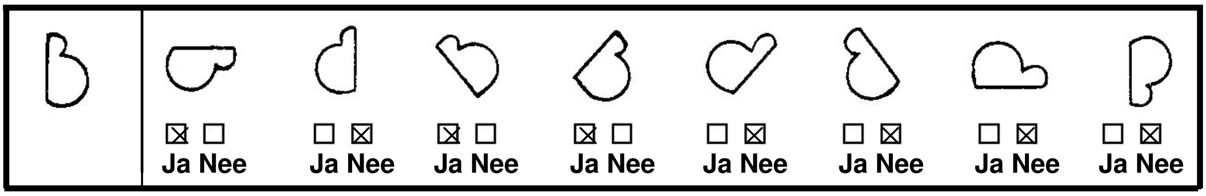


Figure 8. Sample row of items from the adapted Dutch translation of Ekstrom et al.'s (1976) Card Rotations test. The correct answers have been marked. “Ja” (*yes*) has been marked if the card can be rotated to look the same as the leftmost card; “No” (*nee*) has been chosen if this is not the case.

Splitting the tests. The original tests could easily be split into two halves, to be used as a pretest and a posttest. The original versions of the MRT-VK, CRT and PFT already consisted of two parts, both having the same number of items and the same time limit. The Flags test could be divided into two parts by taking the first 16 items as the first half and the last 16 items as the second half. For each of the four tests, the two halves seemed reasonably similar in difficulty. However, since no literature was found on the similarity of the two parts of any of the tests, exact similarity could not be assumed. The use of a control group, however, overcomes this problem.

The first half of each test was used as a pretest and the second half as a posttest. The tests' instructions and practice items were given both at the pretest and at the posttest, in exactly the same form[1]. The characteristics of the thus obtained pretests and posttests are given in Table 3, including the time limits used for each test, which were based on the findings of the pilot study.

Test scores. For all four spatial ability tests, two scores were determined: a speed score and a power score. The speed score was computed following the test's original scoring method (Table 4), only taking into account the items completed within the time limit. After the time limit, the participants were allowed to finish the remaining items at their own pace. The computation of the power score included both the items completed within the time limit and those finished after time was up. This score was determined as the proportion of attempted items that was marked correctly.

[1] One exception was the PFT. After administering the Paper Folding pretest, we discovered that the second practice item closely resembled the seventh test item. To increase comparability of pre- and posttest, in the posttest the second practice item was replaced by one that was highly similar to the seventh posttest item (see Appendix B).

Table 3. *Characteristics of Spatial Ability Pre- and Posttests.*

Test	Number of items	Score formula[a]	Score range	Time limit (min)
Card Rotations test (CRT)	80	number correct – number incorrect	0-80	3[a]
Paper Folding test (PFT)	10	number correct – ¼ * number incorrect	0-10	3[a]
Flags test	16	number correctly marked – number incorrectly marked[b]	0-30	2.5[c]
Mental Rotations test (MRT-VK)	12	number correct	0-12	4[d]

[a]Drawn from the tests' manuals. [b]Unmarked items are not considered. [c]Half of the original test's time limit was taken for half of the items. [d]A 4 min time limit, which, according to the manual (Peters, 1995), can be used instead of the standard time limit of 3 min, was considered to be more appropriate for the current age group.

Tridio test

To be able to measure content-specific effects of the Tridio training and to study the relationship between Tridio performance and spatial ability, a test was developed to measure children's performance in working with Tridio. To determine how Tridio performance could best be operationalized, the existing Tridio exercise sets were analyzed. Six frequently found types of tasks were selected as being the core problem types of Tridio. One of these types can be seen as a combination of two other types: the two steps of going from orthogonal views to a three-dimensional cube building and from a cube building to an isometric view have to be taken at once. This problem type was supposed to be suitable only for children who have considerable experience with Tridio exercises. As at the time of the Tridio pretest the children were assumed to never have worked with Tridio before, it was decided not to include tasks of this type in the test. The other core problem types were considered to be appropriate for 11-year-old Tridio novices. This was confirmed by pilot sessions with two children of approximately this age. The Tridio test, then, was constructed to consist of five subtests, each measuring children's performance on one of these types of problems (see Figure 9 for an example of each problem type):

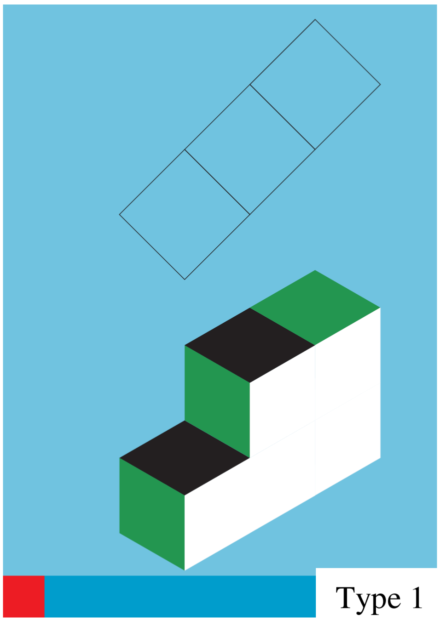
Type 1. The student is given an A4-sized book page on which an isometric view of a cube building is displayed. The task is to construct this cube building by placing cubes on the squares pictured at the top of the page, which have the same size as the cubes.

Type 2. Type 2 tasks are similar to those of Type 1: Again a cube building has to be constructed in accordance with a picture of its isometric view. In this case, however, the cube building is built on a cardboard board. This feature makes the task more difficult than a Type 1 task, since the positions of the cubes on the board have to be taken into account as well.

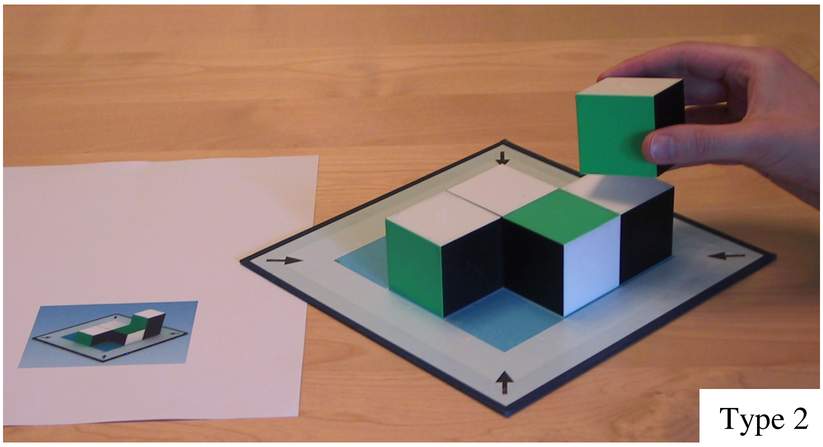
Type 3. The child is given a book page similar to those of Type 1 problems, but now some of the cube edges in the isometric view are uncolored, whereas some of the squares in the top picture do have a color. The colors of the squares indicate the colors of the cube building when viewed from the top. The task is to place cubes on the squares and to place mosaic pieces on the uncolored pieces of the isometric view picture – in which they fit exactly – in order to create a cube building and its isometric view.

Type 4. A cube building on the cardboard board is shown. The student has to construct a two-dimensional mosaic pattern corresponding to the isometric view of this building.

Type 5. The student is given a book page containing pictures of three orthogonal views of a cube building: a top view and two adjacent side views. Also, a board is displayed diagonally on the page, with a blue and a green arrow pointing to the sides of which the side views are given. By placing cubes on this board, the child is supposed to create a cube building that matches the three views.



Type 1



Type 2

Figure 9. Sample problems of each problem type in the Tridio test. In Type 1 problems, the task is to construct the pictured cube building by placing cubes on the squares. Type 2 tasks require the child to create a cube building on the board, corresponding to the isometric picture (continued on next page).

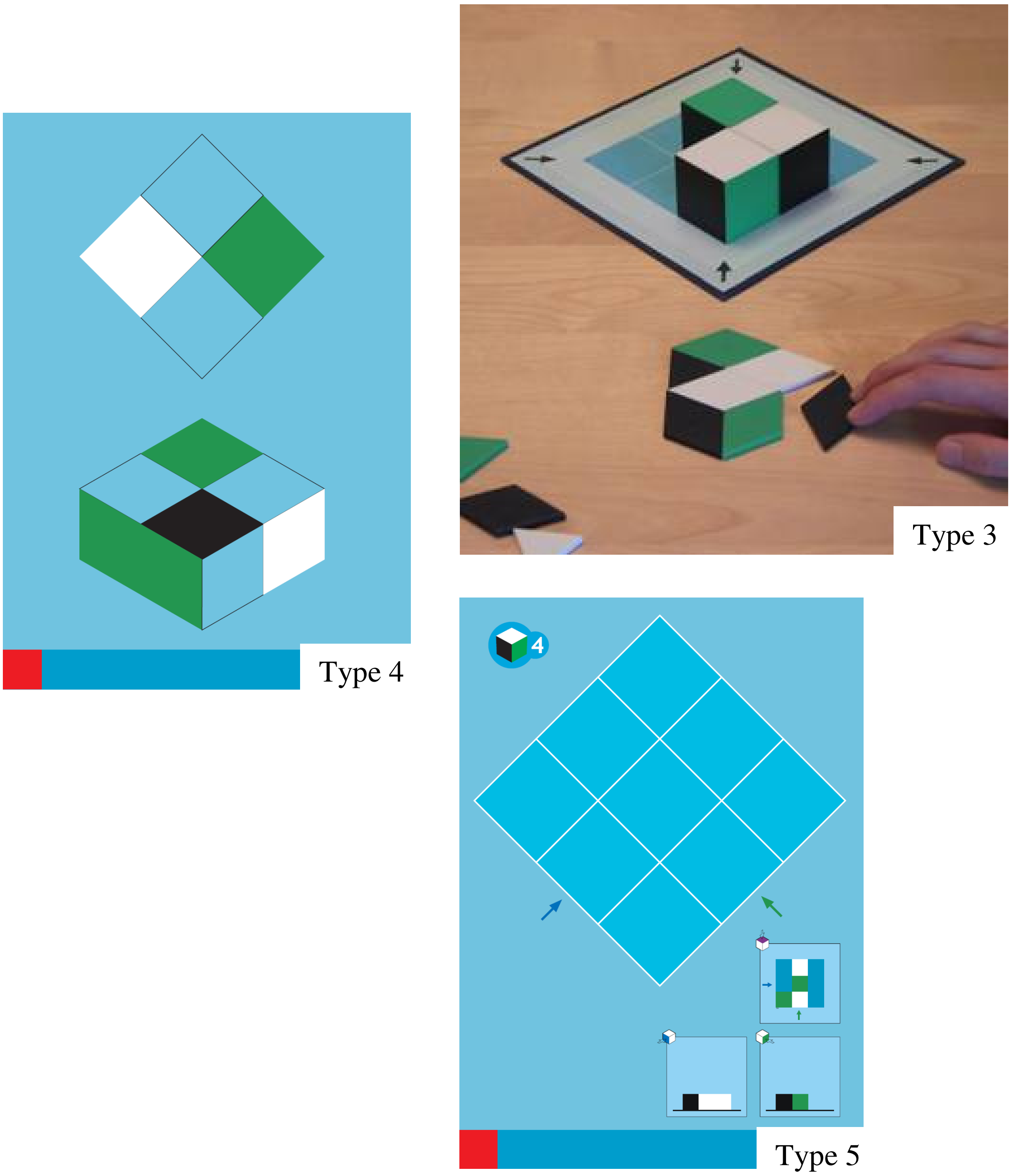


Figure 9 (continued). In Type 3, the task is to put mosaic pieces in the empty spaces in the bottom picture and cubes on the squares in the upper picture, such as to create a cube building and an isometric view that look the same (continued on next page). In problems of Type 4, the student has to create a mosaic pattern representing the isometric view of the presented cube building. In Type 5 tasks, the child is asked to construct a cube building on the pictured board, corresponding to the pictures of two orthogonal side views (displayed at the bottom) and the top view (the upper picture). A blue and a green arrow indicate the positions of the side views and the orientation of the top view. On the top left of the page, the number of cubes to be used is given.

In the Tridio exercise sets, the problems of Type 5 were also found in another form: using the cardboard board instead of a board displayed on a book page. This makes the task unnecessarily complicated, as in the absence of the blue and green arrows confusion about the orientation of the different views can easily arise. Therefore, it was decided to use the book page version of these assignments.

Scoring method. A scoring method based on judging the degree of correctness of completed Tridio assignments would probably be arbitrary, because it is unclear what aspects of a child's solution are most important. Therefore, it was decided to measure performance on each subtest by determining the number of correctly completed assignments within a time limit. For this method to produce reasonable variation between children, time limits had to be long enough for highly performing children to be able to finish a fair number of tasks (approximately 5 on each subtest). Also, all problems of a certain type had to take similar amounts of time to be solved. The number of cubes or mosaic pieces involved in a certain assignment can be seen as reasonable predictor of the time it takes to solve it. Obviously, it takes longer to construct cube buildings or mosaic patterns when they consist of more pieces. In addition, problems involving larger buildings or patterns usually are more difficult, thus requiring more time.

Exercises. From the existing Tridio exercise sets (Productief B.V., 2003, 2005, 2006a, 2006b, 2006c, 2006d, 2006e), several problems of each type were selected, each involving roughly the same number of cubes or mosaic pieces and thus expectedly requiring similar amounts of time to be solved. Problems were selected that were considered to be easy enough to allow inexperienced students to solve them and, at the same time, difficult enough to be able to produce some variation between children. In cases where the exercise sets did not contain enough appropriate exercises, additional problems were created. For the problems of Type 4, the material delivered to the student consists of cube buildings constructed by the experimenter. For each of these tasks, the form of the cube building was prespecified, but the positions of the colors were not, as this would inhibit fast construction of the cube buildings during test administration.

In this way, for each subtest a set of tasks was composed. These sets were pilot-tested with a 10-year-old and a 12-year-old with no previous Tridio experience. The tasks were found to be suitable for these children.

Subtest characteristics. The five subtests appeared in the Tridio test in the order in which the problem types were mentioned before (i.e., Subtest 1 contained Type 1 tasks, Subtest consisted of Type 2 tasks, etc.). In this way, the subtests were ordered by increasing difficulty, as was found in the pilot sessions. Also, similar problem types (Type 1 and 2, and Type 3 and 4) followed each other, which probably decreased the amount of instruction needed.

From the results of the pilot tests, time limits for each subtest were determined that would allow for a fair amount of exercises to be completed by high-performance students (one of the pilot children could be considered as highly performing). Subtests 1 and 2 were allotted 2 min time; for Subtests 3, 4, and 5, the time limit was set to 3 min. The number of assignments in each subtest was chosen to be high enough for any child to be unable to complete them all within the given time limit. Subtests 1, 2, and 3 consisted of 8 assignments each, whereas Subtests 4 and 5 both had 10.

Each subtest started with the task that was considered to be easiest. The remaining tasks were randomly ordered in advance. Thus, the order of the assignments within each subtest was the same for all participants, which made test scores comparable. An easy practice problem was included for every subtest. For each subtest (except for Subtest 4), the assignments were put together in a folder, each on a separate page, starting with the practice problem.

Posttest. The tasks for the Tridio posttest were adapted from those of the pretest. For Type 1 and Type 3 assignments, the mirror image was taken and the edge colors were systematically exchanged. Problems of Type 2 and 4 were only mirrored (in the format of Type 2 exercises, changing colors was not possible, while Type 4 problems did not have predefined colors). Type 5 assignments were modified by rotating the top view picture by 180°, adapting the side view pictures accordingly, and exchanging colors. Both the order of the subtests and the order of the tasks within each subtest were kept the same as in the pretest.

Procedure

As shown in Table 4, the experiment consisted of six phases. First, the spatial ability pretests were administered to all children, in two in-class sessions. After that, all children took an individual Tridio pretest. Matched pairs were constructed based on the spatial ability pretest scores. The children were divided into an experimental group (one child of each pair) and a control group (the other child of each pair). The experimental group children were given five Tridio training sessions of half an hour each. After this, all children (experimental group, control group, and non-matched children) completed the spatial ability posttests, also divided over two sessions. This was followed by a Tridio posttest. Due to limited time in the school schedule, this test was only administered to the experimental group. Finally, all children completed a questionnaire on their attitudes towards the tests and Tridio.

Table 4. *Phases of the Experiment.*

Phase	Experimental group	Control group and non-matched children	Period
Spatial ability pretests	*	*	week 1
Tridio pretest	*	*	week 2-3
Tridio training	*		week 3-6
Spatial ability posttests	*	*	week 7
Tridio posttest	*		week 7
Questionnaire	*	*	week 7

Note. The * symbols indicate which participant group(s) was/were involved in each phase of the experiment.

Spatial ability tests

The four spatial ability tests were divided over two sessions, each containing an SR test and a VZ test. The SR tests were given first in both sessions, as these were considered to be easiest. In the first session the CRT and then the PFT were administered, while the second session consisted of the Flags test and the MRT-VK. The order of the tests was the same for pretest and posttest. The two tests of each session were put together in a booklet. The first page of each booklet was a title page, on which the children had to fill out their names; the second page contained general instructions to the test session, which were the same as in the pilot study and were repeated for each test session.

The spatial ability tests were administered in the normal classroom settings. The test sessions took place in the mornings of two consecutive days, except for one class on the pretest, for which the first pretest session was on a Friday and the second session on the next Monday. Each session lasted approximately 40 minutes. While the experimenter (the author) administered the tests, the class teacher was present. At the first pretest session, one of the classes had a substitute teacher. This caused the children of this class to be noisier than during the other sessions in the same class and the sessions in the other classes. At the time of the posttest, the three classes had been combined to two classes, due to illness of one of the teachers. The classroom settings of the posttest, thus, differed from those of the pretest.

The tests were administered using the same procedure as in the pilot study. To be able to compute both speed and power scores, the red pencil method described by Johnson and Meade (1987, 2008) was used, although in this case pens were used instead of red pencils. In the current study, every child was allowed to fully complete each test, to minimize the influence of children's working speed on their power scores. The model items used in the pilot test were again employed in illustrating the test instructions. For the MRT-VK, models of the example items were shown in the same way as was done for the MRT-JM in the pilot study. At the end of each test session, the children were asked not to discuss the test with the children in the other classes.

Three children were absent during one or both of the pretest sessions. To two of them the missed test sessions could be administered later in the same week. The other child was excluded from the study, because she was absent the entire week. On the posttest sessions, all students were present, except for one child on the first session. This child took the missed session later that week.

Matching procedure

To control for several variables (specified below) assumed to be important, and in this way to enhance the power of the study (Cohen, 1988), the participants were combined into matched pairs. In the matching procedure three criteria were used, in the following order: score on spatial ability pretests, class, and gender. The score on the spatial ability pretests was seen as the most important criterion, as was done in Brinkmann's (1966) study. The idea behind this is that children's level of spatial ability may be related to their room for improvement, and thus to the extent to which their spatial ability can be enhanced. Alternatively, the spatial ability tests may be seen as aptitude tests in addition to being an achievement test, thus being able to predict future learning. As there were multiple classes involved in the study, children would possibly differ with respect to subjects treated in class or references to the experiment made by the teacher. To control for these differences, class was used as the second matching criterion. Another reason for matching on class is that the first pretest session's administration was quite disorderly in one of the classes. Therefore, it would not be very appropriate to compare pretest scores of children from this class to those of children from the other classes. Gender was added as a third criterion. As was previously indicated, gender is regarded as an important factor in studies on spatial abilities. Boys and girls often differ in their spatial ability. However, as the score on the spatial ability pretests was already used as the first matching criterion, this effect of gender did not have to be controlled for. But to answer the research question whether the effect of Tridio on spatial ability differs by gender, same-gender pairs would be useful.

To be able to make pairs on the basis of pretest scores, a combined pretest score had to be determined. This was done by first standardizing all eight pretest scores (i.e., four speed scores and four power scores), and then computing the median of the obtained *Z*-scores for each child. The median was chosen instead of the mean, because this measure of centre is more resistant to outliers than is the mean. In this context, this means that the median does not give extra weight to extreme scores compared to less extreme scores, whereas the mean does. As a compromise, the mean *Z*-score was also taken into account: Children whose mean *Z*-scores were too far apart were not matched to each other.

It was not possible to create enough pairs matching on all three criteria. Therefore, we decided to use the criteria in the order described above, first matching on combined pretest score and then, if possible, on class and gender. This matching procedure resulted in 25 pairs. The median *Z*-scores of paired children differed at most 0.15 from each other. The maximum difference between mean *Z*-scores was 0.35. Of the pairs, 22 consisted of two children from the same class, whereas in 3 pairs the children were from different classes. The matching on gender was less successful: Ten of the pairs were same-gender pairs, while in 15 of them the pairmates had different genders.

From each pair, one of the children was randomly assigned to the experimental condition and the other one to the control condition.

Tridio test

Because of the tactile nature of the Tridio tasks and the transience of its solutions, in-class administration of the Tridio test was not feasible, since this would not allow for checking the correctness of all children's answers. Also, cheating would probably be a problem. Therefore, the Tridio pretests and posttests were administered individually, in a separate room in the school. The children were taken out of their classes one by one to do the test.

Instructions. At the start of the Tridio test sessions, the children were told they would be doing a test containing five types of assignments, and that this test would not influence their school grades. They were instructed for each type to correctly complete as many assignments as possible within a certain time limit. It was explained that it would not be possible to complete all tasks, so it would not be a problem if they did not manage to do this.

Before starting each subtest, the practice assignment was shown to the child and it was explained what had to be done. Then, the child was asked to make this assignment. When problems arose, the child was helped in solving them. Mistakes were corrected. It was made sure the student understood the task. Next, the most important features of the problem type were briefly repeated. Participants were instructed to state they were ready whenever they would have finished an assignment, and not to turn the page to the next exercise until the experimenter said so. For Subtest 4, in which the tasks were not printed on book pages, the children were told that when they said they were ready, the experimenter would quickly construct the next cube building.

The students were instructed not discuss the test with other participants.

Test administration. Each child was observed while working on the assignments. When the child reported being ready, the experimenter asked if it was sure about its answer. This was done because, as had become clear in the pilot sessions, children easily overlook minor mistakes in their solutions, such as one cube edge having a wrong color. Explicitly asking if they were sure encouraged the children to have a second look at their answer and to detect possible errors. Several children actually discovered mistakes after this question was asked. Children were allowed to change their answer if they wished. When they were satisfied with their solution, they were told to go on with the next problem (or, in Subtest 4, the experimenter built the next cube building). Without the child seeing it, the experimenter wrote down whether the answers were correct.

All test sessions were videotaped (except in cases where the child or its parents refused this), to allow for a qualitative analysis to be conducted in a future study. Also, this created the possibility of rechecking the answers.

The time was kept with a kitchen timer, which beeped when time was up. If at that moment the child had nearly finished the current problem, it was asked to complete it. Otherwise, the child had to stop. For children who were not being filmed, the amount of extra time they needed to finish the last assignment was recorded.

To check children's previous experience with Tridio, they were asked if they had ever worked with the material before. All children answered “No” to this question.

Time schedule. The time schedule for the individual Tridio pretest sessions was determined by first dividing the available time into blocks of sessions in which children of one particular class would be tested. Subsequently, each class's students were randomly divided over the sessions reserved for that class. For the posttest sessions, a semi-random order was used, sometimes exchanging test sessions to accommodate to the classes' teaching schedules. If a child was absent on the day it was supposed to be tested, its session was swapped with that of another child yet to be tested, so that no time was lost. In this way, none of the children missed their test session.

Scoring. The score on each of the Tridio subtests was determined by the number of tasks correctly completed within the time limit. The timing onset was not very precise: Some children slowly turned the page to the first problem when they were instructed to start, others quickly did so; sometimes something went wrong in starting the timer. Therefore, it seemed fair to include in the score those exercises that had been finished just after time was up. This can be seen as a sort of rounding: When an exercise is correctly completed just after the time limit has been reached, more than half of it has probably been done within the allotted time, in which case the number of correctly completed exercises can be rounded upwards. It was decided that tasks that were finished within an extra time of at most 5% of the time limit would be considered on time. Thus, for Subtests 1 and 2 all problems that were completed within 6 seconds after time was up were included in the score, whereas for Subtests 3, 4, and 5, assignments were allowed to be finished at most 9 seconds beyond the time limit. Whether an exercise had been completed within this 5% extra time, was decided by watching the test session videos, or, for the participants who were not filmed, using the recorded amount of extra time needed.

The subtest scores were not summed up to obtain a single Tridio test score, because, since mean scores varied between subtests, this would not give equal weights to all scores. Instead, the test scores were analyzed separately for each subtest.

Tridio training

Pilot study. To get an idea of what a Tridio training would look like, a pilot study was conducted with two children, an 8-year-old boy and a 12-year-old girl. In two individual sessions the children worked on several Tridio exercises. From this pilot study, it became clear that children sometimes need help in arriving at the correct solution to an exercise. Furthermore, at times the children in the pilot study needed assistance in judging the correctness of their answers. Sometimes it was appropriate to skip exercises, whereas at other times a sequence of exercises slowly increasing in difficulty was more suitable, depending on the degree to which the child understood the type of exercises. From this, we concluded that the Tridio training had to be an adaptive one, in which the sequence of exercises would depend on the child's needs. This would give the opportunity of thoroughly practicing exercises with children who experience many problems with them. On the other hand, faster children would be able to go through easier exercises more quickly, and to continue with more challenging ones, which would likely keep them motivated.

Training sessions. For the training to be adaptive and to allow for help and assistance, the Tridio training had to consist of individual sessions with a trainer. This corresponds to the current use of Tridio in some primary schools, as examined by the telephone interviews mentioned in the Introduction section, held with 11 primary school teachers. Although most schools (8) use Tridio as a self-working in-class activity, some (3) treat it in individual sessions with a remedial teacher. One teacher remarked that Tridio was not very suitable to work on without support, because children found it difficult or tended to skip exercises.

Another lesson learned from the pilot study is that the training had to take at least 2 hours to cover all core types of Tridio exercises. To be on the safe side, the total duration of the training was decided to be 2.5 hours. Because long training sessions would probably be detrimental to children's motivation, and because repetition would probably improve the effectiveness of the training (Baenninger and Newcombe, 1989), the 2.5 hours were split up into five sessions of 30 minutes each.

Exercises. In the Tridio training sessions, all problem types that were part of the Tridio tests were treated, accompanied by several other types of exercises found in the Tridio exercise sets. Some extra problem types not present in the original exercise sets were added, because they were assumed to be helpful in understanding some other task types. The exercise types appearing in the Tridio training are listed in Appendix C.

The order in which the exercise types appeared in the training was based on increasing difficulty. For the exercise types that were considered most important, repetition was built in: Exercises of these types appeared in two consecutive training sessions. Furthermore, exercise types involving only isometric views were treated separately from problem types dealing with orthogonal views. The exact distribution of the exercise types over the training sessions is given in Appendix C.

The exercises for the training sessions were taken from the existing Tridio exercise sets, and whenever necessary, extra exercises were developed. The tasks of each type were ordered by increasing difficulty. The exercises included in the training were different from those appearing in the Tridio pretest and posttest.

Adaptiveness. As was mentioned before, the training was decided to be an adaptive one, in which the actual exercises presented to a child and their sequence depend on the child's performance level as perceived during the training. This adaptiveness was achieved in two ways: by the possibility of skipping exercises and by the availability of extra exercises to work on when having time left after completing the mandatory exercises.

To allow for a standardized training procedure, a set of rules was used in deciding whether an exercise could be skipped. When a child had completed an exercise correctly, quickly, and without help, it was allowed to skip the next one and to continue with the exercise after that. Exercises could only be skipped when they were of the same type as the previous one. Some exercises could not be skipped, because they were considered to be major steps in the exercise sequence. If a child experienced problems with an exercise just after one had been skipped, it was instructed to go back to the previously skipped exercise. Decisions about skipping exercises were always made by the trainer (the author): The child was just given the next exercise to be made.

For highly performing children, the training included additional, more challenging exercise types (the optional exercise types in Appendix C). These types of exercises required a thorough understanding of the more basic Tridio exercises, such as the ones appearing in the Tridio test, and thus were only presented to children having time left from these exercises. Moreover, for the mandatory exercise types, the most difficult exercises were optional ones, which were only presented to children who had time left.

Training session procedure. Just like the Tridio tests, the training sessions took place in a separate room in the school. The children were taken out of their classes one by one. They were told that this time the exercises would not be a test and that they were allowed to ask questions whenever they wanted. For each exercise type, it was explained what the child had to do. Then, the participant worked through the exercises, possibly skipping some of them. When problems arose, the trainer helped the child by giving hints. When an exercise had been completed, the student was told whether the solution was correct. In the case of an incorrect solution, the child was asked to try to discover what was wrong and to correct this. The children did not continue to the next exercise until they had correctly solved the current one.

It was recorded which exercises were performed. For exercise types that were repeated in a later session, more than half of the exercises were made in the first session, whereas in the second session the child completed the remaining (mandatory) tasks of this type.

To create the possibility of performing strategy analyses in a later study, all Tridio training sessions were videotaped (except for one child who refused). The children were instructed to think aloud. Some children extensively described their reasoning process, while others forgot or refused to do this, even after repeatedly being reminded of it. This resulted in differences between children's training situations: Since thinking aloud takes time, children who did not do so probably were able to complete more exercises within the allotted time. To improve on the similarity of children's training sessions, it was decided to leave out the thinking aloud procedure in the fourth and fifth training sessions.

Training schedule. For each experimental group participant, five training sessions were scheduled. As much as possible, each child's sessions were divided roughly evenly over the total duration of the 3 week training phase (i.e., so that the amount of time between each pair of successive training sessions was roughly equal, ranging from 2 to 6 days). To control for time-of-day effects, the different training sessions for each child were planned on different times of the day: some sessions in the morning, and others in the afternoon; some just after coming to school, others just before going home. For the children who had had their pretest early in the pretest phase, the training sessions were planned early in the training phase, to ensure similar time durations between pretest and training for all children.

When a child was absent during one of its training sessions, another child's training session was scheduled instead, so that no time was lost. In all cases, it was possible to catch up the missed session the next day.

Questionnaire

To measure their attitudes towards the spatial ability tests and towards working with Tridio, the children were asked to complete a questionnaire. Two different questionnaires were constructed, one for the experimental participants and one for the control children, who had not participated in the Tridio training and posttest. Most questions were multiple-choice questions, on which the different answer options often represented different points on an attitude scale. For example, for the question “Did you like working with Tridio?” the children could choose between four answers: “I liked it very much.”, “I liked it”, “I didn't like it so much”, and “I didn't like it at all”. For most questions, the children were also asked to write down why they chose a particular answer (full questionnaires can be obtained from the author). The questionnaires were filled out in class.

School performance data

To be able to control for school performance in examining the relationship between Tridio performance and spatial ability, some school performance data were obtained from the teachers. As a measure of general school performance, we acquired the grades of the children's second school reports (three reports are made every year), which they got just after the spatial ability pretests had been administered. This report's grades are based almost exclusively on children's school performance prior to the experiment. From these grades, a general school performance score was determined by calculating a weighted average. The grades were weighed in such a way that the overall score was determined equally by four component scores: reading, language (writing), math, and world orientation (geography, history, and biology). In addition to this weighted school grade average, we wanted use a specific measure of mathematics performance, since this is especially related to spatial ability (e.g., Guay & McDaniel, 1977; Tracy, 1987). For this aim, the children's latest math scores of the Cito Leerlingvolgsysteem (Student Following System, a Dutch national testing system for monitoring students' learning; Janssen & Engelen, 2002) were collected. Of science performance, also found to be related to spatial ability (e.g., Tracy, 1987, 1990), no separate measure was available.

Results

The data were analyzed using different statistical methods. All outliers identified by the 1.5 × *IQR* rule[2] were examined. Some of these outliers could be eliminated, because they could be expected to be caused by unusual circumstances (e.g., a child performing a lot worse on the posttest than on the pretest). The assumption of normality was checked by performing Kolmogorov-Smirnov tests. In cases where the normality assumption was violated or outliers

[2] *IQR* is the interquartile range: Q_3 (the 3rd quartile or 75th percentile) minus Q_1 (the 1st quartile or 25th percentile). The 1.5×*IQR* rule flags all cases lying more than 1.5×*IQR* above Q_3 or more than 1.5×*IQR* below Q_1 as possible outliers (Tukey, 1977).

could not be eliminated, a nonparametric test was used instead of a parametric one. For all statistical tests, an alpha level of .05 was used.

As was described in the Method section, for each of the spatial ability tests, in addition to the standard speed score, a power score was computed. The results obtained with the power scores, however, hardly differed from those found for the speed scores. To be concise, then, we decided to only report results pertaining to the speed scores (the standard way of scoring), leaving out the power scores data.

Relation between Tridio performance and spatial ability

The relation between children's Tridio performance and their spatial ability was studied by computing partial correlations between Tridio pretest scores and spatial ability pretest scores, controlling for children's math and general school performance as measured by their Leerlingvolgsysteem (LVS) math score and their weighted grade average, respectively. The obtained partial correlation coefficients are displayed in Table 5. All correlations were positive. However, correlations were significant only for Subtests 1 and 5, and for Subtest 4 with regard to the MRT-VK score.

Table 5. *Partial Correlations of Tridio Pretest Scores with Spatial Ability Pretest Scores Controlling for Math and General School Performance (n = 57).*

	Tridio test score				
Spatial ability test score	Subtest 1	Subtest 2	Subtest 3	Subtest 4	Subtest 5
SR tests					
Card Rotations test	.40**	.10	.23	.20	.34*
Flags test	.40**	.10	.01	.16	.33*
VZ tests					
Paper Folding test	.36**	.10	.16	.05	.32*
Mental Rotations test	.29*	.06	.23	.37**	.28*

Note. Pearson's r correlations were used. Leerlingvolgsysteem (LVS; Student Following System) math score and average grade were controlled for. Two of the 59 participants were excluded from the calculations, because no average grades and LVS math test scores were available for them. SR = spatial relations; VZ = visualization.
*p < .05. **p < .01. (two-tailed)

To further examine the apparent differences between the Tridio subtests, partial correlations between the subtests were computed, again controlling for math and general school performance. Significant correlations were found between Subtest 1 and 2 ($r = .38$, $p < .01$), between Subtest 2 and 4 ($r = .33$, $p = .01$), and between Subtest 3 and 4 ($r = .29$, $p = .03$).

Spatial ability scores

Gain scores. For the speed scores of all four spatial ability tests, for each participant a gain score was computed by subtracting its pretest score from its posttest score. Table 6 displays the means and standard deviations of the pretest, posttest, and gain scores, separately for the matched pairs' experimental and control group participants. The scores all increased from pretest to posttest, except for the experimental group's MRT-VK score. The gains were significant for the two SR tests, for both the experimental group (CRT: $t(24) = 3.84$, $p < .01$; Flags test: $t(23) = 6.79$, $p < .01$) and the control group (CRT: $t(24) = 2.25$, $p = .03$; Flags test: $t(24) = 4.02$, $p < .01$).

Gain score differences. To determine whether children's spatial ability was affected by working with Tridio, i.e., whether the training produced transfer effects, we subtracted the gain scores of each control group child from those of the experimental group child in the same pair. The thus obtained differences were analyzed using *t* tests, or Wilcoxon signed rank tests in cases were outliers could not be eliminated. None of the gain score differences were significantly higher than 0.

For comparability, effect sizes of the gain score differences were all computed using the formula for Wilcoxon signed rank tests: $r = Z / 2n$ (Field & Hole, 2003). In this formula, r is the measure of effect size (Cohen, 1988), Z is the Z-statistic of the Wilcoxon signed rank test, and n is the number of pairs. For the two VZ tests, slightly negative effect sizes were found (PFT: T[3] = 192.5, $r = -.02$; MRT-VK: $T = 120.5$, $r = -.08$), which indicates that for these tests gain scores were higher in the control group than in the experimental group. On the SR tests, we found positive effects (CRT: $T = 178.5$, $r = .12$; Flags test: $T = 144.5$, $r = .11$). These effect sizes can be considered small, since they are close to Cohen's (1988) benchmark of .1.

[3] Sum of positive ranks of the Wilcoxon signed rank test.

Table 6. *Means (Standard Deviations) of Spatial Ability Pretest, Posttest and Gain Scores of Matched Pairs' Experimental and Control Group Participants.*

		Experimental group			Control group		
Test score[a]	*n*[b]	Pretest	Posttest	Gain[c]	Pretest	Posttest	Gain[c]
SR tests							
Card Rotations test	25	37.7	45.4	7.7**	39.7	43.9	4.2*
(range 0-80)		(13.0)	(11.6)	(10.1)	(12.5)	(13.5)	(9.3)
Flags test	24	14.1	21.3	7.2**	13.6	19.0	5.4**
(range 0-30)		(6.4)	(6.9)	(5.2)	(7.2)	(8.8)	(6.6)
VZ tests							
Paper Folding test	25	3.6	4.3	0.6	3.5	4.1	0.6
(range 0-10)		(2.4)	(1.7)	(2.1)	(2.3)	(2.4)	(2.2)
Mental Rotations test	23	5.3	5.0	-0.3	5.6	5.8	0.2
(range 0-12)		(2.8)	(3.0)	(1.9)	(3.2)	(3.0)	(2.4)

Note. SR = spatial relations; VZ = visualization.
[a]For each test, the range of possible scores is given. [b]There were 25 pairs, but for some tests one or two outlying pairs were eliminated. [c]Posttest score – pretest score. Gains significantly higher than 0 have been marked.
*$p < .05$. **$p < .01$. (one-tailed).

Within-pair correlations. To evaluate the success of the matching procedure, we looked at the correlations between the gain scores of paired children. These correlations should be positive and high (above .30) for a matched pairs design to have higher power over a non-matched control group design (Cohen, 1988). However, the within-pair correlations of gain scores were rather low and some of them even negative (CRT: r = -.02; Flags test: r = .27; PFT: r = -.16; MRT-VK: r = .27).

Tridio test scores

To study the content-specific effects of the Tridio training, the experimental group's scores on the Tridio pretest and posttest were compared. The subtest scores were not normally distributed, except for Subtest 4. For simplicity, the non-parametric Wilcoxon signed rank test was used for all five subtests. Descriptive and test statistics are reported in Table 7. On each of the Tridio subtests, the posttest score was significantly higher than the pretest score. Thus, as expected, the children's Tridio performance improved from pretest to posttest. Effect sizes of these improvements, given in the rightmost column of Table 7, were computed using Field & Hole's (2003) previously discussed formula for *r*. Since the obtained effect sizes are all close to or higher than .5, they can be considered as large (for *r* effect sizes, .5 is defined as large, comparable to .8 for *d* effect sizes; Cohen, 1988).

Table 7. *Experimental Group Pre- and Posttest Scores on Tridio Subtests (n = 25).*

	Pretest		Posttest		Difference	
Tridio test score[a]	*M* (*SD*)	*Mdn* (*IQR*)	*M* (*SD*)	*Mdn* (*IQR*)	*T*	*r*[b]
Subtest 1	3.6	4.0	5.1	5.0	231**	.58
(range 0-8)	(0.9)	(1.0)	(1.2)	(2.0)		
Subtest 2	3.5	3.0	4.6	4.0	211.5**	.48
(range 0-8)	(1.0)	(1.0)	(1.1)	(1.0)		
Subtest 3	2.4	2.0	4.4	4.0	272.5**	.58
(range 0-8)	(1.0)	(1.0)	(1.5)	(1.5)		
Subtest 4	2.4	2.0	6.1	6.0	300**	.61
(range 0-10)	(2.0)	(3.5)	(2.1)	(2.5)		
Subtest 5	1.8	1.0	4.6	5.0	316**	.59
(range 0-10)	(1.6)	(1.0)	(2.1)	(3.0)		

Note. Since the data were not normally distributed, medians and interquartile ranges (i.e., 75th percentile – 25th percentile) were reported as measures of centre and spread in addition to means and standard deviations, and a Wilcoxon signed rank test was used for comparing pretest and posttest scores. *IQR* = interquartile range.
[a]For each subtest, the range of possible scores is given. [b]Effect size, computed by $z / \sqrt{(2n)}$ (Field & Hole, 2003).
*$p < .05$. **$p < .01$. (one-tailed)

Gender differences

To examine the differences between boys' and girls' spatial ability, the pretest scores of all children were analyzed. Table 8 shows the means and standard deviations of these scores by gender. On all tests, boys scored higher than girls. However, this difference was only significant for the MRT-VK, $t(44.6) = 3.26$, $p < .01$). Cohen's (1988) *d* effect sizes of the gender differences are reported in the rightmost column of table 10. According to Cohen's definitions of low ($d = .2$), medium ($d = .5$), and high ($d = .8$) effect sizes, the effect on the MRT can be considered as very high, the PFT effect size as rather low and the CRT and Flags test effect sizes as medium.

In order to study whether the effect of the Tridio training on spatial ability differed with gender, a 2 × 2 (ExperimentalCondition × Gender) ANOVA was performed on the gain scores of the 50 children in the pairs. In this analysis, the experimental and control group were treated as independent (i.e., non-matched) groups, because most pairs were not matched on gender. The only significant effect found was an ExperimentalCondition × Gender interaction effect on the CRT gain score, $F(1, 46) = 5.91$, $p = .02$. Experimental boys had much higher gains on the CRT ($M = 10.9$, $SD = 9.7$) than control boys ($M = 0.7$, $SD = 9.2$), whereas experimental girls' gain scores ($M = 4.3$, $SD = 9.6$) were lower than those of control girls ($M = 6.9$, $SD = 8.8$).

Table 8. *Means (Standard Deviations) of Boys and Girls' Spatial Ability Pretest Scores; Effect Sizes of Gender Differences.*

Test score	Boys (n = 27)	Girls (n = 32)	d^a
SR tests			
Card Rotations test	41.3 (14.1)	34.8 (11.7)	.51
Flags test	15.4 (7.7)	12.4 (6.1)	.43
VZ tests			
Paper Folding test[b]	4.0 (2.6)	3.3 (1.9)	.28
Mental Rotations test	6.5 (3.5)	4.0 (2.3)	.88**

Note. SR = spatial relations; VZ = visualization.
[a]The effect size d of the gender difference was computed using Cohen's (1988) formula for d_s (p. 66-67). [b]One outlying girl was eliminated ($n_{girls} = 31$).
*$p < .05$. **$p < .01$ (two-tailed).

Questionnaire

Only the answers to the most interesting questions of the questionnaire are reported. In discussing these answers, we distinguish between the experimental group and the control group, in this case composed of both the matched pairs control group and the non-matched participants ($n = 34$).

The children generally liked working with Tridio. Of the experimental group children 87% reported liking working with Tridio (they either "liked it" or "liked it very much"). The control group children only worked with Tridio on the pretest. Of them, 74% liked this Tridio test. Some of these children did not like the test because it was timed or because it was videotaped. The popularity of Tridio was further shown by the answers to the question whether the child would want to work with Tridio again in future. Of the experimental group children who answered this question, 87% reported wanting to work with Tridio again. For the control group, this percentage was somewhat lower, but still fairly high: 70% answered "yes". However, most children did not think of Tridio as a free time activity. Of all children who answered the question, only 9% reported they liked Tridio so much that they would want to work with it at home.

As an explanation of why they thought Tridio is fun, many children mentioned that it is a kind of game or puzzle. One child verbalized this nicely: "You can play and work at the same time". Also, some children were pleased by the use of real objects. One child, for example, wrote down: "You can rotate it and pile it up. You can't do that on a piece of paper." Some children especially liked the 3D-2D connections involved in Tridio: "When you watch closely, it looks the same." Children who did not like working with Tridio reported that it was hard or childish. Remarkably, difficulty and easiness were given as explanations of both liking and disliking Tridio. As a reason for wanting to work with Tridio more often, some children mentioned that it would allow them to skip lessons.

The experimental group children were asked what they thought about the number and length of the Tridio sessions. The number of sessions, including the two test sessions, was OK for 36% of them. 44% reported they would have wanted more sessions, whereas 20% found they had worked with Tridio too often. One child said he wanted to do it all day. Some children did not want to miss lessons; some found that Tridio gets boring when working with it too often. The length of the sessions was found to be OK by most children (65%). Some children thought the sessions were too long (22%) or too short (13%).

General Discussion

In this section, we discuss the outcomes of our main study. First, we study whether Tridio adds to the general curriculum in focusing on spatial ability and thus has the potential of improving it. This is done by looking at the relationships between children's Tridio performance and their spatial ability test scores, controlling for school performance. Then we look at the effects of the Tridio training on children's spatial ability, i.e., the transfer effects of Tridio. Subsequently, the content-specific effects of the training are discussed. We also look at the gender differences in spatial ability and in training effects. Finally, we discuss children's attitudes towards working with Tridio.

Relationship between Tridio performance and spatial ability

When controlling for math and general school performance, a positive relationship was found between children's spatial ability test scores and their performance on Tridio Subtests 1 and 5. In addition, performance on Tridio Subtest 4 was found to be positively related to the score on the MRT-VK. For Subtests 2 and 3, and for Subtest 4 with regard to the other spatial tests, correlations were non-significant. These findings indicate that, as we expected, Tridio provides an additional focus on spatial ability beyond the standard school curriculum, but that not all Tridio activities contribute to this. The activities included in Subtests 1 and 5 add to the standard school curriculum in treating spatial ability, and may thus be helpful in improving it, while this is not the case for Subtests 2 and 3. Subtest 4 may only be helpful in enhancing children's three-dimensional mental rotation skills.

The activities of Subtest 5 clearly differ from those in the other subtests, involving orthogonal instead of isometric views. Apparently, the activity of constructing cube buildings corresponding to orthogonal views is more related to spatial ability than are most activities involving the relationship between a cube building and its isometric view. This may be explained by looking at possible solution strategies. When constructing a cube building in accordance with an isometric view or when laying out the isometric view of a cube building, an analytic strategy can be used instead of a more spatial one. For example, if a mosaic pattern has to be constructed representing the isometric view of a cube building, one may, instead of visualizing the whole structure to be made, construct the mosaic pattern piece by piece, each time comparing it to the model cube building. In contrast, when constructing a cube building from three orthogonal views (two side views and a top view), such a try-and-compare method may be less efficient, since the mapping between the views and the cube building to be constructed is less apparent: By observing the three views, one cannot directly see how the corresponding cube building will look, and simply constructing the cube building view by view often does not work, since, when altering one side of a cube building to make it match the corresponding view, another, earlier completed side may not match its view anymore. To more efficiently solve this type of exercise, the child has to mentally combine the three views to a single cube building, which certainly is a spatial strategy. The activities of Subtest 5, thus, may call for a more spatial strategy than the activities involving isometric views, which may explain the relation between this subtest and spatial ability.

Considering this discussion on strategies, it is remarkable that Subtest 1 is also related to spatial ability, approximately to the same extent as Subtest 5, while the other subtests involving isometric views are not (or not to all spatial test scores, as is the case for Subtest 4). It is not clear why this is the case. Possibly, the fact that Subtest 1 was administered first caused the children to use a different approach to it than to the other subtests, which may have contributed to the difference in their relations with spatial ability. Another explanation may be that more spatial strategies were used on Subtest 1 than on the more difficult Subtests 2-4: According to Glück and Fitting (2003), on more difficult tasks, strategies are more analytic and less holistic (spatial). These explanations may be further examined in a future study.

Although not all Tridio activities appear to have the potential of improving children's spatial ability, by inclusion of activities of the types included in Subtests 1, 4, and 5, the Tridio training still has the possibility of enhancing spatial ability. Also, the other activities included in the training, which were not in the Tridio test and, thus, were not examined for their relation to spatial test scores, may still be helpful in improving spatial ability. Whether improvements were indeed found, is discussed next.

Transfer effects

Contrary to our assumption that Tridio may improve spatial ability, transfer effects of the Tridio training on spatial ability were not found. On both SR tests, both the experimental and the control group significantly improved from pretest to posttest, but the experimental group gains did not differ significantly from the control group ones. On the VZ tests, no significant pretest-posttest improvements were found for either experimental or control group, and differences between experimental and control group gains were again non-significant.

A similar failure to find transfer effects was reported by Shavalier (2004), in a study on spatial training with 4th- to 6th-grade children. However, our finding is inconsistent with the results of several other studies (Alias et al., 2002; Brinkmann, 1966; Kwon, 2003; Lord, 1985;

Sanz de Acedo Lizarraga & García Ganuza, 2003), which did show transfer effects of spatial training with older participants (from 14 years of age). Our results also contradict Ben-Chaim et al.'s (1988) claim that 5th- to 9th-grade (11- to 14-year-old) children's performance on a VZ test increased by working on activities highly similar to the Tridio exercises. However, since Ben-Chaim et al. did not use a control group in their study and did not separate content-specific effects from transfer effects, their results are hard to interpret. Our study gives a clearer view of the effects of a training including isometric and orthogonal views on 5th-grade children's spatial ability.

The fact that we did not find transfer effects may be explained by several factors, which are discussed next.

Power. The failure of the current study to find transfer effects on spatial ability may partly be attributed to a low power. The found effects were much smaller (highest $r = .12$) than the r effect size of .3 roughly corresponding to the estimated d effect size of .5 (Cohen, 1988). On the VZ tests, the effects were even slightly negative, indicating that the gains on these tests were somewhat lower for the experimental group than for the control group. With a higher power, then, we still do not expect positive effects on these tests. The effects of the training on the SR test scores were positive, but small (CRT: $r = .12$; Flags test: $r = .11$), close to Cohen's (1988) benchmark of .1, which is comparable to a d effect size of .2. When entering .2 for the effect size in the G*Power 3 software program (Faul et al., 2007), we found that, concerning effects on the SR tests, the actual power of the study, i.e., the probability of finding significant results, was only .25. In a future study, effects on the factor SR may be detected by drastically increasing the sample size to about 150 pairs, to obtain a power of .80.

Another way to improve the power of the study is to find a better way of matching. From the low within-pair correlations of gain scores, we can conclude that our matching procedure was not very successful: It did not provide much advantage over a non-matched control group design (Cohen, 1988). Apparently, spatial ability pretest scores, together with class and gender, were not sufficiently predictive of children's gain scores. This may in part have been caused by including in the matching procedure both speed and power scores of the pretests: As power scores are not the standard way of scoring spatial tests, they are harder to interpret and their potential in predicting future improvement is less clear than is the case for speed scores. Since the power scores did not provide a useful addition to the current study, in a future study it may be better to use only speed scores for matching. When leaving out power scores, the matching on speed scores can be more accurate, since fewer factors have to be taken into account. Furthermore, the matching procedure might be improved by adding to the matching variables other factors that are expected to be related to spatial improvement, such as measures of general intelligence, school performance or motivational factors.

Thus, the Tridio training as employed in this study might have a small effect on the SR factor, which may be detected using more participants and/or better matching, whereas an effect on VZ is not expected, even when the study's power is increased.

Training content and duration. Another explanation for the fact that we did not find transfer effects may lie in the content and the duration of the training. In the employed Tridio training, there were several optional exercise types that were only presented to children who had time left from the mandatory, more basic, exercises. These optional exercises were rather difficult, and required thorough understanding of some more basic exercise types, such as the ones appearing in the Tridio test. Not all children got to work on the optional exercise types, and some only had very little time to work on them. The optional exercise types, however, seem very interesting for improving spatial ability. One of them, for example, requires the child to imagine how a cube building will look when a cube is moved (type H in Appendix C), while in another exercise type, the child is asked to construct the isometric view that would be seen when looking at a cube building from another angle (type J in Appendix C). Both activities clearly involve mental manipulation: imagining how something would look after performing some transformation on it. Possibly, these exercise types have more potential in improving spatial ability, especially VZ, than have the more basic Tridio exercises, which merely involve translating between two- and three-dimensional representations. A longer training, then, in which every child gets the chance of extensively working on these more difficult exercises, after having had enough practice with the more basic exercise types, can be expected be more effective in improving spatial ability, primarily VZ. It may be interesting to investigate the effects of such a longer training in a future study.

Participants' age. A further factor that may have played a role in our failure to find significant transfer effects is the age of the participants. Our study employed 5th-grade (11-year-old) students. This age group was selected because it most closely matched the curriculum standards, the current use of Tridio, and the recommendations of other researchers proposing similar activities. However, earlier studies in which transfer effects of spatial training were found (Alias et al., 2002; Brinkmann, 1966; Kwon, 2003; Lord, 1985; Sanz de Acedo Lizarraga & García Ganuza, 2003) involved older participants than the current study, ranging from 14 to approximately 21 years of age. The Shavalier (2004) study, in which no transfer effects were found, included children aged 10 to 12 years, similar to our study. The training activities employed in these studies were not very similar to those in the current study. But also in Ben-Chaim et al.'s study with 5th- to 8th-grade students, involving activities highly related to the Tridio exercises, gains in spatial ability scores were found to be higher for older children, with the highest gains found for 7th-graders (13-year-olds). Together, these findings show that spatial abilities – specifically VZ, since most of the aforementioned studies only involved VZ tests – may be more easily improved at higher ages, i.e., from approximately 13 years. Possibly, then, Tridio has more effect on older children's spatial abilities, especially VZ. This may be explained by the use of different strategies: Possibly, younger children find the Tridio exercises more difficult than do older ones, thus using less spatial (more analytic) strategies in solving them (Glück & Fitting, 2003). Following this theory, a training with Tridio may provide less spatial experience to younger children than to older ones.

Visualization and spatial relations. Although non-significant, the effects of the Tridio training on SR were higher than those on VZ, which is contrary to our assumption that Tridio would more likely affect VZ. As suggested before, this could partly have been caused by the age of the participants: VZ may be more easily improved in older participants, while for SR, no such pattern was found.

Another explanation may lie in the fact that SR tasks are easier than VZ tasks. When prior performance on easier tasks is not very high, and thus much improvement is possible, it seems reasonable that on these easier tasks, performance is more easily improved than on more difficult tasks. In our study, pretest scores on both the SR and the VZ tests were not very high: Mean scores were less than half of the maximum possible scores (CRT: M = 38.7, max. 80; Flags test: M = 13.9, max. 30; PFT: M = 3.6, max. 10; MRT: M = 5.4, max. 12), leaving much room for improvement on both SR and VZ. In this case, thus, we may assume that the relatively easy SR was easier to improve than the more difficult VZ.

As mentioned before, by focusing more on the more complex Tridio activities, which were optional in the current training, the training will possibly have more effect on VZ.

Content-specific effects

On all Tridio subtests, experimental group children improved much from pretest to posttest. For these improvements, no comparison with the control group was possible, since the posttest was not administered to the control group. However, the found r effect sizes, ranging from .48 to .61, are comparable to a *d* effect size of at least .8 (Cohen, 1988), which is much larger than the test-retest effect size of approximately .4 found in Baenninger and Newcombe's (1989) meta-analysis on training effects on spatial tests. This indicates that the effects of the Tridio training were so large that they can be assumed to be higher than simple test-retest gains that would have been found for the control group. Thus, by working with Tridio, children's Tridio performance is improved: a content-specific training effect. This result corresponds to our expectations and is consistent with Baenninger and Newcombe's finding that test-specific spatial training is highly effective.

Thus, although working with Tridio was not found to improve children's spatial ability, children do learn something from it: They get better at the Tridio exercises. Maybe, children just learn a trick, a specific way to solve these special kinds of exercises. But, possibly, other abilities than spatial ability might be involved, such as logical reasoning. The potential of Tridio in improving other abilities might be interesting to investigate in a future study.

Gender differences

In accordance with our hypothesis that gender differences will be present on three-dimensional mental rotation tasks, we found that boys outperformed girls on the Vandenberg & Kuse MRT. This result further supports the notion that gender differences on the MRT are already present in children, which was earlier shown by, for example, Geiser et al. (2007), Kerns and Berenbaum (1991), and Vederhus and Krekling (1996). The effect size of this gender difference was very large, d = .88, which is close to the adult effect size of .94 found for the same scoring method in the meta-analysis by Voyer et al. (1995). Thus, for 11-year-olds almost adult gender effects were found.

On the SR tests, contrary to our expectations, gender differences were non-significant, which seems to contradict the results obtained in other studies with children (Kerns & Berenbaum, 1991; Richmond, 1980; Vederhus & Krekling, 1996) However, the effect sizes of the SR gender differences in the current study, d = .51 on the CRT and d = .43 on the Flags test, are comparable to the significant adult effect sizes found in Voyer et al.'s (1998) meta-analysis, varying from .31 to .44. This indicates that on these tests, gender differences might have been present in the current sample, but were not detected due to limited power (caused by a small sample size). Indeed, the power for finding these effects was .49 for the Flags test and .61 the CRT, respectively, both not very high. Thus, although we found no significant gender differences on the SR tests, the effect sizes indicate that gender differences would probably have been significant if a larger sample were used.

As we expected, no significant gender difference was found on the Paper Folding test. This agrees with the results of other studies, both with children (e.g., Shavalier, 2004) and with adults (Voyer et al., 1995).

Gender differences in training effects. A gender difference in training effects was only found on the CRT, as indicated by a significant ExperimentalGroup × Gender interaction effect on gain scores. On this test, experimental boys improved much more than control boys, whereas for the girls, there was more improvement in the control group than in the experimental group. For boys, thus, the effect of the Tridio training on CRT performance was larger than for girls. This runs counter to the experiential explanation of gender differences, from which it is hypothesized that spatial training has more effects on girls than on boys, because boys have had more spatial experience in the past than girls, which leaves less room for improvement (Baenninger & Newcombe, 1989). The findings on the other tests, on which no significant gender differences in training effect were present, also fail to support this experiential hypothesis. Thus, as expected, our results add to the growing body of research contradicting the experiential explanation of gender differences (Baenninger and Newcombe; Ben-Chaim et al., 1988; Casey et al., 2008; Clements et al., 1997; Sanz de Acedo Lizarraga & García Ganuza, 2003; Shavalier, 2004).

The finding that, with regard to the CRT, the effect of the training was larger for boys than for girls, is rather remarkable. We did not encounter such results in earlier studies. However, because none of the other tests in the current study showed such a training effect difference in favor of boys, not even a tendency towards it, the result on the CRT may be seen as a statistical artifact.

Attitudes towards Tridio

From the questionnaire, it was found that children generally like working with Tridio. This agrees with earlier experiences (M. van Herel at Productief B.V., personal communication, September 10, 2007). Thus, introducing Tridio in the school curriculum may generally meet with enthusiasm, and children may be highly motivated to work with it. Some of the experimental group children (20%) found that the amount of training sessions was too high. When planning a future study involving more sessions, motivational issues should be taken into account, possibly by allowing more time between training sessions

Conclusions

Tridio adds to the standard school curriculum in providing practice in spatial ability, but not every single Tridio activity contributes to this. Activities in which children have to construct a cube building from three orthogonal views (Subtest 5) appear to have the potential of being helpful in improving children's spatial ability. For activities dealing with translations between cube buildings and their isometric views (Subtests 1-4), the pattern is less clear: Some of these activities add to the school curriculum, others do not.

The current Tridio training did not produce transfer effects on children's spatial ability. When drastically increasing the number of participants to about 150 and/or improving the matching procedure, significant effects may be found on the factor SR, but probably not on the VZ factor. A longer training, in which the children can spend more time on the more difficult, in our training optional, exercises, is expected to be more effective in improving children's spatial ability, especially in enhancing VZ. Furthermore, the failure to find transfer effects may be related to the age of the participants: Especially the VZ factor of spatial ability may be more easily improved in older participants, from approximately 13 years of age.

In accordance with our expectations, the Tridio training produced content-specific training effects. Children's performance on Tridio activities can greatly be improved by a training with Tridio. Possibly, other abilities than spatial ability are involved in Tridio performance. This may be an interesting subject for a future study.

A significant gender difference favoring boys was found on the MRT-VK, with a close-to-adult-level effect size. On the SR tests, although gender differences were non-significant, effect sizes were comparable to adult gender differences on this factor, which indicates that on these tests gender differences might have been present, but have not been detected because of a small sample size.

As was expected, Tridio was not more effective in girls than in boys. Thus, our findings fail to support the experiential explanation of gender differences.

Practical implications

Although in this study we found no significant effects of Tridio on children's spatial ability, Tridio might still be interesting to use in schools. Children generally like working with Tridio, and they clearly learn something from it, although no transfer effects on spatial ability were found. Whether there is transfer to abilities other than spatial ability might be explored in future studies.

For improving spatial ability, the activities involving orthogonal views may possibly be more effective than the exercises in which children have to translate between an isometric views and a cube building, although this is not exactly clear. Possibly, the more difficult Tridio exercises, such as constructing the isometric view that would be seen when viewing a cube building from another side, or imagining how a cube building would look if a cube were moved, are more effective in improving spatial ability than are the more basic Tridio activities in which the child merely has to translate between two- and three-dimensional representations of cube buildings. For a Tridio training to be more effective in enhancing children's spatial ability, then, it should be aimed more at acquiring understanding of these more difficult exercises, using the basic exercise types primarily as a means to acquire the prior knowledge necessary for the more difficult activities. The Tridio exercises sets may be improved by focusing more on these more difficult exercises, since currently these exercises do not have a very prominent place in the exercise sets and might easily be overlooked by the teachers.

Future research

To further examine the effects of Tridio on spatial ability, several lines of research are possible. First of all, the current study may be replicated with a higher power, which can be achieved by using a much higher number of participants and/or a better matching procedure. With a higher power, significant effects may be found on SR, but probably not on VZ.

Furthermore, one may investigate the effects of Tridio on older participants (from 13 years), for whom larger improvements on VZ can be expected. It is also interesting to examine whether a training focusing more on the more complex Tridio exercises has more effects on spatial abilities, especially VZ. Such a training should probably consist of more sessions than the current one, allowing every child to first practice the basic Tridio exercises, and then to extensively work on the more difficult activities. Motivational issues should be taken into account, possibly by allowing more time between sessions.

Additionally, it may be interesting to further study the role spatial ability plays in solving the different Tridio exercises. This can be done by a correlational study similar to the one we performed with the Tridio subtest scores and the spatial ability test scores, now including all types of Tridio exercises, not just the basic ones. Another way to assess the role of spatial ability in the Tridio activities is to analyze the strategies used by children in solving the exercises. This may, for example, be done by analyzing video-material of children working on Tridio activities, by using a thinking aloud procedure, or by employing questionnaires on strategy use.

Another line of research may be to examine the possibility that working with Tridio enhances other abilities, such as logical reasoning.

References

Alias, M., Black, T. R., & Gray, D. E. (2002). Effects of instruction on spatial visualization ability in civil engineering students. *International Educational Journal, 3*(1), 1-12.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5/6), 327-344.

Barzel, B., Haug, R., Häger, K., & Rabstein, A. (2007). Körper erkunden: Erfahrungen in der Lernwerkstatt [Exploring bodies: Experiences in the learning-working place]. *Mathematik Lehren, 144*, 24-41.

Ben-Chaim, D., Lappan, G., & Houang, R. T. (1988). The effect of instruction on spatial visualization skills of middle school boys and girls. *American Educational Research Journal, 25*(1), 51-71.

Brinkmann, E. H. (1966). Programed instruction as a technique for improving spatial visualization. *Journal of Applied Psychology, 50*(2), 179-184.

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

Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction, 26*, 269-309.

Clements, D. H., & Battista, M. T. (1992). Geometry and spatial reasoning. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 420-464). New York: Macmillan.

Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal, 98*(2), 171-186.

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

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of factor referenced cognitive tests.* Princeton, NJ: Educational Testing Service.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175-191.

Field, A. P., & Hole, G. (2003). *How to design and report experiments.* London: SAGE Publications.

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2007). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences.* doi: 10.1016/j.lindif.2007.09.001

Glück, J., & Fitting, S. (2003). Spatial strategy selection: Interesting incremental information. *International Journal of Testing, 3*(3), 293-308.

Gravemeijer, K., Figueiredo, N., Feijs, E., Van Galen, F., Keijzer, R., & Munk, F. (2007). *Meten en meetkunde in de bovenbouw: Tussendoelen annex leerlijnen bovenbouw basisschool* [Measurement and geometry in the upper grades: Intermediate attainment targets and learning-teaching trajectories for the upper grades in primary school]. Groningen, Netherlands: Wolters-Noordhoff.

Guay, R. B., & McDaniel, E. D. (1977). The relationship between mathematics achievement and spatial abilities among elementary school children. *Journal for Research in Mathematics Education, 8*(3), 211-215.

Halpern, D. F., & Collaer, M. L. (2005). Sex differences in visuospatial abilities: More than meets the eye. In P. Shah & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 170-212). New York: Cambridge University Press.

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

Janssen, J., & Engelen, R. (2002, July). *Verantwoording van de toetsen rekenen-wiskunde 2002* [Justification of the math tests 2002]. Arnhem, Netherlands: Citogroep. Retrieved July 12, 2008, from http://toetswijzer.kennisnet.nl/html/tg/2.pdf

Johnson, E. S., & Meade, A. (1985). *Description of the JM spatial battery.* Unpublished manuscript, University of North Carolina at Chapel Hill.

Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development, 58,* 725-740.

Johnson, E., & Meade, A. (2008). *The JM battery of spatial tests: Instructions for administration - Middle level.* Unpublished manuscript.

Kerns, K. A., & Berenbaum, S. A. (1991). Sex differences in spatial ability in children. *Behavior Genetics, 21*(4), 383-396.

Kwon, O. N. (2003). Fostering spatial visualization ability through web-based virtual reality program and paper-based program. In Goos, G., Hartmanis, J., & van Leeuwen, J. (Series Eds.) & Chung, C.-W., Kim, C.-K. Kim, W., Ling, T.-W., & Song, K.-H. (Vol. Eds.), *Lecture notes in computer science: Vol. 2713. Web and communication* *technologies and internet related social issues - HSI 2003* (pp. 701-706). Berlin, Germany: Springer.

Lappan, G., & Winter, M. J. (1982). Spatial visualization. In L. Silvey & J. R. Smart (Eds.), *Mathematics for the middle grades (5-9)* (pp. 118-129). Reston, VA: The National Council of Teachers of Mathematics.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56*(6), 1479-1498.

Lohman, D. F. (1993, July). *Spatial ability and g.* Paper presented at the first Spearman Seminar, University of Plymouth, UK.

Lord, T. R. (1985). Enhancing the visuo-spatial aptitude of students. *Journal of Research in Science Teaching, 22*(5), 395-405.

National Council of Teachers of Mathematics. (n.d.). *Geometry standard.* Retrieved October 16, 2008, from http://standards.nctm.org/document/appendix/geom.htm

Peters, M. (1995). *Revised Vandenberg & Kuse mental rotations tests: Forms MRT-A to MRT-D* (Technical Report). Guelph, Ontario, Canada: University of Guelph, Department of Psychology.

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28,* 39-58.

Piaget, J., & Inhelder, B. (1956). *The child's conception of space.* London: Routledge & Kegan Paul.

Productief B.V. (2003). *Tridio®: Spatial aptitude for (highly) gifted children from 3 to 5 years of age A.* Rotterdam, Netherlands: Author.

Productief B.V. (2005). *Tridio®: Niveau 3* [Tridio®: Level 3] (2nd ed.). Rotterdam, Netherlands: Author.

Productief B.V. (2006a). *Tridio®: B.* Rotterdam, Netherlands: Author.

Productief B.V. (2006b). *Tridio®: Handleiding ruimtelijk inzicht niveau 1* [Tridio®: Manual spatial ability level 1] (3rd ed.). Rotterdam, Netherlands: Author.

Productief B.V. (2006c). *Tridio®: Handleiding ruimtelijk inzicht niveau 2* [Tridio®: Manual spatial ability level 2] (3rd ed.). Rotterdam, Netherlands: Author.

Productief B.V. (2006d). *Tridio®: Werkboek ruimtelijk inzicht niveau 1* [Tridio®: Workbook spatial ability level 1] (3rd ed.). Rotterdam, Netherlands: Author.

Productief B.V. (2006e). *Tridio®: Werkboek ruimtelijk inzicht niveau 2* [Tridio®: Workbook spatial ability level 2] (3rd ed.). Rotterdam, Netherlands: Author.

Sanz de Acedo Lizarraga, M. L., & García Ganuza, J. M. (2003). Improvement of mental rotation in girls and boys. *Sex Roles, 49*(5/6), 277-286.

Shavalier, M. (2004). The effects of CAD-like software on the spatial ability of middle school students. *Journal of Educational Computing Research, 31*(1), 37-49.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701-703.

Snow, R. E., & Swanson, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology, 43,* 583-626.

Spiegel, H., & Spiegel, J. (2003). Potzklotz: Ein raumgeometrisches Spiel [Potzklotz: A spatial geometric game]. *Die Grundschulzeitschrift, 163,* 50-52.

Thurstone, L. L., & Thurstone, T. G. (1941). *The Chicago tests of primary mental abilities.* Washington, DC: The American Council on Education.

Tracy, D. M. (1987). Toys, spatial ability, and science and mathematics achievement: Are they related? *Sex Roles, 17*(3/4), 115-138.

Tracy, D. M. (1990). Toy-playing behavior, sex-role orientation, spatial ability, and science achievement. *Journal of Research in Science Teaching, 27*(7), 637-649.

Tukey, J. W. (1977). *Exploratory data analysis.* Reading, MA: Addison-Wesley.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599-601.

Van den Heuvel-Panhuizen, M., & Buys, K. (2005). *Young children learn measurement and geometry: A learning-teaching trajectory with intermediate attainment targets for the lower grades in primary school.* Utrecht, Netherlands: Freudenthal Institute, Utrecht University.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*(2), 250-270.

Vederhus, L., & Krekling, S. (1996). Sex differences in visual spatial ability in 9-year-old children. *Intelligence, 23,* 33-43.

Acknowledgements

First of all, I want to thank my supervisors Casper Hulshof and Jan-Maarten Luursema. Thank you both for your everlasting support, your feedback and your humor. I enjoyed working with you very much. Also, thanks to Fenna van Nes at the Freudenthal Institute. Thank you for suggesting the subject of the study in the first place, for your very useful advice on studying children, and for your feedback and continuing interest, even when the study was moving further and further from your research field. To the people at Productief B.V., many thanks for providing the materials for the Tridio test and training. Of course, I also want to thank the teachers, children, and parents of the pilot and research schools. Without your enthusiastic participation this study could not have been performed. Also, thanks to the children and parents participating in the pilot sessions, who helped me a lot in preparing the Tridio test and training. Furthermore, I want to thank Rinke Klein-Entink, Hans Vos and the others at the statistics department for their advice on the procedure of matching participants and on the data-analysis. Jeroen Blok, thank you for your profound interest and your help in constructing the tests' model items and in correcting the tests. And to Jan-Henk Annema, thank you for the helpful discussions and for your help in making the photographs included in this thesis. Finally, I want to thank my family and friends for all their support.

Appendix A. The Tridio® learning material

The Tridio® learning material consists of the following parts:

- Plastic cubes with 4.7 cm edges. The cubes have two black, two white, and two green sides. Opposite sides have the same color, while adjacent sides have different colors.
- Black, white and green plastic mosaic pieces. The mosaic pieces come in two shapes: rhombuses and triangles. All sides of the mosaic pieces are equal: 3.8 cm.
- A cardboard board with a 3 × 3 grid of squares on it, on which the cubes can be placed. The squares on the board have the same size as the cubes. In each corner of the board, there is an arrow pointing at the corner of the grid.

Tridio is sold in work boxes. Each work box contains 8 cubes, 20 rhombuses, 12 triangles, and a board, as shown in Figure A1.

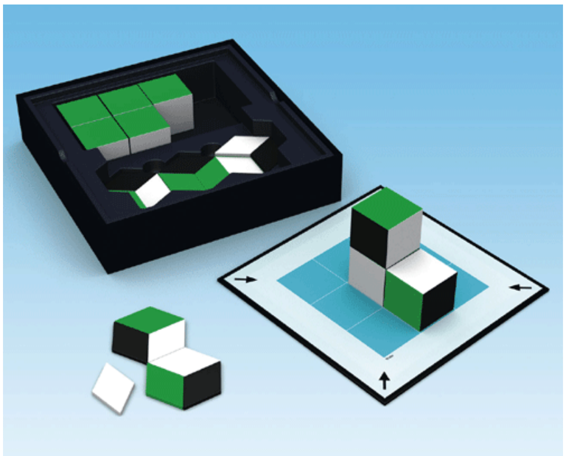


Figure A1. Tridio® work box containing cubes, mosaic pieces and a board.

Appendix B. Paper Folding practice items

Figure B1 displays the two practice items that were added to the instructions of the Paper Folding pretest. In the posttest instructions, the second practice item was replaced by the one shown in figure B2.

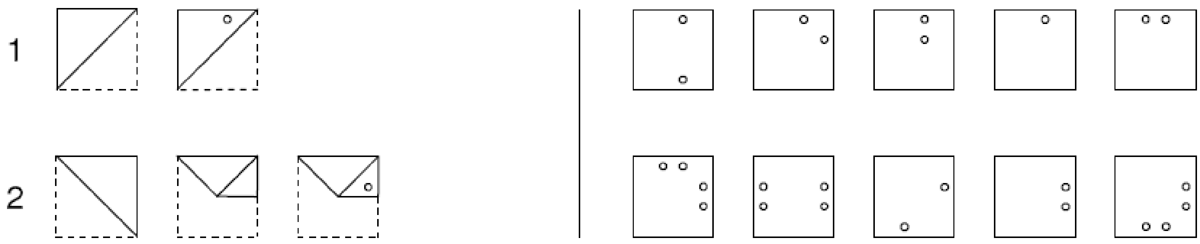


Figure B1. Practice items of the Paper Folding pretest. The correct answer of item 1 is the second option; of item 2 it is the fifth option.

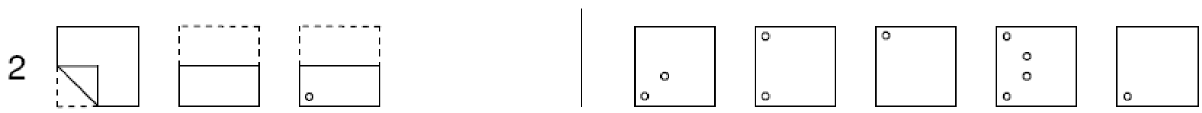


Figure B2. Second practice item of the Paper Folding posttest. The third option is the correct answer.

Appendix C. Exercise types in the Tridio® training

Table C1 lists all exercise types included in the Tridio training, in the order in which they appeared. Of the problem types that were not part of the Tridio test, a sample exercise is displayed in Figure C1. As is indicated in the table, some exercise types were only presented to children who had time left after completing the exercises of the other types. For none of the children there was time to work on exercises of type Q.

Exercise types K and L were not part of the original Tridio exercise sets. These exercise types were added as they were considered to be a good introduction to the exercises of types M, N, O, P, and Q.

Table C1. *Exercise Types Treated in the Tridio Training Sessions, in Order of Appearance.*

	Exercise type	Session(s)[a]
A	Copying a mosaic pattern shown in a picture.	1
B	Copying a picture of an isometric view of a cube building, using mosaic pieces.	1
C	Copying a cube building shown on a picture by placing cubes on a book page (Type 1 of the Tridio test).	1, 2
D	Placing cubes and mosaic pieces on a book page to create a cube building and its isometric view (Type 3 of the Tridio test).	1, 2
F	Copying a cube building shown on a picture by placing cubes on the board (Type 2 of the Tridio test).	1[b], 2, 3
G	Creating a mosaic pattern representing the isometric view of a cube building built on the board (Type 4 of the Tridio test).	1[b], 2, 3
H	First creating a mosaic pattern representing the isometric view of a cube building built on the board (like Type 4 tasks), and then modifying this pattern to represent the isometric view of the cube building that would be obtained when one of the cubes would be moved in a particular way.	1[b], 2[b], 3[b]
I	Completing a partial picture of a cube building's isometric view by placing mosaic pieces in it.	2[b], 3
J	Creating a mosaic pattern representing the isometric view of a cube building built on the board, as it would be seen when viewed from the right, left, or back.	2[b], 3[b]
K[c]	With square pieces, constructing three orthogonal projections of a built cube building: two side views and the top view.	4
L[c]	With square pieces, constructing three orthogonal projections of a cube building shown on a picture: two side views and the top view.	4, 5
M	Constructing a cube building that matches three pictured orthogonal projections: two side views and a top view (Type 5 of the Tridio test).	4, 5
N	Constructing a cube building that matches four orthogonal side view pictures.	4[b], 5
O	By placing mosaic pieces on a picture displaying outlines of every piece, constructing the isometric view of cube building that matches three pictured orthogonal projections: two side views and a top view.	5[b]
P	By placing mosaic pieces on a picture displaying the outline of the pattern to be made, constructing the isometric view of cube building that matches three pictured orthogonal projections: two side views and a top view.	5[b]
Q	By placing mosaic pieces on the table, constructing the isometric view of cube building that matches three pictured orthogonal projections: two side views and a top view.	5[bd]

[a]Number(s) of the session(s) in which the exercise types appeared. [b]Optional exercise type, only presented to students who had time left. [c]Not part of the original Tridio exercise sets. [d]No children had time left for this exercise type.

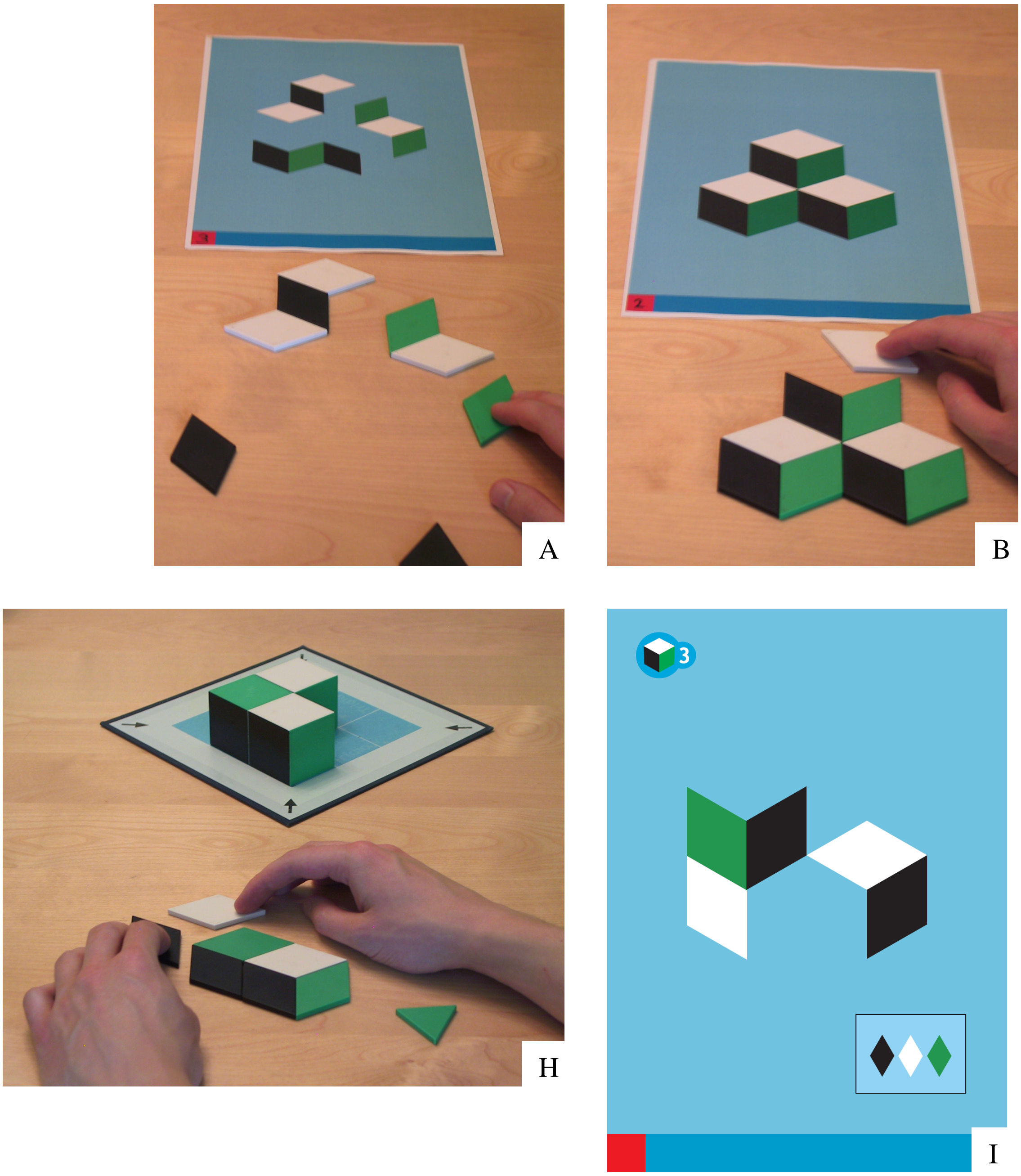


Figure C1. Sample exercises of the Tridio problem types treated in the training sessions, marked with the letters appearing in Table C1. In exercise type A, the task is to copy the mosaic pattern shown in the picture. In type B exercises, the child has to copy the pictured isometric view of a cube building, using mosaic pieces. Type H exercises require the student to first create a mosaic pattern corresponding to the isometric view of a cube building and then to modify this pattern to represent the isometric view of the cube building that would result when a cube would be moved to another place (in the pictured exercise, the child is told that the back cube will be moved to the left corner of the board). In type I exercises, the child is asked to complete the isometric view displayed on the book page, by placing on the picture the mosaic pieces pictured at the bottom right. On the top left, the number of cubes contained in the cube building of which the isometric view is to be created, is indicated (continued on next page).

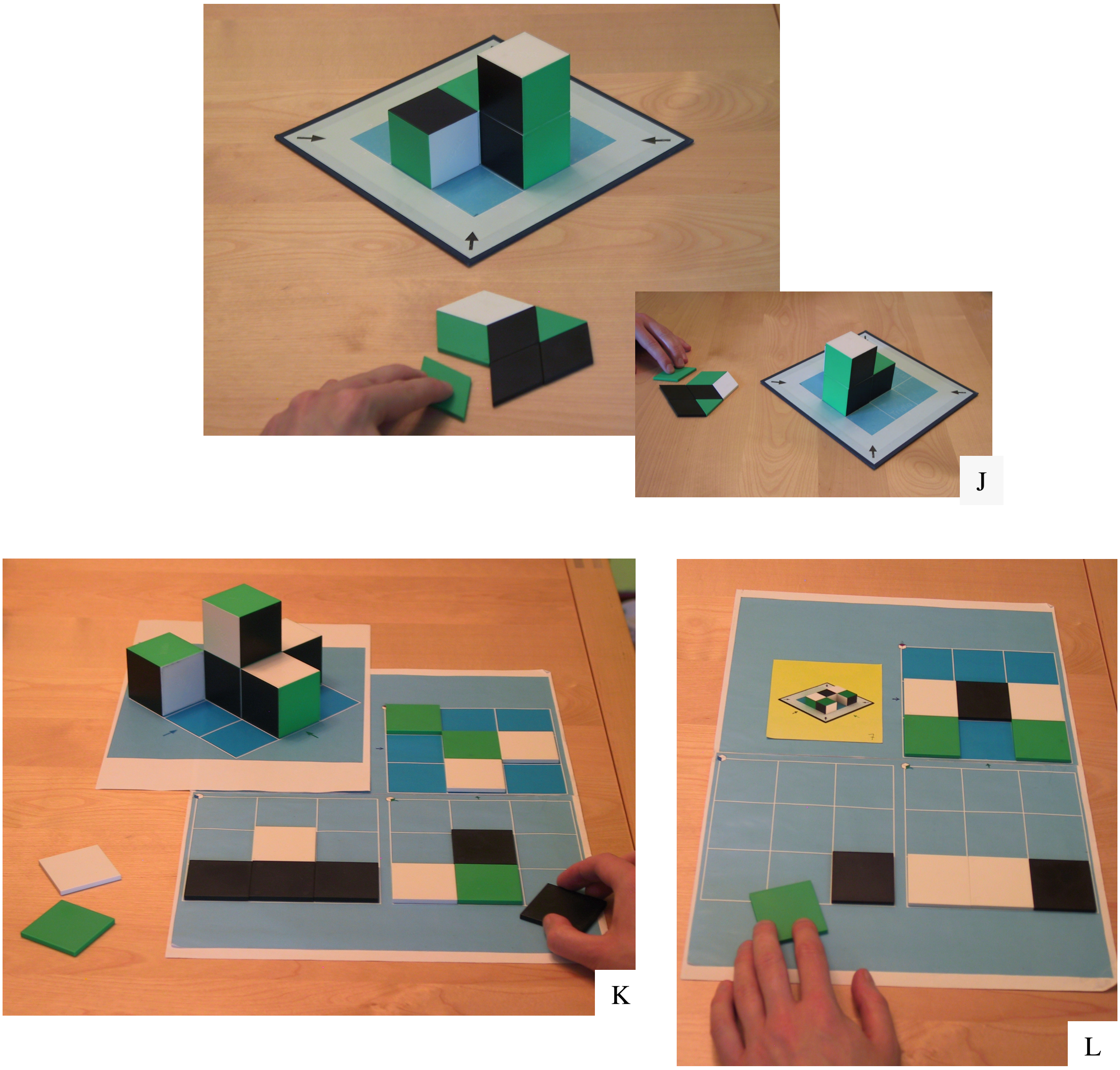


Figure C1 (continued). In exercise type J, the task is to create a mosaic pattern corresponding to the isometric view of the cube building as it would be seen when viewed from another side, in this case from the right. The leftmost picture shows the position of the student while making the exercise, the rightmost picture displays the cube building as viewed from the right side. In exercises of type K, the child has to create two orthogonal side views and an orthogonal top view of the presented cube building, using square pieces. The side views are laid out on the two 3×3 grids at the bottom, indicated with a blue and a green arrow, while the top grid is used for the top view. In type L exercises, two orthogonal side views and a top view are created corresponding to the cube building on the picture (continued on next page).

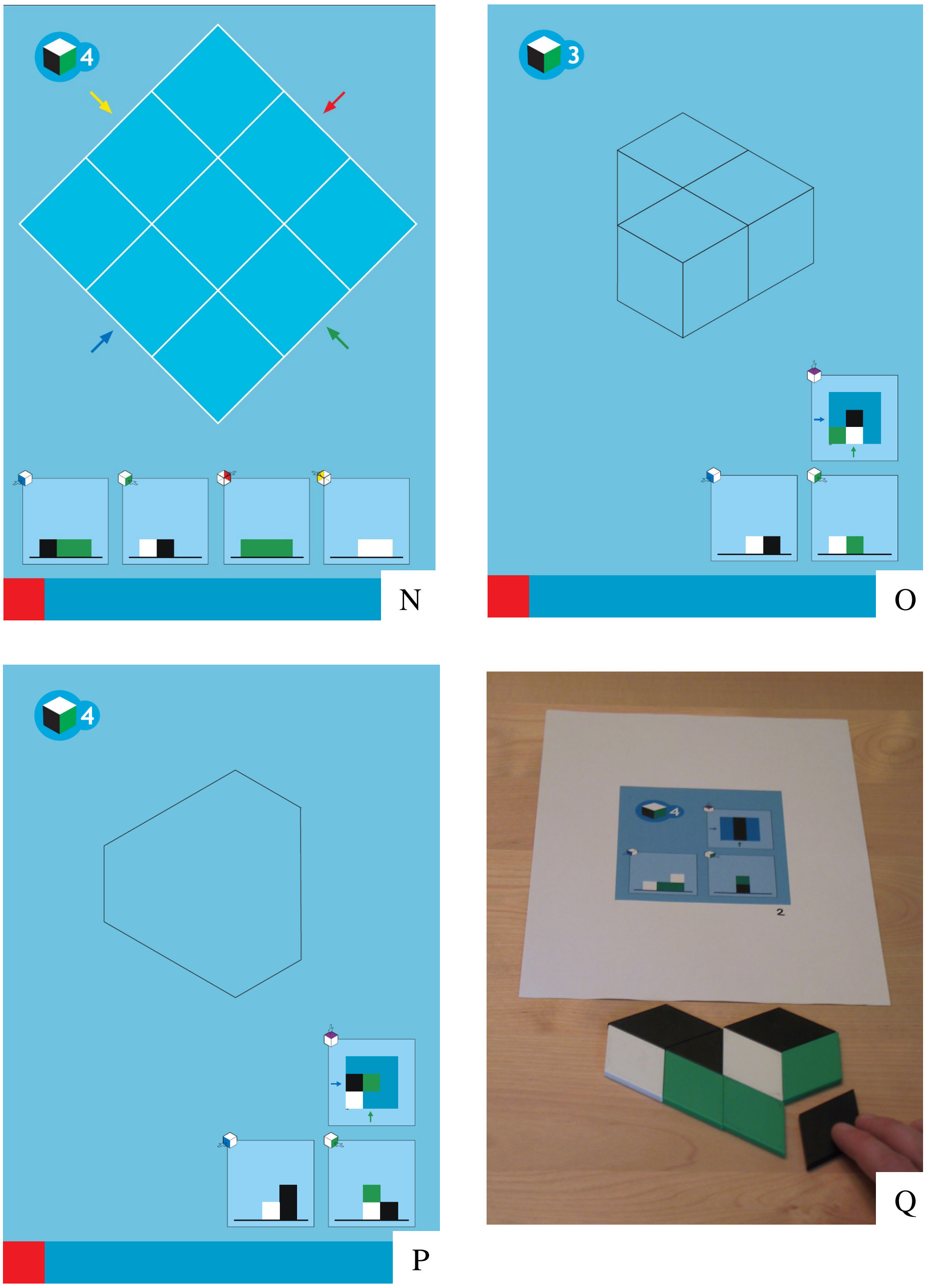


Figure C1 (continued). In type N exercises, the child is required to build a cube building on the board displayed on the book page, corresponding to the four side views. Coloured arrows indicate the positions of the side views. On the top left, the number of cubes to be used is indicated. In exercise type O, the task is to create an isometric picture of the cube building corresponding to the pictures of two orthogonal side views and a top view. The child has to place mosaic pieces within the piece outlines displayed on the page. The number of cubes contained in the cube building is indicated. The task of type P is the same, but in this case only the outline of the whole mosaic pattern to be created is given. In type Q exercises, the task is again the same, but now no outlines of the pattern are shown.