The Coloured Cube Test and The Coloured Mental Rotation Test. Two New Measures of Spatial Ability and Mental Rotation

Nikolay Lütke

Thesis submitted in partial fulfilment of the requirements of the degree of Doctor of Philosophy

London Metropolitan University

July 2020

Abstract

This research introduces two new measures of mental rotation (MR) for 4- to 11-year-old children. Instead of the complex achromatic three-dimensional (3D) cube aggregates used with adults (Shepard & Metzler, 1971), or the flat two-dimensional animals used with children (Quaiser-Pohl, 2003), the new tests uses 3D colourful cubes, either as a standalone, or as a cube aggregate but with fewer elements. The test format is similar to the Raven's Coloured Progressive Matrices Test (RCPM) which also served as a validation tool. The first new test, the Rotated Colour Cube Test (RCCT), consists of multi-coloured single cubes in different orientations. Three age groups of 7- to 10-year-old children (*N*=100) were increasingly successful in identifying cubes, with boys from socio-economic background that did not receive state benefits performing better in the more challenging test sections. While cubes that were different to the target in terms of cube face colour made the test easier, differently oriented cubes increased task difficulty. RCCT and RCPM were correlated, with the RCCT being the easier test. The second new test development, the Coloured Mental Rotation Test (CMRT), investigated differences in set-size, angularity, and axis of rotation of coloured cube aggregates in 4- to 11-year-old children (*N*=80). Several higher-order interactions all involved set-size and showed that 4-cube aggregates were the most economical and best 3D object for children's MR in all age groups. Interestingly, the linear decrease in performance with increasing angularity of 4-cube aggregates was already observed in 4-to 5-year but also still in 10- to 11-year-old boys, as well as in 6- to 7- and 8- to 9-year-old girls. It was concluded that the magical number 4, a capacity limit in attention and short-term memory (Cowan, 2001), can also be observed in MR, due to the Good Gestalt of the 4-cube aggregates.

Contents

Acknowledgments

First of all, I would like to express my gratitude Prof. Dr. Lange-Küttner for her invaluable supervision. Her generosity with her time, her inspirational standards of academic diligence and guidance are qualities that I aspire to. I must also thank Prof. Dr. Kubinger, who is professor for methods, diagnostics, and test development at the University of Vienna, for his insightful feedback during my test development. I am grateful to my family and friends who were always there for me when I needed them.

I must extend my greatest thanks, however, to my wife, Maria. I could not have completed this without her, her belief in me, her support, and her sacrifice. I dedicate this to her with all my love.

1. Theoretical Background

1.1. Geometric Complexity of Drawing Cubes

Cave paintings, which include rotated objects, and depth cues when drawing more distant objects that need to be depicted as reduced in size, are evolutionarily the most ancient purported activities of drawing pictures (Lange-Küttner & Green, 2007; Milbrath, 2005, 2009). Palaeolithic cave paintings are mainly two-dimensional without viewpoint perspective or object volume. It has been argued that modern technical drawings illustrate abstract conceptual knowledge and measurement (Piaget, 1969; Wilder & Green, 1963). Spatial accuracy in drawing predicts memory and learning in STEM subjects (Schwamborn, Mayer, Thillmann, Leopold, & Leutner, 2010).

Children are thought to develop the ability to draw in three dimensions (3D) on a two-dimensional (2D) surface in distinct stages (Luquet, 1927; Piaget & Inhelder, 1956). The first stage is *fortuitous realism* in which children produce scribbles without a recognisable object, these may represent an action such as a rabbit hopping across the page. The second stage is *failed realism* in which the child draws object features, but these are drawn unrelated to each other or extremely simplified. The third stage is *intellectual realism,* in which children define objects in terms of their build and function. With respect to the cube drawing, this can imply that all faces of a cube are drawn in a fold-out style as in diagrammatic drawing (Kosslyn, Heldmeyer, & Locklear, 1977; Mitchelmore, 1978; Morra, 2008). Only in the fourth stage of *visual realism* do children draw a projective image similar to a photograph.

Cube drawings have been extensively studied as cubes are simple enough geometric objects with surfaces in all three spatial dimensions (Cox & Perara, 1998). Perceptual and developmental research has explored children's ability to draw cube characteristics, focussing on geometric properties such as number of cube faces, orientation, occlusion and convergence in 3D perspective (Bremner & Batten, 1991; Bremner, Morse, Hughes, & Andreasen, 2000; Chen & Cook, 1984; Cox, 1986; Cox & Perara, 1998; Deregowski & Strang, 1986; Lange-Küttner & Ebersbach, 2012; Mitchelmore, 1978; V. Moore, 1986; Nicholls & Kennedy, 1995; Toomela, 1999, 2003). Research suggests that there are distinct stages in children's representations of cubes (Chen, Therkelsen, Griffiths, & Therkelson, 1984; Cox & Perara, 1998; Deregowski, 1977; Lange-Küttner & Ebersbach, 2012; Mitchelmore, 1978).

Mitchelmore (1978) compared drawings of cuboids, cylinders, pyramids, and cubes in 7- to 15-year-old children. Results showed that children's ability to draw a cube develops across a sequence of four stages. In *plane schematic* drawings, a single cube face is depicted in an orthogonal shape outline. *Solid schematic* drawings consist of several cube faces, including both visible and hidden features, often not drawn in the correct spatial relation to each other; they are also depicted without depth cues. *Prerealistic* drawings include depth information, seen from a single viewpoint, with visible cube faces shown in the correct position relative to each other. And finally, *realistic* drawings are those in which cubes are drawn with parallel or converging lines to represent edges. Cube drawings were found to provide the most consistent measure of children's drawing ability. Mitchelmore defined cubes as representing the 'purest' regular geometric object, as each edge and cube face are positioned in relation to a coordinate axis system .

Cox and Perara (1998) devised a nine-point scale for assessing cube drawings in 5- to 13-year-old children. Results of 489 children showed a linear trend of age progression through drawing stages, with distinct categories of drawing systems as shown as in Figure 1. In the *first* category, cube drawings consist of a single closed region which children are expected to demonstrate by the age of four to five years. The *second* category, at about six years of age, shows a cube represented by a square that conveys the “squareness” of the entire object as it includes four angular corners. The *third* category that typically shows at age seven is a multi-side configuration comprising impossible views of cube faces, either adjacent or enclosed within one another. Lange-Küttner and Ebersbach (2012) showed that often more than the six sides of the cube are displayed. The shapes may be depicted as rotated on the page. A *fourth* category which emerges at about age 8 consists of just two visible sides, with some spatial correspondence of either the front or side of the cube. The *fifth* category occurs first at about age 9 and shows three squares with some correct spatial correspondence between cube faces. The *sixth* category can emerge at about the same age will show a cube with a flat front, horizontal baseline. Obliques indicate that cube faces point into different spatial depth planes, the first sign of three dimensions on the two-dimensional surface of the page, but the sides are deformed, and the angles of cube sides often incorrect. The *seventh* category emerges at about age 11 and shows a modified baseline that recedes into depth. The *eighth* category at about the same age shows an oblique cube with geometrical precision that can be depicted in three dimensions from any point of view. The *ninth* category emerging at about age 12 shows a converging, visually realistic and optically correct depicted cube in alignment with the viewer, with receding edges converging to a one-, two- or three-point perspective.

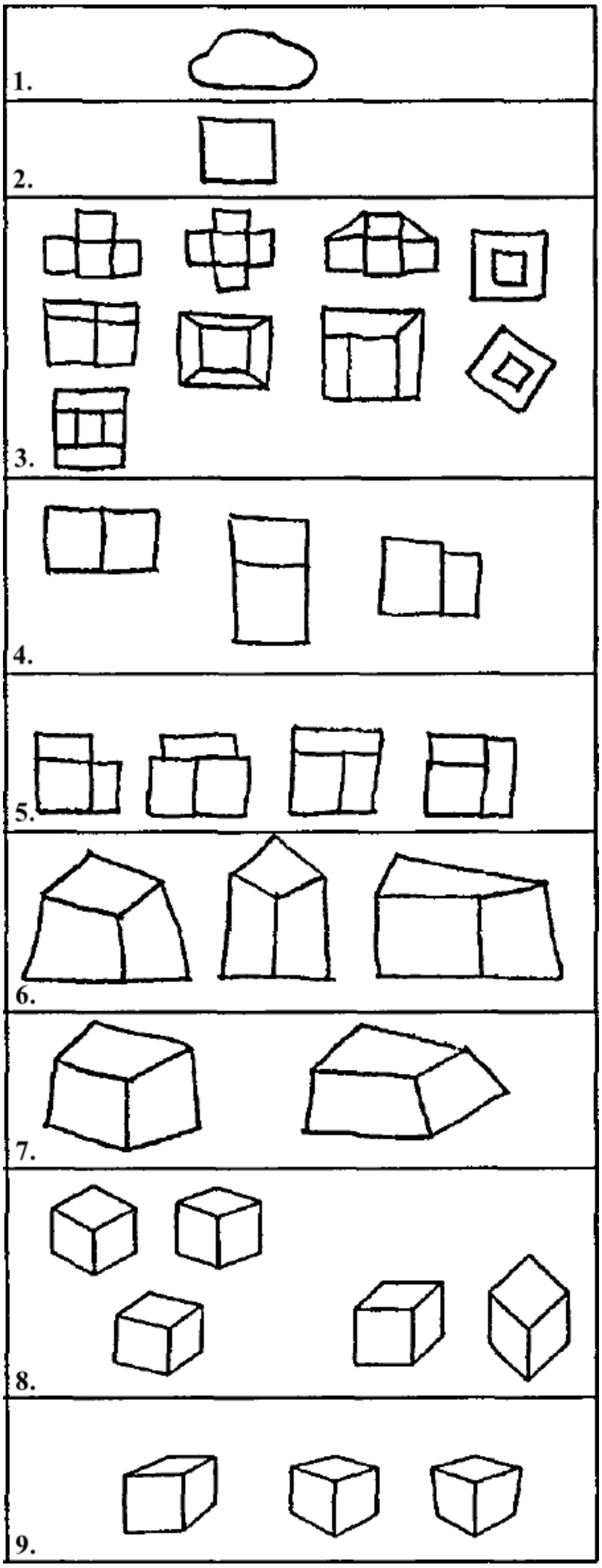


Figure 1 Stage model or cube drawings by Cox and Perara (1998).

Stage models of cube drawing suggest that children below the age of about nine years do not draw cube faces in different depth planes. The projective portrayal of depth information, achieved through perspective lines and occlusion, depends on multiple abilities, such as the working memory necessary to co-ordinate an increase in the number of objects parts (Morra, Moizo, & Scopesi, 1988), drawing differentiated contours and adjusting size to the pictorial spatial context (Lange-Küttner, 2008a, 2009; Lange-Küttner, Kerzmann, & Heckhausen, 2002), and drawing objects from different perspectives in different contexts.

Stages of cube drawing ability were introduced to help educators assess children's development. As drawing requires fine motor skills and technical knowledge, categorising children's ability to draw cubes will not only conceptualise the different types of spatial systems that they use. Identifying categorical differences in children's spatial and mental rotation ability, through the control of specific geometric attributes (e.g. orientation, rotation angle, axis of rotation and number of cubes), would most likely contribute to better understanding of the development of spatial conceptual knowledge.

In summary, cubes' geometric attributes are expected to be suitable for measuring children's ability to mentally transform objects in a three-dimensional space. Cubes' symmetrical features yield a predictable object insofar as it allows them to make inferences about occluded surfaces. Children can understand cubes in an object-specific way where all sides of the cubes are equal in dimensions. However, mentally rotating cubes remains challenging because the visible surface area will change in relation to the viewers' perspective and the object-specific symmetrical orthogonality of angles transforms into a viewer-specific projection.

2. Mental Rotation

2.1. Cubes in Mental Rotation: The Shepard and Metzler Test

Cube drawings were also an important aspect in Shepard and Metzler's (1971) pioneering work on mental rotation. Rather than just rotating a single cube, they asked participants to rotate alphabetic letters and cube aggregates. Mental rotation is the psychological process of spatially changing an object's orientation rapidly in the mind (Shepard & Metzler, 1971). This ability was a crucial landmark in the imagery debate in cognitive psychology (Bar, 2011; Kosslyn, 1994; Kosslyn, Ganis, & Thompson, 2006; Pylyshyn, 1977, 2003) as it demonstrated that people form mental images in their minds rather than just following a verbal command script. Participants were presented with pairs of perspective drawings of 3D cube aggregates and asked to identify whether the second image was either the same or a mirror image. They found a linear relationship between reaction time and the degree of rotation, akin to that found when physically rotating objects – a small rotation of an object takes less time than a large rotation. This suggested that mental image transformations correspond to transformations in the real world (Shepard & Cooper, 1982).

Shepard and Metzler (1971) found that for both depth and picture-plane rotations, reaction times increased in a linear function in relation to angular disparity. But the authors were surprised to find that larger depth rotations around the vertical axis did not show an angularity effect in the reaction times. Picture-plane rotations require the simple rigid rotation of the picture itself on a two-dimensional plane, whereas depth rotations require far more complex transformations of the object's orientation. Object complexity was also shown to have an effect on mental rotation ability, as symmetric similarities between endpoints increased reaction times.

1.2. The Mental Rotation Test

The follow-up experiments to Shepard and Metzler's experiment were so successful (Cooper & Shepard, 1973; Cooperau & Shepard, 1973; Metzler & Shepard, 1974; Tapley & Bryden, 1977), that Vandenberg and Kuse (1978) used these black-and-white two dimensional drawings of 3D cube aggregates to produce the Mental Rotation Test (MRT). While in Shepard and Metzler (1971) experiments, participants were required to make same vs. difference judgements of stimuli presented in pairs, in the MRT, participants were asked to identify two structurally identical, but differently rotated cube aggregates from a multiple-choice selection of four aggregates, see Figure 2. Thus, one could argue that in the Vandenberg and Kuse test, mental rotation was not really necessary, but that the main process that was tested was perceptual mapping between the target and the cube aggregate choices: The response format consisted of two targets and two distracters. Correct responses were only recorded if both targets had been identified. The test contains 20 items, in five sets of four items. Response times increased with larger angularity of the rotated cube aggregate.

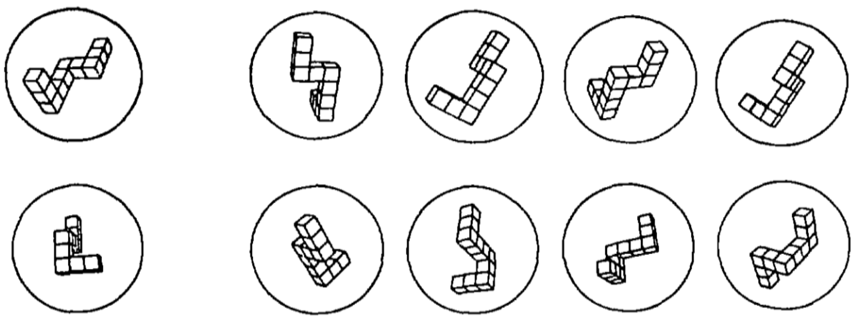


Figure 2 Vandenberg and Kuse's Mental Rotation Test. Correct answers in the top row are items in position 2 and 5, and in the bottom row items 2 and 3.

This test format has become widely applied in research with adults (Geiser, Lehmann, & Eid, 2008; Peters, 2005; Peters et al., 1995; Peters, Manning, & Reimers, 2007; Voyer & Saunders, 2004). One of the reasons for the large and continuously growing body of research into mental rotation are sex differences, in which males typically outperform females (e.g. Lauer, Yhang, & Lourenco, 2019; Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995).

1.3. Spatial Ability and Mental Rotation

Spatial ability is considered to be a central component of human intelligence. It was included as spatial visualisation in Thurstone's (1938) primary mental abilities model, and in Carroll's (1993) model as a broad visual perception factor. Nevertheless, in Carroll's view, spatial ability is not a unitary process but can be divided into a number of distinct forms which include the ability to perceive, analyse, store, and recall visual representations. In their meta-analytic review, Linn and Petersen (1985) also defined spatial ability as a skill in representing, transforming, generating, and recalling symbolic and non-linguistic information. Similar to the meta-analytic review by Voyer et al. (1995), they distinguished between *spatial perception, spatial visualisation* and *mental rotation*. In *spatial perception*, participants are required to process spatial relationships in relation to the orientation of their own bodies. Corballis and Roldan (1975) suggested that in order to solve such a task, participants use processes of symmetry detection to rotate stimuli and achieve visual or gravitational alignment. *Spatial visualisation* is commonly associated with complex manipulations of spatial information which require a sequence of individual steps, which may include *spatial perception* and *mental rotation*, but distinguishes itself because of multiple solution strategies. An example of such tests is the verbal reasoning section of the Differential Aptitudes Test (DAT-V, The Psychological Corporation, 1995). In the most recent meta-analytic review by Lauer et al. (2019) mental rotation was defined as the ability to visualise the rigid transformation and rotation of objects, or object parts, and includes measures that require participants to complete objects through mental translation or rotation of shapes, discriminate between rotated mirrored images, and to solve spatial analogies by visualising rotated objects. This definition of mental rotation is far broader than the one originally proposed by Shepard and Metzler (1971), which focussed on the time it takes to determine whether two simultaneously represented cube aggregates are the same or different. The definition of mental rotation has been widened to include new experimental paradigms, and, most importantly, to include the variety of different solution strategies that have been identified when processing MRT items. These include mental rotation of holistic images, piecemeal rotation, perspective taking, and feature-based, viewpoint-independent strategies (Hegarty, 2018) and hence suggest that a number of different abilities may be applied when solving mental rotation tasks. Lauer et al. (2019) included the results of Study 1 (Lütke & Lange-Küttner, 2015) in their meta-analysis in the categories ‘abstract stimulus type’ and ‘three dimensional properties’, as two categories central to their review of mental rotation studies.

In the following paragraphs, the large body of research inspired by Shepard and Metzler’s original findings will be reviewed, covering the developmental factors that influence mental rotation.

1.4. Developmental Differences

Piaget and Inhelder (1956, 1971) had already acknowledged the role of imagery in developmental psychology. They proposed that children would not be able to demonstrate dynamic imagery before reaching the concrete-operational stage at about age seven. However, studies that followed found that young children could mentally rotate (Kosslyn, Margolis, Barrett, Goldknopf, & Daly, 1990), albeit at a much slower rate than adults (Frick, Hansen, & Newcombe, 2013; Frick & Möhring, 2013; Krüger, Kaiser, Mahler, Bartels, & Krist, 2013; Marmor, 1975, 1977; Schwarzer, Freitag, Buckel, & Lofruthe, 2012).

Piaget and Inhelder thought that children in the preoperational stage until about age seven would only use static imagery. In particular, young children would not understand how changing an object's orientation also changes its features in a coordinated way (dynamic imagery) (Piaget & Inhelder, 1971, p. 120). However, subsequent research demonstrated that this widely accepted assumption was incorrect, and underestimated young children's ability to process rotated objects. The ability to mentally rotate develops already in infancy (e.g. Moore & Johnson, 2008, 2011; Quinn & Liben, 2008), during early childhood and continues to do so into adolescence (e.g.Kail, 1986; Kosslyn et al., 1990).

Marmor (1975) found in a reaction time and accuracy study where children rotated mirror-images of teddy bears around their own axis from 30° to 60° to 120° to 150°, that error rates increased with larger angularity for both 5-year-old and 8-year-old children, whereas only for the older age group, reaction times also increased. In a follow up study using geometric stimuli, Marmor (1977) confirmed a linear increase in reaction times in 4- to 5-year-old children, which suggests that for young children, geometric stimuli may be easier to process than teddy bears which have more irregular contours.

In a forced choice paradigm study of rotated 2D shapes, 4- and 5-year-old's accuracy decreased with the angle of rotation, but 3-year-old's accuracy did not (Frick, Hansen, et al., 2013). However, in a different study, 3-year-olds demonstrated the ability to rotate 2D objects although at very slow speeds of 2500ms, increasing up to 3000ms for larger angles (Krüger et al., 2013). In a Tetris game with dynamic 2D rotated shapes, error rates of 4- and 5-year-olds did not suggest mental rotation ability, but 5-year-olds' response times increased with greater rotational angularity from 2200ms to 3200ms (Frick, Ferrara, & Newcombe, 2013). Interestingly, these developmental reaction time studies demonstrated a transition from static to dynamic imagery already in terms of speed and not just accuracy in very young children albeit at a much younger age, and increasingly so, rather than in one abrupt stage after the onset of operational thought as Piaget assumed.

Moreover, Schwarzer et al. (2012) used the original Shepard and Metzler cube aggregates in a dynamic video film with 9-months-old infants. Importantly for the assumption that mental rotation mirrors motoric real-life object rotation, the results suggested an active motor component insofar as the more mobile crawlers looked longer at the mirror image of cube aggregates in a habituation task than static infants who could only sit. When using simpler letter stimuli, the motor component was found to be important in a linear fashion in 8- to 10-month-olds, with walkers being more likely to distinguish impossible letter rotation outcomes compared to crawlers, belly crawlers, and sitting infants, respectively (Frick & Möhring, 2013). If manual exploration was permitted, even 6-month-olds showed an increase in looking time for impossible rotations in a habituation experiment (Möhring & Frick, 2013). This motoric component was also found in adults with low scores on the Vandenberg and Kuse (1978) test (redrawn version of Peters et al. 1995) as they would gesture more in their explanations regarding differences in the structure of a wooden 3D model of Shepard and Metzler's cube aggregates (Göksun, Goldin-Meadow, Newcombe, & Shipley, 2013).

1.5. Developmental Differences on the MRT

Hoyek, Collet, Fargier, and Guillot (2012) explored whether the MRT (Vandenberg & Kuse, 1978), in which stimuli consist of 10-cube aggregates, was suitable for 7- to 8- and 11- to 12-year-old children, and compared results to nonsense letters. While performance in rotation of letters was three times greater in the older age group, the increase in performance on the MRT did not even double, yet both mental rotation scores showed a significant correlation of r = .42 in 11- to 12-year-olds. Hoyek et al. suggested that Vandenberg and Kuse's cube aggregates were too difficult for school children because of the number of orthogonal turns in the intrinsic spatial axis. They also suggested that the MRT was too difficult for the youngest age group because of (1) the difficulty in encoding abstract geometric stimuli, (2) complex test instructions, and (3) stringent time constraints.

Neuburger, Jansen, Heil, and Quaiser-Pohl (2011) investigated the impact of varying stimulus types in 6- to 9-year-olds and 8- to 12-year-olds, using 2D animal pictures, letters, and 3D cube aggregates. They found that animal pictures were the easiest and cube aggregates the most difficult to complete, supporting the theory that encoding of abstract geometric objects may be too difficult for young children. However, Titze, Jansen, and Heil (2010a) suggested that 8-year-old children were able to successfully complete the MRT with cube aggregates if they had previously been introduced to a simpler picture mental rotation exercise with 2D animals. They also found a sex effect in favour of males for 10-year-olds, but not for younger children.

Besides these recent publications, numerous other studies have investigated the development of mental rotation in children using different stimuli and paradigms, including: images of pandas (Marmor, 1975), letters (Jansen, Schmelter, Kasten, & Heil, 2011), 2D images of humans and animals (Quaiser-Pohl, 2003), sex stereotyped 3D objects (Kaltner & Jansen, 2018), tangible cube aggregates (Bruce & Hawes, 2014), one coloured cube (Lütke & Lange-Küttner, 2015), machine and animal toys (Hirai, Muramatsu, & Nakamura, 2018) and images of cartoon monkeys (Wimmer, Robinson, & Doherty, 2017). While these studies have made valuable contributions to our understanding of the development of mental rotation, the number of studies that used items similar to those of Shepard and Metzler (1971) are still limited (Geiser, Lehmann, Corth, & Eid, 2008; Hoyek et al., 2012; Quaiser-Pohl, Geiser, & Lehmann, 2006; Titze et al., 2010a), especially with children under 6 years of age. The second test development, the Coloured Mental Rotation Test (CMRT) tries to address the outlined issues by using simplified, coloured three-dimensional cube aggregates, in which critical variables (angel of rotation, set-size, axis of rotation and object dimensionality) have been systematically varied, in order to explore impact factors sensitive to the development of spatial ability and mental rotation ability in young children.

1.6. Individual Differences

Sex differences in mental rotation have been widely reported (e.g. Alexander & Evardone, 2008; Astur, Tropp, Sava, Constable, & Markus, 2004; Birenbaum, Kelly, & Levi-Keren, 1994; Butler et al., 2006; Collins & Kimura, 1997; Linn & Petersen, 1985; Maccoby & Jacklin, 1974; Peters, Lehmann, Takahira, Takeuchi, & Jordan, 2006; Peters et al., 2007). In a meta-analysis by Voyer, Voyer and Bryden (1995), men outperformed women on the MRT by nearly one standard deviation. However, no sex differences were found in 7 out of 15 chronometric studies, with an overall small to medium effect size ($d = .37$) which is not exactly convincing evidence but rather indicates that the male advantage is down to chance, or due to differences in experimental paradigm. Follow-up studies with children and adults have tried to identify the reasons for the appearance of sex differences in the MRT.

In a recent systematic meta-analysis (Lauer et al., 2019) on sex differences in the development of mental rotation, numerous task factors of the rotated test items were controlled: 2- vs. 3-dimensionality, mirror vs. non-mirror images, abstract vs concrete shapes, and animate vs. inanimate shapes. They also controlled for other performance factors such as computerized vs. paper presentation, group vs. individual test settings and time limits vs. unlimited time. The main result was an increase with age in the male advantage from a small effect size of .20 at 6 years to a large effect size of .50 at age 14 years. Another important result was that while dimensionality and mirror image were both important factors, when age and gender were added to the regression analysis, only cube dimensionality remained a significant factor in the male advantage (Lauer et al., 2019, p. 546). Also time constraints were not a reason for boys' advantage in mental rotation (see also Heil & Jansen, 2008), although it is often found that adult men decide in a more timely fashion (Glück & Fabrizii, 2010). The Lauer et al. (2019) review of the mental rotation literature emphasised several critical task factors supporting the design parameters for both new tests developments introduced in this thesis.

Why may boys have an advantage when spatially transforming 3D cube aggregates? When children drew two overlapping cubes, during the transition period towards drawing in perspective, boys more often depicted the projective edges of the cubes, while girls were more likely to unfold the cube to display all its six sides (Lange-Küttner & Ebersbach, 2012). The authors argued that boys focused more on the projective appearance of the cubes, while girls were more interested in its design principles. This finding was further supported as mental rotation predicted the ability of girls to draw cubes in 3D volume, whereas for boys the best predictor was the embedded figure test in which participants needed to find a shape's edges embedded in visual noise. Thus, a gender-specific bias towards appearance versus identity (Flavell, Green, Flavell, Watson, & Campione, 1986) may shift boy's attention towards projective edges of cubes which would also support the ability to mentally rotate this object. Kozhevnikov, Kosslyn, and Shephard (2005) termed these two different styles as 'object visualizers' who were more common in females, and 'spatial visualizers' more common in males. In a spatial training program using ambiguous 2D and 3D line drawings, visual attention was directed towards edges, resulting in improved performance for both sexes, highlighting the importance of edge perception which may also impact mental rotation performance.

The age at which sex differences may emerge is still unclear (Hoyek et al., 2012; Neuburger et al., 2011; Quinn & Liben, 2008) and depends on specific task demands. Research suggests that sex differences on the MRT are reduced or eliminated through practice on computer games (Okagaki & Frensch, 1994), through sports activities (Blüchel, Lehmann, Kellner, & Jansen, 2012; Quaiser-Pohl & Lehmann, 2010), by lifting time constraints (Goldstein, Haldane, & Mitchell, 1990; Peters, 2005; Voyer, 2011) and with extensive item-specific practice (Kail, 1986; Kass, Ahlers, & Dugger, 1998) as well as 2D-3D dimensional transformation training (Moreau, 2012; Tzuriel & Egozi, 2007, 2010). Voyer (1995) identified only one study in their meta-analysis which found sex differences in children younger than 10, whereas Linn and Petersen (1985) review did not include any studies with children below the age of 10 years. However, subsequent studies have found sex differences in younger children. Heil and Jansen (2008) found sex differences in 7- to 8-year-olds on a mental rotation task in favour of boys only in regard to accuracy but not speed measures.

These studies suggest that many factors influence the magnitude of sex differences in mental rotation performance and that changes in experimental paradigms will change what a test measures. It is therefore important to identify the main factors that affect mental rotation, enabling researchers to adapt their experimental paradigms and produce more consistent and meaningful results.

Research has shown that socioeconomic status (SES) is associated with a broad array of physiological, cognitive, and socioemotional outcomes in children, with influences already present prior to birth and extending into adulthood (Bradle & Corwyn, 2002). Studies have reported effects of SES on disparities in brain structure, cognitive skills and academic outcomes (Baydar, Brooks-Gunn, & Furstenberg, 1993; Hackman & Farah, 2009; Hackman, Farah, & Meaney, 2010; Neville et al., 2013). Children from a low SES are 1.3 times more likely than children from non-poor backgrounds to experience developmental delays and learning difficulties (Brooks-Gunn & Duncan, 1997). Poorer children also scored between 6- to 13-points lower on standardised IQ, verbal ability, and achievement tests (Smith, Brooks-Gunn, & Klebanaov, 1997), with this poverty effect already present in 3- to 8-year-old children. Studies have shown a significant association between higher childhood SES and higher levels of cognitive functioning in later life (Beck et al., 2018; Richards & Wadsworth, 2004; Singh-Manoux, Richards, & Marmot, 2005; Zhang, Liu, Li, & Xu, 2018), highlighting the importance of developing cognitive tests for young children and early educational intervention plans.

Children with lower SES were shown to be disadvantaged in comparison to their middle class counterparts, falling behind on very early measures of cognitive development such as the Bayley Infant Behaviour Scales (Farah, 2010) and on school readiness tests (Brooks-Gunn & Duncan, 1997). Research by Mezzacappa (2004) indicated that socially disadvantaged 5- to 7-year-old children performed worse on measures of executive attention when trying to process competing demands. Levine, Vasilyeva, Lourenco, Newcombe, and Huttenlocher (2005) found that socioeconomic status (SES), especially in boys, influenced spatial cognition and the development of visuo-spatial memory. Their research examined spatial ability across second- and third-grade children from different SES groups. They found that boys from both middle- and high-SES groups performed better on an aerial map and a 2D mental rotation task than girls from the same SES groups. However, no such sex differences were found between lower-SES groups in both tasks.

Surprisingly and to the best of my knowledge, no further research has investigated the link between SES and spatial ability or mental rotation. Investigating the impact of SES on a child's test performance, using 3D objects similar to the MRT (Vandenberg & Kuse, 1978), will provide new insights into how environmental factors influence spatial ability. It was therefore important to control for SES and fluid intelligence, as measured by eligibility for state financed (free) school meals and by the Raven's Coloured Progressive Matrices Test, while interpreting individual differences in the performance on the new test developments.

1.7. Task Characteristics and Complexity

Test items used in children's mental rotation tasks vary widely. For instance, images of pandas (Marmor, 1975) and monkeys (Wimmer et al., 2017), letters (Jansen et al., 2011), 2D images of humans and animals (Quaiser-Pohl, 2003), sex stereotyped 3D objects (Kaltner & Jansen, 2018), real tangible 3D cube aggregates (Bruce & Hawes, 2014), as well as machine and animal toys (Hirai et al., 2018). The original cube aggregates have also been used in developmental studies, but mostly with children older than 10-years of age (e.g. Geiser, Lehmann, Corth, et al., 2008; Lauer et al., 2019; Quaiser-Pohl et al., 2006; Titze et al., 2010a). When 'adult' cube aggregates were used for mental rotation with 7-year-olds, reliability was reduced to .56 (Carr, Steiner, Kyser, & Biddlecomb, 2008) as most of the children at this age perform at floor level and below chance (Hawes, LeFevre, Xu, & Bruce, 2015). Hawes et al. found significant correlations, between the rotation of 3D cube aggregate and 2D animal shapes, $r = .33$, as well as letters, $r = .38$ at age seven, which indicates that while the hit rate was very low, it was not completely at random.

A widely used approach for studying children's mental rotation ability is the reduction of cognitive load through the simplification of stimuli complexity.

For instance, in a forced-choice paradigm study, outline drawings of human figures provided children with an apparently more socially suitable test item compared to the classic more complex cube aggregates (Estes, 1998). However, when mental rotation performance was measured in terms of increases in reaction time along with angular discrepancy, only 6-year-olds performed akin to adults, while 4-year-olds did not. Moreover, when a hand was used as a mental rotation stimulus this produced an increase in reaction times with increases in rotation from about 3000ms to 4500ms in 5- to 6-year-old children who did not yet attend school, and from about 2000ms to 3500ms in 7-year-old first graders (Krüger & Krist, 2009).

Nevertheless, bodies instead of cube aggregates support women's mental rotation ability (Alexander & Evardone, 2008). Interestingly, a hybrid between cube aggregates and human heads, hands and feet also lowered the cognitive load for adults in comparison to the classic cube aggregates, but only when the body parts were orderly attached and not when they were randomly fixed onto the ends of the aggregates (Krüger, Amorim, & Ebersbach, 2014).

Quaiser-Pohl, Neuburger, Heil, Jansen, and Schmelter (2014) found that measuring mental rotation ability with cube aggregates and time limits was too difficult for second graders (6- to 9-year-olds) but not for fourth graders (8- to 12-year-olds). The aim was to keep the 3D cubes similar to the original stimuli of Shepard and Metzler (1971), but to test children in the multiple choice test format used by Vandenberg and Kuse (1978).

The Rotated Colour Cube Test (RCCT) was designed to depict a single multi-coloured three-dimensional cube and thus simplified the complexity of the geometric cube aggregates (Vandenberg & Kuse, 1978), but not their three-dimensional volume. Similar facilitations were effective in the Three-Mountains-Task that measures the ability to form spatial perspectives when three overlapping mountains (Piaget & Inhelder, 1956) were reduced to a single clearly visible mountain (Liben & Belknap, 1981).

Differences in the characteristics of stimuli, such as using animate objects instead of cubes (Alexander & Evardone, 2008; Neuburger et al., 2011; Rosser, Ensing, & Mazzeo, 1985) facilitated mental rotation in women and children. This indicates that a key difficulty may be related to stimulus identification and encoding (Bialystok, 1989). Research suggests that the mental rotation of an object's encoded image (Jolicoeur, 1988; Moreau, 2012) may be either matched with a more abstract, structural representation (Hyde, 1981), or directly compared with the nearest and most similar stored view (Hedges & Nowell, 1995). Hence, in the initial two sections of the first new test development of the RCCT, the perceptual matching of model and target was assessed as a baseline ability for mental rotation; only thereafter were the model and target cube differently rotated.

A further facilitating factor in mental rotation performance is colour information (Alington, Leaf, & Monaghan, 1992). It was argued that since colour is one of the fundamental properties of an object, it might be perceived pre-attentively similar to other primary properties, such as brightness and line orientation in visual search tasks (Enns & Rensink, 1991; Treisman, 1986) and therefore may provide less able participants with an additional 'processing channel'. Children are especially sensitive to colour signals in early stages of retinal perception, whereas size and orientation features are processed in later processing stages (Donnelly et al., 2007). In order to make cubes and cube aggregates easier to process for children, colour has been added in form of individual cube faces or uniformly coloured cubes within an aggregate in both new test developments (RCCT & CMRT).

Metzler and Shepard (1974) believed that matching aggregate arms was especially difficult, and hence added dots of colour over end points to help participants distinguish starting from end points, which led to a reduction in the difference between reaction times between picture- and depth-plane rotations. Jordan, Wüstenberg, Heinze, Peters, and Jäncke (2002) used aggregates with alternating black and white cubes and found that these were still more difficult than letters and abstract line drawings. Khooshabeh and Hegarty (2010) investigated how colour would influence performance on the MRT, by colouring three cubes within an aggregate. They found that participants with good rotation ability did not benefit from colour, whereas poor rotators benefitted as it helped them to identify individual pieces of the shape in rotation. This implies that colour can indeed facilitate mental rotation, and hence might reduce task complexity sufficiently for young children to successfully mentally rotate cube aggregates in the new Coloured Mental Rotation Test (CMRT). Seven year-old children can represent multiple colours of a cube in the correct spatial location (V. Moore, 1986). Hence, as in the coloured version of the Raven's Standard Progressive Matrices Test for children, coloured cubes and cube aggregates are used in both new versions of the test (RCCT and CMRT). Further, in the RCCT colour incongruency between targets and distracters was gradually reduced, which resulted in increases of task difficulty across test section. The rationale behind this approach was similar to that of a visual search task where increased colour similarities result in a reduction of feature uniqueness between target and distracters (Gerhardstein & Rovee-Collier, 2002; Treisman, 1988; Treisman & Gelade, 1980).

Simplifying task demands by adding colour will also influence what strategy children use to solve mental rotation tasks. Studies have shown that participants use a variety of different strategies to solve mental rotation tasks (Bethell-Fox & Shepard, 1988; Folk & Luce, 1987; Just & Carpenter, 1985; Yuille & Steiger, 1982) and include holistic (global-shape), piecemeal (counting cubes), feature-based (focussing on changes in direction in an aggregate), perspective taking and viewpoint independent strategies (Hegarty, 2018). The global shape strategy was the only strategy positively correlated with accuracy. Yuille and Steiger (1982) suggested that people would spontaneously simplify if additional cubes were added to the original 10-cube aggregates, but below this threshold they would be more likely to use all features.

The debate on the nature of images, the depictive account (Kosslyn, 1994) opposed to the propositional (Pylyshyn, 2002), lead Kozhevnikov et al. (2005) to demonstrate that low-spatial visualisers outperform high-spatial visualisers on tasks requiring a focus on detailed visual properties of stimuli, whereas for mental rotation the opposite was true. This suggests that a holistic approach to mental rotation is beneficial. Hegarty and Kozhevnikov (1999) distinguished between visualisers that consistently used either object-based pictorial or spatial schematic representations while solving mathematical problems, and that object visualisers encode and process images holistically, whereas spatial visualisers process images analytically, in a fragmented fashion (Kozhevnikov et al., 2005).

Stimulus complexity has been shown to impact mental rotation ability with research showing conflicting results. Mental rotation studies with polygons where task difficulty was varied through the number of vertices (Cooper, 1975; Cooper & Podgorny, 1976) revealed no increase in task complexity for objects with more vertices, which suggests the use of a holistic strategy. However, other research found a complexity effect using polygons (Folk & Luce, 1987) or three dimensional cube aggregates (Bethell-Fox & Shepard, 1988; Yuille & Steiger, 1982), supporting a piecemeal strategy. These conflicting results suggest, that while an objects complexity influences mental rotation task difficulty, further variables need to be considered. Krüger et al. (2014) showed that adding further information to cube aggregates such as human body parts would simplify task complexity through spatial embodiment and the projection of their own body's axis onto similar stimuli. Aretz and Wickens (1992) showed that participants switched from holistic to piecemeal strategies when increasing complexity in maps.

A further factor influencing task difficulty is the axis of rotation. Participants performed better at picture-plane compared to vertical rotations (Neuburger, Heuser, Jansen, & Quaiser-Pohl, 2012), and also better at vertical compared to horizontal rotations (Battista & Peters, 2010; Waszak, Drewing, & Mausfeld, 2005). It has also been suggested that depth rotations produce a larger sex difference where boys outperform girls (Neuburger et al., 2011).

Stimulus dimensionality has also been shown to influence mental rotation performance. Aggregates that were distributed in one depth plane (flat) were rotated more quickly than aggregates with features protruding into depth (Bauer & Jolicoeur, 1996). Aggregates were either distributed in one depth-plane (flat on the picture-plane) or had features protruding into depth, see Figure 3. This is an important adaptation in the current context of modifying cube aggregates to children's abilities because the flat 3D aggregates are similar to the 2D animal pictures. The angularity effect was clearly present for both types of aggregates. An advantageous effect of flat aggregates emerged in rotations larger than 90 degrees. They argued that this demonstrated how dimensionality contributes to stimulus complexity, as additional depth information needs to be maintained and encoded when comparing structural properties. Similarly, Metzler and Shepard (1974) argued that 3-D cube aggregates that undergo occlusion or crossings may contribute to task difficulty, which would also apply to depth axis rotations.

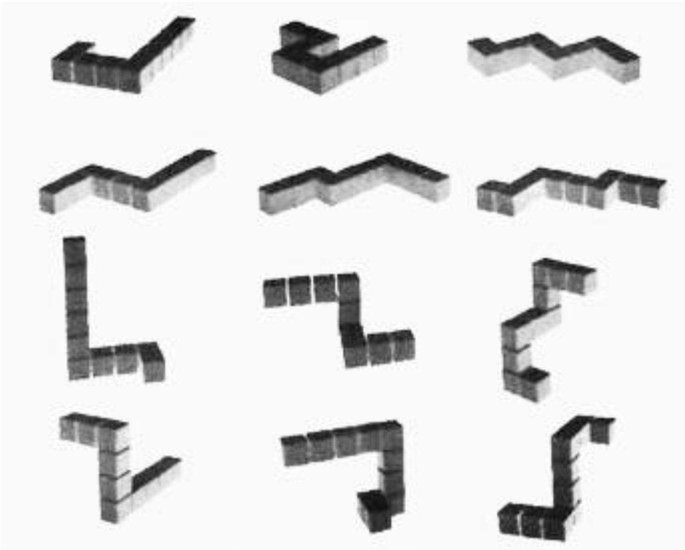


Figure 3 Bauer and Joliceur's (1996, p. 85) experiment on the aggregate depth effect. Flat aggregates in the first upper row are equally sized to those in the third row, and those in the second row are the same as those with 3D depth protrusions in the last row.

In order to measure the effect of dimensionality on children's mental rotation performance, dimensionality was systematically varied in the second test development (CMRT), by repeating each set-size of aggregates that had either flat or protruding elements.

Can the orientation of an object affect task difficulty in spatial tasks? The orientation of an object can be defined as the angle from which an object is viewed at a particular point in time, with objects in a canonical (familiar) or a non-canonical (unfamiliar) orientation (Palmer, Rosch, & Chase, 1981). Recognition and categorisation of canonical objects is faster than those in other orientations (Edelman & Bülthoff, 1992; Jolicoeur, 1985; Tarr, 1995). In a mental rotation study with 18- to 30-year-old participants, Francuz (2014) found that three-dimensional objects presented in a canonical orientation were indeed easier than objects presented in a non-canonical orientation. He argues that since the purpose of the MRT is primarily to determine if two differently rotated images of an object represent the same object, reaction times and accuracy scores should be independent of whether a test stimulus is presented in a canonical or non-canonical perspective. This is a rebuttal to the suggestion of the position of Bülthoff, Edelman, and Tarr (1995) and Cutzu and Edelman (1998) that three-dimensional objects are stored as a collection of simplified snapshots taken from different views, and that angularity judgements could be assessed in terms of the similarity between two dimensional images of a three dimensional object. Following this explanation, the MRT could be defined as a categorisation mechanism rather than rotating an image in mind. Francuz (2014) results showed how an object's initial orientation impacts on mental rotation ability in adults. As there are no studies with children that investigate how task difficulty might be affected by the orientation of a three-dimensional test stimulus, this was controlled for in the first new test development (Study 1) by comparing cubes in canonical perspective with cubes balanced on a corner.

1.8. Training Studies and Dimensionality

Kail (Kail, 1986; Kail & Park, 1990) indicated that with extensive practice children were able to reach adult levels of performance on mental rotation tasks, but the training effect was limited to item-specific features of just one object, with no transfer of mental rotation skills to other objects. This suggested that children stored unique view-specific images of an object without developing an abstract ability to rotate and thus were not able to generalize this ability across other stimuli. It is likely that the mental rotation task can be solved with the storage of visual snapshots, similar to visual priming in children (Lange-Küttner, 2010b; Stupica & Cassidy, 2014). Recent studies on practice and training of mental rotation have focused on dimensionality of the object, in particular on 2D-3D task difficulty (Tzuriel

& Egozi, 2007, 2010) as the visual information processing system finds 2D stimuli easier to process than 3D (Rosser, 1980) and because the degree of rotation is less influential in two dimensions (Bauer & Jolicoeur, 1996; Jolicoeur, Regehr, Smith, & Smith, 1985). Hoyek et al. (2012) investigated whether both 2D letters and 3D cubes were appropriate for the use with children between seven and twelve years. They found no correlation between 2D and 3D test scores on dimensionally different mental rotation tasks in 7- to 8-year-olds, which supports the notion of dimension-specific processing.

Systematic comparisons between 2D letter-like stimuli and the classic 3D cube aggregates confirmed that 7- to 10-year-old children found it easier to process 2D rather than 3D stimuli (Jansen, Schmelter, Quaiser-Pohl, Neuburger, & Heil, 2013), and this difference appeared to increase with age (Hoyek et al., 2012). However, first, 2D stimuli are not accurate representations of real objects. Second, children favour 3D pictures, become progressively more interested in depth depiction and develop their ability to represent three dimensions in their own graphic constructions (Kosslyn, Heldmeyer, & Locklear, 1980; Lange-Küttner, 1994a, 2004, 2009). Girls preferred to unfold cube faces and drew large amounts of surface detail that might have distorted the overall view of the cube, whereas boys favoured keeping the cube’s visual appearance intact (Lange-Küttner & Ebersbach, 2012). Children of kindergarten age are already able to estimate the volume of 3D cubes (Ebersbach, 2009). It would thus be both appropriate and beneficial to measure young children’s mental rotation ability in a test with three-dimensional cube images and aggregates.

What is the difficulty when processing three-dimensional stimuli? Two-dimensional perception requires processing stimuli only within a single plane based on straightforward object similarity judgments, whereas three-dimensional perception requires more complex spatial inferences about visually incomplete, hidden-from-view information, where object features must be interpolated. Superior 2D-3D "dimensionality crossing" (spatial transformations) was identified in males who outperformed women on most mental rotation tasks (Voyer et al., 1995), with corroborative evidence suggesting that occluded parts of cube aggregates were more difficult to process especially for women (Voyer & Hou, 2006). However, training in 2D-3D spatial transformations successfully improved girls' performance (Tzuriel & Egozi, 2007, 2010). In adults, 2D training led only to improvements in 2D tasks, whereas 3D training led to improvements in both 2D and 3D tasks (Moreau, 2012). This clearly demonstrates the specificity of dimensionality in the mental rotation task and how the use of more realistic 3D depictions of objects can be beneficial in terms of the general transfer of mental rotation skills across a wider variety of objects.

A common approach to adapting Shepard and Metzler's (1971) complex three dimensional cube aggregates for use with school children is through a reduction in dimensionality and by changing item characteristics, for instance, using either 2D pictures of humans and animals (Quaiser-Pohl, 2003), or letters (Kosslyn et al., 1990), see Figure 4.

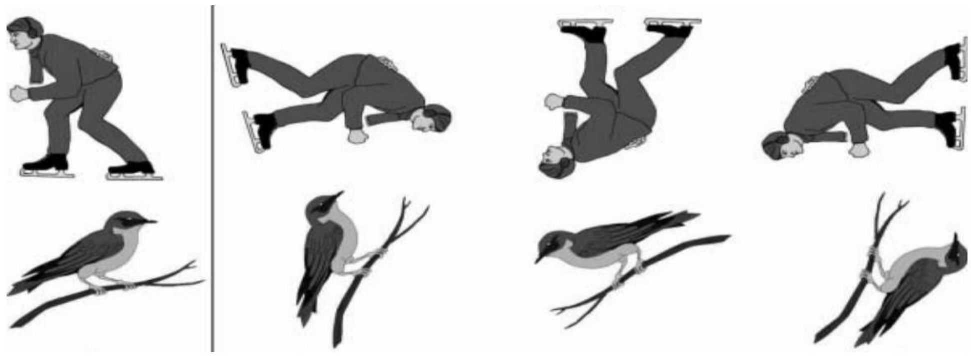


Figure 4 Quaiser-Pohl's (2003) Picture rotations test.

To summarise, children's ability to draw two-dimensional images of three-dimensional cubes has been extensively studied (Kosslyn et al., 1977; Luquet, 1927; Piaget & Inhelder, 1956). Cubes are simple enough and familiar geometric objects with surfaces that extend into all three spatial dimensions (Cox & Perara, 1998) and hence are well suited to measure the development of spatial ability and mental rotation. The most widely used measure of spatial ability for adults, the Mental Rotation Test (Vandenberg & Kuse, 1978), uses complex cube aggregates which might be too difficult to process for young children (Hoyek et al., 2012). This lead to the development of new simplified, mostly 2D tests for children (Blüchel et al., 2012; Bruce & Hawes, 2014; Hawes, Moss, Caswell, & Poliszczuk, 2015; Iachini, Ruggiero, Bartolo, Rapuano, & Ruotolo, 2019; Jansen et al., 2013; Marmor, 1975; Neuburger et al., 2011; Perrucci, Agnoli, & Albiero, 2008; Quaiser-Pohl, 2003). Reducing objects to two dimensions might have been successful in simplifying test item complexity and overall task demands for young children, but a systematic analysis of whether sacrificing the third dimension was necessary to achieve this aim is still necessary. Similar to the stage approach of children's ability to draw three dimensional cubes, their spatial ability to visually process and manipulate cube images needs to be broken down into its defining three-dimensional geometric characteristics. In other words, researching how a cube's orientation, colour, and differences in angularity and axis of rotation influence task difficulty will contribute to a better understanding of children's development of spatial ability in relation to the fundamental features of three-dimensional objects.

The first study investigated whether the omission of the third dimension is necessary and offers a new test where complexity was reduced, but without resorting to images of two-dimensional objects. The Rotated Colour Cube Test (RCCT) was designed to reduce task complexity without sacrificing its three-dimensional properties. In the first version of the test, the Shepard and Metzler's classic cube aggregates were simplified to one single coloured 3D cube, and in the second version of the test, the number of cubes was gradually and systematically increased.

2. Study 1: Single Coloured Cubes

2.1. Test Development

In the first study, the aim was to develop a test which drastically simplified the Vandenberg and Kuse cube aggregates. Because it was shown that cube aggregates were too difficult compared to single animals depicted in two dimensions until age 10, and the intention was to test children as young as six years, the three-dimensional properties of the original aggregates were maintained while reducing the aggregate size to one cube.

In the pilot test development, a diagrammatic fold-out cube that showed each coloured side of the cube, also those not visible, was added to each test slide, as children might have wondered about the hidden sides of the cube, see Figure 5. However, in the pilot, it turned out that children were unsure what the fold-out cube should represent. The interpretation at the time was that they did not see the connection between the 2D fold-out and the 3D cube images, confirming earlier research (Kosslyn et al., 1977; Mitchelmore, 1978; Morra, 2008) that not all children would draw diagrammatic cubes. Amongst the questions they had about the fold-out cube in the upper right corner, only one child enquired about hidden cube faces, in Figure 5, these are the green, the orange and the violet cube faces. Thus, the conclusion was that the fold-out depiction of the cube did not help children to find the correct answer.

The factor that might have prevented children from intuitively interpreting the ‘cross’ as a fold-out cube would have been that the hidden colours were not present in any of the distractor cubes. However, using only three coloured cube-faces in RCCT and asking children to only perform picture-plane rotations, while systematically increasing colour congruence was identified as an appropriate reduction in task complexity compared to the original 3D MRT. As the new RCCT test development was conceptualised as a non-verbal measure of spatial ability similar to the Ravens Coloured Progressive Matricies test (RCPM), and since the additional fold-out cube lead to children asking the examiner for additional instructions, this test version was not used in Study 1.

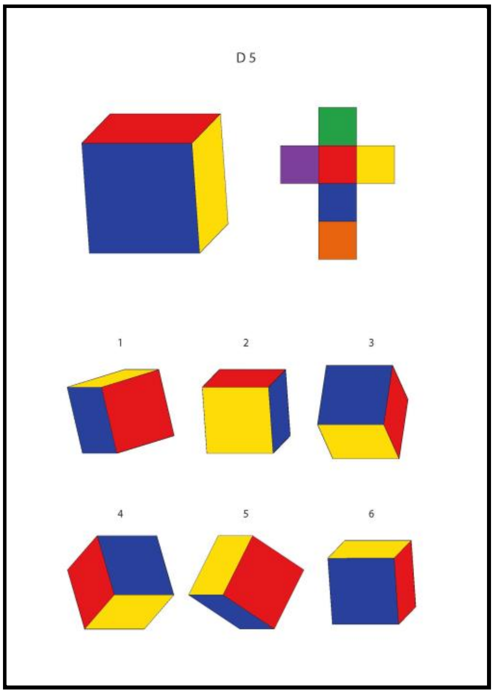


Figure 5 Test page with a later omitted diagrammatic cube depiction.

Thus, another version was developed which consisted of identical cube images to those used in the pilot version of the RCCT, but in which colour congruence between distractors and target cubes was still not systematically controlled (Lütke, 2009). A sample of 52 children between 6- to 8-years were tested. A univariate one-way analysis with the (RCCT) as dependent variable by 3 (Age) by 2 (Sex) as between-subjects factors revealed a significant difference between age groups, $F(2, 46) = 7.046$, $p = .002$, $\eta^2 = .23$. Post-hoc pairwise comparisons revealed that 5- to 6-year-olds ($M = 63.5\%$) performed marginally worse than 7-year-olds (*M* = 73.9%) and significantly worse than 8- to 10-year-olds (*M* = 80.4%). There was no significant difference between the two oldest age groups. This showed that only 5- to 6-year-olds performed significantly worse than older age groups.

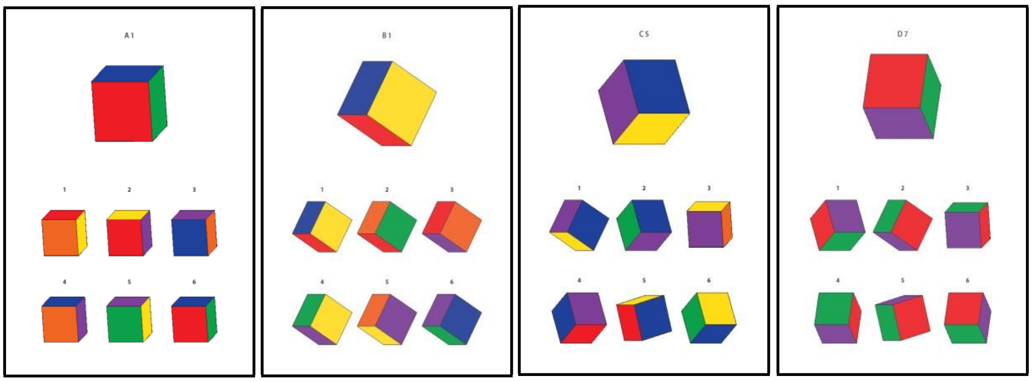
A two-way 2 (Test) by 3 (Age) by 2 (Sex) analysis of variance revealed a significant difference between the RCCT and Raven's Coloured progressive Matrices test, $F(1, 46) = 6.169$, $p < .05$, $\eta^2 = .017$. The RCCT (*M* = 72.6%) was easier than RCPM (*M* = 65.7%).

While these results were promising, the new test was still further developed. It was designed to increase in task difficulty more systematically and allowed several test parameters such as (1) number of distractors, (2) orientation of cube, (3) variation of the number of differently rotated distractors, (4) colour congruence, and (5) differences between orientation of target and model cube, to be controlled, see Appendix Table A1 for further detail.

In the new test, see Figure 6, firstly, task difficulty was gradually increased by varying the rotation of the model cube from an upright canonical position to balancing it on one of its corners without rotating the target and distracters. Thereafter, differences in rotation and colour between the model, target and distracters were gradually introduced. The initial two test sections (A & B), investigated whether young children could successfully encode and appropriately respond to test items by perceptual matching the identical 3D cube images in the identical rotational angle. Only thereafter were more complex spatial manipulations required in sections C and D.

Secondly, 7- to 10-year-old children participated as research suggests that differences in mental rotation, specific to this test format, emerge during this age range (Geiser, Lehmann, Corth, et al., 2008; Geiser, Lehmann, & Eid, 2008; Johnson & Meade, 1987; Titze, Jansen, & Heil, 2010b; Vederhus & Krekling, 1996). As both time limits (Glück & Fabrizii, 2010; Voyer et al.,

1995) and response format (Glück & Fabrizii, 2010) influence mental rotation performance and can lead to children solving questions quickly rather than accurately, time restrictions were removed. A response configuration was adapted similar to that of the RCPM, which requires participants to identify one target from six to eight distracters.



(A) (B) (C) (D)

Figure 6 **Examples of Task Sheets.** ***RCCT A***: differently coloured cubes identical in orientation (the correct test cube is on lower row furthest to the right). ***RCCT B:*** differently coloured cubes identical in orientation, but in a non-canonical view (correct cube is in the upper row, furthest to the left). ***RCCT C***: rotational variance between distracter cubes (the correct cube is in the upper row, furthest to the left). ***RCCT D:*** –rotational variance between distracter cubes, but all cubes have the same colours (the correct cube is in the lower row in the middle).

Thirdly, children also completed the RCPM as it is a standardised test used to measure non-verbal reasoning. For children, the RCPM is one of the purest measures of fluid intelligence. The RCPM first appeared in 1947 as a variation form the Raven's Standard Progressive Matrices Test, specifically created for testing 5- to 10-year-old children. In a meta-analysis by Vijver (1997) of cross-cultural intelligence test scores, the RCPM was the second most used test after the Wechsler Intelligence Scales for children. The aim was to control children's fluid intelligence, but also to compare the two non-verbal tests with each other for shared variance.

The general assumption is that there are no sex differences in the RCPM scores. This inference was first made by Raven (1939, p. 30) who noted that in the standardization sample, there were no sex difference between boys and girls up to the age of 14 years, both in the mean and the variance of scores. Eysenck (1981, p. 41) also noted equal scores between the sexes for children and adults. Jensen (1998, p. 541) concluded that there was no consistent discrepancy between female and male scores in the Raven's Standard or Coloured Progressive Matrices Tests. Also, Court's (1983) literature review which summarized 118 studies confirmed that no sex differences in performance were found. However, contrary to these findings, a more recent meta-analysis by Lynn and Irwing (2004), found that boys performed significantly better than girls on the RCPM ($d = 0.21$). As one of the most widely used measures of intelligence in children, the RCPM provides an important objective measurement in comparison to other mental rotation tests.

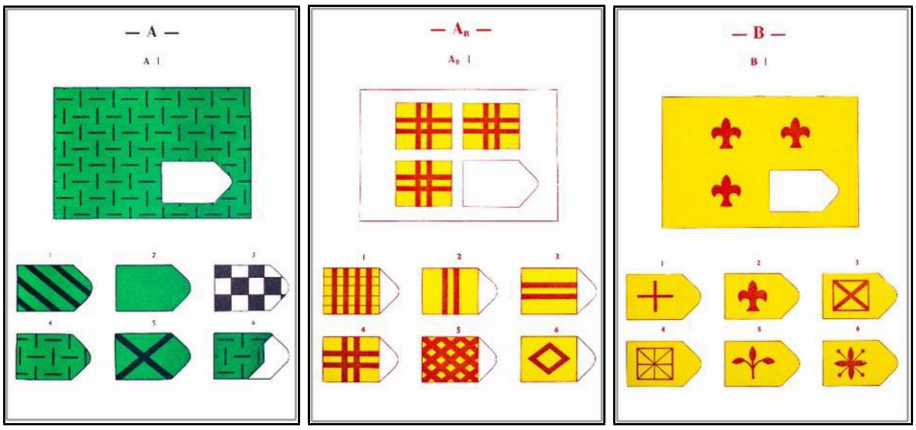
The RCPM has been identified to measure four distinct ability factors: (1) simple continuous pattern completions, (2) discrete pattern completion, (3) continuity and reconstruction of simple and complex structures, and (4) reasoning by analogy (Corman & Budoff, 1974). Lehmann, Quaiser-Pohl, and Jansen (2014) also found a strong correlation between the RCPM test and 2D mental rotation, $r = .55$, $p < .001$, and working memory, $r = .57$, $p < .001$. As mental rotation tasks can also be solved through analytic and feature-based strategies (Hegarty, 2018), the RCCT scores are expected to correlate with those of the RCPM.

SES was controlled in the current sample of children by considering whether the London council, or their parents, paid for their school meals and included as a factor in the statistical analyses. Boys from a lower SES background were expected to perform worse than boys from families who were able to pay for their school meals.

The hypotheses for Study 1 predict that (1) especially in the youngest age group, children would perform better on cubes with a more familiar canonical orientation than on cubes balanced on a corner, similar to children's developmental progression through stages when drawing cubes (Cox & Perara, 1998; Luquet, 1927); (2) Colour congruence between targets and distractors would increase task difficulty, similar to visual search tasks, where targets with a singular defining feature such as colour (e.g. a green T among blue T distractors), pop-out, but shared colours make visual search more difficult and result in a time-consuming serial search. (Treisman & Gelade, 1980) and (3) as in perceptual matching tasks, identically rotated distractor cubes would be easier to discriminate than identifying target cubes amongst differently rotated distractors; (4) children will perform better on test items with fewer response options, particularly on more difficult questions; (5) children from a more financially stable SES background would perform better on the RCCT.

2.2. Raven's Coloured Progressive Matrices Test

The Raven Coloured Progressive Matrices Test (RCPM) is a non-verbal spatial reasoning tests that uses two-dimensional patterns and shapes, with a similar response format as the RCCT. The RCPM is designed to measure the ability to reason by analogy and to form perceptual relations akin to Spearman's g. The RCPM is made up of a sequence of three sections (i.e. A, AB, B); see the examples in Figure 7. Pattern fragments require integration into a larger systematic context (Raven & Court, 1998).



(A) **(AB)** **(B)**

Figure 7 Examples of Test Sheets. *A*: Identification of a patch in a continuous pattern (correct item is on lower row furthest to the left). *AB:* Identification of a patch of a discrete pattern (correct item is lower row furthest to the left). *B:* Identification of a patch in a continuous pattern with discrete items (correct item is upper row in the middle).

The RCPM consists of 36 such individual matching-to-sample tests. Each page depicts a task that offers a context with a fragment left blank in the bottom right corner of the pattern. A 2 by 3 matrix of fragments below shows one target and five distracter fragments. Participants are required to find the fragment from this set of 6 alternatives that best completes the pattern.

2.3. Method

2.3.1. Participants

Participants (N = 100) were 51 boys and 49 girls from a school in West London. Children come from a wide variety of different ethnic backgrounds. Parental consent was obtained, and children were informed that they were free to withdraw from the study at any time they wished. There were n = 47 children on state financed school meals, and n = 53 children on parent financed school meals, with a total sample of *N* = 100 children, age means are listed in Table 1. Age labels are defined by the mean age (years; months) as 7-year-olds (*M* = 7;5), 8-year-olds (*M* = 8;5) and 9-year-olds (*M* = 9;8) throughout the manuscript, sometimes overriding means of age rantes in subgroups. This variable was used as a between-subjects variable of socio-economic status (SES) in addition to sex.

Table 1

Participant Numbers, Mean and Standard Deviation of Age Groups

	State Financed School Meals (SFM)								
	Boys				***Girl***				**Total**
Age groups	n	*M*	*Min*	*Max*	n	*M*	*Min*	*Max*	
7-years	8	7;5	6;10	7;11	9	7;5	7;0	7;10	17
8-years	11	8;2	8;1	8;11	5	8;4	8;0	8;11	16
9-years	6	9;9	9;0	10;2	8	9;5	9;1	10;1	14
Total	25				22				47
	Parent-Financed School Meals (PSM)								
7-years	7	7;5	7;3	7;11	5	7;5	7;2	7;11	13
8-years	8	8;5	8;3	8;10	13	8;5	8;0	8;10	21
9-years	11	10;0	9;2	10;7	9	9;8	9;0	10;3	20
Total	26				27				53
Total Sex	51				49				100

Note. Years; months

In order to test whether fluid intelligence was dependent on socio-economic status, t-tests were conducted on RCPM scores, see Table 2. Only boys in the two older age groups differed significantly in fluid intelligence, but not the girls, with significantly better scores in the parent-financed school meal groups for the 8-year-old and the 9-year-old group. The analysis of the mental rotations test was therefore controlled by sex, age and whether school meals were state financed or paid for by parents (SES). In a second analysis, the RCPM scores were included as a covariate in order to establish the shared variance between fluid intelligence and mental rotation. In a third analysis, the correlations between RCCT and RCPM were analysed in order to validate the new test on mental rotation.

Table 2

RCPM scores by Socio-Economic Status, Age and Sex in per cent

Sex	State school meal		Parent school meal			
	M	*SD*	*M*	*SD*	*t*	*p*
7-years	61.44	15.21	70.83	14.96	-1.649	.111
Girls	62.04	14.16	77.78	16.55	-1.881	.084
Boys	60.76	17.28	65.87	12.60	-0.645	.530
8-years	**58.68**	**15.28**	**73.81**	**13.74**	**-3.162**	**.003****
Girls	60.00	15.91	70.94	14.37	-1.407	.178
Boys	**58.08**	**15.74**	**78.47**	**12.04**	**-3.062**	**.007****
9-years	**71.03**	**13.02**	**81.39**	**13.09**	**-2.275**	**.030***
Girls	70.49	14.08	74.38	12.71	-0.600	.558
Boys	**71.76**	**12.72**	**87.12**	**10.78**	**-2.641**	**.019***

Note. * $p < .05$. ** $p < .01$. Statistical effects set in bold were significant.

2.3.2. Apparatus and Material

Rotated Colour Cube Test (RCCT). The RCCT is a non-verbal task with 3D images of coloured cubes, which were digitally produced with Adobe Illustrator. Three cube faces are visible at all times with each face showing one of six distinct colours (i.e. yellow, orange, red, green, blue, and purple). In the following paragraphs, the rationale for including different types of rotations and increasing levels of colour congruence is explained.

The RCCT is composed of 36 pages, which are split into of four equally sized sections A, B, C and D. Task difficulty was gradually increased through using differently rotated cubes and through the colour similarity of cube faces. The prediction was that increases in similarity between target cube and distracter cubes should add to task difficulty because this produces a loss of perceptual discriminability as in visual search tasks (Gerhardstein & Rovee-Collier, 2002; Treisman, 1988; Treisman & Gelade, 1980).

Similarly, cube drawing studies have shown that children progress through a number of distinct stages (Cox & Perara, 1998; Luquet, 1927) and that variations in the orientation between of depictions of three dimensional test items increase task complexity (Ruthsatz, Neuburger, Rahe, Jansen, & Quaiser-Pohl, 2017).

Table 3 *Properties of the Test, Target, and Distractors Cubes*

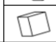
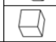
Level	Page	Cubes	Section	Rotation
			RCCT A	
I	**A1**	6	100 % distracters with 1 colour identical to test cube	
	A2	8		
II	**A3**	6	100 % distracters with 2 colours identical to test cube	
	A4	8		
III	**A5**	6	50% distracters with 3 colours identical to test cube	
	A6	8		
IV	**A7**	6	100 % distracters with 3 colours identical to test cube	
	A8	8		
			RCCT B	
I	**B1**	6	100 % distracters with 1 colour identical to test cube	
	B2	8		
II	**B3**	6	100 % distracters with 2 colours identical to test cube	
	B4	8		
III	**B5**	6	50% distracters with 3 colours identical to test cube	
	B6	8		
IV	**B7**	6	100 % distracters with 3 colours identical to test cube	
	B8	8		

Note. Cubes = Number of distractors and target cubes per page; Rotation = cube perspective for test, target, and distractor ubes

In **RCCT A**, target and distracter cubes are displayed in the canonical perspective, as if standing on a flat surface. In **RCCT B**, target and distracter cubes are displayed standing in a familiar and in a physically impossible position on one corner. Besides using differently rotated cube perspectives, both sections A and B have identical subsection levels (I-IV) which increase in complexity by systematically increasing distractor colour congruency. This allowed to investigate how the cubes' orientation might affect task difficulty while keeping other factors such as colour congruence and distractors number variation constant, see Table 3.

In **RCCT C and D**, the number of differently rotated Distractors per test page is increased across subsection levels (I-IV). Test and Target cubes are initially presented in an identical orientation (Levels I-II), and thereafter in different orientations (Levels III-IV). While the types of rotated cube perspectives per test page are identical in both section C and D, colour congruency is increased in Levels I-IV only in section C. In section D only three colours in total are used per test page, see Table 4.

Table 4 *Properties of the Test, Target, and Distractors Cubes*

Level	Page	Cubes	Section	Rotation
			RCCT C	
I	**C1**	6	100 % distracters with 1 colour identical to test cube	
	C2	8		
II	**C3**	6	100 % distracters with 2 colours identical to test cube	
	C4	8		
III*	**C5**	6	100% distracters with 2 colours identical to test cube	
	C6	8		
IV*	**C7**	6	50 % distracters with 3 colours identical to test cube	
	C8	8		
			RCCT D	
I	**D1**	6	100% distracters with 3 colours identical to test cube	
	D2	8		
II	**D3**	6	100% distracters with 3 colours identical to test cube	
	D4	8		
III*	**D5**	6	100% distracters with 3 colours identical to test cube	
	D6	8		
IV*	**D7**	6	100 % distracters with 3 colours identical to test cube	
	D8	8		

Note. Cubes = Number of Distractors and Target Cubes per page; Rotation = Cube perspective for Test, Target, and Distractor Cubes; * Indicates that the orientation of the Test and Target cubes differ.

Sections A and B are identical in terms of the variance of colour congruence across subtests (Levels I-V), attributing any changes in task difficulty between the two section to the variation of the two cube orientations. Similarly, sections C and D are identical in the terms of the number of differently rotated cubes used, attributing any variation in task difficulty between the two sections to changes in colour congruency.

Each section had four trials with two levels of distracter numbers, that is, first six, then eight distracters. Trials gradually increased in difficulty through colour congruency by having distracters with one, two or three colours in common with the target cube. Only section D had distracters with all three colours identical with the target.

Gradually increasing task difficulty over the four RCCT Test sections (A-D) provided the framework for testing object identification and object rotation. The initial two sections (RCCT A-B) measure perceptual matching ability. The target orientation was changed from a canonical cube view (RCCT A) to that of a more unusual view of a cube balanced on a corner (RCCT B), because children prefer objects in a view that is functional (Davis, 1985). For instance, children draw a car from the side and a house from the front because this is where they enter the object. They would find a cube sitting flat on the ground more familiar than a cube balancing on one corner, and hence probably easier to compare.

The following two test sections (RCCT C-D) measure a more complex perceptual matching where distracters no longer have a uniform orientation, that is, the target as well as the distracter cubes vary in orientation. In section C, colour similarity of the distracters was gradually increased, see Table 4, but in section D the target and the distracter cubes were similar in colour in all trials. In these two sections the model and the target cube had different orientations in Levels III and IV.

Test booklet. A booklet with 36 A4 sized pages that all followed the same layout was used for testing. Each page showed one enlarged model cube on the top and two rows of up to 8 smaller cubes below, see Figure 6, page 32. The first four pages of the booklet were for practice only, with one example from each category (A, B, C and D) and thus included two simple perceptual matching and two mental rotation tasks. The participants' task was to identify, verbally and/or through pointing, which of the cubes was identical to the target cube on top of the page.

2.3.3. Procedure

Children were tested individually in a quiet, familiar setting at their school. They could choose a sticker as a reward after completing the test. Answers were recorded by the researcher on a response sheet during the session. Scores were added up by two researchers independently. No disagreement was found.

Test instructions. In the warm-up phase children were asked, "Do you want to play a game?" and were then shown a physical model of a coloured cube and asked, "Do you know what this is?" All children responded positively with the answer 'This is a cube' or 'This is a dice'. Thereafter, children first solved four practice trials in which the experimenter pointed at the enlarged target cube and asked the participant, "Which cube is the same as this one?"

After the participant had correctly answered the first two practice questions identifying the same cube as the target cube, the child was then tested with two practice questions that involved mental rotation. At this point, the researcher showed a physical cube model, turned it slightly and said, "These sides are turned". The researcher then pointed towards the 3D cube illustrations on the cube panel. The practice questions were repeated until the child could identify the correct cube image. Children then proceeded to the task proper. Initially it was considered adding a fold-out cube in the upper right corner of each page, but this proved to be too difficult in trial periods.

2.4. Results Study 1

The two main ANOVA's for the RCCT and the Raven, respectively, were followed by two more ANOVA's used as control measures for the RCCT test design. The latter compared the within-section level of difficulty, and the effect the number of distractors had on task difficulty.

Because boys showed significantly better Raven scores in the parent-financed school meal groups for the 8-year-old and 9-year-old group, in a further analysis, the RCPM scores were included as a covariate in order to establish the shared variance between fluid intelligence and spatial ability. The final analysis consists of the correlations between RCCT and RCPM in order to validate and compare task difficulty of the new test on mental rotation.

2.4.1. Rotated Colour Cube Test (RCCT)

Accurate performance was computed in per cent for each section of the Rotated Colour Cube Test (RCCT). In cases where the Mauchley's test of Sphericity was violated, the degrees of freedom were adjusted with Greenhouse-Geisser. Statistically significant effects were followed up with post-hoc tests. In the first part of this section, individual and age differences in the RCCT, in the second part individual and age differences on the RCPM, and the third part compares the RCCT and the RCPM overall scores, are reported.

A 4 (Sections) by 3 (Age) by 2 (Sex) by 2 (School meal type, FSM) analysis of variance was carried out, with repeated measurement for the RCCT sections. Group means are listed in Table 5 and statistical effects in Table 6.

Table 5

RCCT scores by age group and sex (Accuracy in per cent)

Sex	7-years	8-years	9-years	Total
		Section A		
		State School Meals		
Girls	88.89 (17.05)	95.00 (06.85)	93.75 (09.45)	92.05 (12.53)
Boys	85.94 (18.22)	89.77 (09.39)	93.75 (06.85)	89.50 (12.33)
		Parent School Meals		
Girls	97.50 (05.59)	94.23 (08.25)	95.83 (06.25)	95.37 (07.06)
Boys	92.86 (09.83)	96.88 (05.79)	98.86 (03.77)	96.63 (06.67)
		Section B		
		State School Meals		
Girls	81.94 (18.87)	95.00 (06.85)	89.06 (10.43)	87.50 (14.43)
Boys	92.19 (13.26)	87.50 (14.79)	93.75 (10.46)	90.50 (13.15)
		Parent School Meals		
Girls	92.50 (06.85)	89.42 (10.01)	93.06 (06.59)	91.20 (08.36)
Boys	91.07 (06.10)	96.88 (05.79)	100.00 (0.00)	96.64 (05.65)
		Section C		
		State School Meals		
Girls	79.17 (06.25)	80.00 (6.85)	82.81 (09.30)	80.67 (07.45)
Boys	78.13 (11.08)	82.95 (14.00)	87.50 (13.69)	82.50 (13.01)
		Parent School Meals		
Girls	77.50 (05.59)	83.65 (15.63)	87.50 (00.00)	83.80 (11.40)
Boys	80.36 (09.84)	93.75 (06.68)	88.64 (08.76)	87.98 (09.67)
		Section D		
		State School Meals		
Girls	44.44 (21.75)	57.50 (11.18)	45.31 (22.10)	47.73 (19.91)
Boys	39.06 (21.59)	40.91 (19.44)	41.67 (15.14)	40.50 (18.50)
		Parent School Meals		
Girls	40.00 (16.30)	47.12 (15.44)	44.44 (12.67)	44.91 (14.40)
Boys	46.43 (15.67)	65.63 (19.76)	71.59 (19.44)	62.98 (20.76)

There was a significant main effect for the RCCT sections, $F(2.23, 195.96) = 330.56$, $p < .001$, $\eta^2 = .79$, with a very large effect size. Pairwise post-hoc comparisons (*ps* < .001, two-tailed) confirmed that simple perceptual matching in RCCT A (M = 93.6%) and RCCT B (M = 91.9%) differed significantly from identification of the more difficult rotated targets in RCCT

C (*M* = 83.5%) and RCCT D (*M* = 48.7%), but not from each other. This demonstrated that the step from a canonical orientation in RCCT A to a more unusual position, balanced on one corner in RCCT B did not increase task difficulty for children. In RCCT C and D, test items and distracters were rotated at different angles.

The introduction of individually rotated model and distracter cubes in the test panel in section C led to a decrease in performance of about ten per cent, and the removal of distinctive and unique cube colours in section D led to an even more pronounced drop from about 80% to 40%, see Figure 8.

Table 6

Analysis of variance for RCCT performance with Age, Sex and School Meal type as between subject variables

Source	SS	*df*	*F*	*p*	η^2
Within-Subject Effects					
RCCT	**121229.74**	**2.227**	**330.559**	**.000*****	**.79**
RCCT*Sex	436.56	2.227	1.190	.309	.01
RCCT*Age	624.37	4.454	0.851	.504	.02
RCCT*FSM	251.22	2.227	0.685	.520	.01
RCCT*Sex*Age	586.104	4.454	0.799	.539	.02
RCCT*Sex*FSM	**2371.39**	**2.227**	**6.466**	**.001*****	**.07**
RCCT*Age*FSM	841.18	4.454	1.147	.337	.03
RCCT*Sex*Age*FSM	506.37	4.454	0.690	.615	.02
Between-Subjects Effects					
Sex	582.73	1.000	2.040	.157	.02
Age	**2701.44**	**2.000**	**4.729**	**.011***	**.10**
FSM	**2287.70**	**1.000**	**8.010**	**.006****	**.08**
Sex*Age	429.21	2.000	0.751	.475	.02
Sex*FSM	**1595.04**	**1.000**	**5.584**	**.020***	**.06**
Age*FSM	134.65	2.000	0.236	.790	.01
Sex*Age*FSM	922.15	2.000	1.614	.205	.04

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. RCCT = Rotated Cube sections; FSM = school meal type; Age = Age groups. * $p < .05$. ** $p < .01$. *** $p < .001$. Statistical effects set in bold were significant.

Furthermore, there were two significant main effects for FSM, $F(1, 88) = 8.01$, $p < .01$, $\eta^2 = .08$, and age groups, $F(2, 88) = 4.73$, $p < .05$, $\eta^2 = .10$. Children with state school meals (M = 76.9%) performed overall significantly worse than children with parent financed school meals (M = 81.9%). Multiple comparisons of age differences showed that 7-year-old children (M = 75.5%), differed from 9-year-olds (M = 81.7%), but no other comparisons were significant (no figure).

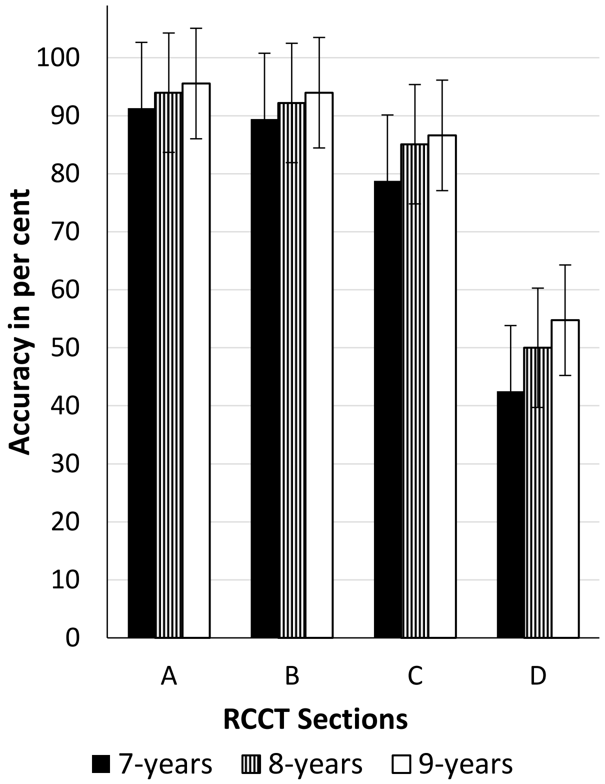


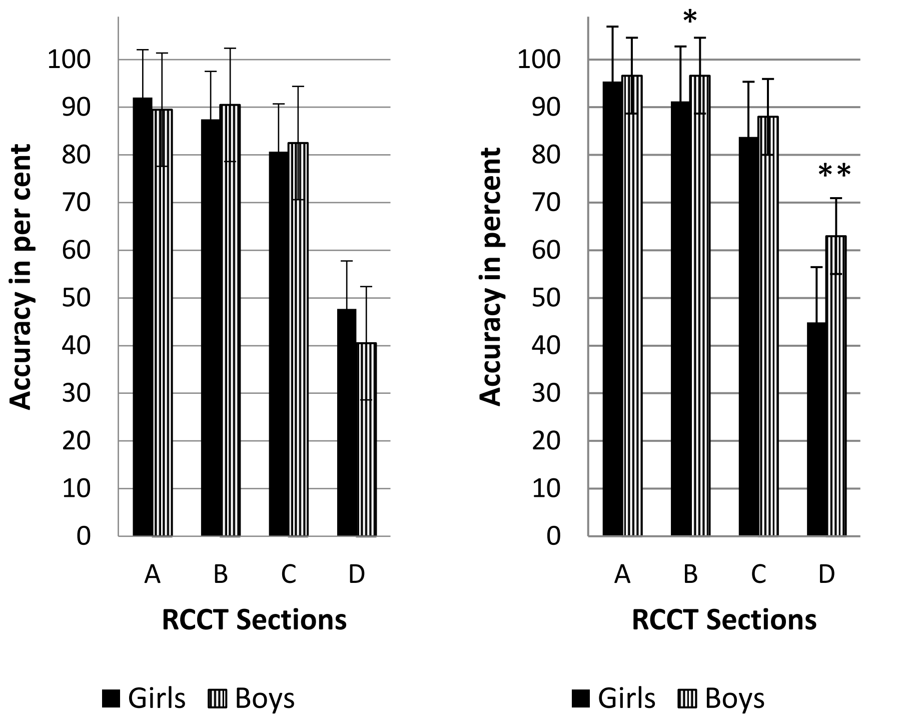
Figure 8 Performance decrease across RCCT sections per age group

There was also a significant two-way interaction of sex and FSM, $F(1, 88) = 5.58$, $p < .05$, $\eta^2 = .06$, and these factors interacted significantly in a three-way interaction with the RCCT sections, $F(2.23, 195.96) = 6.47$, $p < .01$, $\eta^2 = .07$.

In order to investigate the three-way interaction between RCCT sections, sex and FSM, the analysis of variance was re-run with a split sample analysis for SES in order to consider how boys and girls from different SES differed in their performance on individual RCCT sections.

In the state-funded free school meal group, there were neither sex differences, *ps* > .61, nor age differences, *ps* > .34 in performance. But when parents were able to pay for school meals, boys (*M* = 85.2%) outperformed girls (*M* = 78.9 %) and this difference was highly significant, $F(1, 47) = 14.27$, $p < .001$, $\eta^2 = .23$. Furthermore, only in this high SES group, the age difference was significant, $F(2, 47) = 6.07$, $p < .01$, $\eta^2 = .21$. Post-hoc t-tests for independent samples were run for boys and girls, for each RCCT section. In each of the two halves of the RCCT, always in the second, more difficult section (RCCT B and D), a significant sex difference was found: In RCCT A, the mean performance of boys and girls did not significantly differ (boys *M* = 96.6%; girls *M* = 95.4%), $p > .51$, but in RCCT B where the cubes had a non-canonical orientation, boys performed significantly better (*M* = 96.6%) than girls (*M* = 91.2%), t (51) = -2.76, $p < .01$.

Likewise, in the RCCT C, boys (*M* = 87.98%) and girls (*M* = 83.80%) did not differ significantly, $p > .16$, but in RCCT D where all RCCT rotated cubes had the same colour, boys (*M* = 63%) performed significantly better than girls (*M* = 44.9%), t (51) = -3.69, $p < .001$, see Figure 9, right. Thus, boys from a relatively higher socio-economic background excelled both in the more challenging perceptual matching and mental rotation task. This confirmed that these two tasks measure related abilities.



A

State school meals

B

Parent school meals

Figure 9 Sex differences in RCCT section performance for (A) children receiving state financed school meals versus (B) children whose parents financed their school meals (B). * = $p < .01$, ** = $p < .001$.

2.4.2. Raven's Coloured Progressive Matrices Test (RCPM)

Accuracy was computed in per cent correct for each type of RCPM section A, AB, and B, see Table 7. As with the RCCT, a 3 (Section) by 3 (Age) by 2 (Sex) by 2 (School meal type) analysis of variance with repeated measurement for each section was conducted.

Table 7

RCPM scores by Age Group and Sex (Accuracy in %)

Sex	7-years	8-years	9-years	Total
Section A				
State School Meals				
Girls	74.07 (08.78)	65.00 (10.87)	78.13 (13.32)	73.49 (11.68)
Boys	77.08 (13.91)	70.46 (14.61)	75.00 (11.79)	73.67 (13.54)
Parent School Meals				
Girls	81.67 (13.69)	78.21 (11.56)	81.48 (12.35)	79.94 (11.84)
Boys	73.81 (10.13)	76.04 (14.39)	86.36 (10.72)	79.80 (12.73)
Section AB				
State School Meals				
Girls	62.03 (21.70)	65.00 (16.03)	75.00 (17.25)	67.42 (19.06)
Boys	58.33 (20.89)	57.58 (25.67)	72.22 (21.52)	61.33 (23.18)
Parent School Meals				
Girls	81.67 (19.00)	70.51 (16.53)	81.48 (17.57)	76.23 (17.55)
Boys	67.86 (16.96)	86.46 (18.87)	89.39 (13.99)	82.69 (18.24)
Section B				
State School Meals				
Girls	50.00 (17.68)	50.00 (22.82)	58.33 (23.57)	53.03 (20.50)
Boys	46.88 (23.54)	46.21 (19.14)	68.06 (23.22)	51.67 (22.69)
Parent School Meals				
Girls	70.00 (24.00)	64.10 (19.95)	60.18 (16.55)	63.89 (19.20)
Boys	55.95 (15.00)	72.92 (07.39)	85.61 (13.99)	73.72 (17.27)

There was a significant main effect for the factor RCPM section, $F(1.86, 163.93) = 43.11$, $p < .001$, $\eta^2 = .33$, with a very large effect size, see Table 8. Pairwise post-hoc comparisons (*ps* < .001) confirmed significant differences between RCPM A (*M* = 76.4%), RCPM AB (*M* = 72.3%) and RCPM B (*M* = 61%). Sections became increasingly more difficult as the RCPM progressed.

Furthermore, there were two significant between-subject effects. The effect for FSM, $F(1, 88) = 16.51$, $p < .001$, $\eta^2 = .16$, showed that children with state school meals (*M* = 63.9%) performed overall significantly worse than children with parent financed school meals (*M* = 75.8%). The post-hoc tests of the age effect, $F(2, 88) = 4.50$, $p < .05$, $\eta^2 = .09$, showed performance of both the 7-year-old and 8-year-old children was worse compared to 9-year-olds.

Table 8

Analysis of variance for RCPM performance with Age, Sex and School meal type

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subject Effects			
RCPM	**12270.34**	**1.863**	**43.109**	**.000****	**.33**
RCPM*Sex	267.71	1.863	0.941	.387	.01
RCPM*Age	799.40	3.726	1.404	.237	.03
RCPM*FSM	**1083.69**	**1.863**	**3.807**	**.027***	**.04**
RCPM*Sex*Age	1262.67	3.726	2.218	.074	.05
RCPM*Sex*FSM	426.85	1.863	1.500	.227	.02
RCPM*Age*FSM	271.72	3.726	0.477	.739	.01
RCPM*Sex*Age*FSM	593.38	3.726	1.042	.384	.02
		Between-Subjects Effects			
Sex	79.64	1.000	0.134	.715	.00
Age	**5330.51**	**2.000**	**4.499**	**.014 ***	**.09**
FSM	**9780.36**	**1.000**	**16.511**	**.000****	**.16**
Sex*Age	2134.50	2.000	1.802	.171	.04
Sex*FSM	202.71	1.000	0.342	.560	.00
Age*FSM	515.22	2.000	0.435	.649	.01
Sex*Age*FSM	1628.43	2.000	1.375	.258	.03

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. RCPM = Ravens Coloured Progressive Matrices; FSM = School meal type; Age = Age groups. * $p < .05$. *** $p < .001$. Statistical effects set in bold were significant.

There was a significant two-way interaction of RCPM sections and FSM, $F(1.86, 163.93) = 3.807$, $p < .05$, $\eta^2 = .04$, that interacted neither with age, $p > .74$, nor with sex, $p > .23$. T-tests for independent samples were run to compare children with state school meals and parent financed school meals per RCPM section. They revealed a significant difference between children below and above the poverty line in each section, see Figure 10, and this increased the further children progressed in the Raven test, section A: $t(98) = -2.5$, $p < .05$, section AB: $t(98) = -3.83$, $p < .001$, section B: $t(98) = -4.04$, $p < .001$.

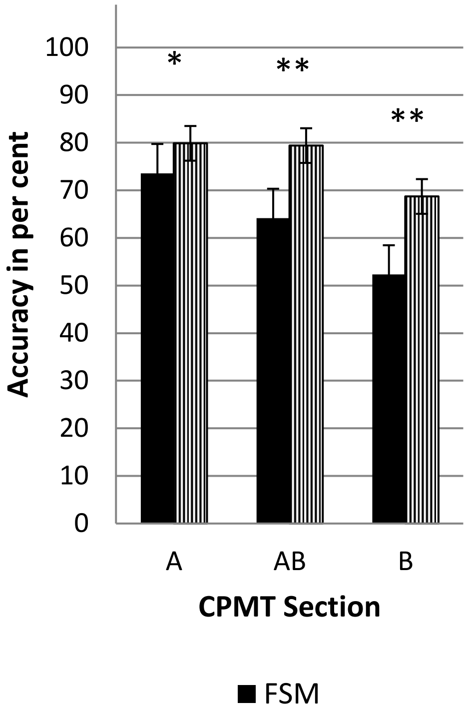


Figure 10 Increasing effects of socio-economic status in the sections of the Coloured Raven Progressive Matrices Test. * = p < .05, ** = p < .001.

2.4.3. RCCT Sections Split and Number of Distracter Control

This analysis was run to test if the RCCT design increased in difficulty within test sections. The first control involved comparing the first 4 test sheets in each set (easier ones) compared with the last 4 (harder ones). The second control involved comparing the number of distractors per test page (6 vs 8) per subsection. Only new effects were reported; the preserved interactions are mentioned but not explained again.

Control 1: RCCT Sections Split (4 Easy vs, 4 Hard Questions).

A 4 (section) by 2 (first 4 vs. last 4 questions per section, E/ H) by 3 (age) by 2 (sex) by 2 (school meal type) analysis of variance with repeated measurement was carried out, see Table 9 for the statistical effects.

Table 9

Analysis of variance for RCCT performance with Age, Sex, and financed school meal type, comparing the first four questions with the latter per section.

Source	**SS**	***df***	***F***	***p***	***η²***
	Within-Subject Effects				
RCCT	**242459.48**	**2.227**	**330.559**	**.000*****	**.79**
RCCT*Sex	873.12	3.000	1.190	.314	.01
RCCT*Age	1248.74	6.000	0.851	.531	.02
RCCT*FSM	502.44	3.000	0.685	.562	.01
RCCT*Sex*Age	1172.21	6.000	0.799	.571	.02
RCCT*Sex*FSM	**4742.78**	**3.000**	**6.466**	**.000*****	**.07**
RCCT*Age*FSM	1682.36	6.000	1.147	.336	.03
RCCT*Sex*Age*FSM	1012.75	6.000	0.690	.658	.02
E/H	**96655.77**	**1.000**	**432.938**	**.000*****	**.83**
E/H*Sex	360.52	1.000	1.615	.207	.02
E/H*Age	742.97	2.000	1.664	.195	.04
E/H *FSM	790.78	1.000	3.542	.063	.04
E/H*Sex*Age	597.99	2.000	1.339	.267	.03
E/H*Sex*FSM	146.13	1.000	0.655	.421	.01
E/H*Age*FSM	232.70	2.000	0.521	.596	.01
E/H*Sex*Age*FSM	**2064.58**	**2.000**	**4.624**	**.012***	**.10**
RCCT*E/H	**75754.49**	**2.430**	**86.123**	**.000*****	**.50**
RCCT*E/H*Sex	40.03	3.000	0.046	.987	.00
RCCT*E/H*Age	2777.28	6.000	1.579	.153	.03
RCCT*E/H*FSM	888.46	3.000	1.010	.389	.01
RCCT*E/H*Sex*Age	1620.90	6.000	0.921	.480	.02
RCCT*E/H*Sex*FSM	366.89	3.000	0.417	.741	.00
RCCT*E/H*Age*FSM	2053.21	6.000	1.167	.324	.03
RCCT*E/H*Sex*Age*FSM	638.36	6.000	0.363	.902	.01
	Between-Subjects Effects				
Sex	1165.47	1.000	2.040	.157	.02
Age	**5402.87**	**2.000**	**4.729**	**.011***	**.10**
FSM	**4575.41**	**1.000**	**8.010**	**.006****	**.08**
Sex*Age	858.43	2.000	0.751	.475	.02
Sex*FSM	**3190.08**	**1.000**	**5.584**	**.020***	**.06**
Age*FSM	269.30	2.000	0.236	.790	.01
Sex*Age*FSM	1844.09	2.000	1.614	.205	.04

Note. *Degrees of Freedom were corrected with Greenhouse-Geisser. RCCT = Rotated Cube sections; E/H = first 4 questions compared to last 4 questions of each section; FSM = school meal type; Age = Age groups. * $p < .05$. ** $p < .01$. *** $p < .001$. Statistical effects set in bold were significant.*

There was a significant main effect for the factor E/ H (easy/ hard), $F(1, 88) = 423.93$, $p < .001$, $\eta^2 = .83$, showing that the first 4 questions ($M = 90.9\%$) were easier than the later 4 questions per section ($M = 68.0\%$).

This effect also interacted with the RCCT sections, $F(3, 264) = 4.62$, $p < .001$, $\eta^2 = .49$. Post-hoc paired-samples t-tests showed that the gap between the first 4 and second 4 test sheets increased over the tests, as can be seen from the increasing t-values per section in Table 10.

Table 10

Paired sample t-tests comparing the difference between the easier first 4 question of each section with the hardest last 4 questions

RCCT Pages	*M*	*SD*	95% CI	*t*	*df*	*p*
A1 to 4 - A5 to 8	6.50	16.90	3.15 - 9.85	3.85	99	0.000
B1 to 4 - B5 to 8	5.75	19.09	1.96 - 9.54	3.01	99	0.003
C1 to 4 - C5 to 8	24.75	22.89	20.21 - 29.29	10.81	99	0.000
D1 to 4 - D5 to 8	55.25	32.62	48.78 - 61.72	16.94	99	0.000

The largest differences were found in sections C and D, where the model cube and target cube were differently rotated, see Figure 11. In RCCT C there was a drop in performance of 24.8% from the easier 4 (E4 $M = 96.3\%$) to the harder 4 questions (H4 $M = 71.5\%$) sets, whereas in section D the drop in performance was over twice as high at 55.3% (E4 $M = 76.8\%$, H4 $M = 21.5\%$).

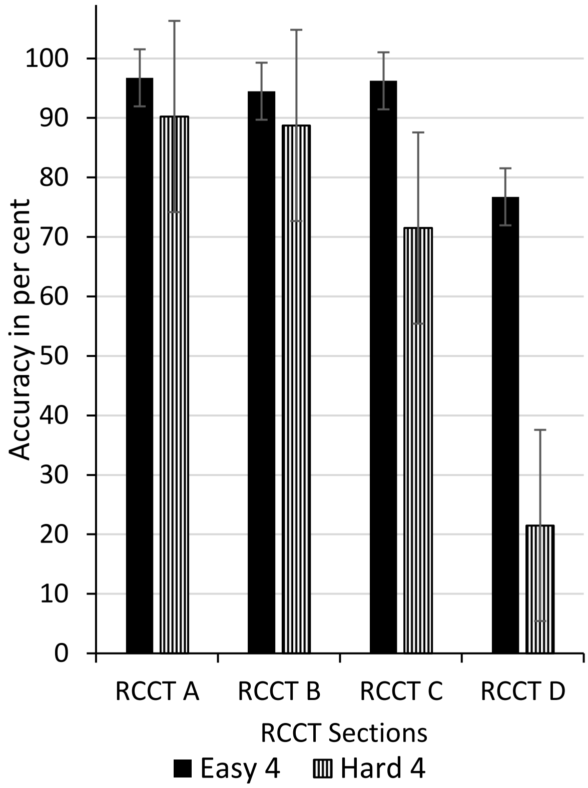


Figure 11 Differences in performance between the first 4 (easy) and last 4 (hard) test items per section.

There was also a significant four-way interaction of E/H, sex, age and FSM, $F(2, 88) = 4.62$, $p < .05$, $\eta^2 = .10$, see Figure 12. In order to disentangle this interaction further, the model was re-run as a split sample by sex, which revealed a significant statistical effect of task difficulty only for boys, $F(2, 45) = 3.86$, $p < .05$, $\eta^2 = .15$, but not in girls. Post-hoc comparisons showed that 8-year-old boys in the FSM group performed significantly better on the first 4 test items than on the last 4 test items (E4 M = 88.6%, H4 M = 61.9%).

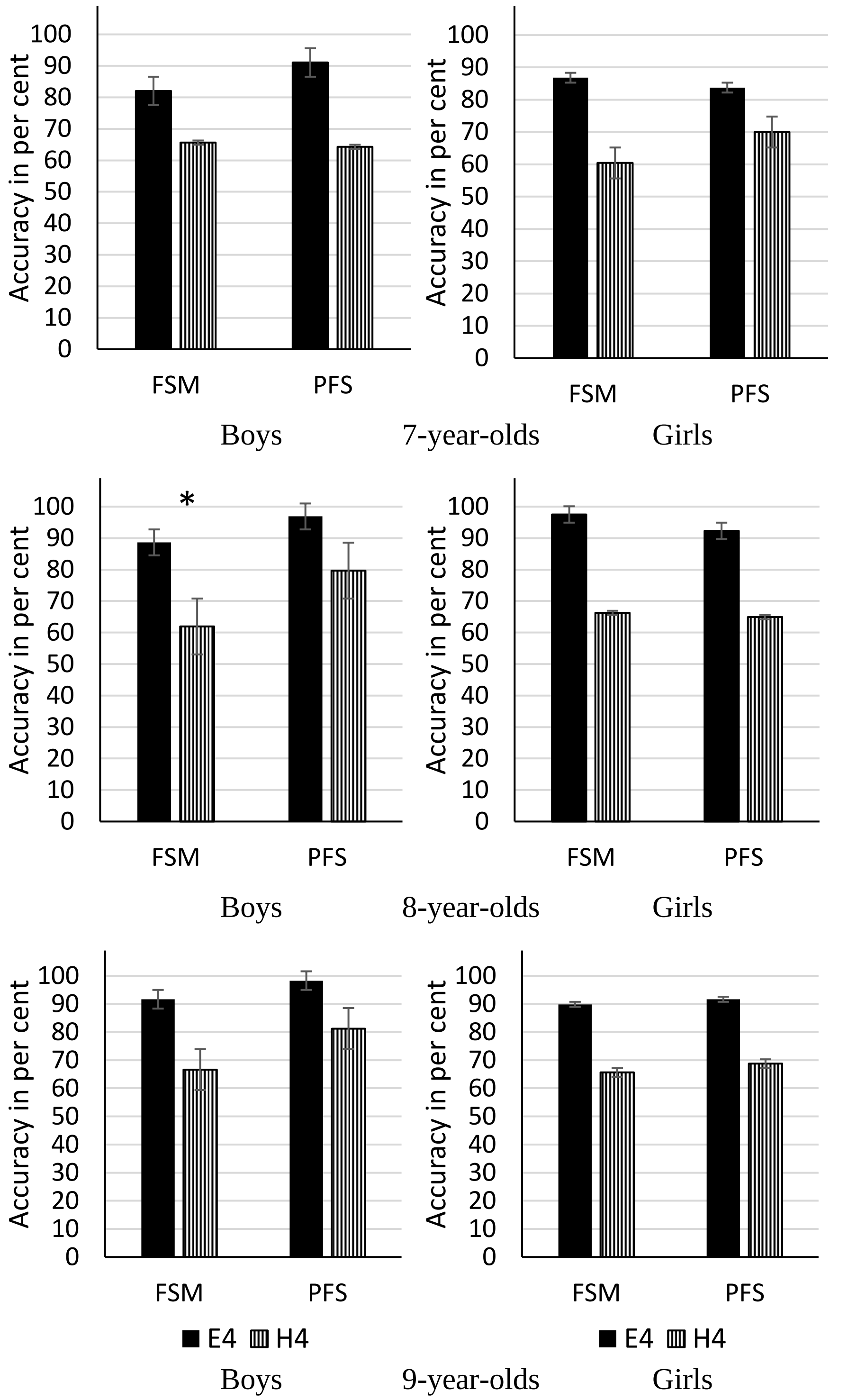


Figure 12 State (FSM) and parent (PSM) school meals by sex by easy vs. hard RCCT questions. * $p < .05$

Control 2: Questions with 6- compared to 8 test items per page

A 4 (Sections) by 2 (Item number) by 3 (Age) by 2 (Sex) by 2 (School meal type) analysis of variance was carried out, see Table 11.

There was a significant main effect for number of distractors, $F(1, 88) = 4.24$, $p < .05$, $\eta^2 = .05$, demonstrating that questions with 6 options (M = 80.7%) are easier than those with 8 response options (M = 78.2%). This effect also interacted with RCCT sections, $F(2.63, 231.23) = 6.23$, $p \leq .001$, $\eta^2 = .02$, see Figure 13. As expected, post-hoc paired-samples t-tests found that the questions with 6 response options were easier in section A, $t(99) = 2.57$, $p < .05$, section B, $t(99) = 2.83$, $p < .01$, and in section D, $t(99) = 2.25$, $p < .05$. Only in section C were questions with 8 response options easier, $t(99) = -2.05$, $p < .05$

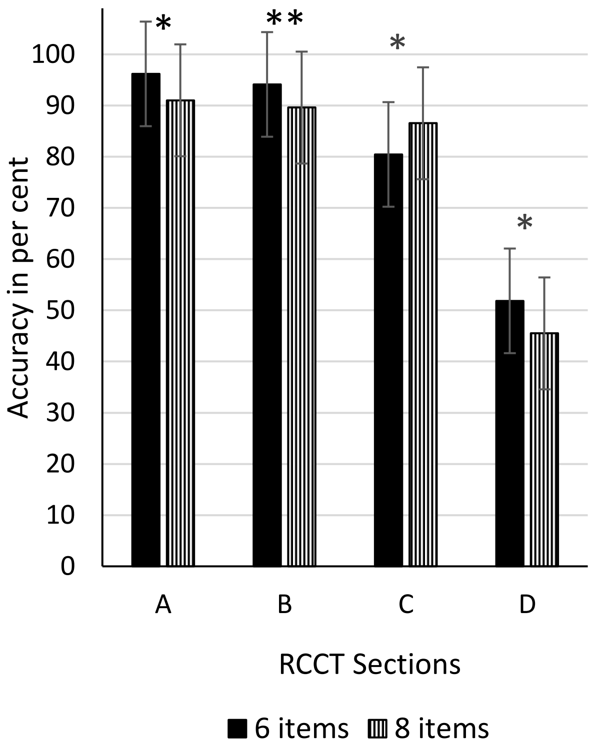


Figure 13 Differences in performance on RCCT sections with 6 and 8 response options. * = $p < .05$, ** = $p < .001$.

Table 11

Analysis of variance for RCCT performance with Age, Sex, and financed school meal type, comparing the first four questions with the latter per section.

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subject Effects			
RCCT	**242459.48**	**2.227**	**330.559**	**.000****	**.79**
RCCT*Sex	873.13	3.000	1.190	.314	.01
RCCT*Age	1248.74	6.000	0.851	.531	.02
RCCT*FSM	502.44	3.000	0.685	.562	.01
RCCT*Sex*Age	1172.21	6.000	0.799	.571	.02
RCCT*Sex*FSM	**4742.78**	**3.000**	**6.466**	**.000****	**.07**
RCCT*Age*FSM	1682.36	6.000	1.147	.336	.03
RCCT*Sex*Age*FSM	1012.75	6.000	0.690	.658	.02
6.vs.8	**1140.11**	**1.000**	**4.237**	**.043**	**.05**
6.vs.8*Sex	27.39	1.000	0.102	.750	.00
6.vs.8*Age	1070.89	2.000	1.990	.143	.04
6.vs.8*FSM	265.40	1.000	0.986	.323	.01
6.vs.8*Sex*Age	150.67	2.000	0.280	.756	.01
6.vs.8*Sex*FSM	150.04	1.000	0.558	.457	.01
6.vs.8*Age*FSM	73.03	2.000	0.136	.873	.00
6.vs.8*Sex*Age*FSM	780.82	2.000	1.451	.240	.03
RCCT*6.vs.8	**4589.51**	**2.628**	**6.230**	**.001***	**.07**
RCCT*6.vs.8*Sex	1321.83	3.000	1.794	.149	.02
RCCT*6.vs.8*Age	693.40	6.000	0.471	.830	.01
RCCT*6.vs.8*FSM	171.26	3.000	0.232	.874	.00
RCCT*6.vs.8*Sex*Age	349.99	6.000	0.238	.964	.01
RCCT*6.vs.8*Sex*FSM	125.53	3.000	0.170	.916	.00
RCCT*6.vs.8*Age*FSM	1694.75	6.000	1.150	.334	.03
RCCT*6.vs.8*Sex*Age*FSM	932.97	6.000	0.633	.704	.01
		Between-Subjects Effects			
Sex	1165.47	1.000	2.040	.157	.02
Age	**5402.87**	**2.000**	**4.729**	**.011***	**.10**
FSM	**4575.41**	**1.000**	**8.010**	**.006**	**.08**
Sex*Age	858.43	2.000	0.751	.475	.02
Sex*FSM	**3190.07**	**1.000**	**5.584**	**.020**	**.06**
Age*FSM	269.30	2.000	0.236	.790	.01
Sex*Age*FSM	1844.09	2.000	1.614	.205	.04

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. RCCT = Rotated Cube sections; 6.vs.8 = number aggregates per page; FSM = school meal type; Age = Age groups. * $p < .05$. ** $p < .01$. ** $p < .001$. Statistical effects set in bold were significant

2.4.4. Control of the Mental Rotation with Fluid Intelligence

In order to control the effect of fluid intelligence (RCPM) on Mental Rotation (MR), the analysis of the RCCT was repeated with the Raven test score as a covariate, (Lehmann, Quaiser-Pohl, & Jansen, 2014; Quaiser-Pohl et al., 2014). If an effect was significant in the previous analysis, but is no longer significant after controlling for the RCPM, this change could then be attributed to differences in fluid intelligence. Likewise, if an effect becomes significant that was not significant before, control for fluid intelligence reveals an effect that was suppressed before by this variable.

In particular, because of the relatively high number of children from a poor socioeconomic background, the control of fluid intelligence will show whether better performance in mental rotation of boys above the poverty line is due higher fluid intelligence. In statistical terms, this would be the case if the three-way interaction of RCCT, sex and socio-economic status would no longer be significant.

A repeated measures ANCOVA of 4 (Sections) by 3 (Age) by 2 (Sex) by 2 (School meal type, FSM) with the RCPM as a covariate showed the statistical effects that are listed in Table 12.

The most compelling effect of the covariate was that the main effect of socioeconomic background (FSM) was no longer significant, p = .502, instead the Raven score as a covariate was significant, $F(1, 87) = 36.55$, $p < .001$, $\eta^2 = .30$. This showed that differences due to socio-economic status in the general performance level were completely explained by fluid intelligence, but the main effect of age, $F(2, 87) = 4.33$, $p < .05$, $\eta^2 = .09$, was not affected. This is somewhat surprising because one would have expected that fluid intelligence would replace the age but not the SES effect as intelligence increases with age.

Table 12

Analysis of variance for RCCT performance with Age, Sex and School Meal type as between subject variables and RCPM scores as a co-variant.

Source	**SS**	***df***	***F***	***p***	***η²***
	Within-Subject Effects				
RCCT	**11735.52**	**2.303**	**33.594**	**.000*****	**.28**
RCCT*RCPM	**1881.47**	**3.000**	**5.386**	**.001****	**.06**
RCCT*Age	872.62	6.000	1.249	.282	.03
RCCT*Sex	407.56	**3.000**	1.167	.323	.01
RCCT*FSM	19.78	**3.000**	0.057	.982	.00
RCCT*Sex*Age	455.50	**3.000**	0.652	.689	.02
RCCT*Sex*FSM	**2109.47**	**3.000**	**6.039**	**.001****	**.07**
RCCT*Age*FSM	918.52	6.000	1.315	.251	.03
RCCT*Sex*Age*FSM	339.36	6.000	0.486	.819	.01
	Between-Subjects Effects				
RCPM	**7439.51**	**1.000**	**36.549**	**.000*****	**.30**
Sex	430.71	1.000	2.117	.149	.02
Age	**1763.34**	**2.000**	**4.334**	**.016***	**.09**
FSM	92.47	1.000	0.455	.502	.01
Sex*Age	119.10	2.000	0.293	.747	.01
Sex*FSM	**1189.82**	**1.000**	**5.849**	**.018***	**.06**
Age*FSM	253.75	2.000	0.624	.538	.01
Sex*Age*FSM	419.02	2.000	1.030	.361	.02

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. RCCT = Rotated Cube sections; RCPM = Raven's Coloured Progressive Matrices Test; FSM = school meal type; Age = Age groups. * $p < .05$. ** $p < .01$. *** $p < .001$. Statistical effects set in bold were significant.

In order to explore the relationship between the Raven and the RCCT further, a curvefit analysis was run, revealing significant linear, $F(1, 98) = 64.45$, $p < .001$, quadratic, $F(2, 97) = 32.05$, $p < .001$, and cubic, $F(3, 96) = 21.27$, $p < .001$, trends. The higher the RCCT score, the better the accuracy of the RCPM. The plotted data in Figure 14 reveal that indeed the linear fit with the highest F-value seemed to provide the best fit for the data.

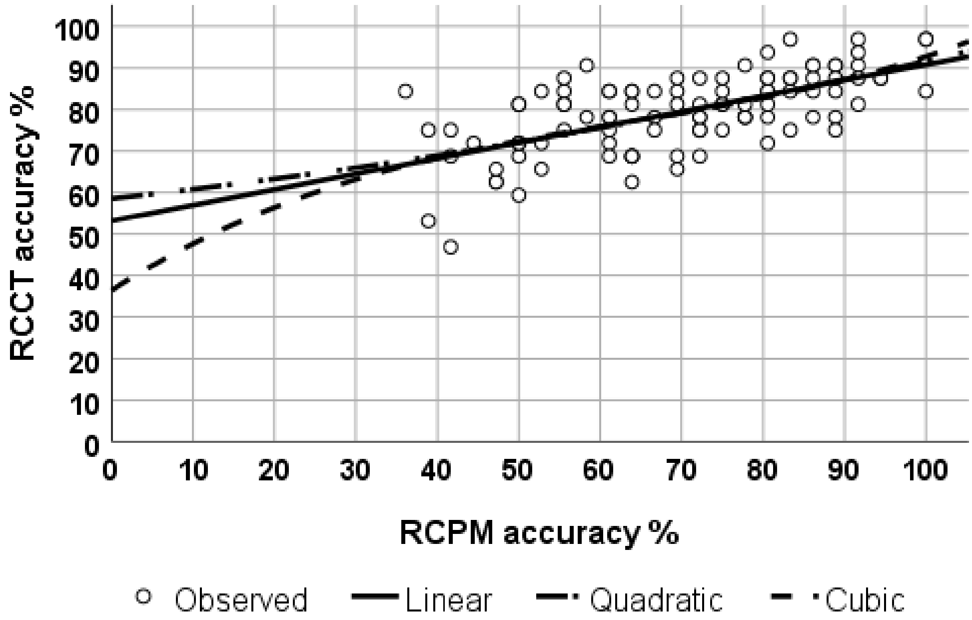


Figure 14 Scatterplot for the RCPM by RCCT

However, the two-way interaction between socioeconomic background and sex of the children stayed significant, $F(1, 87) = 5.85$, $p < .05$, $\eta^2 = .06$, and this again interacted significantly in a three-way interaction with the RCCT, $F(3, 261) = 6.039$, $p < .05$, $\eta^2 = .07$. This showed that fluid intelligence was not the final determining factor for the effect that boys from a more affluent background performed better on the harder mental rotation tasks.

With regards to the remaining within-subject factors, there was again a significant main effect for the RCCT sections, $F(2.30, 200.34) = 33.59$, $p < .001$, $\eta^2 = .28$, that showed that simple perceptual matching in the sections A and B were significantly easier than identification of the rotated targets in sections C and D, independently of the fluid intelligence of children.

Another new effect was the significant two-way interaction between the RCCT sections and the Raven scores, $F(3, 261) = 8.386$, $p \leq .001$, $\eta^2 = .06$. The follow-up analysis of the interaction effect revealed decreasing correlations with fluid intelligence over RCCT sections A ($r = .45$, $p < .001$), B (r = .42, p < .001) and C (r = .36, p < .001) with the Raven score, but the highest correlation occurred with the most difficult RCCT section D (r = .54, p < .001) which showed that fluid intelligence was the most required in the most challenging cube rotation.

2.4.5. Comparison of the RCCT and the RCPM

Because the RCCT was a new test development, the established RCPM was used for cross-validation. The RCCT and RCPM scores were significantly correlated, with r = .52, p = .004 for the 7-year-old children, r = .60, p < .001 for the 8-year-old children and r = .72, p <.001 for the 9-year-old children. These correlations were significant and increased with age.

A confirmatory correlational analysis with the unstandardized residuals of both variables after controlling for the impact of age, sex, and FSM, revealed a highly significant correlation between the RCCT and RCPM (r =. 54, p <. 001). The correlations increased in each age group with r = .53, p = .003 for 7-year-old, r = .54, p < .001, 8-year-old, and r = .63, p <.001 for 9-year-old children. The correlation between the RCCT and the RCPM when controlled for age, sex and FSM are nearly identical in the 7- and 8-year-olds but increased in the 9-year-olds. This shows an increasing and substantial correlation between the RCCT and the RCPM independently of individual differences.

A more comprehensive analysis was conducted of the variance on the two *overall scores* of the RCCT and the RCPM, respectively, that allowed for a direct comparison. A 2 (RCCT vs. RCPM) by 3 (age) by 2 (sex) by 2 (school meal type) analysis of variance revealed no significant sex differences, all *ps* > .16, see Table 13. This demonstrated that sex differences were limited to the more difficult RCCT sections and did not appear when *overall scores* were used. Furthermore, a significant within-subject main effect, F(1, 88) =

60.71, $p < .001$, indicated that on average the RCPM ($M = 69.9\%$) was more difficult than the RCCT ($M = 79.5\%$). This difference showed a comparably large effect size of $\eta^2 = .41$, while all other significant effects sizes were smaller, $\eta^2 < .16$.

Table 13

Analysis of variance for RCCT vs. RCPM performance with Age, Sex and State School Meals

Source	*SS*	*df*	MS	*F*	*p*	η^2
			Between-Subjects Effects			
Sex	148.30	1.000	148.30	.745	.390	.01
Age	**1874.83**	**2.000**	**937.41**	**4.710**	**.011 ***	**.10**
FSM	**3281.50**	**1.000**	**3281.51**	**16.489**	**.000****	**.16**
Sex*Age	644.28	2.000	322.14	1.619	.204	.04
Sex*FSM	397.31	1.000	397.31	1.996	.161	.02
Age*FSM	84.79	2.000	42.40	.213	.809	.01
Sex*Age*FSM	666.78	2.000	333.39	1.675	.193	.04
			Within-Subject Effects			
Test	**4240.60**	**1.000**	**4240.60**	**60.712**	**.000****	**.41**
Test*Sex	23.93	1.000	23.93	.343	.560	.04
Test*Age	**577.37**	**2.000**	**288.68**	**4.133**	**.019***	**.09**
Test*FSM	**550.54**	**1.000**	**550.54**	**7.882**	**.006****	**.08**
Test*Sex*Age	174.53	2.000	87.26	1.249	.292	.03
Test*Sex*FSM	69.02	1.000	69.02	.988	.323	.01
Test*Age*FSM	120.61	2.000	60.30	.863	.425	.02
Test*Sex*Age*FSM	106.54	2.000	53.27	.763	.469	.02

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. Test = Test type (RCCT vs. RCPM); FSM = school meal type; Age = Age groups. * $p < .05$. ** $p < .01$. *** $p < .001$. Statistical effects set in bold were significant.

There was a significant main effect for age groups that interacted in a two-way interaction with the two tests, $F(2, 88) = 4.133$, $p < .05$, $\eta^2 = .86$, see Figure 15, with an even larger effect size. Test scores on the RCCT increased with age, from 7-year-olds ($M = 75.2\%$), to 8-year-olds ($M = 80.2\%$) and to 9-year-old children ($M = 82.5\%$), by 7.3%. Similarly, test scores on the

RCPM also increased with age, from 7-year-olds (*M* = 65.3%), to 8-year-olds (*M* = 67.3%) and to 9-year-old children (*M* = 77.1%), by 11.8%. Post-hoc t-tests for paired samples revealed a significant difference between the two test scores in all three age groups, with the RCCT scores always significantly higher than the RCPM scores, *ps* < .01. However, Figure 15 shows that the difference between the test performance reduced with increasing age, and vice versa, the correlations between the two tests increased with age (see the first paragraph of this part of the report).

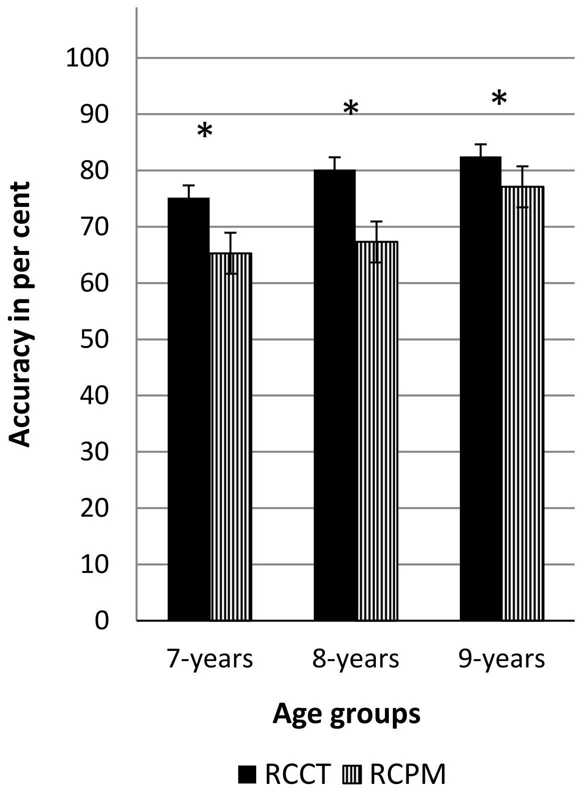


Figure 15 Age differences in RCCT and RCPM. * = $p < .01$.

The significant between-subjects effects for FSM $F(1, 88) = 16.49$ $p < .001$, $\eta^2 = .16$, showed that children receiving state school meals (*M* = 70.4%) scored lower than children with parent financed school meals (*M* = 78.6%), but importantly, a significant two-way interaction revealed that this varied depending on the test, $F(1, 88) = 7.88$, $p < .05$, $\eta^2 = .08$.

To test where the difference was located, the analysis of variance was run again with a split sample for FSM. It showed that children with state school meals showed much better performance on the RCCT ($M = 76.9\%$) than on the RCPM ($M = 63.9\%$). However, a two-way interaction showed that this difference became smaller with age, $F(1, 41) = 3.56$, $p < .05$, $\eta^2 = .15$, 7-year-olds (RCCT: $M = 73.7\%$, RCPM: $M = 61.4\%$), to 8-year-olds (RCCT: $M = 77.3\%$, RCPM: $M = 58.7\%$) and 9-year-olds (RCCT: $M = 78.4\%$, RCPM: $M = 71\%$), as Raven scores were relatively improved in the older children on state school meals. No other statistical effects were significant, *ps* > .08. Thus, the RCCT was especially fair to younger low SES children.

In contrast, the sample of children with parent financed school meals showed the same disparity between tests, RCCT: $M = 81.9\%$; RCPM: $M = 75.8\%$, $F(1, 47) = 13.53$, $p < .001$, $\eta^2 = .22$, but no reduction with age, $F(1, 47) = .78$, $\eta^2 = .03$, *ns*. Instead, there was a main effect of age, $F(2, 47) = 3.66$, $p < .05$, $\eta^2 = .16$, which interacted with sex, $F(2, 47) = 3.93$, $p < .05$, $\eta^2 = .14$. Girls' test scores were similar at 7-years $M = 77.3\%$, 8-years $M = 74.7\%$ and 9-years $M = 77.3\%$ and showed no improvement, whereas boys' performance increased with age 7-years $M = 71.9\%$, 8-years $M = 83.4\%$, 9-years $M = 88.5\%$.

2.5. Discussion Study: Rotated Colour Cube Test

2.5.1. The Rotated Coloured Cube Test

This study introduces a new test, namely the Rotated Colour Cube Test (RCCT) as a measure of perceptual matching and mental rotation in children, using coloured three-dimensional cube illustrations. This new test is not a mental rotation *experiment* where angularity of the rotated stimulus and reaction times are measured (Shepard & Metzler, 1971), but a mental rotation *test* similar those for adults by Vandenberg and Kuse (1978) and Peters et al. (1995). Models, targets, and distracters were differently rotated from each other, so even if participants were able to use perceptual matching strategies, the RCCT can still be classified as also measuring mental rotation.

A single cube is a useful reduction of complexity for children in comparison to the cube aggregates of Shepard and Metzler (1971), but it does not lend itself to spatial rotation in the same way as a cube aggregate because a single cube lacks a clear three-dimensional extension that protrudes into space, comparable to a vector. Cube aggregates have extensions which have been compared to pictures of gymnasts with outstretched limbs (Alexander & Evardone, 2008). Instead, cube faces were distinguished through the use of colour and by rotating the cubes in various directions.

In the present study 7- to 10-year-old children were presented with a three-dimensional multi-coloured cube as a target and asked to find the correct matching cube out of 6 or 8 alternatives. The first part of the test measured children's perceptual matching abilities and the second part of the test measured more complex spatial and mental rotation abilities. Children were found to perform best in identifying the same cube amongst distracters, compared to identifying a rotated cube amongst unique distracters. Identifying a rotated cube amongst rotated similar distracters of the same colours was overall the most difficult task. Additionally, an interaction between sex differences and low vs. high SES background revealed that boys with more resourceful parents, were more likely to successfully complete more difficult items. Mental rotation performance significantly correlated with children's performance on the RCPM which supported the validity and reliability of the test.

In comparison to the original stimuli used by Shepard and Metzler (1971) and other 2D-tests, the RCCT's three dimensional properties were preserved making the test more difficult. But task difficulty was decreased by adding item-distinguishing colour and through the reduction of set-size, as the complex cube aggregates were reduced to a single cube.

The test format was conceived similarly to the RCPM, with a response panel of a target and several distracters, and by increasing task difficulty in the test sections. This provided an assessment baseline for item identification of a model object in a canonical view (section A) and in a rotated position (section B) via colour to make sure that children had an intact object concept. For instance, an object concept is not self-understanding in cube drawings as children often draw just one side, or if they draw more than one side, these multiple cube faces are not integrated (e.g. Lange-Küttner & Ebersbach, 2012). Colour was used to distinguish between distracters as it is such an important feature for children, that for instance, the Raven test for children only exists in colour, whereas the version for adults is in black-and- white. In the RCCT, children could identify the target by finding the correct spatial configuration of the coloured cube faces. Moreover, colour similarity of the cube distracters was increased during trials in each of the first three sections, see Table A1 in the Appendix which made the target less discriminable from the distracters.

As expected, similar to the RCPM, the new RCCT became more difficult over the four test sections. There was no significant performance decrease in perceptual matching between the first two sections A and B where the only difference was the overall cube orientation, except that high SES boys performed better than high SES girls in identifying a cube in a non-canonical position. The main difference in performance arose between simple perceptual identification in RCCT sections A and B, and the more challenging target cube identification amongst individually rotated distracter cubes in sections C and D. In section C, colour similarity was increased between targets and distracters, whereas in section D colour was completely removed as a distinctive feature (see Table A1 in the Appendix). The results suggest that performance in sections C and in section D decreased due to an increase in the number of rotated targets and distracters as well as due to a reduction of colour saliency. Indeed, the study showed that it is not orientation of *one* object as such that is difficult for children, but differences in orientation between targets and distracters.

Performance deteriorated particularly in section D where no unique object colours were available to distinguish between distracter cubes. In fact, increasing colour congruency between target and distracter cubes to a level where all colours were the same and cubes varied only in orientation, resulted in the most pronounced decrease in performance. This result is especially noteworthy because in both sections C and D, the models and the targets were different in regard to their rotations (see Appendix, Column 'Rotation', Levels 5-8). In short, distinct unique colours between distracters were particularly helpful as a visual cue in narrowing down attention towards the rotated target. In conclusion, while different orientations of distracters made the RCCT more difficult, different colours of distracters had the opposite effect and made the RCCT easier because colour facilitated clearer discrimination between target and distracters.

Solving a three-dimensional mental rotation task involves the ability to maintain representations of relevant object attributes and their interrelation, while at the same time rotating mental images (Kaufman, 2007). As with adults, object colour can be more salient and important than object location (Hyun & Luck, 2007). Integration of the cube faces was easier when objects were different: Differences in the target cube's orientation and even between individually rotated distracters was not especially difficult as long as the distracters' distinctive object colours were available as a cue. This may be somewhat counterintuitive as mental rotation is a spatial ability. However, colour is a feature that defines the cube's internal structure based on the cube's face colour location and is not a spatial cue about the location of the cube. Because the cube had only changed orientation and not position, the object-place binding (Lange-Küttner, 2008b, 2013) remains intact in mental rotation. In short, rotated objects stay in place. This in turn suggests that feature integration plays an important role in mental rotation.

As in previous studies on visual memory and spatial ability (Lange-Küttner, 2010a; Levine et al., 2005), a pronounced impact of sex and SES was already present in school children: Boys from a higher social-economic background performed better than girls from the same relatively advantaged background in sections B (non-canonical cube orientation) and D (lack of unique distracter colour) of the test, while there was no difference between boys and girls from a low social-economic background. Even after controlling the influence of fluid intelligence, this effect remained significant, confirming that this gender effect was not due to variations in intelligence. This preserved MR interaction effect is even more remarkable given the overall shared variance of fluid intelligence with SES. The 'gearing up' of the more privileged boys indicated that they were more likely to rise to a challenge (Lange-Küttner, 2012; Lange-Küttner & Green, 2007). This sex by task difficulty effect was particularly apparent when colour cues were no longer available in section D, as only the upper middle-class boys had a success rate of above 60% while everybody else was below 48%. It could well be that these boys developed more responsiveness and attention towards the less obvious cube features such as contour and line orientation (Enns & Rensink, 1991; Hystegge, Heim, Zettelmeyer, & Lange-Kuttner, 2012; Lange-Küttner et al., 2002; Treisman, 1986). Similarly, sex differences on mathematics achievement tests are only found on more difficult items, possibly as the content is higher in measuring spatial ability (Hedges & Nowell, 1995; Hyde, Fennema, & Lamon, 1990).

An alternative explanation for SES differences in ability is prior engagement in activities that promote the development of spatial skills. Children from a lower SES group may not readily have access to Lego, playing video games or completing puzzles, which all correlate with spatial ability (Dorval & Pepin, 1986; Serbin, Zelkowitz, Doyle, Gold, & Wheaton, 1990; Subrahmanyam & Greenfield, 1994). Similarly the freedom to explore their environment correlates with sex differences, as boys spend more time exploring their environment (Entwisle, Alexander, & Olson, 1994). While the promotion of spatial skill through activities might be an important factor in explaining the RCCT results, it does not allow to distinguish between the availability of these activities through a biologically driven inclination or because cultural norms make such resources more readily available to boys (Lange-Küttner, Korte, & Stamouli, 2019).

2.5.2. Mental Rotation and Fluid Intelligence

The results showed that in general, the Rotated Colour Cube Test (RCCT) was easier that the Raven's Coloured Progressive Matrices Test (RCPM), but this difference was more pronounced in the younger age group of 7- year-olds than in the 9-year-old children, especially when from a lower SES.

The most compelling yet also sobering effect was that after controlling performance for fluid intelligence, children from a poorer socio-economic background no longer performed significantly worse than children from families with a more stable financial background. This demonstrated that differences in attainment previously attributed to socio-economic background, were in fact a result of differences in fluid intelligence. Previous research highlighted the adverse impact poverty may have on children in terms of standardised IQ scores (Smith et al., 1997) and measures of spatial ability (Levine et al., 2005).

A surprising result is that age related performance differences were not significantly influenced by differences in fluid intelligence, a finding that contradicts developmental research that has found cognitive ability to increase with age in relation to short term memory capacity (Dempster, 1981) and reasoning ability (Raven & Court, 1998; Wechsler, 1981).

Nevertheless, both the two-way interaction between sex and socioeconomic status and the three-way interaction between RCCT sections, sex and socio-economic status remained significant, highlighting the disproportionately positive impact of a wealthier socio-economic background already has on 6- to 10-year-old boys' spatial and mental rotation ability especially in the more challenging parts of the tests.

It could be argued, that the RCCT was easier than the RCPM since the spatial reasoning component was less complex. While in the RCCT single three-dimensional cubes were used, in the RCPM both continuous and discrete patterns had to be completed. That is, even if multiple coloured cube faces had to be perceptually integrated and distinguished against competing distracters, the RCCT sections did not require the formation of a logical sequence of visual pattern fragments which may require executive attention (Jones, Rothbart, & Posner, 2003), or operational intelligence according to Piaget (Inhelder & Piaget, 1958). This might explain why younger children performed comparatively well in the RCCT except in the last section where colour salience was removed. Therefore, this new test would lend itself to measuring object processing in even younger age groups.

However, besides these differences in the two tests the significant correlations also suggested strong similarities. The RCPMT test was identified as measuring both fluid intelligence and spatial ability (Guttman, 1974).

Similarly, it was proposed that spatial ability tests load considerably on g (Ullstadius, Carlstedt, & Gustafsson, 2004). The significant correlations in each age group could also be a result of a similar response format involving a perceptual discrimination between target and multiple distracters in both tests.

In agreement with previous research (Eysenck & Kamin, 1981; Raven, 1939) no sex differences on the RCPM were found, whereas sex differences in the RCCT only emerged in the more difficult sections B (perceptual matching) and D (mental rotation) for children from a poorer social-economic background. A likely influence is the difference in dimensionality between the two tests. The RCPM only includes 2D items, whereas the RCCT only consists of 3D items. As the visual information system is sensitive to dimensionality and finds processing of three-dimensional information more difficult (Jansen et al., 2013), which may account for the results of only finding sex differences in the three- dimensional RCCT and not in the RCPM. Hence, it could be argued that it is important to have a distinct 3D mental rotation test designed specifically for children, for instance, in order to assess abilities in maths, science and the arts from an early age.

The unexpected result of a drop in the levels of performance for children from a low social-economic background on the RCPM support the findings of a study by Aziz and Farooqi (1991) in which children from a lower social economic background also performed worse on the RCPM compared to those from more affluent background. The RCPM has often been viewed as an ethically and culturally fair measure of intellectual functioning (Anderson, Kern, & Cook, 1968; Jensen, 1974; Kaplan & Sccuzzo, 1996; Valencia, 1984). However, in the current study this was not the case, which contradicts expectations of culture-fair tests of the American Educational Research Association (1999). The disadvantage of low socio-economic status on both the RCCT and RCPM highlights the need for developing tasks that may help to identify children who can benefit from early intervention strategies. Item-specific practice (Kail, 1986; Kass et al., 1998) as well as 2D-3D dimensional transformation training (Moreau, 2012; Tzuriel & Egozi, 2007, 2010) might contribute to rebalancing poverty induced inequalities in cognitive abilities (Noble et al., 2015). However, while the research sample was relatively large with N= 100, cells were unequal to some degree, though not dramatically so (Howell, 1992, section 13.9, pp. 409). There were 51 boys and 49 girls, so sex was nearly perfectly balanced. Likewise, for SES there were 47 children on state school meals and 53 children whose parents could afford to pay for school meals. But the smaller cells of the interactions (see Table 1) were unequal, with subsamples between 5 to 13. This means that there could have been an element of chance in the obtained significances of some interactions. However, because the interactions related largely to the more difficult sections B and D, it is believed that the interaction of task difficulty with the individual differences was genuine.

The newly developed RCCT measures children's spatial ability through perceptual matching and mental rotation tasks. It distinguishes itself from other available tests for children by preserving and simplifying Shepard and Metzler's (1971) three dimensional geometric properties and by providing a modified test specifically adapted for young children. The RCCT results show that young children can solve tasks with three-dimensional cube illustrations, but increasingly struggle when supportive colour information is reduced. The current study demonstrated a reduction in task complexity without resorting to 2D images, and that varying distracter similarity and colour salience were effective means of adjusting task difficulty. In their meta-analysis Voyer et al. (1995) argued that since spatial ability is not a unitary concept, each spatial test might provide a distinct operational definition of abilities. As the MRT is widely used with adults, it was important to create a simplified MRT test for younger children that preserves its three-dimensions as well as the use of geometric stimuli. This aims to bridge the gap between the classic complex three-dimensional cube aggregates used for adults (Shepard & Metzler, 1971) and simplified two-dimensional versions for children.

3. Study 2: Coloured Cube Aggregates

3.1. Test Development Introduction

The first test (RCCT) was designed to measure both object recognition and mental rotation in 6- to 10-year-old children, using single, three-dimensional, coloured cube images (Lütke & Lange-Küttner, 2015). It established an important baseline measure, demonstrating that young children could indeed process three-dimensional geometric stimuli, based on the cube’s perspective and distinctively coloured cube faces. Children averaged a high success rate of 90.1% correct responses in the object identification sections (A & B, see Figure 16). In later sections, where more difficult object identification tasks and mental rotation were measured, performance decreased further by 8.3% (C), followed by the most pronounced drop of 43% (D) with the removal of distinctive colours that made each cube unique in terms of colour (target and distracter cubes). Because children performed close to a ceiling effect on easier questions of the one-cube test (RCCT), children were expected to also be able to rotate larger cube aggregates. As mental capacity increases with age, aggregate set-size was systematically increased from four to six cube aggregates.

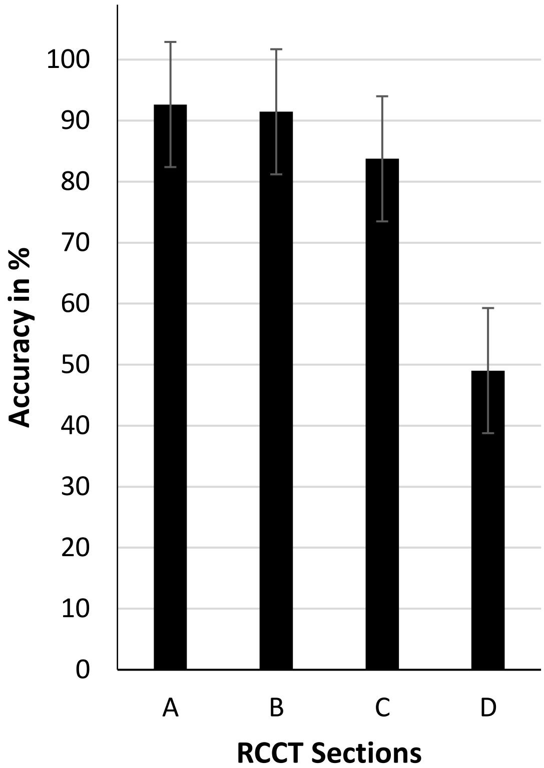


Figure 16 Decreases in task difficulty across RCCT sections

Moreover, again coloured stimuli were used as an adaptation towards children's perceptual preferences, similar to the coloured version of the Ravens Progressive Matrices test for children. The Coloured Mental Rotation Test (CMRT) is also based on a format akin to the Raven Coloured Progressive Matrices test (RCPM), in which children are encouraged to identify one correct target from a selection of six items.

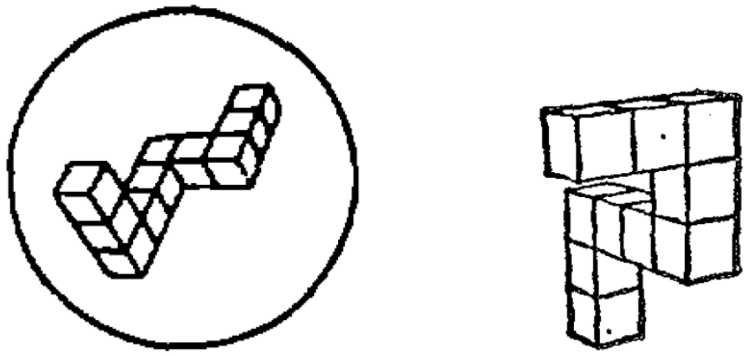
In short, the aim of the new test development (CMRT) was to produce a mental rotation test that: (1) has a more gradual increase in task difficulty across subtests; (2) modifies task difficulty though systematically varying critical object features (set-size, angularity, axis of rotation, aggregate dimensionality and colour congruence); and (3) bridges the gap between the established MRT for adults and a geometrically similar test for children. In comparison to the RCCT, which only uses single cubes, the new CMRT includes cube aggregates that gradually become larger in set-size.

In the current study it is expected, as in previous research, that children as young as 4 -to 5-year-olds can mentally rotate cube aggregates, that their ability improves with age, and that boys will outperform girls. With regards to the design factors the hypotheses are as follows: (1) Task difficulty increases with the number of cubes used per aggregate (set-size); (2) smaller rotations are easier than larger rotations (angularity); (3) picture-plane rotations are easier than in-depth-plane-rotations (rotational axes); (4) flatly aligned aggregates will be easier to rotate than those with elements protruding into depth.

Furthermore, predicted on the basis of the recent meta-analysis (Lauer et al., 2019) in which 3D rotation tasks produced larger gender differences than 2D tasks, in the current study the male advantage should show in the comparison between the Raven test which uses two-dimensional stimuli and the 3D colour cube test, with a performance difference observable in girls but not in boys.

3.2. Test Development Geometric Properties

Shepard and Metzler (1971) and Vandenberg and Kuse (1978) used computer generated 2D drawings of 3D cube aggregates, see Figure 17. As the original images were no longer accessible and only deteriorated copies were available, Peters et al. (1995) produced a redrawn version of the MRT which resulted in similar sex differences as the original MRT version. This latest MRT version by Peters is now widely used to measure mental rotation ability (Battista & Peters, 2010; Thompson, Nuerk, Moeller, & Kadosh, 2013; Titze, Heil, & Jansen, 2008; Voyer, Rodgers, & McCormick, 2004).



(A) (B)

Figure 17 Examples of: (A) Shepard and Metzler's (1971) and (B) Vandenberg and Kuse's (1978) test items. Note that the authors reduced cube size in relation to how far an aggregate element is distanced from the observer

Peters et al. (1995) MRT version provided better image quality in terms of pixelation and image sharpness, however it did not reproduce accurate image depth, because unlike in the original cube aggregates of Shepard and Metzler, the individual cubes were identically sized, see Figure 18, even if they were supposed to appear smaller with increasing distance from the foreground, diminishing in size in background.

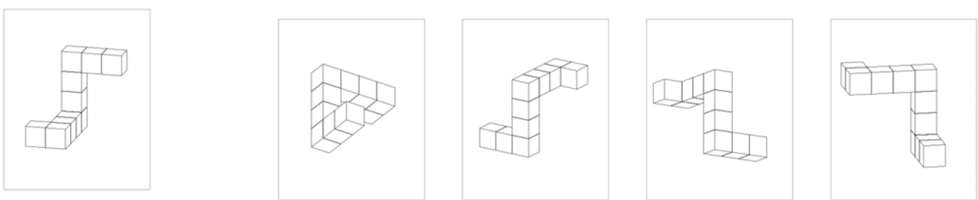


Figure 18 Peters (1995) redrawn MRT test

Using a two-point perspective and a horizon level, objects closest to the observer should be proportionately larger in size than those further away, providing the illusion of depth, see Figure 19.

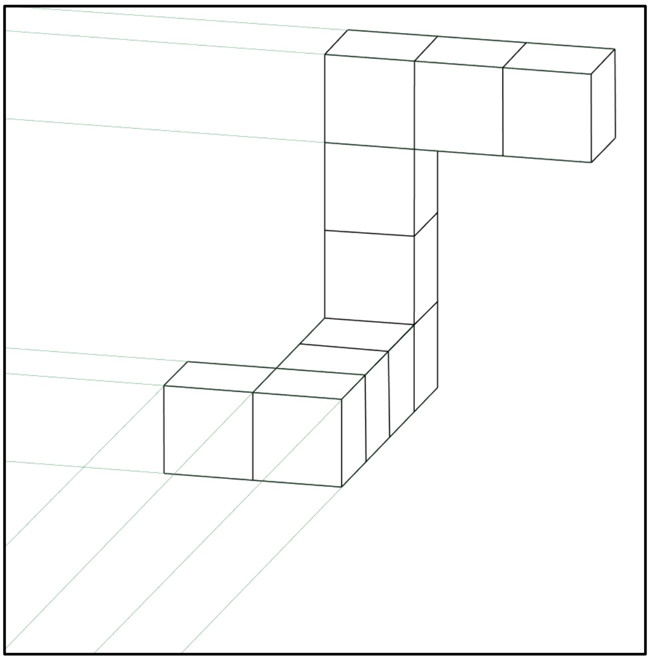


Figure 19 Enlarged image with improved image sharpness, but missing depth perspective (Peters et al., 1995). The cube closest to the observer is identical in size to the cube further away, making the cubes farther away appear larger in size.

A two-dimensional geometric drawing of a three-dimensional shape is a "compromise between appearance and structure" (Freeman, 1986). While the front of the cube aggregates provides the observer with the *true shape* and structure of the cube face, an oblique perspective edge signals a *change in direction* but not a change in structure. Side lengths are defined in relation to perspective lines and hence provide *consistent proportion* as in the real cube. The constant bottom line of the cube provides a *resting position* of an object signalling the stability of the object under gravity. As Peters et al (1995) aggregates did not follow these structural principles, the CMRT was newly designed using Adobe Illustrator, a vector-based design program. This allows 2D images of 3D objects to be drawn more realistically and with greater precision. Vectors are defined mathematically in a three-dimensional space and so the resulting images can be altered in size and follow precise perspective lines with no deterioration in image quality.

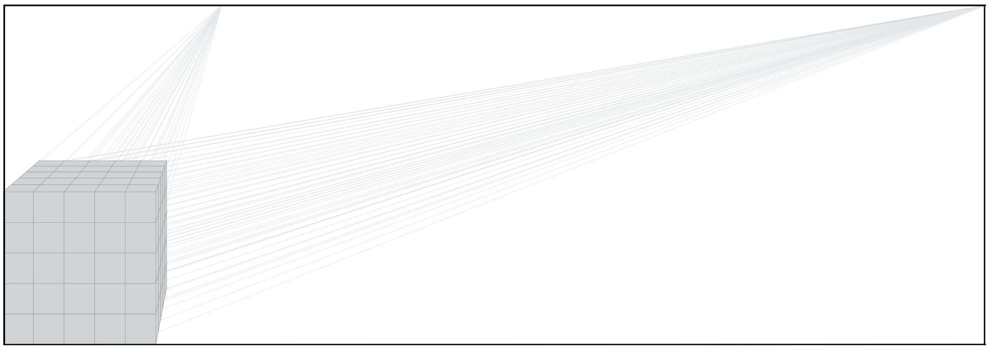


Figure 20 The basic cube used for cube aggregate derivations. CMRT 2-point depth perspective with horizon line.

Over 5000 vector-based individually placed anchor points define the spatial relations of two 5x5x5 cube aggregates consisting of 125 individual cubes each, see Figures 20 and 21. These aggregates were used to define the cube aggregates in the CMRT. It was important to use a vector-based graphics program, which works much like drawing lines by hand with a ruler, as at the time of design, no graphic suite provided the necessary precision to control all necessary design parameters. Although the design element of the test was very substantial, it was felt to be necessary in order to precisely specify angle of rotation and the realistic perspective of cube sizes. Similarly, the test page layout, colours, and task booklet were all created in Adobe Illustrator.

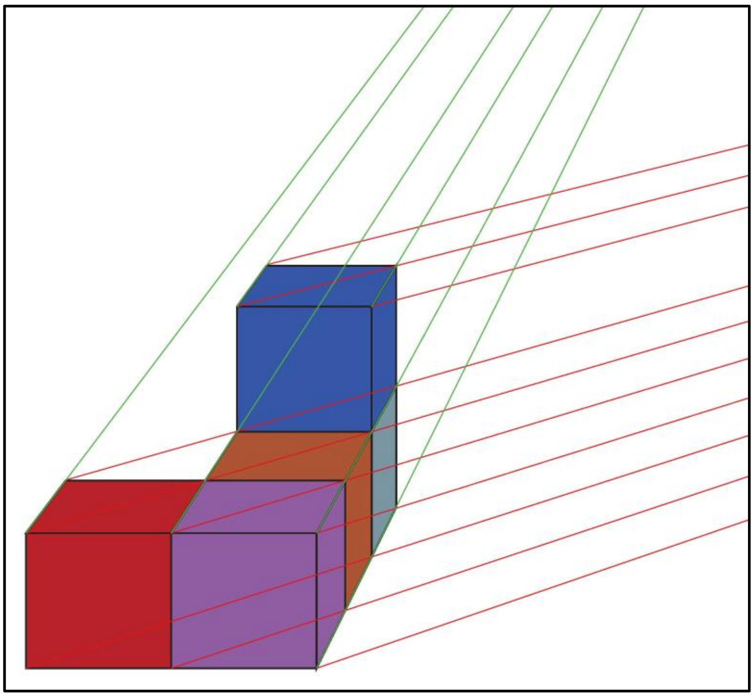


Figure 21 Enlarged aggregate, cube depth is defined by two perspective lines (red and green guide-lines).

Because information processing capacity increases with age, cube aggregate set-size was systematically increased. Compared to the 'adult size' of ten cube elements per aggregate, the aggregate size was reduced to four, five and six cube elements. The previous study used one rotated cube, but this material did not produce a gradual increase in task difficulty unless colour was removed as a cue (Lütke & Lange-Küttner, 2015).

Two-cube aggregates were not used in the test as this would have produced a straight object unlike the angular cube aggregates used with adults (see Figure 2, Metzler & Shepard, 1974; Vandenberg & Kuse, 1978). Three cubes could have generated an L shape object more similar to the angular adult cube aggregates but would still have fallen short of multiple 3D protrusions used in the MRT. Also, recent research with infants in the second year of life used a Z shape, rather than an L shape, to match the adult cube aggregates (Lauer, Udelson, Jeon, & Lourenco, 2015). Nevertheless, L-shapes were more difficult to rotate than concrete stimuli such as a hand or a face (Iachini et al., 2019). The design of 4-cube aggregates, although small, concisely emulated the much larger 10-cube aggregates for adults in terms of 3D geometric complexity.

In order to measure the effect of dimensionality on children's mental rotation ability (Bauer & Jolicoeur, 1996), two identically sized but differently designed types of cube aggregates were compared. One type was *flat* insofar as cubes were distributed in one depth-plane, and the other had *protruding* elements with cubes distributed in multiple depth planes, see Figure 22.

Why were not the same age ranges examined in both studies? Study 1 revealed that children in the 7-year-old age group already performed close to ceiling level with a 91.3% success rate on the easiest questions, and confirmed that children at this age range were already able to process 3D cube images (Lütke & Lange-Küttner, 2015). Also, in a study by Hawes, LeFevre, et al. (2015) published in the same year, 5-year-olds performed above chance level on a mental rotation task using real-life, tangible 3D cube aggregates. Therefore, a younger age group was included in the second study.

Socio-economic status of the children's family context could not be included in Study 2. Universal free school meals were introduced from 2015 onwards for all children up to grade 2, corresponding to 6- to 7-years of age, therefore, for Study 2, government data about the number of children in schools from families receiving financial state subsidies were no longer available.

3.3. Method

3.3.1. Participants

Power analysis with G*Power version 3.1.9.2. showed that with four age groups as a between-subject factor and four repeated measurements for angularity (mental rotation), an effect size of $\eta^2 = .25$, power $(1 - \beta) = .95$ and an α level of .05, a sample size of $N = 76$ is required. Thus, a gender-balanced sample of $N = 80$, with 40 boys and 40 girls from schools in West London was recruited. The sizes of groups, means and standard deviations of children's age in years and months as well as the number of boys and girls are reported in Table 14.

Parental consent was obtained, and children were informed that they were free to withdraw from the study at any time they wished. Data were anonymised at source, with only sex and date of birth of children registered.

Table 14

Age Groups (Mean in Years; Months)

	Boys			***Girls***			
Age groups	n	*M*	*SD*	n	*M*	*SD*	**Total**
4-5 years	10	5;4	0;3	10	5;3	0;4	20
6-7 years	10	6;9	0;5	10	7;1	0;6	20
8-9 years	10	9;1	0;7	10	8;9	0;6	20
10-11 years	10	11;1	0;5	10	10;9	0;5	20
Total	40			40			80

Note. *M* = Mean, SD = Standard deviation

3.3.2. Apparatus and Material

Coloured Mental Rotation Test (CMRT). The were 36 A4 sized pages in which tasks increased in complexity across sections (A-F). Prior to the test trials, participants were presented with 3 practice questions, with one example form each section. The number of cubes per aggregate increases from 4-6 cubes (see Figure 22), that is, there are three aggregate set-sizes in total. Each set of cube aggregates is presented once connected in a one-dimensional flat depth-plane (A, C, E) and once as an aggregate where some cubes protrude into depth (B, D, F).

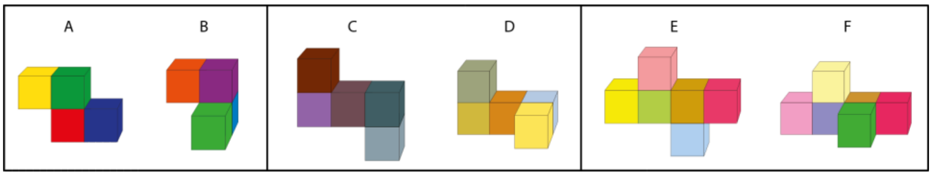
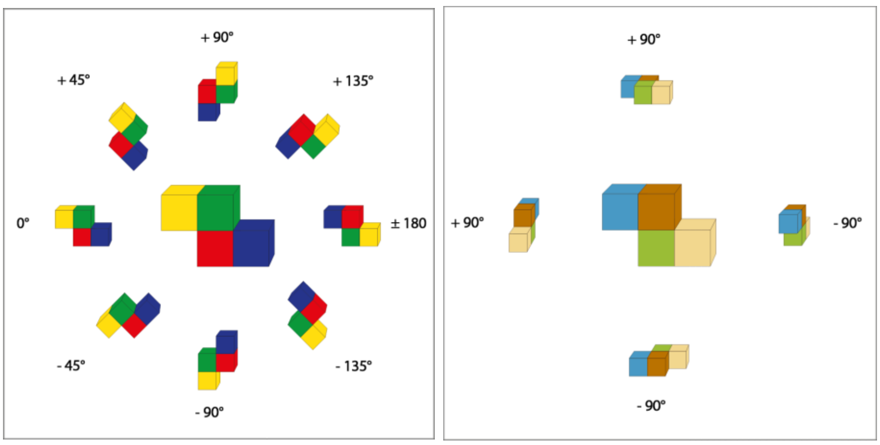


Figure 22 Coloured edge points in subtests A-F

The task was to find the rotated target amongst distracters. Angularity increased within each section on each page in a sequential manner. The first four pages in each section showed rotated targets in small steps within the range of 45°-180° first for a flat cube (Figure 22 A, C, E) followed by the cube with protruding elements (Figure 22 B, D, F), respectively. The last two pages in each section showed targets with 90° rotations around the x- and y-axis, respectively, see Figure 23B.



A B

Figure 23 Figure A shows picture rotations in 45° steps. Figure B shows 45° rotations around the y- and x-axis.

Focussing on aggregate end points has been reported as a strategy to identify targets (Krüger et al., 2014). Performance increased when human body part cues (heads, feet) were attached at the end of cube aggregates, with central elements of the aggregates being neutral cubes. Hence, due to the sensitivity to aggregate endpoints, the cube colours on each page were systematically varied, see Figure 24. Initially only two distracter aggregates have the same endpoint colours as the target, however, this gradually increases to five distracters on each page within each subtests A-F. This reduces the possible impact of analytic endpoint strategy and supports a more holistic approach.

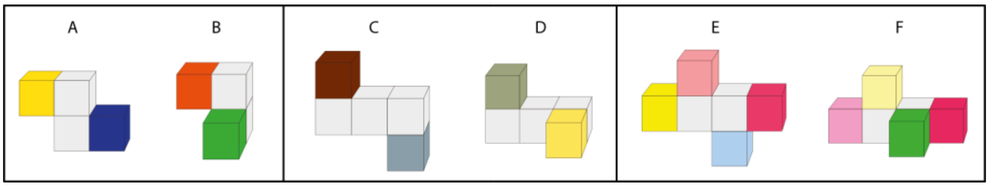


Figure 24 Coloured edge points in subtests A-F

Adobe illustrator is as it is a vector-based graphics program, which allows mathematical precision when specifying design elements such as cube size, rotation, and orientation, see Figure 25. A one-point perspective with an additional focus point on the perspective horizon, provided the spatial framework for designing the CMRT cube aggregates.

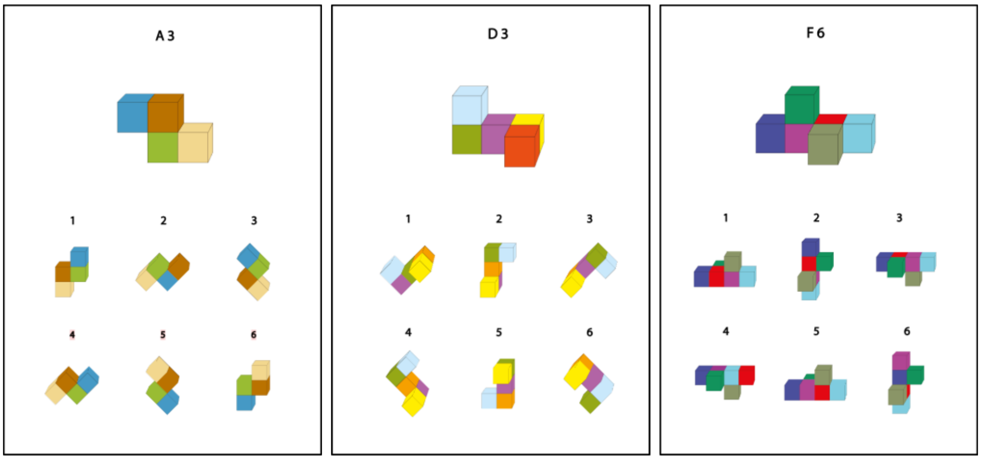


Figure 25 Coloured Mental Rotation Test with 4-, 5- and 6-cube aggregates. Correct answers: A3 = 4, D3 = 3, F6 = 5.

In summary, features that lead to the greater task complexity are: (1) increases in set-size, (2) degree of rotation, (3) depth dimensionality of cube aggregates, (4) location of cube colours and levels of endpoint similarity between the target and distracter cubes. The latter one was not explicitly tested, but instead it was assumed that the gradual and controlled disappearance of this cue would lead to a gradually increased difficulty. Colours were selected on the basis that similar colours should not repeat within a section (Lange-Küttner & Küttner, 2015).

3.3.3. Procedure

Children were tested individually in a quiet, familiar setting at their school. They received feedback on example questions prior to the test proper. During the test, participants did not receive any feedback, but were only encouraged to take their time to look at all possible answers. After finishing, they could choose a sticker as a reward. Answers were recorded by the researcher on a test response sheet. Scores were added up by two researchers independently. No disagreement was found.

Task instructions. In the warm-up phase children were asked, “Do you want to play a game?” and were then presented with the task booklet, and asked, “Do you know what this is?” Thereafter, children first solved three practice trials in which the experimenter pointed at the enlarged target aggregate and asked the participant, “Which cube is the same as this one?” The practice questions were repeated until the child could identify the correct aggregate. Children then proceeded to the task proper.

3.4. Results Study 2

The report of the results begins with an analysis of variance for picture-plane rotations (45°, 90°, 135° and 180°), in which aggregates with different number of cubes (size) and differently attached cube aggregates (flat vs. protruding) are compared. The same analysis is then conducted for different types of rotations around 90^{o} axis. In order to control the shared variance between fluid intelligence and mental rotation, both models are re-run with the Raven as a covariate. The correlations between the CMRT and the RCPM were also analysed in order to validate and compare task difficulty of this new test on spatial ability and mental rotation.

When the Mauchly’s test of sphericity showed a significant violation of the normal distribution, degrees of freedom were adjusted with Greenhouse-Geisser. Post-hoc tests were conducted within the model (Bonferroni-corrected) or as pairwise or independent samples t-tests, when interactions were significant. When the factor angularity was significant, polynomial trends from within the model are reported. The report of the ANOVA begins with the between-subject group effects followed by the within-subject task effects and interactions.

3.4.1. CMRT Picture-Plane Rotations (45°, 90°, 135° and 180°)

A 3 (set-size) by 2 (aggregate-depth) by 4 (angularity) by 4 (age groups) by 2 (sex) analysis of variance was carried out, with repeated measures for set-size, angularity, and aggregate-depth. The between-subject factors are age and sex. The first CMRT factor, angularity (45º, 90º, 135º and 180º), has four levels: The angle of picture-plane rotations is increased by 45° increments up to 180°. The second CMRT factor, axis of rotation, compares 90° rotations around three distinct axes: Picture-plane and two depth planes (x- and y-axis). The third factor, set-size, defines the number of cubes per aggregates and has three levels: Aggregates consist of four, five and six cubes. The fourth factor, aggregate-depth, has 2 levels: Aggregates distributed into one- and two-depth-planes. The group means for the age groups are listed in Table 15 for each section of the test.

The statistical results are listed in Table 16. A highly significant and reliable main effect for sex $F(1, 72) = 15.49$, $p < +.001$, $\eta^2 = .93$, showed that in general boys ($M = 77.3\%$) outperformed girls ($M = 62.8\%$).

There was also a large and highly significant main effect for age groups $F(3, 72) = 34.32$, $p < .001$, $\eta^2 = .55$, as performance increased with age (4- to 5 years $M = 39.0\%$; 6- to 7-years $M = 72.7\%$; 8- to 9-years $M = 81.5\%$; 10- to 11-years $M = 87.0\%$). Post-hoc comparisons showed that performance of 4- to 5-year-olds differed significantly from all other age groups, and that performance of 6- to 7-year-olds also differed significantly from 10- to 11-year-olds.

Table 15

CMRT picture-plane rotation by Age Group (Accuracy in per cent)

Age	4-5 years	6-7 years	8-9 years	10-11 years	**Average**
			Section A		
45°	85.0 (36.6)	100.0 (0.0)	100.0 (0.0)	95.0 (22.4)	95.0 (14.8)
90°	55.0 (51.0)	65.0 (48.9)	90.0 (30.8)	90.0 (30.8)	75.0 (40.4)
135°	40.0 (50.1)	65.0 (48.9)	55.0 (51.0)	90.0 (30.8)	62.5 (45.3)
180°	25.0 (44.4)	50.0 (51.3)	60.0 (50.3)	75.0 (44.4)	52.5 (47.6)
Average	51.3 (45.6)	70.0 (37.4)	76.3 (33.0)	87.5 (32.1)	71.3 (37.0)
			Section B		
45°	70.0 (47.0)	90.0 (30.8)	100.0 (0.0)	100.0 (0.0)	90.0 (19.5)
90°	45.0 (51.0)	80.0 (41.0)	80.0 (41.0)	100.0 (0.0)	76.3 (33.3)
135°	60.0 (50.3)	90.0 (30.8)	85.0 (36.6)	100.0 (0.0)	83.8 (29.4)
180°	45.0 (51.0)	65.0 (48.9)	85.0 (36.6)	95.0 (22.4)	72.5 (39.7)
Average	55.0 (49.8)	81.3 (37.9)	87.5 (28.6)	98.6 (05.6)	80.6 (30.5)
			Section C		
45°	45.0 (51.0)	75.0 (44.4)	95.0 (22.4)	85.0 (36.6)	75.0 (38.6)
90°	50.0 (51.3)	80.0 (41.0)	85.0 (36.6)	80.0 (41.0)	73.8 (42.5)
135°	15.0 (36.6)	60.0 (50.3)	70.0 (47.0)	75.0 (44.4)	55.0 (44.6)
180°	15.0 (36.6)	75.0 (44.4)	85.0 (36.6)	100.0 (0.0)	68.8 (29.4)
Average	31.3 (43.9)	72.5 (45.0)	83.6 (35.7)	85.0 (30.5)	68.1 (38.8)
			Section D		
45°	20.0 (41.0)	45.0 (51.0)	80.0 (41.0)	85.0 (36.6)	57.5 (42.4)
90°	35.0 (48.9)	65.0 (48.9)	85.0 (36.6)	80.0 (41.0)	66.3 (43.9)
135°	50.0 (51.3)	90.0 (30.8)	75.0 (44.4)	90.0 (30.8)	76.3 (39.3)
180°	20.0 (41.0)	80.0 (41.0)	90.0 (30.8)	90.0 (30.8)	70.0 (35.9)
Average	31.3 (45.6)	70.0 (43.0)	82.5 (38.2)	86.3 (24.8)	67.5 (40.4)
			Section E		
45°	55.0 (51.0)	80.0 (41.0)	90.0 (30.8)	85.0 (36.6)	77.5 (39.9)
90°	25.0 (44.4)	70.0 (47.0)	60.0 (50.3)	85.0 (36.6)	60.0 (44.6)
135°	35.0 (48.9)	75.0 (44.4)	85.0 (36.6)	90.0 (30.8)	71.3 (40.2)
180°	35.0 (48.9)	65.0 (48.9)	85.0 (36.6)	85.0 (36.6)	67.5 (42.8)
Average	37.5 (48.3)	72.5 (45.4)	80.0 (38.6)	86.3 (35.2)	69.1 (41.9)
			Section F		
45°	35.0 (48.9)	65.0 (48.9)	75.0 (44.4)	80.0 (41.0)	63.8 (45.8)
90°	30.0 (47.0)	65.0 (48.9)	70.0 (47.0)	75.0 (44.4)	60.0 (46.9)
135°	20.0 (41.0)	75.0 (44.4)	85.0 (36.6)	75.0 (44.4)	63.8 (41.6)
180°	25.0 (44.4)	75.0 (44.4)	85.0 (36.6)	85.0 (36.6)	67.5 (40.5)
Average	27.5 (45.4)	70.0 (45.7)	78.8 (41.1)	78.7 (41.6)	63.8 (43.7)

The two factors sex and age groups interacted significantly, $F(3, 72) = 3.37$, $p < .05$, $\eta^2 = .12$. To investigate this two-way interaction further, the ANOVA was re-run as a split sample analysis for each age group. Sex differences were observed in 4- to 5- and 8- to 9-year-old age groups, in both of which boys performed significantly better than girls, see Figure 26.

The within-subject effects showed that set-size was significant as a main effect, $F(2, 144) = 9.80$, $p < .001$, $\eta^2 = .12$, but also in multiple interactions. Performance decreased in relation to cube aggregate size, from 4 cubes (M = 75.9%), to 5 cubes (M = 67.8%), to 6 cubes (M = 66.4%). Pairwise post-hoc comparisons indicated that only aggregates with 4 cubes were significantly easier than both 5 and 6 cube aggregates.

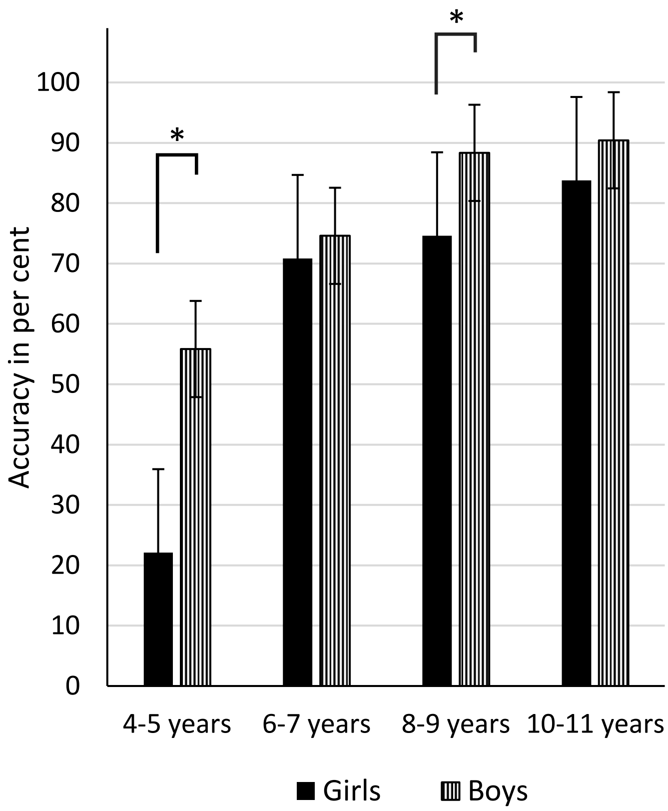


Figure 26	Accuracy per age group and sex

Table 16

Analysis of variance for performance on picture-plane rotational differences, depth distribution and set-size, with age and sex as between subject variables

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subject Effects			
Rota	**27807.29**	**2.670**	**5.743**	**.001****	**.07**
Rota*Sex	98.96	3.000	0.020	.996	.00
Rota*Age	**27796.88**	**9.000**	**1.914**	**.051(*)**	**.07**
Rota*Sex*Age	15255.21	9.000	1.050	.401	.04
Size	**33885.42**	**2.000**	**9.798**	**.000*****	**.12**
Size*Sex	3447.92	2.000	0.997	.372	.01
Size*Age	**27156.25**	**6.000**	**2.617**	**.019***	**.10**
Size*Sex*Age	9843.75	6.000	0.949	.462	.04
Dep	630.21	1.000	0.424	.517	.01
Dep*Sex	15.05.208	1.000	1.012	.318	.01
Dep*Age	1765.63	3.000	0.396	.756	.02
Dep*Sex*Age	4473.96	3.000	1.003	.397	.04
Size*Dep	**18010.42**	**2.000**	**5.430**	**.005****	**.07**
Size*Dep*Sex	822.92	2.000	0.248	.781	.00
Size*Dep*Age	2281.25	6.000	0.229	.967	.01
Size*Dep*Sex*Age	3385.42	6.000	0.340	.915	.01
Size*Rota	**58364.58**	**4.625**	**8.541**	**.000*****	**.11**
Size*Rota*Sex	5635.42	6.000	0.825	.551	.01
Size*Rota*Age	31593.75	18.000	1.541	.072	.06
Size*Rota*Sex*Age	**39072.92**	**18.000**	**1.906**	**.014***	**.07**
Dep*Rota	**39765.63**	**3.000**	**10.419**	**.000*****	**.13**
Dep*Rota*Sex	1807.29	3.000	0.474	.701	.01
Dep*Rota*Age	11255.21	9.000	0.983	.455	.04
Dep*Rota*Sex*Age	86.30.208	9.000	0.754	.659	.03
Size*Dep*Rota	**19156.25**	**6.000**	**2.399**	**.027***	**.03**
Size*Dep*Rota*Sex	2177.08	6.000	0.273	.950	.00
Size*Dep*Rota*Age	17885.42	18.000	0.747	.762	.03
Size*Dep*Rota*Sex*Age	25947.92	18.000	1.083	.366	.04
		Between-Subjects Effects			
Sex	**81000.00**	**1.000**	**15.486**	**.000*****	**.93**
Age	**465222.22**	**3.000**	**34.32**	**.000*****	**.55**
Sex*Age	**50222.22**	**3.000**	**3.373**	**.023***	**.12**

Note. Degrees of Freedom were corrected with Greenhouse Geisser. Rota = Picture-plane angular differences; Age = Age groups; Size = Number of cubes per Aggregate; Dep = Aggregate depth distribution. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant statistical effects are set in bold.

Set-size and age groups interacted significantly, $F(6, 144) = 2.62$, $p < .05$, $\eta^2 = .10$, see Figure 27. The split sample analysis by age groups showed that only in 4- to 5-year-olds did differences in set-size significantly increase task difficulty. Pairwise post-hoc comparisons between set-sizes revealed that in 4- to 5-year-old children, 4-cube aggregates (M = 53.1%) were significantly easier than both 5-cube (M = 31.3%) and 6-cube-aggregates (M = 32.5%). A clearly graded effect of task difficulty according to set-size only showed in the 10- to 11-year-old children.

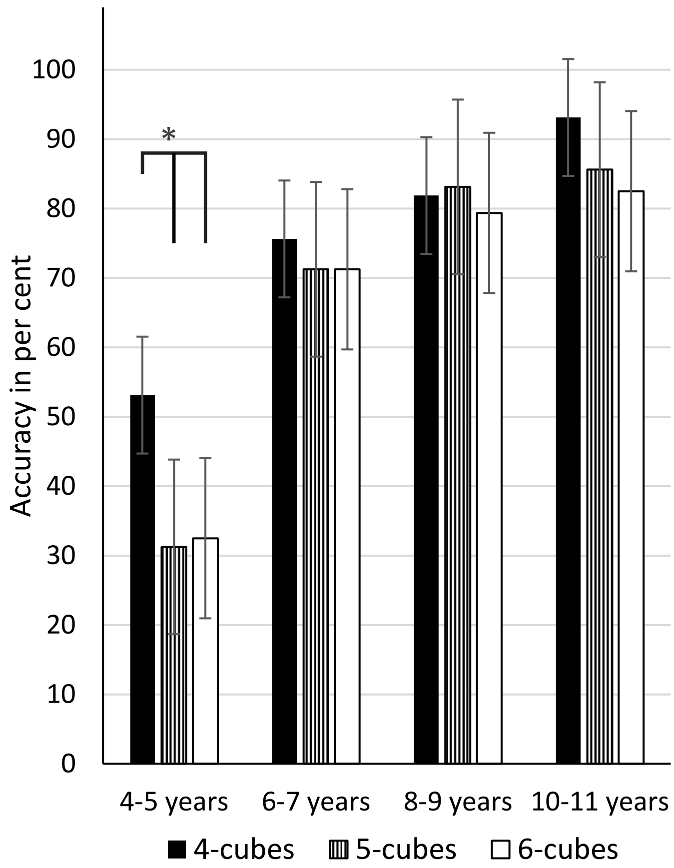


Figure 27 Development of the CMRT scores per age group and set-size.

The two-way interaction between set-size and aggregate-depth, $F(2, 144) = 5.43$, $p < .01$, $\eta^2 = .07$, showed that performance involving the smallest 4-

cube aggregates was significantly better if these had protruding cubes (*M* = 80.6 %) compared to those without (*M* = 71.3%), see Figure 28. This was not the case for larger 5- and 6-cube aggregates. The 4-cube aggregate with protruding cubes will be called the Good Gestalt cube in the following text.

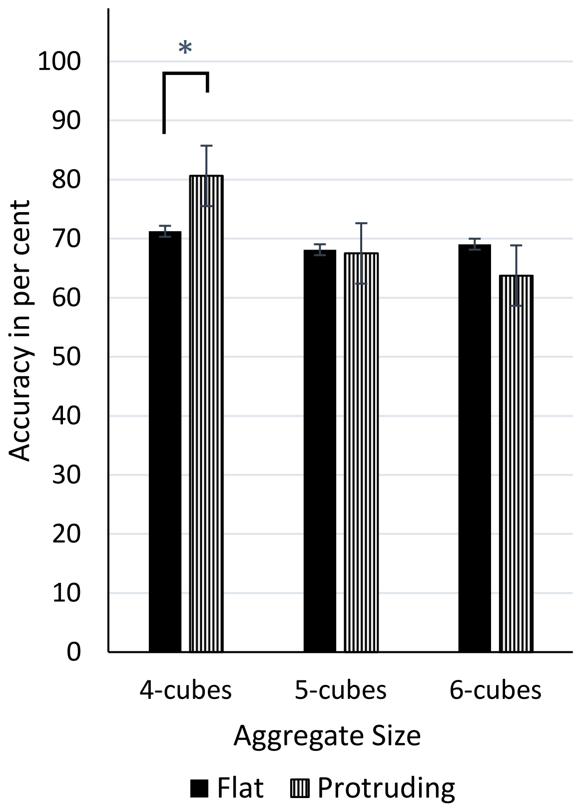


Figure 28 Development of the CMRT scores per set-size and protrusion.

An important finding for the mental rotation test was the significant effect for angularity, *F*(2.67, 192.22) = 5.74, $p = .001$, $\eta^2 = .07$, showing a performance decrease with increases in angularity. Polynomial contrasts of the angularity effect showed a significant linear trend (1, 72) = 9.80, $p < .05$, $\eta^2 = .12$ (no Figure). Post-hoc comparisons showed that only 45º rotations (*M* = 76.5%) differed significantly from 90º (*M* = 68.5%), 135º (*M* = 68.8%) and 180º (*M* = 66.5%) rotations. This showed that only the smallest 45º rotation is significantly easier than larger rotations. For larger rotations, other factors such as set-size and Good Gestalt had a greater impact on task difficulty than degree of rotation in children, see the following paragraphs.

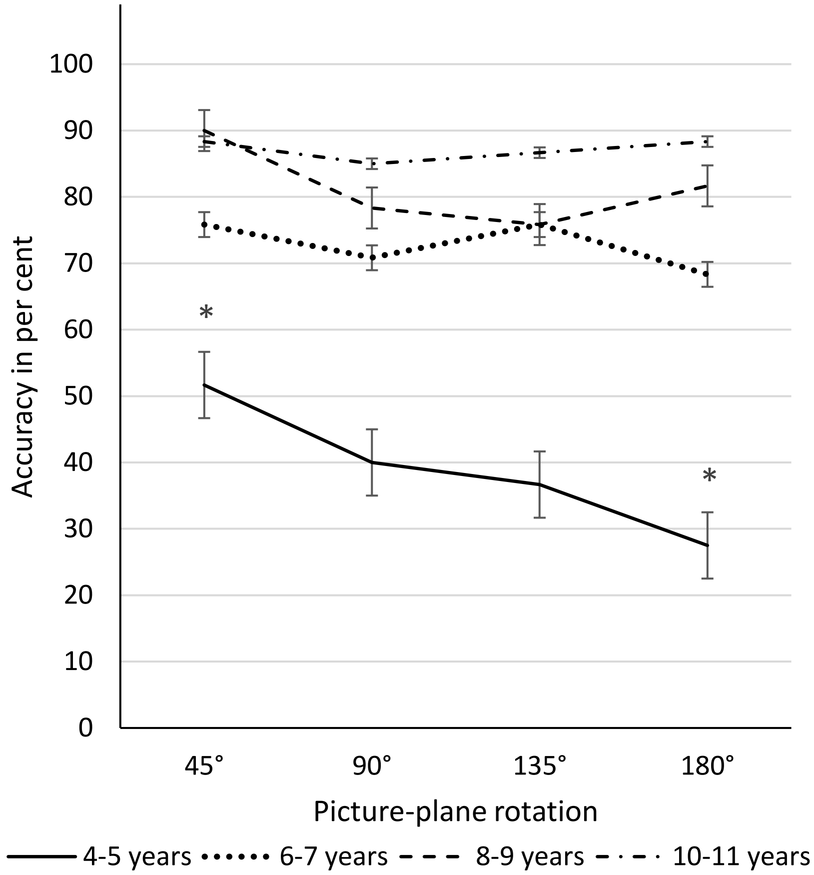


Figure 29 Performance by picture-plane rotations and age groups.

There was a marginally significant two-way interaction between angularity and age groups, $F(8.01, 192.22) = 1.91$, $p = .051$, $\eta^2 = .07$, see Figure 29. Polynomial contrasts of the angularity effect showed a significant linear trend (3, 72) = 3.00, $p < .05$, $\eta^2 = .11$. The split sample analysis by age groups showed the predicted angularity effect only in the youngest 4- to 5-year-olds (45º *M* = 51.7%; 90º *M* = 40%; 135º *M* = 36.7%; 180º *M* = 27.5%), while the older children kept their performance level. Post-hoc comparisons of the performance levels of the 4- to 5-year-old children showed that accuracy on the 45º rotations significantly differed from 180º rotations.

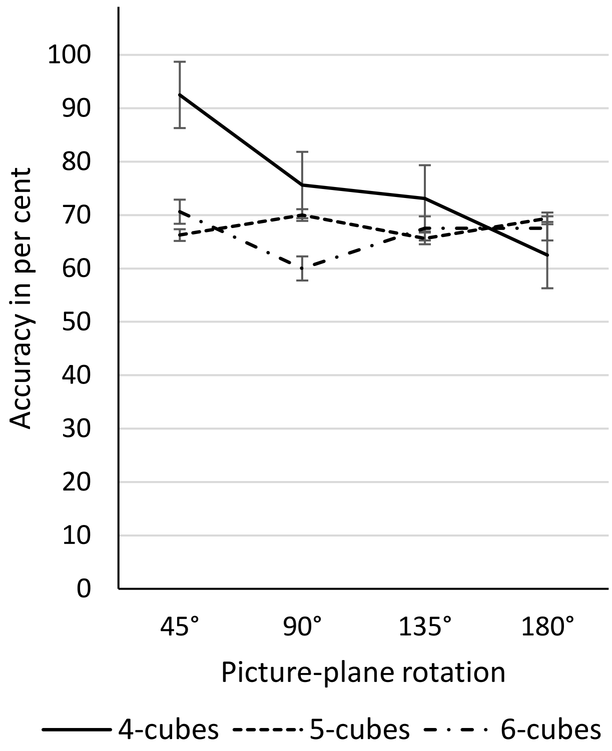


Figure 30 Performance by angularity and set-size.

There was a significant two-way interaction between set-size and angularity *F*(4.63, 333.02) = 8.54, p < .001, see Figure 30. Polynomial contrasts of the angularity effect showed a significant linear trend (1, 72) = 29.21, $p < .001$, $\eta^2 = .29$. A linear decrease in performance occurred on 4-cube aggregates over increases in angularity (45º *M* = 95.0%; 90º *M* = 75.0%; 135º *M* = 62.5%; 180º *M* = 52.5%), whereas for larger 5- (45º *M* = 75.0%; 90º *M* = 73.5%; 135º *M* = 55.0%; 180º *M* = 68.8%) and 6-cube aggregates (45º *M* = 77.5%; 90º *M* = 60.0%; 135º *M* = 71.3%; 180º *M* = 67.5%), performance was lower but remained constant.

There was a significant interaction between angularity and aggregate-depth, $F(3, 216) = 10.42$, $p < .001$, $\eta^2 = .13$. Polynomial contrasts showed a significant linear trend (1, 72) = 20.37, $p < .001$, $\eta^2 = .21$. Performance only decreased with greater angularity for aggregates without protruding cubes, see Figure 31. This confirms that the mental rotation tests for children using just flat pictures in previous research (Jansen et al., 2011; Marmor, 1975; Quaiser-Pohl, 2003) were appropriate.

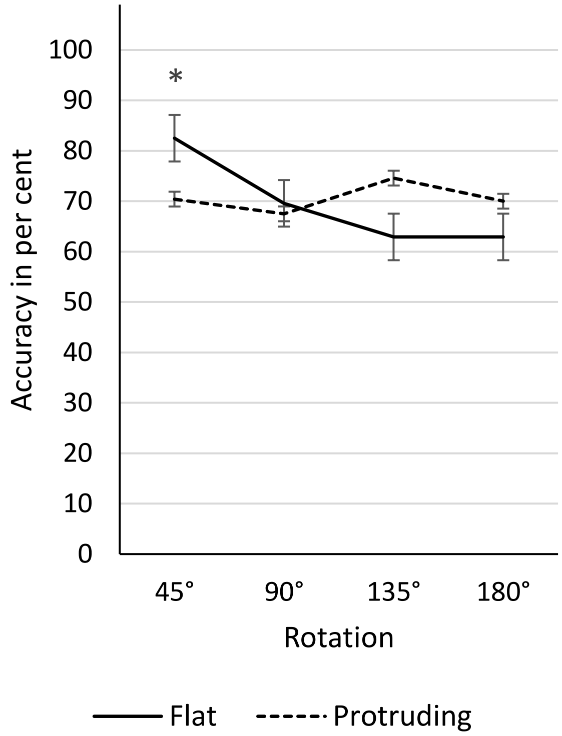
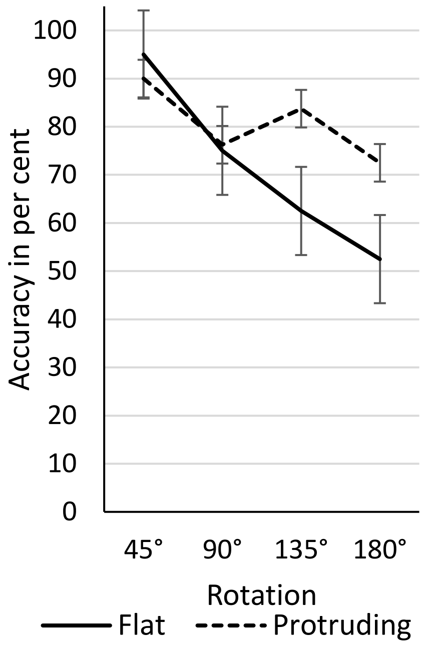
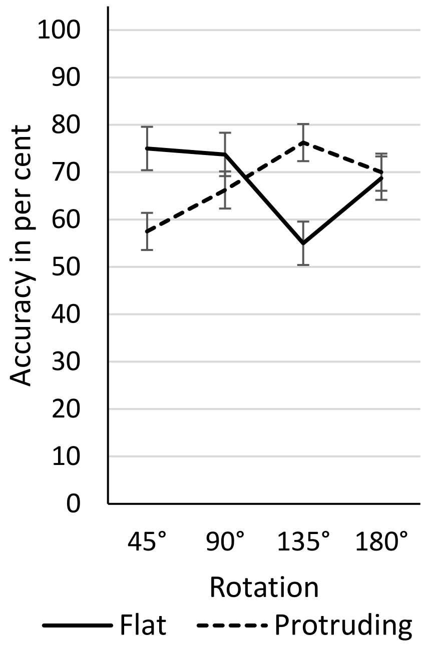


Figure 31 CMRT scores by picture-plane rotations and aggregate depth.

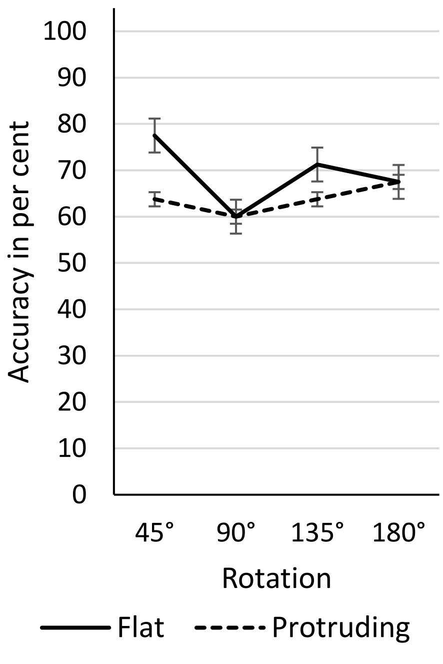
There was a significant three-way interaction between set-size, aggregate-depth, and angularity, $F(6, 432) = 2.40$, $p < .05$, $\eta^2 = .03$, see Figure 32. Polynomial contrasts showed a significant linear decrease only for 4-cube aggregates which were flat without protruding elements. Polynomial contrasts showed a significant cubic trend (1, 72) = 4.06, $p < .05$, $\eta^2 = .05$.



(A) 4-cube aggregate



(B) 5-cube aggregates



(C) 6-cube aggregates

Figure 32 Linear decrease of performance with increase in angularity.

The significant four-way interaction between set-size, angularity, sex and age groups, $F(18, 432) = 1.91$, $p < .05$, $\eta^2 = .07$, was analysed further with a split sample ANOVA by Sex and Age groups.

For girls, only 6- to 7- and 8- to 9-year-olds showed a significant graded effect for task difficult with 4-cube aggregates, see Figure 33 A-D. The task appeared to be too difficult for very young girls in the 4- to 5-year-old age group, only when rotating larger 5- to 6-cube aggregates with performance below chance ($M < 16.7\%$).

In contrast, for boys, mental rotation of the Good Gestalt 4-cube aggregate occurred in 4- to 5- and 10- to 11-year olds, see Figure 34 A-D. Surprisingly, the 4- to 5-year-old boys were already more than 90% correct for the easiest 4-cube aggregate, and then showed the strongest mental rotation decrease with increasing angularity, see Figure 34A. In the older age groups, this decrease was less pronounced, see Figure 34 B-D. Thus, the boys appeared to be ahead of the girls by about one year. Moreover, the angularity effect was still visible in the oldest age group of boys at a very high accuracy level.

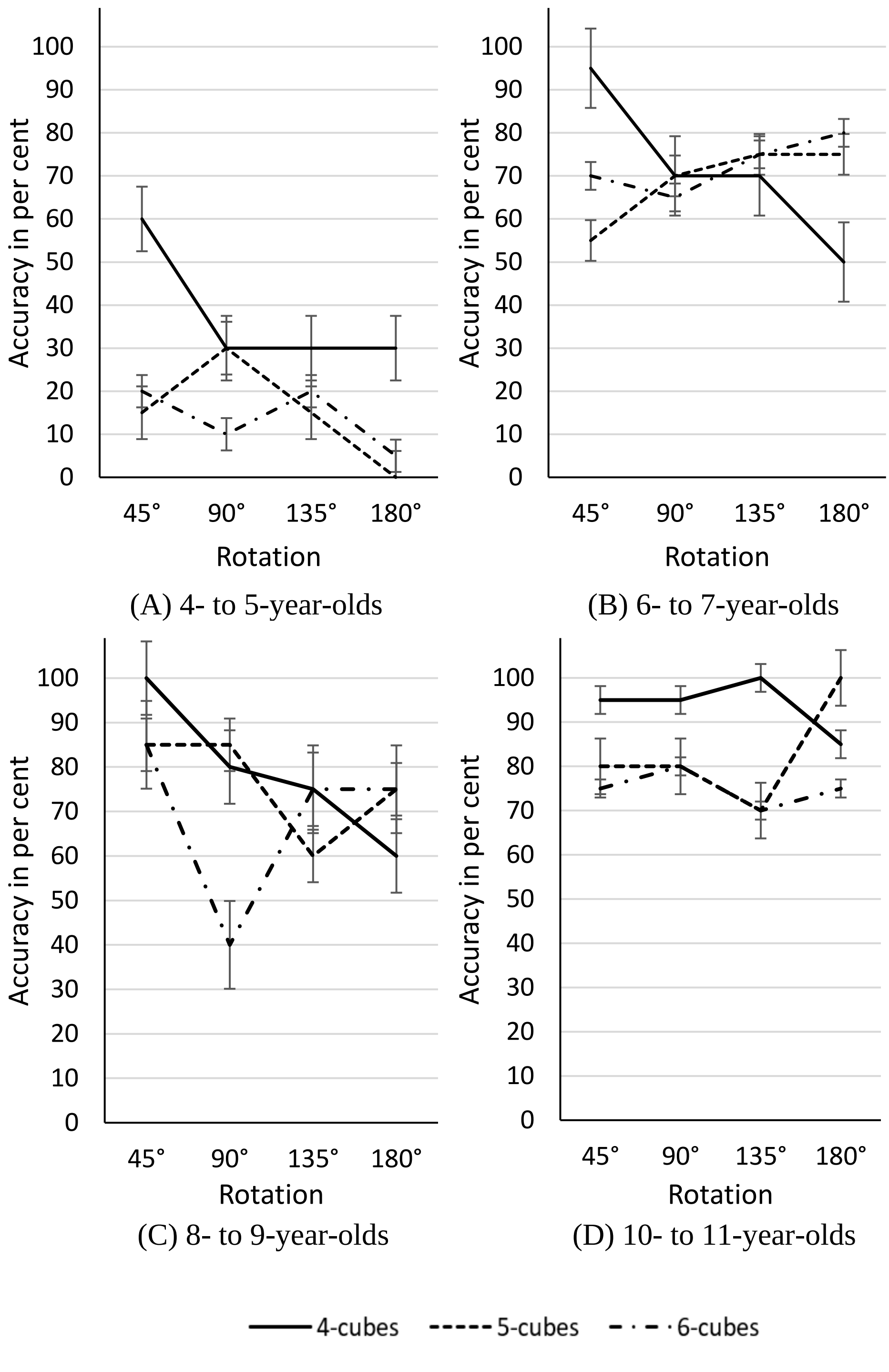
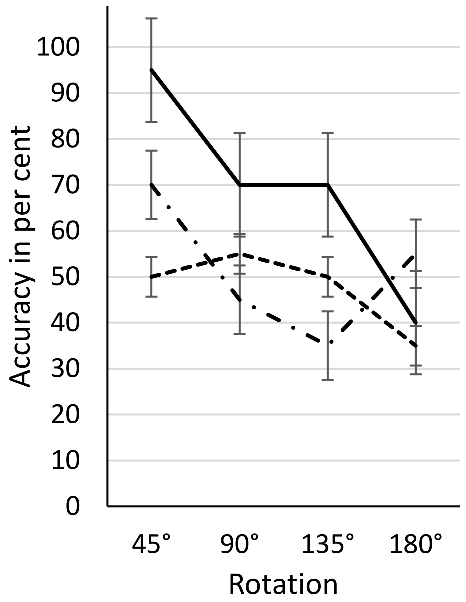
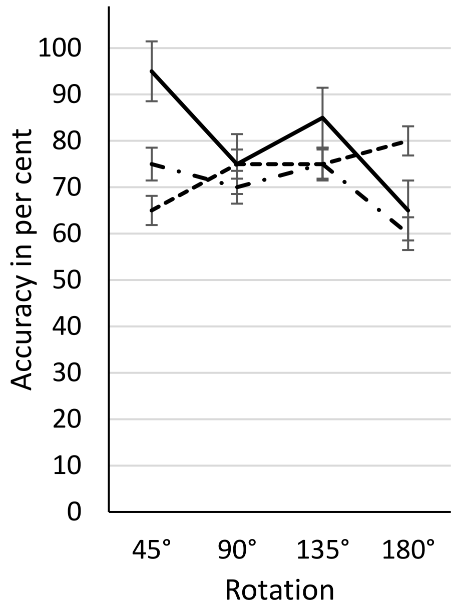


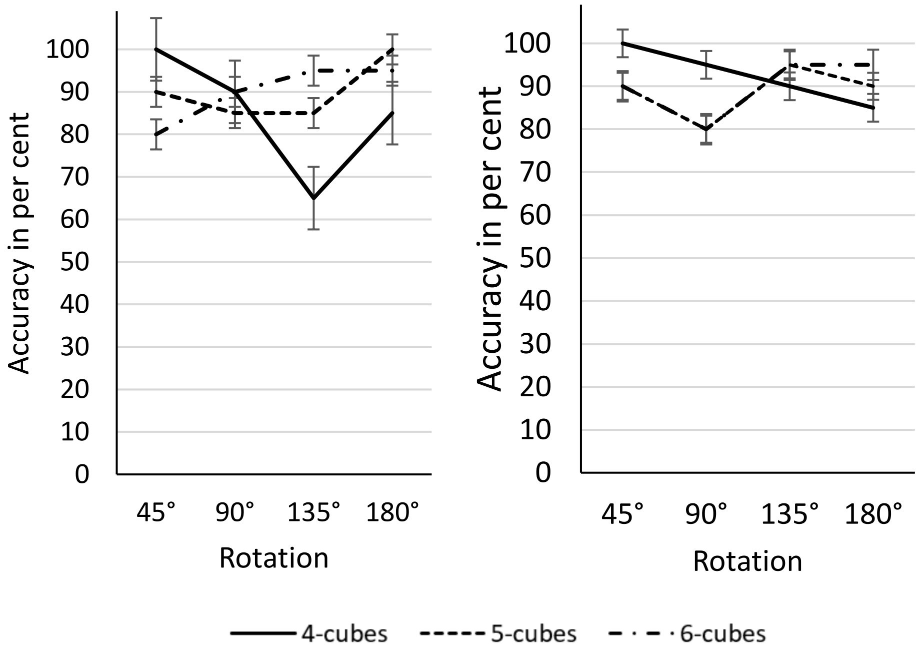
Figure 33 Girls' development of the CMRT by age and cube aggregate size.



(A) 4- to 5-year-olds



(B) 6- to 7-year-olds



(C) 8- to 9-year-olds (D) 10- to 11-year-olds

Figure 34 Boys development of the CMRT by Sex, Age groups and Aggregate size).

3.4.2. CMRT 90º Picture-plane, Y-axis, and X-axis Rotations

So far, the reduction in accuracy resulting from increases in angularity have been explored in relation to set-size and protrusion of elements into 3D (aggregate-depth). The following analysis investigates whether the rotational axes play a role in addition to: Comparing 90° rotations around the vertical, the horizontal and the picture-plane axis, see Figure 23 B on page 82.

A 3 (set-size) by 2 (aggregate-depth) by 3 (rotational axis) by 4 (age group) by 2 (sex) analysis of variance was carried out, with repeated measures for set-size, aggregate-depth, and axis. Again, the between-subject group effects are reported before the within-subject task effects. Mean scores per rotation and age groups can be found in Table 17.

There was also a large and highly significant main effect for age groups with a larger effect size, $F(3, 72) = 29.54$, $p < .001$, $\eta^2 = .55$, as performance increased with age. Post-hoc comparisons indicated that performance of 4- to 5-year-olds (*M* = 31.9%) was significantly lower than all other age groups. Performance of 6- to 7-year-olds (*M* = 61.9%) differed significantly from 10- to 11-year-olds (*M* = 79.7%). The two factors sex and age groups interacted two-way, $F(3, 72) = 3.19$, $p < .05$, $\eta^2 = .12$. The split sample analysis for each age group showed significant differences in two age groups in favour of boys, in 4- to 5-year-olds (girls *M* = 16.1%; boys *M* = 47.8%) with the girls performing below chance and a smaller difference in 8- to 9-year-olds (girls *M* = 60.6%; boys *M* = 81.1%).

The statistical results are shown in an overview in Table 18. A significant main effect for sex, but this time with a smaller effect size $F(1, 72) = 15.43$, $p < .001$, $\eta^2 = .18$, showed that boys (*M* = 68.6%) outperformed girls (*M* = 53.6%) in this selective task comparison.

Table 17

CMRT 90° axis rotations by Age Group (Accuracy in per cent)

Age	4-5 years	6-7 years	8-9 years	10-11 years	**Total**
			Section A		
90° picture	55.0 (51.0)	65.0 (48.9)	90.0 (30.8)	90.0 (30.8)	75.0 (40.4)
90° y-axis	60.0 (50.3)	95.0 (22.4)	80.0 (41.0)	100.0 (0.0)	83.8 (28.4)
90° x-axis	25.0 (44.4)	45.0 (51.0)	55.0 (51.0)	70.0 (47.0)	48.8 (48.4)
Total	48.6 (48.6)	68.3 (41.0)	75.0 (41.0)	86.7 (25.9)	17.3 (39.1)
			Section B		
90° picture	45.0 (51.0)	80.0 (41.0)	80.0 (41.0)	100.0 (0.0)	76.3 (33.3)
90° y-axis	40.0 (50.3)	70.0 (47.0)	70.0 (47.0)	70.0 (47.0)	62.5 (47.8)
90° x-axis	20.0 (41.0)	15.0 (36.6)	25.0 (44.4)	50.0 (51.3)	27.5 (43.4)
Total	35.0 (47.4)	55.0 (41.6)	58.3 (44.2)	73.3 (32.8)	55.4 (41.5)
			Section C		
90° picture	50.0 (51.3)	80.0 (41.0)	85.0 (36.6)	80.0 (41.0)	73.8 (42.5)
90° y-axis	30.0 (47.0)	90.0 (30.8)	75.0 (44.4)	90.0 (30.8)	71.3 (38.3)
90° x-axis	30.0 (47.0)	65.0 (48.9)	55.0 (51.0)	70.0 (47.0)	55.0 (48.5)
Total	36.7 (48.4)	78.3 (40.3)	71.7 (44.0)	80.0 (39.6)	66.7 (43.1)
			Section D		
90° picture	35.0 (48.9)	65.0 (48.9)	85.0 (36.6)	80.0 (41.0)	66.3 (43.9)
90° y-axis	20.0 (41.0)	65.0 (48.9)	75.0 (44.4)	85.0 (36.6)	61.3 (42.8)
90° x-axis	30.0 (47.0)	45.0 (51.0)	65.0 (48.9)	65.0 (48.9)	51.3 (49.0)
Total	28.3 (45.7)	58.3 (49.6)	75.0 (43.3)	76.7 (42.2)	59.6 (45.2)
			Section E		
90° picture	25.0 (44.4)	70.0 (47.0)	60.0 (50.3)	85.0 (36.6)	60.0 (44.6)
90° y-axis	30.0 (47.0)	75.0 (44.4)	95.0 (22.4)	80.0 (41.0)	70.0 (38.)
90° x-axis	20.0 (41.0)	70.0 (47.0)	75.0 (44.4)	90.0 (30.8)	63.8 (40.8)
Total	25.0 (44.2)	71.7 (46.2)	76.7 (39.0)	85.0 (36.2)	64.6 (41.4)
			Section F		
90° picture	30.0 (47.0)	65.0 (48.9)	70.0 (47.0)	75.0 (44.4)	60.0 (46.9)
90° y-axis	15.0 (36.6)	35.0 (48.9)	75.0 (44.4)	80.0 (41.0)	51.3 (42.8)
90° x-axis	15.0 (36.6)	20.0 (41.0)	60.0 (50.3)	75.0 (44.4)	42.5 (43.1)
Total	20.0 (40.1)	40.0 (46.3)	68.3 (47.2)	76.7 (43.3)	51.3 (44.2)

Table 18

ANOVA for performance on 90° rotational differences on varying axis, depth distribution and set-size, with age and sex as between subject variables

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subjects Effects			
Rota	**122263.89**	**2.000**	**38.219**	**.000*** **	**.35**
Rota*Sex	125.00	2.000	0.039	.962	.00
Rota*Age	14402.78	6.000	1.501	.182	.06
Rota*Sex*Age	3986.11	6.000	0.415	.868	.02
Size	7513.89	2.000	2.167	.118	.03
Size*Sex	7625.00	2.000	2.199	.115	.03
Size*Age	**25652.78**	**6.000**	**2.466**	**.027***	**.09**
Size*Sex*Age	20652.78	6.000	1.985	.071	.08
Depth	**46694.44**	**1.000**	**25.131**	**.000*** **	**.26**
Depth*Sex	694.44	1.000	0.374	.543	.01
Depth*Age	12750.00	3.000	2.287	.086	.09
Depth*Sex*Age	4972.22	3.000	0.892	.450	.04
Size*Depth	3347.22	2.000	0.835	.436	.01
Size*Depth*Sex	5180.56	2.000	1.293	.278	.02
Size*Depth*Age	10041.67	6.000	0.835	.545	.03
Size*Depth*Sex*Age	3986.11	6.000	0.332	.919	.01
Size*Rota	**49069.44**	**4.000**	**7.714**	**.000*** **	**.10**
Size*Rota*Sex	4625.00	4.000	0.727	.574	.01
Size*Rota*Age	27597.22	12.000	1.446	.144	.06
Size*Rota*Sex*Age	16263.89	12.000	0.852	.596	.03
Depth*Rota	**15680.56**	**2.000**	**6.009**	**.003** **	**.08**
Depth*Rota*Sex	97.22	2.000	0.037	.963	.00
Depth*Rota*Age	10208.33	6.000	1.304	.259	.05
Depth*Rota*Sex*Age	10569.44	6.000	1.350	.239	.05
Size*Depth*Rota	9402.78	4.000	1.660	.159	.02
Size*Depth*Rota*Sex	**17402.78**	**4.000**	**3.073**	**.017***	**.04**
Size*Depth*Rota*Age	11375.00	12.000	0.669	.780	.03
Size*Depth*Rota*Sex*Age	29597.22	12.000	1.742	.058	.07
		Between-Subjects Effects			
Sex	**81000.00**	**1.000**	**15.429**	**.000*** **	**.18**
Age	**155074.07**	**3.000**	**29.538**	**.000*** **	**.55**
Sex*Age	**16740.74**	**3.000**	**3.189**	**.029**	**.12**

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. Rota = Picture-plane angular differences; Age = Age groups; Size = Number of cubes per Aggregate; Depth = Aggregate depth distribution. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant statistical effects are set in bold.

The within-subject effects showed a significant main effect for aggregate-depth $F(1, 72) = 25.13$, $p < .001$, $\eta^2 = .26$, as performance on flat aggregates (M = 66.8%) was more accurate than on aggregates with protruding elements (M = 55.4%).

There was a significant two-way interaction of set-size and age groups, $F(6, 144) = 2.47$, $p < .05$, $\eta^2 = .09$. The split sample analysis by age groups showed that the expected decrease in performance over increases in set-size was only significant in 4- to 5-year-olds (4-cubes M = 40.8%; 5-cubes M = 32.5%; 6-cubes M = 22.5%). However, this difference had disappeared completely in the 10- to 11-year-old children who were accurate independently of set-size as accuracy was at a very high level of around 80%, see Figure 35.

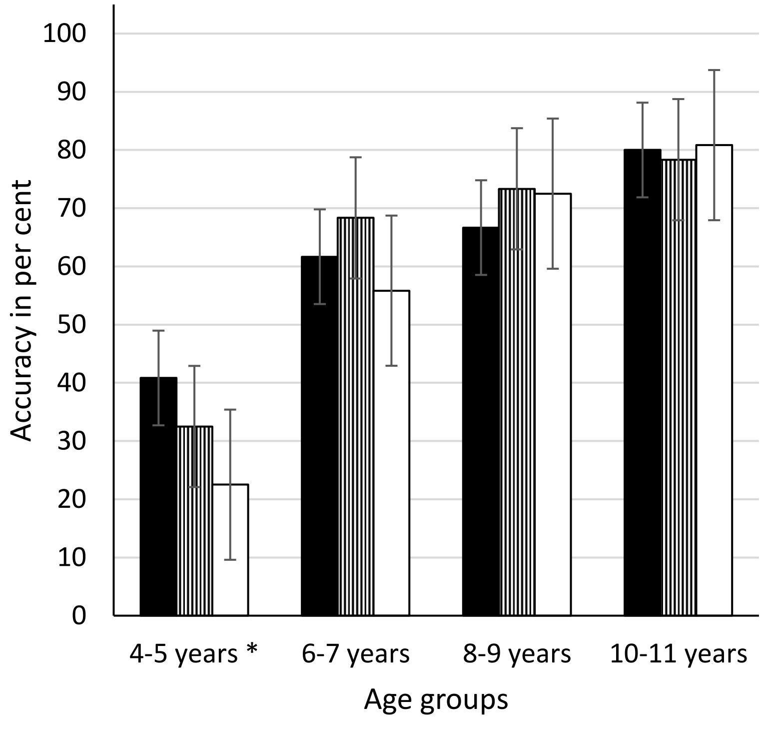


Figure 35 CMRT scores per age group and set-size.

Importantly, and this is the new effect, there was a significant effect for rotational axis with a medium effect size $F(2, 144) = 38.22$, $p < .001$, $\eta^2 = .35$. Ninety-degree rotations in the picture-plane (*M* = 68.5%) and around the y-axis (*M* = 66.7%) were both significantly easier than rotations around the x-axis (*M* = 48.1%) in a Cartesian coordinate system (no Figure).

There was also a significant two-way interaction of rotational axis and aggregate-depth, $F(2, 144) = 6.00$, $p < .01$, $\eta^2 = .08$, see Figure 36. Post-hoc comparisons showed that in rotations in the picture-plane, both flat (*M* = 69.6%) and protruding (*M* = 67.5%) cube aggregates did not significantly differ from each other, but in both y- and x-axis rotations it did matter whether the cube aggregates were flat or protruding as flat aggregates were easier to rotate for children.

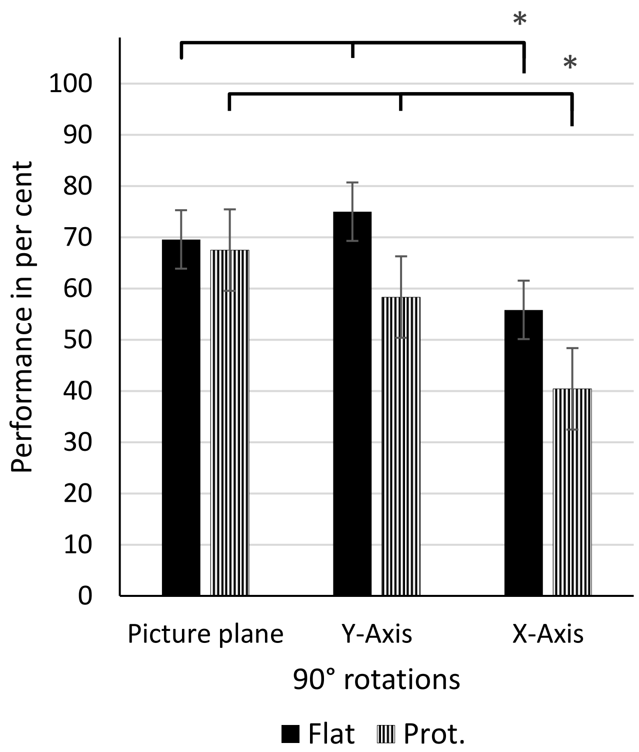


Figure 36 Development of the CMRT scores per rotational axis and aggregate depth. Prot. = Protruding.

There was another two-way interaction of rotational axis, this time with set-size, $F(4, 288) = 7.71$, $p < .001$, $\eta^2 = .10$, as performance gradually decreased over set-sizes for the picture-plane rotations and the rotation around the y-axis, but not the x-axis, see Figure 37.

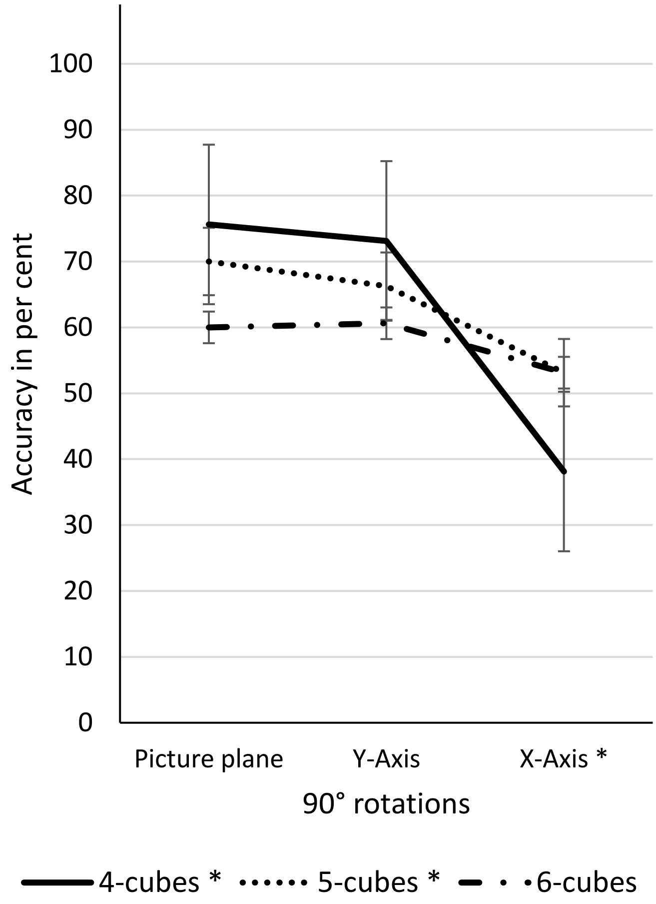
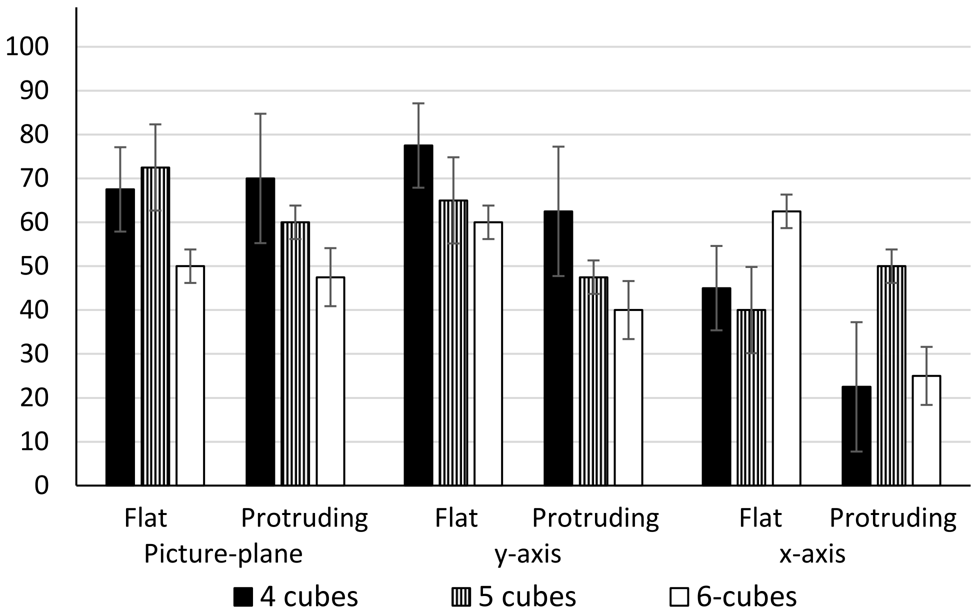
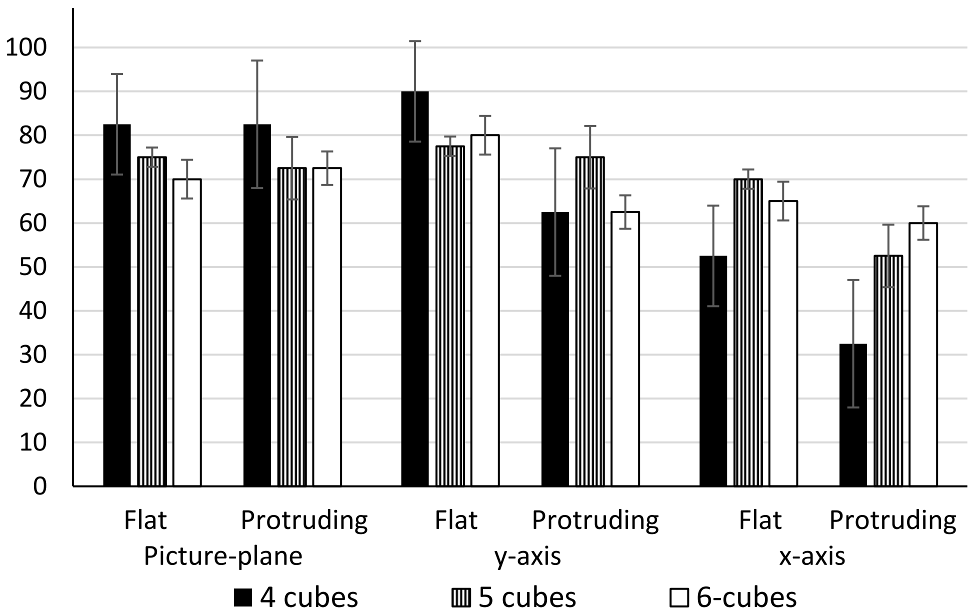


Figure 37 CMRT scores by rotational axis and set-size.

There was a significant four-way interaction of set-size, rotational axis, aggregate-depth and sex, $F(4, 288) = 3.07$, $p < .05$, $\eta^2 = .02$. This four-way interaction was investigated further by re-running the ANOVA as a split sample analysis by sex, revealing that the three-way interaction only became significant for girls.



(A) Girls



(B) Boys

Figure 38 CMRT scores by 90° axis of rotation and aggregate depth for 4-, 5- and 6-cube aggregates

3.4.3. Control of Mental Rotation with Fluid Intelligence

In order to investigate the effect of fluid intelligence (RCPM) on picture-plane rotations and on 90º rotations around the x- and y-axis, the analyses above were repeated with the Raven test score as a covariate. Effects that are no longer significant can be attributed to fluid intelligence. (Lehmann et al., 2014; Quaiser-Pohl et al., 2014). If an effect was significant in the previous analysis but failed to reach significance after controlling for the RCPM, the change could then be attributed to differences in fluid intelligence. Similarly, if a new effect becomes significant, the control for fluid intelligence indicates it was previously suppressed by this variable

3.4.4. Picture-Plane Rotations (Angularity)

A 4 (angularity) by 3 (set-size) by 2 (aggregate-depth) by 4 (age groups) by 2 (sex) analysis of variance was carried out, with repeated measures for angularity, set-size and aggregate-depth, and RCPM as a covariate. The between-subject factors are again age and sex. The statistical results are listed in Table 19. The report of the ANCOVA begins with the between-subject group effects followed by the within-subject task effects and interactions.

Table 19

Picture-plane *rotation ANCOVA for, with protrusion and set-size, with age and sex as between subject variables, and RCPM scores as a co-variant.*

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subjects Effects			
Rota	**26223.80**	**2.677**	**5.629**	**.002****	**.07**
Rota*RCPM	**17861.42**	**3.000**	**3.834**	**.010****	**.05**
Rota*Sex	824.58	3.000	0.177	.911	.00
Rota*Age	13465.23	9.000	0.963	.466	.04
Rota*Sex*Age	14204.93	9.000	1.016	.428	.04
Size	**12030.50**	**2.000**	**3.557**	**.031***	**.05**
Size*RCPM	8858.05	2.000	2.619	.076	.04
Size*Sex	1809.03	2.000	0.535	.587	.01
Size*Age	**25319.53**	**6.000**	**2.495**	**.025***	**.10**
Size*Sex*Age	8726.01	6.000	0.860	.526	.04
Depth	529.44	1.000	0.354	.554	.01
Depth*RCPM	801.80	1.000	0.536	.467	.01
Depth*Sex	863.09	1.000	0.577	.450	.01
Depth*Age	622.99	3.000	0.139	.937	.01
Depth*Sex*Age	4371.23	3.000	0.974	.410	.04
Size*Depth	**10285.73**	**2.000**	**3.137**	**.046***	**.04**
Size*Depth*RCPM	6034.16	2.000	1.840	.162	.03
Size*Depth*Sex	418.68	2.000	0.128	.880	.00
Size*Depth*Age	6270.11	6.000	0.637	.700	.03
Size*Depth*Sex*Age	4004.30	6.000	0.407	.874	.02
Size*Rota	11736.19	6.000	1.711	.117	.02
Size*Rota*RCPM	5123.23	6.000	0.747	.612	.01
Size*Rota*Sex	5347.04	6.000	0.780	.586	.01
Size*Rota*Age	28076.36	18.000	1.365	.145	.06
Size*Rota*Sex*Age	**39565.78**	**18.000**	**1.923**	**.013***	**.08**
Depth*Rota	**16404.48**	**3.000**	**4.379**	**.005****	**.06**
Depth*Rota*RCPM	8807.33	3.000	2.351	.073	.03
Depth*Rota*Sex	4156.26	3.000	1.109	.346	.02
Depth*Rota*Age	4522.78	9.000	0.402	.932	.02
Size*Rota*Sex*Age	9297.55	9.000	0.827	.592	.03
Size*Depth*Rota	12607.79	6.000	1.581	.151	.02
Size*Depth*Rota*RCPM	8466.02	6.000	1.061	.385	.02
Size*Depth*Rota*Sex	3275.07	6.000	0.411	.872	.01
Size*Depth*Rota*Age	17619.97	18.000	0.736	.773	.03
Size*Depth*Rota*Sex*Age	27064.18	18.000	1.131	.319	.05
		Between-Subjects Effects			
RCPM	**227407.57**	**1.000**	**67.144**	**.000*****	**.49**
Sex	**29548.54**	**1.000**	**8.724**	**.004****	**.11**
Age	**72666.99**	**3.000**	**7.152**	**.000*****	**.23**
Sex*Age	**39753.15**	**3.000**	**3.912**	**.022***	**.12**

Note. Rota = Picture-plane angularity; Age = Age groups; Size = Number of cubes per Aggregate; Dep = Aggregate depth distribution. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant statistical effects are set in bold.

As in Study 1, the Raven score as a covariate was highly significant $F(1, 71) = 67.14$, $p < .001$, $\eta^2 = .48$. indicating that part of the differences in the general performance level could be explained by fluid intelligence, and as before, sex and age differences stayed significant.

The main effect of angularity was still significant, $F(3, 213) = 5.63$, $p < .01$, $\eta^2 = .07$ but a new significant two-way interaction of angularity with RCPM, $F(3, 213) = 3.83$, $p \leq .01$, $\eta^2 = .05$, suggested that differences picture-plane rotations are accounted for by variance in fluid intelligence.

In order to explore the relationship between angularity (picture-plane rotations) further, a curvefit analysis was run, revealing significant linear, $F(1, 78) = 198.78$, $p < .001$, quadratic, $F(2, 77) = 105.14$, $p < .001$, and cubic, $F(3, 76) = 78.16$, $p < .001$, trends, see Figure 39. The linear fit with the highest F-value provided the best fit for the data, showing that higher CMRT scores would also result in higher RCPM scores.

The main effect set-size, $F(3, 142) = 3.56$, $p < .05$, $\eta^2 = .05$, the two-way interactions set-size by age, $F(3, 142) = 2.50$, $p < .05$, $\eta^2 = .10$, set-size by protrusion, $F(2, 142) = 3.14$, $p < .05$, $\eta^2 = .04$, and protrusion by angularity, $F(3, 213) = 4.38$, $p < .01$, $\eta^2 = .06$, and the four-way interaction set-size by angularity by sex by age, $F(18, 426) = 1.92$, $p < .05$, $\eta^2 = .08$, all remained significant, confirming that even after controlling for fluid intelligence, young 4- to 5-year-old boys showed a mental rotation effect a year earlier than girls.

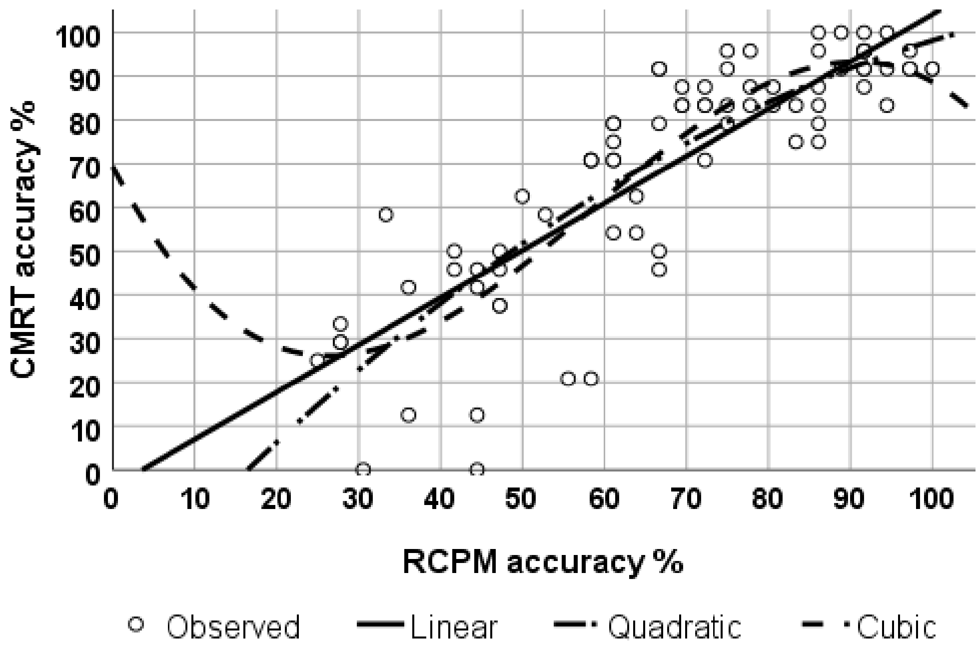


Figure 39 Scatterplot for RCPM and picture-plane rotations

The main effect of angularity by age (*p* = .466), was no longer significant. The within-subject factors set-size by angularity (*p* = .117) and set-size by protrusion by angularity (*p* = .151), were also no longer significant, suggesting that in particular the combination of the two factors set-size and angularity were influenced by levels of fluid intelligence.

3.4.5. 90º Axis Rotations (Picture-Plane, X-, and Y-Axis)

A 3 (set-size) by 2 (aggregate-depth) by 3 (rotational axis) by 4 (age groups) by 2 (sex) analysis of variance was carried out, with repeated measures for axis, set-size, and aggregate-depth and the RCPM as a covariate. The between-subject factors are age and sex. The statistical results are listed in Table 20. The report of the ANCOVA begins with the between-subject group effects followed by the within-subject task effects and interactions.

Table 20

90° axis rotation ANCOVA with aggregate depth and set-size, with age and sex as between subject variables and RCPM scores as a co-variant.

Source	*SS*	*df*	*F*	*p*	η^2
		Within-Subjects Effects			
Rota	**12683.46**	**2.000**	**3.956**	**.021***	**.05**
Rota*RCPM	2691.02	2.000	0.839	.434	.01
Rota*Sex	589.87	2.000	0.184	.832	.00
Rota*Age	16608.01	6.000	1.727	.119	.07
Rota*Sex*Age	3349.65	6.000	0.348	.910	.02
Size	**13295.65**	**2.000**	**3.942**	**.022***	**.05**
Size*RCPM	10210.7	2.000	3.028	.052	.04
Size*Sex	3418.82	2.000	1.014	.365	.01
Size*Age	9646.34	6.000	0.953	.459	.04
Size*Sex*Age	19603.42	6.000	1.938	.079	.08
Depth	**23108.462**	**1.000**	**13.556**	**.000*****	**.16**
Depth*RCPM	**12742.044**	**1.000**	**7.475**	**.008****	**.10**
Depth*Sex	37.459	1.000	0.022	.883	.00
Depth*Age	**17545.560**	**3.000**	**3.431**	**.022***	**.18**
Depth*Sex*Age	4784.485	3.000	0.936	.428	.04
Size*Depth	2477.58	2.000	0.613	.543	.01
Size*Depth*RCPM	1559.41	2.000	0.386	.681	.01
Size*Depth*Sex	4573.40	2.000	1.131	.325	.02
Size*Depth*Age	11578.42	6.000	0.955	.458	.04
Size*Depth*Sex*Age	3756.26	6.000	0.310	.931	.01
Size*Rota	1450.68	4.000	0.227	.923	.00
Size*Rota*RCPM	3509.57	4.000	0.548	.700	.01
Size*Rota*Sex	5233.43	4.000	0.818	.515	.01
Size*Rota*Age	26553.89	12.000	1.383	.173	.06
Size*Rota*Sex*Age	15694.02	12.000	0.817	.633	.03
Depth*Rota	2776.18	2.000	1.058	.350	.02
Depth*Rota*RCPM	1520.96	2.000	0.579	.562	.01
Depth*Rota*Sex	240.54	2.000	0.092	.912	.00
Depth*Rota*Age	11688.10	6.000	1.484	.188	.06
Depth*Rota*Sex*Age	10219.04	6.000	1.298	.262	.05
Size*Depth*Rota	2898.54	4.000	0.507	.730	.01
Size*Depth*Rota*RCPM	2013.15	4.000	0.352	.842	.01
Size*Depth*Rota*Sex	**16048.87**	**4.000**	**2.808**	**.026**	**.04**
Size*Depth*Rota*Age	11525.09	12.000	0.672	.778	.03
Size*Depth*Rota*Sex*Age	28911.52	12.000	1.686	.069	.07
		Between-Subjects Effects			
RCPM	**170654.62**	**1.000**	**58.436**	**.000*****	**.45**
Sex	**25066.82**	**1.000**	**8.583**	**.005****	**.11**
Age	**34113.82**	**3.000**	**3.894**	**.012***	**.14**
Sex*Age	**25662.41**	**3.000**	**2.929**	**.039***	**.11**

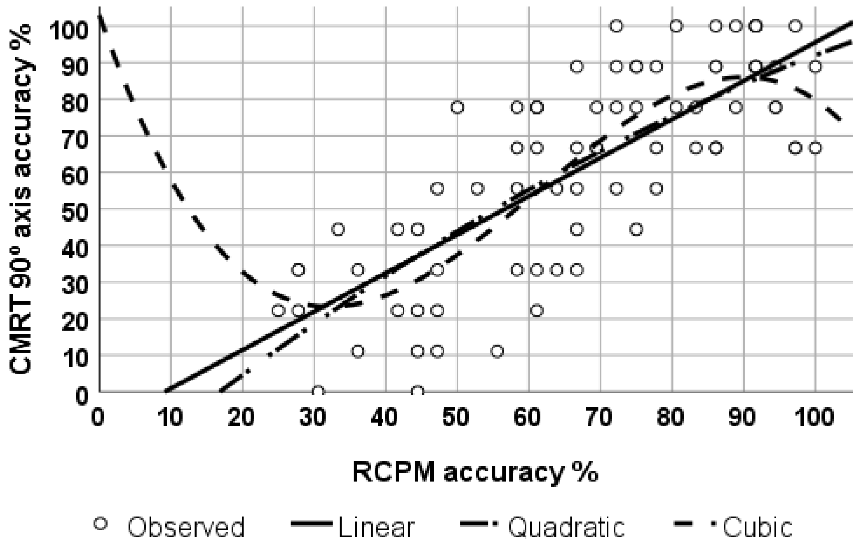
Note. Rota = 90° axis rotations; RCPM = Raven's Coloured Progressive Matrices; Size = Number of cubes per Aggregate; Depth = Aggregate depth distribution. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant statistical effects are set in bold

The Raven covariate was significant as a main effect, $F(1, 71) = 58.44$, $p < .001$, $\eta^2 = .45$, indicating that the general performance level was explained by fluid intelligence. The other between-subject factors sex, $F(1, 71) = 8.58$, $p < .01$, $\eta^2 = .11$, age, $F(3, 71) = 3.89$, $p < .05$, $\eta^2 = .14$, and the sex by age interaction, $F(3, 71) = 2.93$, $p < .05$, $\eta^2 = .45$, and the main effects for axis of rotation, $F(2, 142) = 3.96$, $p < .05$, $\eta^2 = .05$ all remained significant.

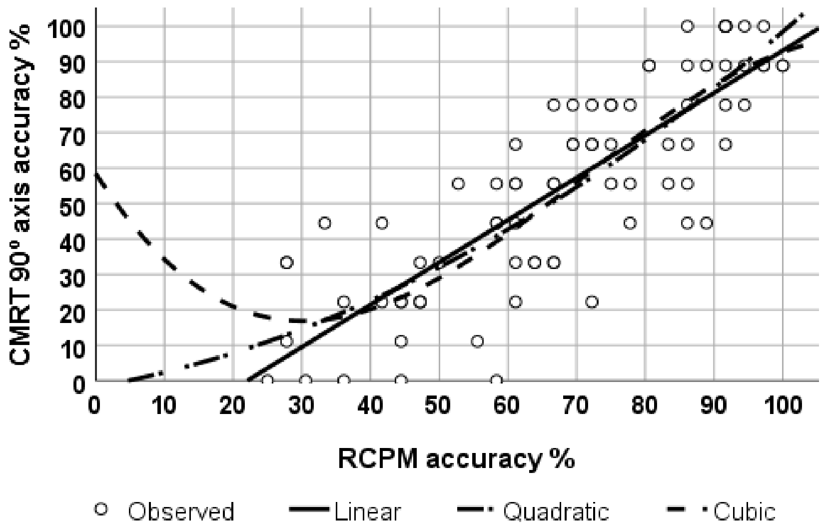
When controlling for fluid intelligence the two-way interactions set-size by age group ($p = .459$) and set-size by axis of rotation ($p = .932$) both no longer reached significance. However set-size as a main effect became significant, $F(2, 142) = 3.94$, $p < .05$, $\eta^2 = .05$, in which aggregates with 4-cubes (M = 62.3%), were easier than those with 5-cubes (M = 63.1%), but surprisingly aggregates with 6-cubes were the easiest (M = 57.9%).

The main effect for protrusion remained significant, $F(2, 71) = 13.56$, $p < .001$, $\eta^2 = .95$, but this effect also produced a new two-way interaction with the RCPM, $F(1, 71) = 7.48$ $p < .01$, $\eta^2 = .10$, suggesting that better performance on flat aggregates can be explained by individual differences in fluid intelligence.

In order to explore the relationship between protrusion and the RCPM further, a curve fit analysis was run for both protruding and flat aggregates. For protruding aggregates, significant linear, $F(1, 78) = 173.06$, $p < .001$, quadratic, $F(3, 77) = 88.96$, $p < .001$, and cubic, $F(3, 76) = 60.39$, $p < .001$, trends were found. For flat aggregates, significant linear, $F(1, 78) = 115.64$, $p < .001$, quadratic, $F(2, 77) = 58.10$, $p < .001$, and cubic, $F(3, 76) = 44.05$, $p < .001$, trends were also found, see Figures 40A and B. In all three trends, the F-values of the prediction of 90^o axis rotations by fluid intelligence are higher for the aggregates with protruding elements than for the flat cube aggregates, showing more impact. The scatterplots illustrate this in terms of a slightly tighter fit of individual values around the trend lines in Figure 40A with less floor effects. This suggests that children excelling at rotating protruding aggregates were also more likely to have higher levels of fluid intelligence, or vice versa.



(A)



(B)

Figure 40 RCCT Scatterplots: (A) flat aggregates 90-degree rotations. (B) protruding aggregates 90° rotations

There was also a new two-way interaction between protrusion and age groups, $F(3, 71) = 3.43$, $p < .05$, $\eta^2 = .13$. The the youngest 4- to 5-year-old age group performed at the same low level for flat or aggregates with protruding elements. In all other age groups, protruding aggregates were more difficult to process than flat aggregates, see Figure 41. A re-run split file by age groups ANCOVA, revealed that only in the 6- to 7- year-olds flat aggregates were significantly easier, $F(1, 17) = 7.94$, $p < .05$, $\eta^2 = .32$.

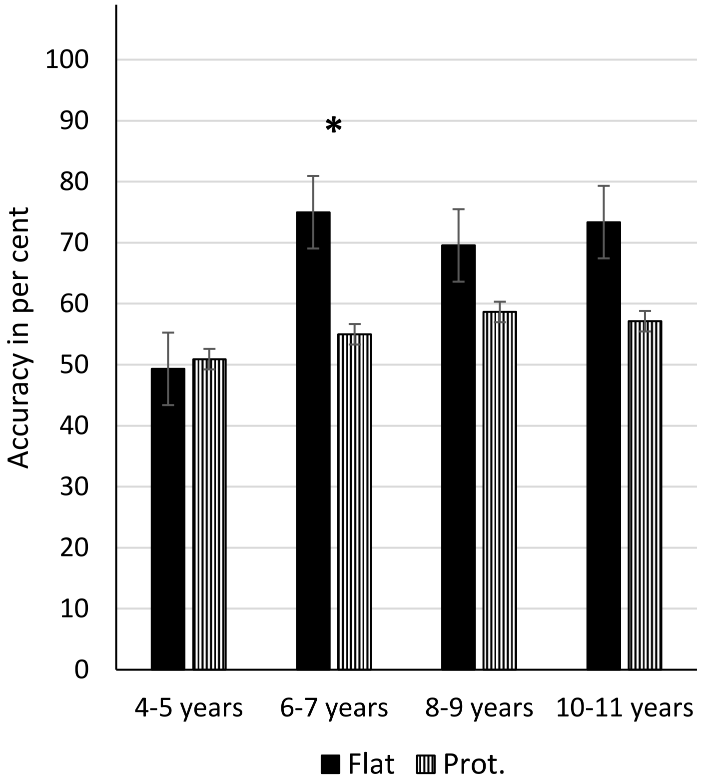


Figure 41 CMRT Accuracy by protrusion and age groups.

However, the four-way interaction between set-size, protrusion, axis of rotation, and sex, $F(4, 284) = 2.81$, $p < .05$, $\eta^2 = .04$, remained significant. It confirmed that for girls aggregate depth was the most important factor influencing task difficulty in picture-plane and y-axis rotations, and that x-axis rotations were generally the most difficult.

When controlling for fluid intelligence, the two-way interactions of set-size by axis of rotation (p = .923) and of protrusion by axis of rotation (p = .350), no longer reached significance. In toto, the controls for individual differences in fluid intelligence demonstrate a newly emerging visual discriminability of dimensional objects on paper in school-aged children.

3.4.6. Raven Coloured Progressive Matrices Test (RCPM)

Accuracy was computed in per cent correct for each type of RCPM section A, AB, and B, see Table 21. As with the RCCT, a 3 (section) by 3 (age) by 2 (sex) analysis of variance with repeated measurement for each section was conducted.

Table 21

ANOVA for performance RCPM subtests with age and sex as between subject variables

Source	*SS*	*df*	*F*	*p*	*η²*
		Within-Subjects Effects			
Subtest	**5988.43**	**2.000**	**27.128**	**.000****	**.274**
Subtest*Sex	48.61	2.000	0.22	.803	.003
Subtest*Age	**6405.09**	**6.000**	**9.672**	**.000****	**.287**
Subtest*Sex*Age	**1756.94**	**6.000**	**2.653**	**.018***	**.100**
		Between-Subjects Effects			
Sex	**3190.10**	**1.000**	**3.961**	**.050***	**.052**
Age	**49449.94**	**3.000**	**20.466**	**.000****	**.460**
Sex*Age	493.92	3.000	0.204	.893	.008

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. Subtest = RCPM sections; Age = Age groups. * p < .05. ** p < .001. Significant statistical effects are set in bold.

A significant main effect for sex, $F(1, 72) = 3.96$, $p = .05$, $\eta^2 = .05$, showed that boys (M = 71%) outperformed girls (M = 63.7%). There was also a main effect for age groups $F(3, 72) = 20.47$, $p < .001$, $\eta^2 = .46$, with performance increasing with age. As could be expected, 4- to 5-year-olds showed the lowest score (M = 46.3%) followed by 6- to 7-year-olds (M = 63.8%), 8- to

9-year-olds (*M* = 74%) and 10- to 11-year-olds (*M* = 85.3%). Post-hoc comparisons showed that the performance of 4- to 5-year-olds differed significantly from all other age groups, and that the performance of 6- to 7-year-olds also differed significantly from 10- to 11-year-olds.

The RCPM sections also revealed a significant main effect, $F(2, 144) = 27.13$, $p < .001$, $\eta^2 = .27$. Sections became more difficult over the progression of the RCPM. Pairwise post-hoc comparisons ($p_s < .001$) confirmed significant differences between both the RCPM A (*M* = 71.6%) and RCPM AB (*M* = 70.1%), only when compared to the RCPM B (*M* = 60.3%).

There was a significant two-way interaction of RCPM sections and age groups, $F(6, 144) = 9.67$, $p < .001$, $\eta^2 = .29$, see Figure 42. Post-hoc comparisons showed that 4- to 5-year-olds performed significantly differently on all subtests with scores decreasing over subtests (A *M* = 60.8%, AB *M* = 46.7%, B *M* = 31.3%); 6- to 7-year-olds performed significantly better in both sections A (*M* = 68.8%) and AB (*M* = 66.7%) compared to B (*M* = 55.8%); there was no significant difference in performance between subtests for 8- to 9-year-olds; and finally, 10- to 11-year-olds performed significantly worse in section A (*M* = 80.4%) compared to AB (*M* = 90.4%). Overall, the younger children produced steeper decreases across sub-sections.

The significant three-way interaction between RCPM subtests, sex and age groups, $F(6, 144) = 2.65$, $p < .05$, $\eta^2 = .10$ was investigated further with a split-sample ANOVA for sex and age groups.

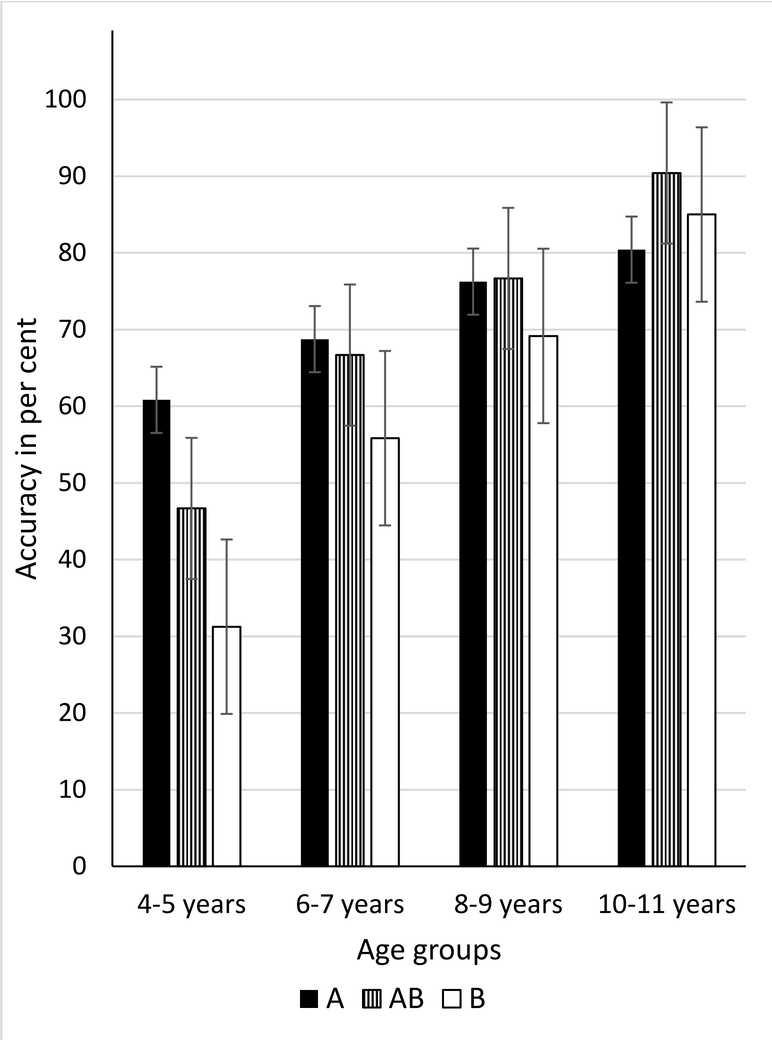
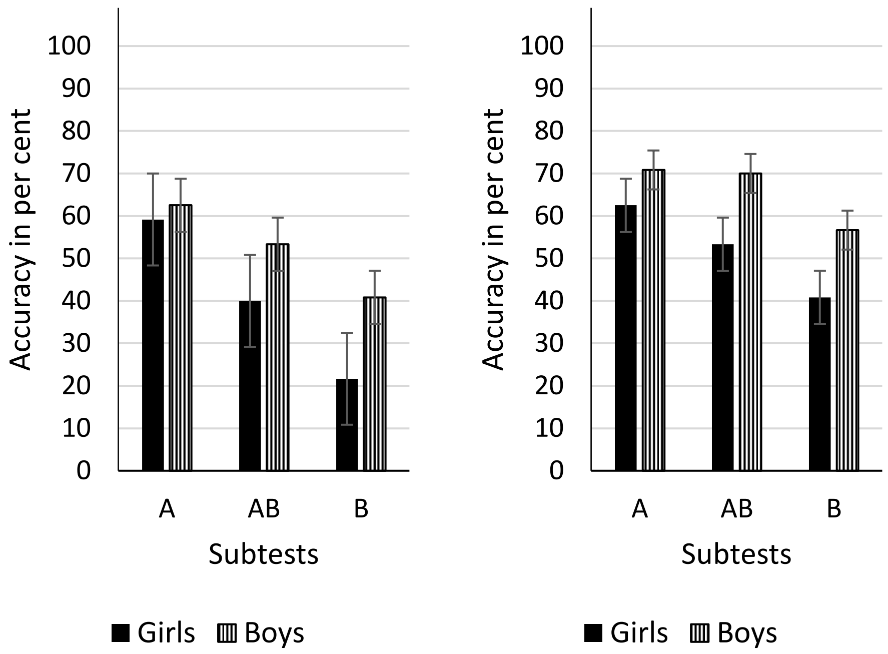
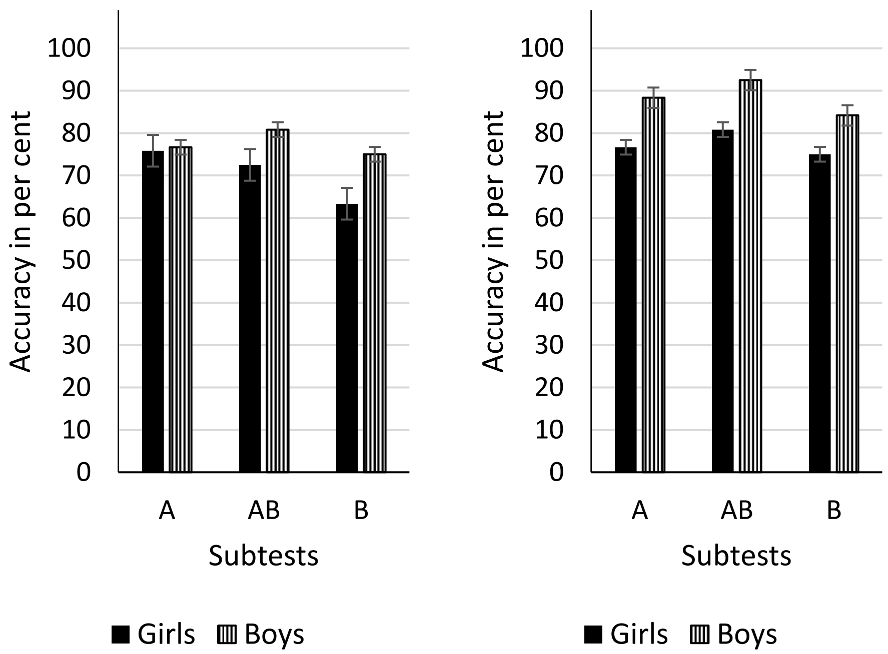


Figure *42* RCPM scores by Age group

Four- to 5-year-old girls showed decreases in performance over subtests, which differed significantly from each other (Subtest A *M* = 59.2%, AB *M* = 40% and B *M* = 21.7%). Performance for 7- to 8- (subtest A *M* = 66.7%, AB *M* = 63.3% and B *M* = 55%) and 9- to 10-year-old (subtest A *M* = 75.8%, AB *M* = 72.5% and B *M* = 63.3%) girls decreased over subtests but did not reach significant levels. Performance for girls in the 10- to 11-year-old age group was the highest in the second subtest, but no subtest differed significantly from each other (subtest A *M* = 88.3%, AB *M* = 92.5% and B *M* = 84.2%)



(A) 4- to 5-year-olds (B) 6- to 7-year-olds



(A) 4- to 5-year-olds (B) 6- to 7-year-olds

Figure 43 RCPM subtest scores by sex and age groups

Performance for boys in the 4- to 5-year-old age group decreased over subtests, with only the first and last subtest differing significantly from each other (subtest A *M* = 62.5%, B *M* = 53.3% and C *M* = 40.8%). Performance for boys in the 6- to 7-year-old age group decreased over subtests, in which only subtest A and AB both differed significantly from subtest B (subtest A *M* = 70.8%, AB *M* = 70% and B *M* = 56.7%). Children performed best in subtest AB in both 8- to 9- (subtest A *M* = 76.7%, AB *M* = 80.8% and B *M* = 75%) and 10- to 11-year-old (subtest A *M* = 76.7%, AB *M* = 80.8% and B *M* = 75%) age groups, but this difference was not significant, see Figure 43.

3.4.7. Comparison of the CMRT and RCPM

As the CMRT was a new test development, the established RCPM was used for cross-validation. The scores for the CMRT and the RCPM were highly correlated with each other, $r = .87$, $p < .001$. The CMRT and RCPM scores per age group were also significantly correlated, with $r = .69$, $p = .001$ for the 4- to 5-year-olds, $r = .87$, $p < .001$ for the 6- to 7-year-olds, $r = .71$, $p < .001$ for the 8- to 9-year-olds and $r = .77$, $p < .001$ for 10- to 11-year-olds. A confirmatory correlational analysis with unstandardized residuals resulting from controlling the impact of age and sex in regression analyses, revealed a similar, still strong substantial correlation between the RCCT and RCPM, r =. 72, $p < .001$. The correlations per age group were $r = .62$, $p < .01$ for the 4- to 5-year-olds, $r = .87$, $p < .001$ for the 6- to 7-year-olds, $r = .64$, $p < .01$ for the 8- to 9-year-olds and $r = .76$, $p < .001$ for 10- to 11-year-olds. They were still substantial but no longer gradually increased with age.

A 2 (Test: CMRT vs. RCPM) by 3 (age) by 2 (sex) analysis of variance was carried out, the statistical effects are listed in Table 22.

Results showed a significant main effect for sex, $F(1, 72) = 13.13$, $p = .001$, $\eta^2 = .15$. Boys (M = 72.7%) performed significantly better than girls (M = 61.5%).

Table 22

ANOVA for performance on the CMRT and RCPM with age and sex as between subject variables

Source	*SS*	*df*	*F*	*p*	*η²*
	Within-Subjects Effects				
Test	**271.22**	**1.000**	**4.413**	**0.039***	**0.058**
Test*Sex	**491.99**	**1.000**	**8.006**	**0.006****	**0.100**
Test*Age	**1073.56**	**3.000**	**5.823**	**0.001****	**0.231**
Test*Sex*Age	**648.11**	**3.000**	**3.516**	**0.019***	**0.128**
	Between-Subjects Effects				
Sex	**5031.15**	**1.000**	**13.131**	**0.001****	**0.154**
Age	**44155.29**	**3.000**	**38.414**	**0.000*****	**0.615**
Sex*Age	2370.10	3.000	2.062	0.113	0.079

Note. Degrees of Freedom were corrected with Greenhouse-Geisser. Test = Test type; Age = Age groups. * $p < .05$. ** $p < .01$. *** $p < .001$. Significant statistical effects are set in bold.

There was also a main effect for age groups, $F(3, 72) = 38.41$, $p < .001$, $\eta^2 = .62$. Performance increased with age. Post-hoc pairwise comparisons showed that 4- to 5-year-olds (M = 40.8%) performed significantly worse than 6- to 7-year-olds (M = 66.2%), 8- to 9-year-olds (M = 76.7%) and 10- to 11-year-olds (M = 84.9%). Moreover, 10- to 11-year-olds performed significantly better than 6- to 7-year-olds, see Figure 44. The interaction between sex and age was not significant.

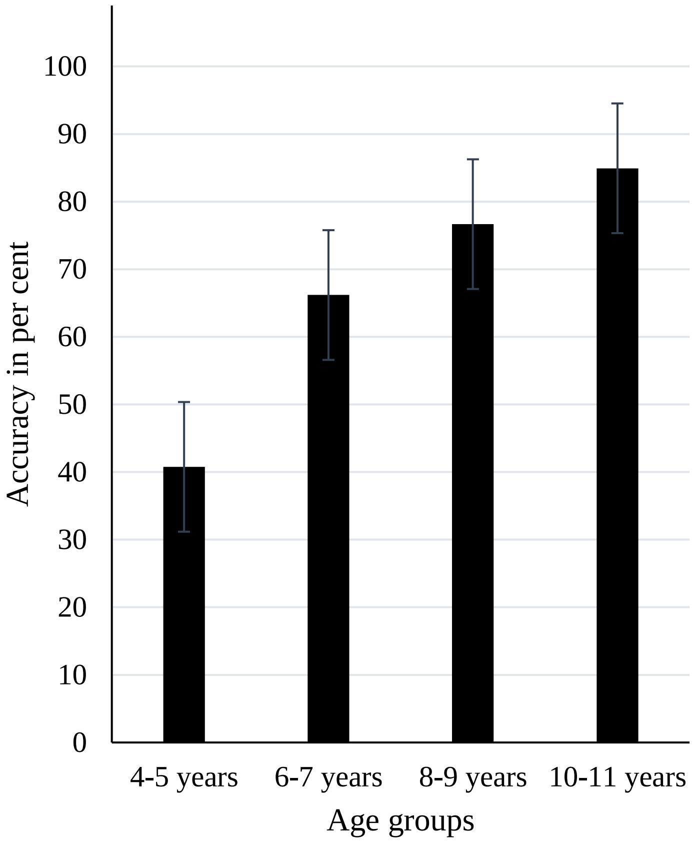


Figure 44 Test scores by age groups

There was a significant main effect for Test type, $F(1, 72) = 4.41$, $p < .05$, $\eta^2 = .06$. Performance in the CMRT ($M = 65.8\%$) was slightly easier than on the RCPM ($M = 68.4\%$).

A significant interaction for test type by sex, $F(1, 72) = 8.006$, $p < .01$, $\eta^2 = .1$. Post-hoc tests showed that girls performed significantly worse on the CMRT ($M = 58.5\%$) compared to the RCPM ($M = 64.6\%$), whereas boys performed similarly on the CMRT ($M = 73.2\%$) and RCPM ($M = 72.3\%$), see Figure 45.

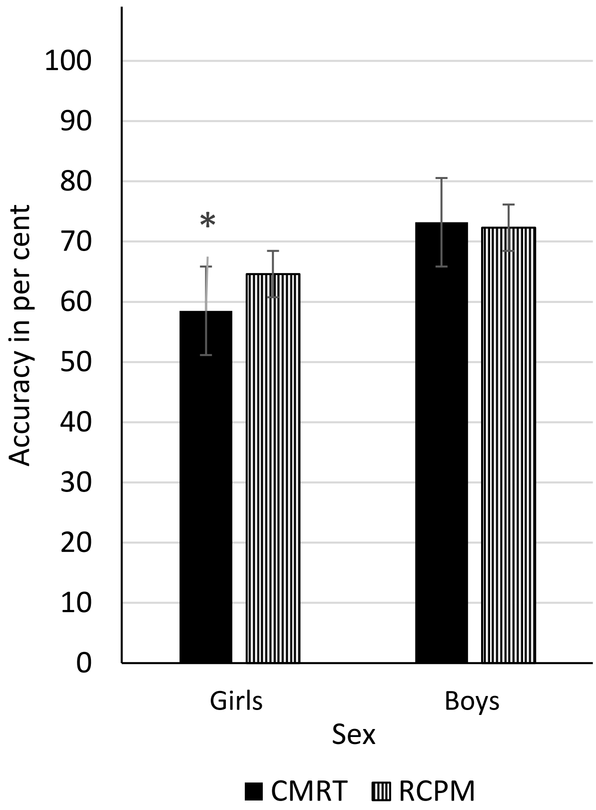
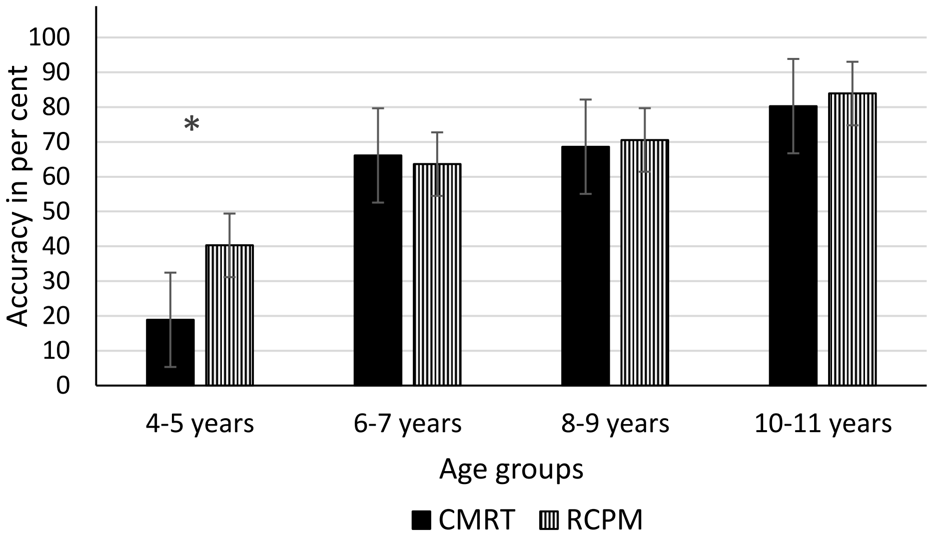
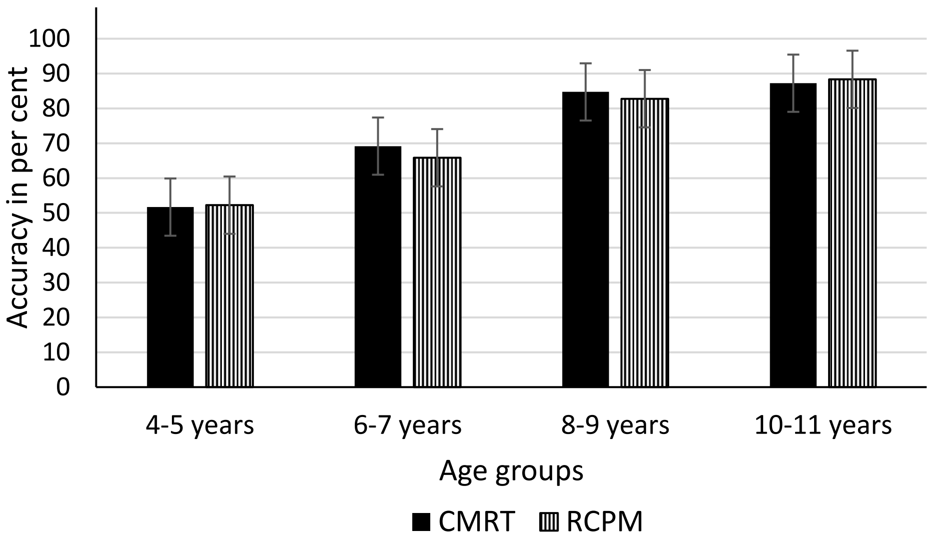


Figure 45 Test type scores by sex

There was a significant interaction for test type by age groups, $F(3, 72) = 5.823$, $p = .001$, $\eta^2 = .23$. Post-hoc comparisons showed that only 4- to 5-year-olds performed significantly worse on the CMRT (M = 35.3%) compared to the RCPM (M = 46.3%), while in the older age groups, there was no significant difference, as also indicated by the high correlations reported at the beginning of this section. Because of a significant three-way interaction between test-type, sex and age groups, $F(3, 72) = 648.11$, $p < .05$, $\eta^2 = .13$, the ANOVA was re-run as a split sample for sex and age groups which revealed that only girls in the 4- to 5-year-old group performed significantly worse on the CMRT (M = 18.8%) compared to the RCPM (M = 40.3%), see Figure 46 A, but boys did not, see Figure 46 B.



(A) Girls



(B) Boys

Figure 46 CMRT and RCPM by age and sex

3.5. Discussion Study 2

3.5.1. Goals and hypothesis

The aim of this research was to create another new test, similar to the Mental Rotation Test for adults (Vandenberg & Kuse, 1978), specifically for 4- to 11-year-old children. The new Coloured Mental Rotation Test (CMRT) was simplified through the reduction of cubes per aggregate and by the inclusion of colour, a modification of two critical performance factors in comparison to the Mental Rotation Test. Time constraints for responses were also removed (see also, Frick, Hansen & Newcombe, 2013; Hawes, LeFevre, Xu & Bruce, 2015; Quaiser-Pohl, 2003).

The systematic variation of the performance factors aggregate complexity in subtests (A-F) as well as angle and dimension of rotation allowed a direct comparison of how these might influence young children's spatial ability and the ability to mentally rotate. The hypotheses were that task difficulty would be influenced by (1) the angle of rotation, (2) the number of cubes per aggregates, (3) aggregate dimensionality (comparing flat and protruding aggregates) and (4) the axis of rotation (comparing picture-plane, vertical and horizontal rotations).

3.5.2. Optimal Gestalt of 4-Cube Aggregates

Statistical analysis revealed several significant effects and interactions: (1) Four-cube aggregates were easier to rotate than 5- and 6-cube aggregates, indicating a simple main effect for complexity in terms of aggregate size; (2) Four to 5-year-olds performed better on MR tasks with 4-cube aggregates compared to larger aggregates, suggesting that already very young children's ability to mentally rotate is sensitive to aggregate size; (3) Only 4-cube aggregates were easier to rotate if they had protruding elements that point into multiple depth-planes, demonstrating that additional depth information can indeed facilitate mental rotation, but only if the object is not too complex in terms of set-size; (4) Only 4-cube aggregates distributed in one depth-plane (flat) displayed the clearest linear decrease in accuracy over increases in angularity; (5) the graded effect for task difficulty was only present on 4-cube aggregates for 6- to 7- and 8- to 9-year-old girls, whereas for boys the effect was already present in the younger 4- to 5-year-olds and in the oldest 10- to 11-year-old age group. Collectively these results clearly demonstrate that especially the simplest 4-cube aggregates are most receptive to the variation of performance factors and present the optimal Gestalt for measuring developmental differences in mental rotation ability in children as young as 4-years and up to 10-11-years of age.

A possible explanation for the sensitivity of 4-cube aggregates in relation to performance factors in the CMRT can be drawn from research on working memory. The classical achromatic MRT (Shepard & Metzler, 1971) differs from the CMRT in which aggregates consist of individually coloured cubes. It hence could be argued that children were encouraged to use analytical strategies such as verbal labelling while identifying the sequence and location of coloured cubes within an aggregate. From a working memory perspective, this processing of individual elements would place high demands on working memory which may explain differences in performance between age groups. Halford, Cowan, and Andrews (2007) suggested that for adults there is a central working memory capacity limit about 4 chunks, and similarly, representations in reasoning are limited to four interrelated variables (Halford, Baker, McCredden, & Bain, 2005), both of which can predict mistakes in reasoning and thinking.

Rensink (2001) showed that in visual serial presentations, if the pauses between one item and another (interstimulus interval ISI) are 120ms in time, the capacity is five to six items, but when the ISI is 360ms, capacity is three to four items. The conclusion, then, could be that the capacity limit is time-based and not item-based (Lange-Küttner, 2012). Moreover, the number of features within each object needs to be considered in addition to the number of items (Oberauer & Eichenberger, 2013), for instance, changes in colour and shape enhanced, while orientation and size changes decreased visual memory in adults. Individual features of an object are easier to remember than combinations of two features as an additional binding process increases the cognitive load (Cowan, Blume, & Saults, 2013). However, the effect of multiple features in an object does not multiply the cognitive load, instead, a hierarchy of features unfolds like a folding fan with colour being consistently the easiest, followed by a black marker, orientation and length (Hardman & Cowan, 2015). This showed differential weights of features pointing towards perceptual saliency as marker values of object dimensions.

In the current study, the number of features corresponded to the number coloured cubes in the cube aggregates. Young children can already demonstrate combinatorial visual processes when combining colours into one square when drawing a cube (V. Moore, 1986) or shapes into an outline figure (Lange-Küttner, 1994b; Lange-Küttner et al., 2002). If an object has just one colour, colour labelling can help children (Cowan, AuBuchon, Gilchrist, Ricker, & Saults, 2011), however, the cube aggregates in the current study were multi-coloured.

Lange-Küttner and Küttner (2015) found that 7- and 9-year old children could remember 4 items in the first set of trials and up to ceiling level during repetitions, but memory deterioration occurred as soon as a new set was to be remembered that varied in colour and shape. This suggests that it is the object change which limits capacity. Indeed, Riggs, McTaggart, Simpson, and Freeman (2006) found a very low capacity estimate of 1.52 for 5-year-olds, 2.89 items for 7-year-olds and 3.83 items for 10-year-olds in a visual memory task that involved changing object colour.

This assumption is supported by findings that working memory capacity accounts for a substantial proportion of the variance in reasoning (Kane et al., 2004) and intelligence (Cowan et al., 2005). Younger children have a brain system that can retain only a limited number of items in active form compared to older children (Burtis, 1982; Cowan, 2001; Pascual-Leone & Smith, 1969). Increases in working memory capacity during elementary school years are particularly pronounced, and subsequently lesser between those years and adulthood (Cowan et al., 2011). Cowan (2001) found differences in the storage capacity in terms of item number for 7-year-olds (about 1.5) compared to older children and adults (about 3.5). At first glance, the capacity limit of 1.5 chunks seems to contradict the CMRT results, in which 4- to 5-year-olds were already able to process 4-cube aggregates, exceeding Cowan's storage capacity. However, comparing continuously visible cube aggregates requires less working memory than recall tasks. In order to correctly identify the matching aggregate, children only needed to realise that the orientation of an aggregate shape can be changed, and subsequently, identify the new orientation and then process the sequence of coloured cubes in line in the new orientation. For example, a cube aggregate rotated by 180° in the picture-plane would have the reverse order of coloured cubes per aggregate, see Appendix, RMRT, Question A4, answer 5. To explain this process with reference to Baddeley's (2000) multicomponent model, the spatial relation of aggregate orientation would be supported by colour information, either in a verbal format using colour labels by the phonological loop, or via a graphic design impression by the visuo-spatial sketchpad. Both these subsystems feed into and get input from into the short-lived episodic buffer which binds information into integrated chunks. In this model, the central executive provides the processing power to combine features into a lasting single integrated representation. It could therefore be argued that Cowen's capacity limit only applies to the lasting representations which are the result of mental operations, while in the current study only short-lived episodic binding was necessary as all information needed for decision-making was constantly available and no memory representation needed to be formed. These conclusions are however speculative, as children were only tested on the CMRT and RCPM with no other measures of working memory or strategy use. Nevertheless, further support for this hypothesis comes from Lachmann and Van Leeuwen (2008) who found that figural goodness reduces the central processing load.

An alternative explanation for the persistent prevalence of significant results with 4-cube aggregates could be drawn from "Good Gestalt" principles and the simplicity-complexity dimension (Palmer, 1991). Garner (1974) in his theory of rotation and reflection suggested that figural goodness depended on the number of transformational variants, in which good figures had less variants than poor figures. Garner and Clement (1963) elegantly demonstrated this using 5 dot patterns within a 3x3 matrix at 0°, 90°, 180° and 270° rotations (picture-plane) and 4 types of reflections (horizontal, vertical left and right) where patterns rated as "good" had fewer transformational variants. Transformational variants were defined as differences between the position of the 5 dots between the original figure and after being rotated or reflected. This is in line with Gestalt Psychologists definition of bilateral symmetry (Koffka, 1935; Wertheimer, 1923), which was later proven to play an important role in shape perception (Machilsen, Pauwels, & Wagemans, 2009). The smallest 4-cube aggregate in the CMRT fits this definition of an optimal Gestalt, compared to larger aggregates. Also, a Good Gestalt aggregate has the smallest number of transformational variants, yet all the properties of the larger aggregates. In the current study, the 4-cube aggregate has cube elements pointing into three-dimensional space. Thus, the 4-cube aggregate also provides the best fit in terms of the simplicity-complexity dimension, both with respect to aggregate size and in relation to aggregate dimensionality, as 4 cubes provide the minimum size where aggregates with protruding cubes can be assembled, see section A and B of the CMRT in the appendix.

The number of meaningful and significant results in relation to the Good Gestalt 4-cube aggregate was particularly striking. The statistical interactions with set-size showed how 4-cube aggregates already enabled 4- to 5-year-old children to mentally rotate geometric objects by producing a clear angularity effect, in which rotations with larger angles were more difficult than smaller angles, similar to results produced by adults on the MRT (Vandenberg & Kuse, 1978), see Figure 34A. The Good Gestalt 4-cube aggregate also revealed how this angularity effect is particularly sensitive to task difficulty, as only this smallest aggregate showed performance levels akin to the classic mental rotation effect for angularity, see Figure 30. It revealed an angularity effect in 6- to 7 and 8- to 9-year-old girls, see Figure 33 B and C. Consistent with previous research on a male advantage in mental rotation, an angularity effect was already present in 4- to 5-year-old boys, but interestingly this effect was also present in 10- to 11-year-old age group, although at higher accuracy levels. This suggests that the Good Gestalt may play a larger role in mental rotation even at higher capacity processing levels.

A further alternative explanation of the persistent angularity effect in relation to the Good Gestalt 4-cube aggregate could be attributed to use of strategy. The Good Gestalt aggregate was sufficiently and economically proportioned, resulting in a reduction in the required cognitive processing load and hence may have enabled the use of a holistic strategy. In contrast larger cube aggregates may have prompted an analytic piecemeal strategy as the increased number of relevant object properties (cube elements) would make the child use a computational rather than a figural approach.

3.5.3. Axis of Rotation

A further contributing factor to task difficulty is axis of rotation. Picture-plane rotations were only significantly easier than horizontal depth rotations (x-axis), but not compared to vertical depth (y-axis) rotations. A possible explanation is that children are more familiar with vertical rotations in everyday life and hence benefit from a practise effect. Humans for example rotate around a vertical axis more frequently while standing or dancing than a horizontal axis. However, the finding that picture-plane rotations do not differ significantly from horizontal rotations is surprising, especially with previous research attributing increases in task difficulty to the occlusion of object parts during rotation (Voyer & Hou, 2006). Both vertical and horizontal rotations produce occlusion, however only horizontal rotations differed from picture-plane and vertical rotations. Interestingly, there was also an interaction between axis of rotation and aggregate depth, in which only aggregates with protruding elements were more difficult to rotate than flat aggregates, but only for both depth rotations (vertical and horizontal). A possible explanation is that it is more difficult to identify an intrinsic axis of rotation for aggregates with protruding elements compared to flat aggregates, especially for depth plane rotations, see Figure 46 B and D.

Similar results were observed by Courbois (2000) who showed that a salient intrinsic axis made it easier for 5- and 8-year-old children to rotate abstract line drawings. This result in relation to three-dimensional cube aggregates is important as it allows future test designs to control for mental rotation performance markers of 3D cube aggregates differentiating between picture-plane and two depth rotations.

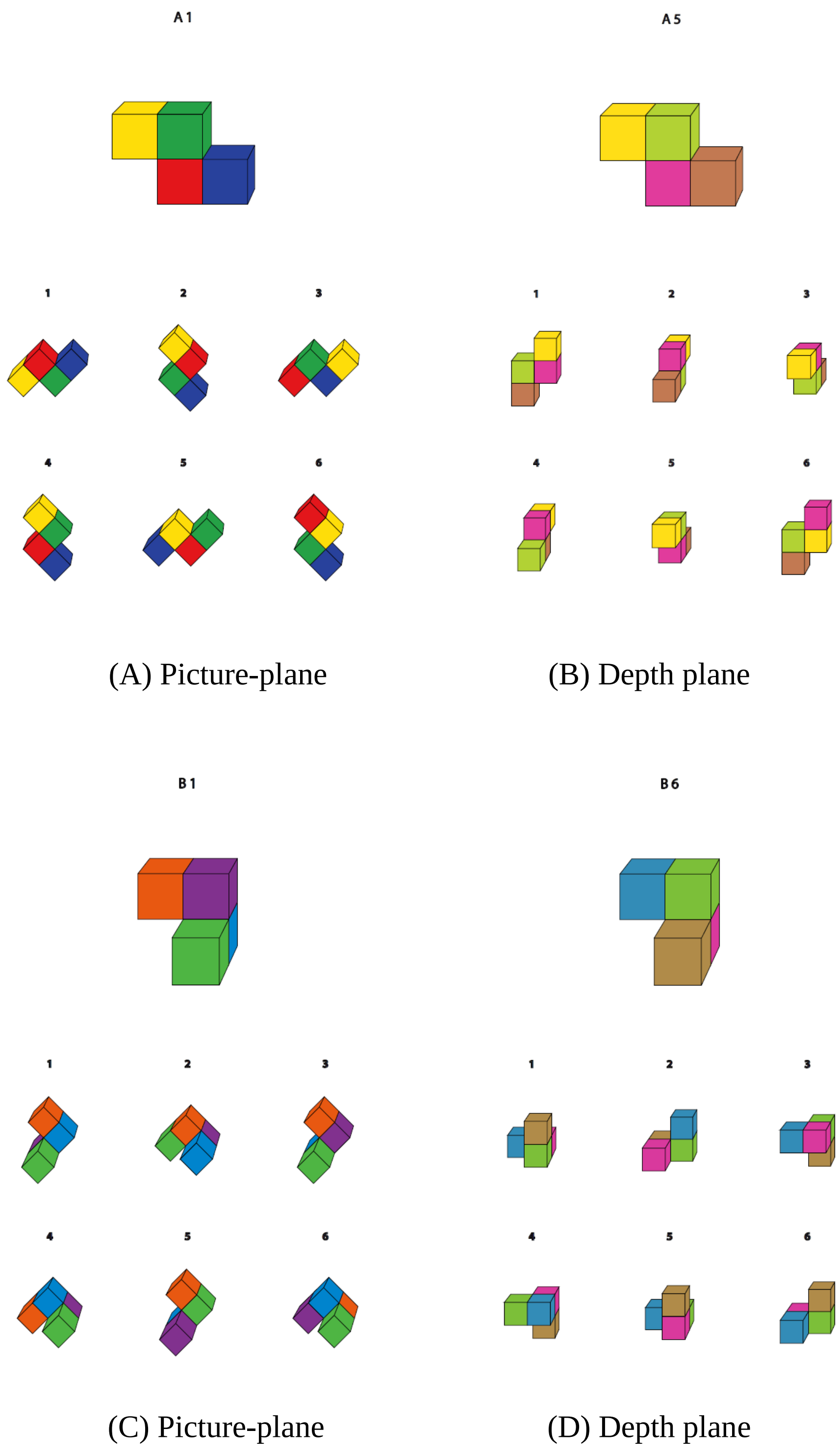


Figure 46 Flat (A, B) and protruding (C, D) 4-cube aggregates.

3.5.4. Individual Differences

The exact onset of sex differences still seems unclear and dependent on the paradigm used. As very young children were unable to complete classical mental rotation tests, reports on the earliest observation of mental rotation ability come from novelty looking paradigms (Quinn & Liben, 2008, 2014; Schwarzer et al., 2012) which have found a male advantage in infants as young as 3- to 4-months. However, the complex cognitive processes involved in classical mental rotation tasks require further higher order abilities such as language processing, encoding, memory, reasoning and more complex response formation (Hoyek et al., 2012), and hence may not be equivalent to infant preferential looking.

In the current study, an early sex difference in favour of boys was already present in 4- to 5-year-olds. This clearly shows that sex differences in tasks using similar geometric stimuli as in the MRT do exist at an early age. These individual differences support previous research on the development of mental rotation ability in children as young as 4 years of age (Frick, Hansen, et al., 2013; Marmor, 1977). This result was robust when controlled for by fluid intelligence using the Raven test.

Collectively these results show that task characteristics such as angle of rotation, aggregate size, axis of rotation and aggregate depth plane distribution, all impacted the ability to perform spatial transformations in young children, and the ability to perform mental rotation, especially in 4- to 5-year-olds. The present research systematically analysed some of the basic attributes which affect task difficulty when processing geometric shapes and when mentally rotating geometric stimuli akin to the Mental Rotation Test (Vandenberg & Kuse, 1978).

A limitation of the current test versions is that they were in booklet and not in a computerized form, so precise time measurement was not possible. Lack of a time constraint may have been the reason for the absence of a main effect of sex in mental rotation in the RCCT. Lange- Küttner and Ebersbach (2012) found that boys were comparably more efficient in mental rotation decision making, as they came to more correct conclusions within a set time. The efficient boys with shorter MRT reaction times were also more likely to draw two occluded cubes, whereas this was not the case for girls. Girls worked at their own steady pace independently of the task at hand, but in 6- to 9-year-old boys, mental rotation reaction times were already task-specific. Although reaction times in the RCCT or CMRT were not measured, as this was seen as problematic for very young children, a new computerized test version that would allow precise time measurements has already been developed. Pairing the new chronometric test version with an eye-tracker could provide interesting insights into what strategies have been used to solve mental rotation tasks and how different strategies effect response time measures.

3.5.5. Mental Rotation and Fluid Intelligence

In general, as in Study 1, the Coloured Mental Rotation Test (CMRT) was slightly easier than the Raven's Coloured Progressive Matrices Test (RCPM) with post hoc tests only showing a significant difference between tests in 4- to 5-year-old girls (CMRT M = 18.8%, RCPM M 46.3%). However, in 6- to 7-year-olds girls this gap had closed as there was no significant difference between the CMRT and the RCPM. suggesting that developmental intervention, possibly affected through entering formal schooling and related training, allowed girls to greatly improve their mental rotation performance.

Similar to Study 1 (Lütke & Lange-Küttner, 2015), the two tests also showed a lot of similarities as demonstrated by the significant correlations. This may have been caused by their similar response format and the common affordance of the visual modality. Moreover, Guttman (1974) suggested that the RCPM measures both spatial ability and fluid intelligence, and Ullstadius et al. (2004) stated that spatial ability tests load considerably on *g.* In short, the significant correlations in older age groups could be attributed to similar changes in development in both mental rotation ability and fluid intelligence, and that the two test response formats are similar in terms of perceptual discrimination between target and distracters.

For picture-plane rotations, the control with fluid intelligence as a covariate supported the majority of the important results from the main ANOVA. The interaction of aggregate size by age groups demonstrated that especially for the youngest 4- to 5-year olds the smallest aggregate size was indeed significantly easier than larger aggregates, and hence provided a suitable reduction in task difficulty. The interaction between aggregate size and aggregate depth, also remained significant, demonstrating that the smallest 4 cube aggregate was significantly easier to rotate when it had protruding cube elements compared to equivalently sized flat aggregates. Most importantly, the four-way interaction between, aggregate size, picture-plane rotations, sex and age remained significant, further strengthening the result that the 4-cube aggregate was conducive to show with its optimal minimalist but informative Gestalt that 4- to 5-year and 10- to 11-year-old boys, and 6- to 7-year and 8- to 9- year old girls, displayed performance decrease in accuracy along with increases in angularity typically reported in MRT studies.

Overall, for 90^{o} rotations around different axis (picture-plane, x-, and y-axis), the control for fluid intelligence resulted in more effect changes. Fluid intelligence appeared to have supressed the main effect for aggregate size in rotations around different types of axis, showing a surprising new effect of larger aggregates facilitating depth plane rotations. Thus, the control for fluid intelligence laid bare that mental imagery can be supported by larger objects. A larger object advantage was, for instance, also found in infants' object tracking as they needed to learn to track smaller object while larger objects were easier (Rosander & von Hofsten, 2002) and in adults' mental rotation advantage for life-size objects (Kaltner & Jansen, 2018). There was also a new effect for aggregate depth and age groups which showed that 4- to 5-year-olds performed similarly independent of aggregate depth, although on a low level, whereas for all older children, flat aggregates were always easier than aggregates with protruding elements in 90^{o} rotations. This result justifies the wide use of 2D mental rotation tests with children this age as they guarantee a satisfactory test performance level. Future research needs to show whether the easier rotation of flat objects occurs because children become better able to visually discriminate between depictions of 2D and 3D shapes. This argument was discussed earlier, that is, it should be more difficult to actually identify the axis of rotation when objects have protruding elements. Children may become better able to visually discriminate between 2D and 3D shapes because (1) once they are in primary school, they learn about geometry and aim to draw 3D objects on a 2D paper surface (Kosslyn, Heldmeyer & Locklear, 1980), and (2) they need to give up their dependence on haptic sensory input as a source of information, as primary school children who are not blind are still supported at this age when shapes and letters are printed in 3D providing haptic information (Permana, 2019; Permana, Sarwanto, & Rintayati, 2018). Another explanation would be that while 4- 5 year-old children were above chance (16.7%) in their 90^{o} rotations as they show an accuracy level of around 50%, to achieve an accuracy level of around 70% like the older children does requires more skill. Children might not improve as fast for 90° rotations around different axes for objects with elements protruding into 3D space, as this requires children to simultaneously process spatial information in more than one depth plane. In any case, the important aspect of this result is that the control for fluid intelligence did enhance and not eliminate the impact of object dimensionality in children's mental rotation.

Overall, the control with the Raven as a covariate revealed that the statistical effects for picture plane rotations were less affected by levels of fluid intelligence than for 90° rotations around different axis. Picture-plane rotations require the simple rigid rotation of the picture itself on a two-dimensional plane, whereas depth rotations require far more complex transformations of the object's orientation (Shepard & Cooper, 1982)

4. General Discussion

The number of available Mental Rotation tests for young children, especially those that use similar geometric stimuli as the influential MRT (Vandenberg & Kuse, 1978), is still very limited. It was therefore important to create novel measures that can bridge the gap between established mental rotation tests for adults and mental rotation measures for 4- to 11-year-old children. The Coloured Cube Test (Lütke & Lange-Küttner, 2015) which assesses perceptual matching and mental rotation, provided an important baseline measure in determining children's ability to successfully process three-dimensional stimuli. The RCCT established that primary school children were able to process the orientation of a single cube, which was reflected in no difference being found between performance in section A, in which all cubes are in a canonical orientation, compared to section B, where all cubes balanced on one corner, see Appendix, RCCT section A & B. The main difference in task difficulty arose amongst individually rotated cubes in sections C and D, in which target cubes had a different orientation than model and distracter cubes. Further, task difficulty was added to section C and D by gradually removing distinctive colours as a distinctive feature. Children's performance did not gradually decrease in the initial three test sections although the cube was presented standing in an unusual perspective in section B and was rotated in section C. In contrast, in section D performance did decrease sharply, most likely because no unique object colours were available to distinguish between distracter cubes. In short, only the same coloured cubes made visual discrimination and object rotation difficult. These results suggest that colour cues can be more important than object location, which supports Hyun and Luck's (2007) findings with adults. It was concluded that the single-cube rotation was too easy for children.

Hence, two categorical baseline measures - object rotation and object set-size - were formative in the design of the second test development (CMRT). As colour distinctiveness proved to be too easy as a cue for young children on the RCCT and as differently orientated cubes increased task difficulty, the CMRT was designed to focus on the geometric properties. Similarly, reducing the size of the cube aggregates to a single cube was too much of a concession to the developmental status of young children as a near ceiling effect was obtained.

The second mental rotation test (CMRT) systematically investigated the fundamental properties (i.e. size, dimensionality, axis of rotation) of cube aggregates and how these influenced mental rotation performance in different age groups. It was revealed that a 4-cube aggregate in particular provided a much more sensitive approach to measuring the development of children's mental rotation ability when using objects in three-dimensional space. Moreover, the second mental rotation test was improved because the degree of angularity was systematically varied when designing the target in amongst distracters in the response options which resulted in the more classical measurement of mental rotation in terms of decreased accuracy over increased angularity.

From the multitude of higher order interactions generated by the 4-cube aggregate it can be concluded that Cowan's (2001) magical number four also exists in mental rotation. This assertion is further strengthened by the fact that the Good Gestalt mental rotation ability was not demonstrated in larger aggregates – not even in the older age groups - but remained constant even with 4-cues distributed in multiple depth planes.

The concept of apperception in relation to a computationally economical or good object has not been developed in previous research. It could hence be argued that the Gute Gestalt 4-cube aggregate is sufficiently and computationally economically proportioned so that it can be perceptually processed in high-level vision prior to entering the observer's consciousness. According to Wundt's "Psychophysical process", three temporal distinct stages exist in the observer interaction with its external environment (Wundt, 1899, 1900). First is *perception* in which an object is unconsciously detected by entering the field of vision. Subsequently, during *apperception* the object enters the observer's attention. And finally, the observer interacts willingly with the stimulus. These temporal distinct processing stages range from an early unlimited capacity, but fragile bottom-up representation, to limited durable structured cognition. It has further been argued that prior to processing object features, spatiotemporal information allows the creation of an object file (Kahneman, Treisman, & Gibbs, 1992) after which feature information is processed by attention-dependent mechanisms (Pylyshyn, 1994). The concept of apperception could also be interpreted in relation to visual attention, with its varieties of sustained attention, divided attention, selective, shared attention and focused attention (Schweizer, Moosbrugger, & Goldhammer, 2005). Nevertheless, none of these types of visual attention conceptualises the figural aspects of attention, and none of these types of visual attention integrates Gestalt principles. Good Gestalt attention reduces the effort and brings out talent in children because of the lower cognitive load and more holistic perception enables more imaginative processes. It was therefore important to establish an optimal figural complexity for cognitive processes involved in mental rotation. A good object conceptualisation was also put forward by research into children's canonical drawings (Davis, 1985;

Hodgson, 2002), but this research focused on the functional aspects rather than on the computationally economic ‘Good Gestalt’ aspects of an object (see also Lange-Küttner, 1994).

In the mental imagery debate, Kosslyn (1994) has suggested that mental rotation involves the visuo-spatial process of forming mental images which are then transformed in the visual buffer. These images can be divided into visual information, that include surface details such as colour and brightness, and spatial properties such as orientation, geometric shape, and depth information. It is therefore important to understand the nature of these images and how their complexity influences mental rotation ability. Both the RCCT and the CMRT have provided evidence on the impact of specific object features, colour, complexity, orientation, and types of rotational axis on task difficulty. By using geometric shapes, rather than simplified images, a systematic comparison of such geometric properties and their impact on the development of the ability to mentally rotate was possible.

A limitation of this study is that no chronometric measurements were taken that would have allowed a more classical comparison to children (Frick, Ferrara, et al., 2013; Jansen, Kaltner, & Memmert, 2017; Quaiser-Pohl et al., 2014) and adult response time data (Peters et al., 2007; Vandenberg & Kuse, 1978; Voyer, 2011). Even though a decrease in accuracy with increases in angularity was observed in the youngest and oldest age groups of the CMRT, it does not rule out that that other strategies than visualising the mental rotation of an object might have been used. Nevertheless, a number of different strategies have been attributed to solving mental rotation tasks (Hegarty, 2018) and the definition of what defines mental rotation have widened to include a growing catalogue of mental rotation tests and experimental paradigms (Lauer et al., 2019; Linn & Petersen, 1985; Voyer et al., 1995). Moreover, to mention a limitation, as children were not measured on another mental rotation test, performance on the new tests could not be compared with their performance on established mental rotation tests. Nevertheless, both the RCCT and CMRT were conceptualised similarly to the Raven Progressive Matrices Test's response format which was tested in addition and allowed extensive controls and test validation. A computerised version of the second test is in development that will allow to measure reaction times in future studies.

The limited number of available tests that currently measure mental rotation in the age groups tested, especially with an active response format using geometric stimuli similar to those for adults (MRT) made the current research studies particularly useful to systematically investigate the parameters that lead to an increase in task difficulty. The results in this study show how set-size, dimensionality and axis of rotation can all be adapted to measure developmental differences in mental rotation. The results from this study can be used to develop new shorter forms of an optimal mental rotation test that is sensitive to young children's mental rotation ability. The results will also be useful to help poorer rotators develop new strategies, as even young children seemed to enjoy trying to solve these "puzzles".

References

Alexander, G. M., & Evardone, M. (2008). Blocks and bodies: Sex differences in a novel version of the mental rotations test. *Hormones and Behavior, 53*(1), 177-184. doi:10.1016/j.yhbeh.2007.09.014

Alington, D. E., Leaf, R. C., & Monaghan, J. R. (1992). Effects of stimulus color, pattern, and practice on sex differences in mental rotations task performance. *Journal of Psychology: Interdisciplinary and Applied, 126*(5), 539-553. doi:10.1080/00223980.1992.10543387

American Educational Research, A., American Psychological, A., National Council on Measurement in, E., Joint Committee on Standards for, E., & Psychological, T. (2014). *Standards for educational and psychological testing*.

Anderson, H. E., Kern, F. E., & Cook, C. (1968). Sex, brain damage, and race effects in the progressive matrices with retarded populations. *The Journal of Social Psychology, 76*(2), 207-211. doi:10.1080/00224545.1968.9933614

Aretz, A. J., & Wickens, C. D. (1992). The mental rotation of map displays. *Human Performance, 5*(4), 303-328. doi:10.1207/s15327043hup0504_3

Astur, R. S., Tropp, J., Sava, S., Constable, R. T., & Markus, E. J. (2004). Sex differences and correlations in a virtual morris water task, a virtual radial arm maze, and mental rotation. *Behavioural Brain Research, 151*(1-2), 103-115. doi:10.1016/j.bbr.2003.08.024

Aziz, S., & Farooqi, G. N. (1991). Gender, age, grade and socio-economic status differences on raven coloured progressive matrices. *Pakistan Journal of Psychological Research, 6*, 3-4. doi:10.1016/0010-0285(75)90003-1

Baddeley, A. D. (2000). The episodic buffer: a new component of working memory? *Trends in cognitive sciences, 4*(11), 417-423. doi:10.1016/S1364-6613(00)01538-2

Bar, M. (2011). *Predictions in the brain: using our past to generate a future*. Oxford, Oxford University Press.

Battista, C., & Peters, M. (2010). Ecological aspects of mental rotation around the vertical and horizontal axis. *Journal of Individual Differences, 31*(2), 110-113. doi:10.1027/1614-0001/a000020

Bauer, B., & Jolicoeur, P. (1996). Stimulus dimensionality effects in mental rotation. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 82. doi:10.1037/0096-1523.22.1.82

Baydar, N., Brooks-Gunn, J., & Furstenberg, F. F. (1993). Early warning signs of functional illiteracy: predictors in childhood and adolescence. *Child Development, 64*(3), 815-829. doi:10.2307/1131220

Beck, A., Franz, C. E., Xian, H., Vuoksimaa, E., Tu, X., Reynolds, C. A., . . . Kremen, W. S. (2018). Mediators of the effect of childhood socioeconomic status on late midlife cognitive abilities: a four decade longitudinal study. *Innovation in Aging, 2*(1). doi:10.1093/geroni/igy003

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 12. doi:10.1037/0096-1523.14.1.12

Bialystok, E. (1989). Children's mental rotations of abstract displays. *Journal of Experimental Child Psychology, 47*(1), 47-71. doi:10.1016/0022-0965(89)90062-3

Birenbaum, M., Kelly, A. E., & Levi-Keren, M. (1994). Stimulus features and sex differences in mental rotation test performance. *Intelligence, 19*(1), 51-64. doi:10.1016/0160-2896(94)90053-1

Blüchel, M., Lehmann, J., Kellner, J., & Jansen, P. (2012). The improvement in mental rotation performance in primary school-aged children after a two-week motor-training. *Educational Psychology, 33*(1), 75-86. doi:10.1080/01443410.2012.70761

Bradle, R., & Corwyn, R. (2002). Socioeconomic status and child development. *Annual Review of Psychology, 53*, 371-399. doi:10.1146/annurev.psych.53.100901.135233

Bremner, J. G., & Batten, A. (1991). Sensitivity to viewpoint in children's drawings of objects and relations between objects. *Journal of Experimental Child Psychology, 52*(3), 375-394. doi:10.1016/0022-0965(91)90070-9

Bremner, J. G., Morse, R., Hughes, S., & Andreasen, G. (2000). Relations between drawing cubes and copying line diagrams of cubes in 7- to 10-year-old children. *Child Development, 71*(3), 621-634. doi:10.1111/1467-8624.00171

Brooks-Gunn, J., & Duncan, G. J. (1997). The effects of poverty on children. *Future of Children, 7*(2), 55-71. doi:10.1037/e555172013-001

Bruce, C. D., & Hawes, Z. (2014). The role of 2D and 3D mental rotation in mathematics for young children: what is it? Why does it matter? And what can we do about it? *ZDM, 47*(3), 331-343. doi:10.1007/s11858-014-0637-4

Bülthoff, H. H., Edelman, S. Y., & Tarr, M. J. (1995). How are three-dimensional objects represented in the brain? *Cereb Cortex, 5*(3), 247-260. doi:10.1093/cercor/5.3.247

Burtis, P. J. (1982). Capacity increase and chunking in the development of short-term memory. *Journal of Experimental Child Psychology, 34*(3), 387-413. doi:10.1016/0022-0965(82)90068-6

Butler, T., Imperato-McGinley, J., Pan, H., Voyer, D., Cordero, J., Zhu, Y. S., . . . Silbersweig, D. (2006). Sex differences in mental rotation: top-down versus bottom-up processing. *NeuroImage, 32*(1), 445-456. doi:10.1016/j.neuroimage.2006.03.030

Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18*(1), 61-75. doi:10.1016/j.lindif.2007.04.005

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

Chen, M. J., & Cook, M. (1984). Representational drawings of solid objects by young children. *Perception, 13*(4), 377-385. doi:10.1068/p130377

Chen, M. J., Therkelsen, M., Griffiths, K., & Therkelson, M. (1984). Representational drawings of solid objects: a longitudinal study. *Visual Arts Research, 10*(1), 27-31. doi:10.1348/026151005X27605

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience, 111*(4), 845. doi:10.1037/0735-7044.111.4.845

Cooper, L. A. (1975). Mental rotation of random two-dimensional shapes. *Cognitive Psychology, 7*(1), 20-43. doi:10.1016/0010-0285(75)90003-1

Cooper, L. A., & Podgorny, P. (1976). Mental transformations and visual comparison processes: effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance, 2*(4), 503-514. doi:10.1037/0096-1523.2.4.503

Cooper, L. A., & Shepard, R. N. (1973). Chronometric studies of the rotation of mental images. In *Visual information processing.* (pp. 72-121). Oxford, England: Academic.

Cooperau, L. A., & Shepard, R. N. (1973). The time required to prepare for a rotated stimulus. *Memory & Cognition, 1*(3), 246-250. doi:10.3758/BF03198104

Corballis, M. C., & Roldan, C. E. (1975). Detection of symmetry as a function of angular orientation. *Journal of experimental psychology. Human perception and performance, 13*, 221-230. doi:10.1037//0096-1523.1.3.221

Corman, L., & Budoff, M. (1974). Factor sructures of retarded and nonretarded children on raven's progressive matrices. *Educational and Psychological Measurement, 34*(2), 407-412. doi:10.1177/001316447403400226

Corporation, T. P. (1995). *DAT for selection-verbal reasoning (European adaptation)*. Orlando, FL: Harcourt Brace and Company.

Courbois, Y. (2000). The role of stimulus axis salience in children's ability to mentally rotate unfamiliar figures. *The European Journal of Cognitive Psychology, 12*(2), 261-269. doi:10.1080/095414400382154

Court, J. H. (1983). Sex differences in performance on raven's progressive matrices: a review. *Alberta Journal of Educational Research, 29*(1), 54-74.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral Brain Sciences, 24*, 87-185. doi: 10.1017/S0140525X01003922

Cowan, N., -, E. E. F., Scott Saults, J., Morey, C. C., Mattox, S., Hismjatullina, A., & Conway, A. R. (2005). On the capacity of attention: its estimation and its role in working memory and cognitive aptitudes. *Cognitive Psychology, 51*(1), 42-100. doi:10.1016/j.cogpsych.2004.12.001

Cowan, N., AuBuchon, A. M., Gilchrist, A. L., Ricker, T. J., & Saults, J. S. (2011). Age differences in visual working memory capacity: not based on encoding limitations. *Developmental Science, 14*(5), 1066-1074. doi:10.1111/j.1467-7687.2011.01060.x

Cowan, N., Blume, C. L., & Saults, J. S. (2013). Attention to attributes and objects in working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 39*(3), 731-747. doi:10.1037/a0029687

Cox, M. V. (1986). Cubes are difficult things to draw. *British Journal of Developmental Psychology, 4*(4), 341-345. doi:10.1111/j.2044-835X.1986.tb01029.x

Cox, M. V., & Perara, J. (1998). Children's observational drawings: a nine-point scale for scoring drawings of a cube. *Educational Psychology, 18*(3), 309-317. doi:10.1080/0144341980180304

Cutzu, F., & Edelman, S. (1998). Representation of object similarity in human vision: psychophysics and a computational model. *Vision Research, 38*(15), 2229-2257. doi:10.1016/S0042-6989(97)00186-7

Davis, A. (1985). Conflict between canonicality and array-specificity in young children's drawings. *British Journal of Developmental Psychology, 3*(4), 363-372. doi:10.1111/j.2044-835X.1985.tb00988.x

Dempster, F. N. (1981). Memory span: Sources of individual and developmental differences. *Psychological Bulletin, 89*(1), 63-100. doi:10.1037/0033-2909.89.1.63

Deregowski, J. B. (1977). Pictures symbols and frames of reference. In G. Butterworth (Ed.), *The child's representation of the world* (pp. 219-236). New York: Plenum Press.

Deregowski, J. B., & Strang, P. (1986). On the drawing of a cube and its derivatives. *British Journal of Developmental Psychology, 4*(4), 323-330. doi:10.1111/j.2044-835X.1986.tb01026.x

Donnelly, N., Cave, K. R., Greenway, R., Hadwin, J. A., Stevenson, J., & Sonuga-Barke, E. (2007). Visual search in children and adults: top-down and bottom-up mechanisms. *The Quarterly Journal of*

Experimental Psychology, 60(1), 120-136. doi:10.1080/17470210600625362

Dorval, M., & Pepin, M. (1986). Effect of playing a video game on a measure of spatial visualization. *Perceptual and Motor Skills, 62*(1), 159-162.

Ebersbach, M. (2009). Achieving a new dimension: children integrate three stimulus dimensions in volume estimations. *Developmental Psychology, 45*(3), 877. doi:10.1037/a0014616

Edelman, S., & Bülthoff, H. H. (1992). Orientation dependence in the recognition of familiar and novel views of three-dimensional objects. *Vision Research, 32*(12), 2385-2400. doi:10.1016/0042-6989(92)90102-o

Enns, J. T., & Rensink, R. A. (1991). Preattentive recovery of three-dimensional orientation from line drawings. *Psychological Review, 98*(3), 335-351. doi:10.1037/0033-295x.98.3.335

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (1994). The gender gap in math: Its possible origins in neighborhood effects. *American Sociological Review, 59*(6), 822-838.

Estes, D. (1998). Young children's awareness of their mental activity: The case of mental rotation. *Child Development, 69*(5), 1345-1360. doi:10.1111/j.1467-8624.1998.tb06216.x

Eysenck, H. J., & Kamin, L. J. (1981). *The intelligence controversy*: Wiley.

Farah, M. J. (2010). Mind, brain, and education in socioeconomic context. *The Developmental Relations among Mind, Brain and Education,* 243-256. doi:10.1007/978-90-481-3666-7_11

Flavell, J. H., Green, F. L., Flavell, E. R., Watson, M. W., & Campione, J. C. (1986). Development of knowledge about the appearance-reality distinction. *Monographs of the Society for Research in Child Development, 51*(1), 1-69. doi:10.2307/1165866

Folk, M. D., & Luce, R. D. (1987). Effects of stimulus complexity on mental rotation rate of polygons. *Journal of Experimental Psychology: Human Perception and Performance, 13*(3), 395. doi:10.1037/0096-1523.13.3.395

Francuz, P. (2014). The effect of three-dimensional imaging of well-known objects on time and accuracy of mental rotation. *Health Psychology Report, 2*(2), 90-98. doi:10.5114/hpr.2014.43915

Freeman, N. H. (1986). How should a cube be drawn? *British Journal of Developmental Psychology, 4*(4), 317-322. doi:10.1111/j.2044-835X.1986.tb01025.x

Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3(1/2) and 5(1/2) years of age. *Cognitive Processing*. doi:10.1007/s10339-012-0534-0

Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development, 28*(4), 386-399. doi:10.1016/j.cogdev.2013.06.002

Frick, A., & Möhring, W. (2013). Mental object rotation and motor development in 8- and 10-month-old infants. *Journal of Experimental Child Psychology, 115*(4), 708-720. doi:0.1016/j.jecp.2013.04.001

Garner, W. R. (1974). *The Processing of Information and Structure*. Oxford, England: Lawrence Erlbaum.

Garner, W. R., & Clement, D. E. (1963). Goodness of pattern and pattern uncertainty. *Journal of Verbal Learning and Verbal Behavior, 2*(5-6), 446-452. doi:10.1016/S0022-5371(63)80046-8

Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences, 18*(4), 419-429. doi:10.1016/j.lindif.2007.09.001

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence, 36*(6), 556-563. doi:10.1016/j.intell.2007.12.003

Gerhardstein, P., & Rovee-Collier, C. (2002). The development of visual search in infants and very young children. *Journal of Experimental Child Psychology, 81*(2), 194-215. doi:10.1006/jecp.2001.2649

Glück, J., & Fabrizii, C. (2010). Gender differences in the mental rotations test are partly explained by response format. *Journal of Individual Differences, 31*(2), 106-109. doi:10.1027/1614-0001/a000019

Göksun, T., Goldin-Meadow, S., Newcombe, N. S., & Shipley, T. (2013). Individual differences in mental rotation: what does gesture tell us? *Cognitive Processing, 14*(2), 153-162. doi:10.1007/s10339-013-0549-1

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: the role of performance factors. *Memory and Cognition, 18*(5), 546-550. doi:10.3758/bf03198487

Guttman, R. (1974). Genetic analysis of analytical spatial ability: raven's progressive matrices. *Behavior Genetics, 4*(3), 274-284. doi:10.1007/BF01074160

Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in cognitive sciences, 13*(2), 65-73. doi:10.1016/j.tics.2008.11.003

Hackman, D. A., Farah, M. J., & Meaney, M. J. (2010). Socioeconomic status and the brain: mechanistic insights from human and animal research. *Nature Reviews Neuroscience, 11*(9), 651-659. doi:10.1038/nrn2897

Halford, G. S., Baker, R., McCredden, J. E., & Bain, J. D. (2005). How many variables can humans process? *Pscyhological Science, 16*(1), 70-76. doi:10.1111/j.0956-7976.2005.00782.x

Halford, G. S., Cowan, N., & Andrews, G. (2007). Separating cognitive capacity from knowledge: A new hypothesis. *Trends in cognitive sciences, 11*(6), 236-242. doi:10.1016/j.tics.2007.04.001

Hardman, K. O., & Cowan, N. (2015). Remembering complex objects in visual working memory: do capacity limits restrict objects or features? *J Exp Psychol Learn Mem Cogn, 41*(2), 325-347. doi:10.1037/xlm0000031

Hawes, Z., LeFevre, J. A., Xu, C., & Bruce, C. D. (2015). Mental rotation with tangible three-dimensional objects: a new measure sensitive to developmental differences in 4- to 8-year-old children. *Mind, Brain, and Education, 9*(1), 10-18. doi:10.1111/mbe.12051

Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4*. doi:10.1016/j.tine.2015.05.001

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

Hegarty, M. (2018). Ability and sex differences in spatial thinking: what does the mental rotation test really measure? *Pychometric Bulletin and Review, 25*(3). doi:10.3758/s13423-017-1347-z

Hegarty, M., & Kozhevnikov, M. (1999). Types of visual–spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*(4), 684.

Heil, M., & Jansen, P. (2008). Gender differences in math and mental rotation accuracy but not in mental rotation speed in 8 years old children. *European Journal of Developmental Science 2*(1), 190-196. doi:10.3233/DEV-2008-21212

Hirai, M., Muramatsu, Y., & Nakamura, M. (2018). Role of the embodied cognition process in perspective-taking ability during childhood. *Child Development, 0*(0). doi:10.1111/cdev.13172

Hodgson, D. (2002). Canonical perspective and typical features in children's drawings: a neuroscientific appraisal. *British Journal of Developmental Psychology, 20*(4), 565-579. doi:10.1348/026151002760390864

Howell, D. D. (1992). *Statistical methods for Psychology* (3rd ed.). Belmont, CA: Wadsworth.

Hoyek, N., Collet, C., Fargier, P., & Guillot, A. (2012). The use of the vandenberg and kuse mental rotation test in children. *Journal of Individual Differences, 33*(1), 62. doi:10.1027/1614-0001/a000063

Hyde, J. S. (1981). How large are cognitive gender differences? *American Psychologist, 36*(8), 892-901. doi:10.1037/0003-066x.36.8.892

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155. doi:10.1037/0033-2909.107.2.139

Hystegge, L., Heim, S., Zettelmeyer, E., & Lange-Kuttner, C. (2012). Gender-specific contribution of a visual cognition network to reading abilities. *British Journal of Psychology*(103(1)), 117-128. doi:10.1111/j.-2044-8295.2011.02050.x

Hyun, J. S., & Luck, S. J. (2007). Visual working memory as the substrate for mental rotation. *Psychonomic Bulletin and Review, 14*(1), 154-158. doi:10.3758/bf03194043

Iachini, T., Ruggiero, G., Bartolo, A., Rapuano, M., & Ruotolo, F. (2019). The effect of body-related stimuli on mental rotation in children, young and elderly adults. *Scientific Reports, 9*(1), 1169. doi:10.1038/s41598-018-37729-7

Inhelder, B., & Piaget, J. (1958). *The growth of logical thinking from childhood to adolescence: An essay on the construction of formal operational structures*. New York: Basic Books.

Jansen, P., Kaltner, S., & Memmert, D. (2017). Approaching behavior reduces gender differences in the mental rotation performance. *Psychological Research, 81*(6), 1192-1200. doi:10.1007/s00426-016-0817-7

Jansen, P., Schmelter, A., Kasten, L., & Heil, M. (2011). Impaired mental rotation performance in overweight children. *Appetite, 56*(3), 766-769. doi:10.1016/j.appet.2011.02.021

Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., & Heil, M. (2013). Mental rotation performance in primary school age children: Are there gender differences in chronometric tests? *Cognitive Development, 28*(1), 51-62. doi:10.1016/j.cogdev.2012.08.005

Jensen, A. R. (1974). How biased are culture-loaded tests? *Genetic Psychology Monographs, 90*(2), 185-255.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Johnson, E. S., & Meade, A. C. (1987). Developmental patterns of spatial ability: An early sex difference. *Child Development*, 725-740. doi:10.2307/1130210

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289-303. doi:10.3758/BF03202498

Jolicoeur, P. (1988). Mental rotation and the identification of disoriented objects. *Canadian Journal of Psychology/Revue canadienne de psychologie, 42*(4), 461. doi:10.1037/h0084200

Jolicoeur, P., Regehr, S., Smith, L. B. J. P., & Smith, G. N. (1985). Mental rotation of representations of two-dimensional and three-dimensional objects. *Canadian Journal of Psychology/Revue canadienne de psychologie, 39*(1), 100. doi:10.1016/s0013-4694(97)89048-0

Jones, L. B., Rothbart, M. K., & Posner, M. I. (2003). Development of executive attention in preschool children. *Developmental Science, 6*(5), 498-504. doi:10.1111/1467-7687.00307

Jordan, K., Wüstenberg, T., Heinze, H. J., Peters, M., & Jäncke, L. (2002). Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia, 40*(13), 2397-2408. doi:10.1016/S0028-3932(02)00076-3

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92*(2), 137. doi:10.1037/0033-295X.92.2.137

Kahneman, D., Treisman, A. M., & Gibbs, B. J. (1992). The reviewing of object files: object-specific integration of information. *Cognitive Psychology, 24*(2), 175-219. doi:10.1016/0010-0285(92)90007-o

Kail, R. (1986). The impact of extended practice on rate of mental rotation. *Journal of Experimental Child Psychology*, *42*(3), 378-391. doi:10.1016/0022-0965(86)90032-9

Kail, R., & Park, Y. (1990). Impact of practice on speed of mental rotation. *Journal of Experimental Child Psychology*, *49*(2), 227-244. doi:10.1016/0022-0965(90)90056-e

Kaltner, S., & Jansen, P. (2018). Sex of human stimulus matters: female and male stimuli are processed differently in mental rotation tasks. *Journal of Cognitive Psychology, 30*(8), 854-862. doi:10.1080/20445911.2018.1530677

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O. A.-O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology, 133*(2), 189-217. doi:10.1037/0096-3445.133.2.189.

Kaplan, R. M., & Sccuzzo, D. P. (1996). *Psychological Testing: Principles, Applications, and Issues*. Pacific Grove, CA: Brooks/ Cole.

Kass, S. J., Ahlers, R. H., & Dugger, M. (1998). Eliminating gender differences through practice in an applied visual spatial task. *Human Performance, 11*(4), 337-349. doi:10.1207/s15327043hup1104_3

Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35*(3), 211-223. doi:10.1016/j.intell.2006.07.009

Khooshabeh, P., & Hegarty, M. (2010). *Representations of shape during mental rotation*. Paper presented at the AAAI SPring Symposium Series, CA: Palo Alto.

Koffka, K. (1935). *Principles of Gestalt psychology*. London: Routledge & Kegan Paul.

Kosslyn, S. M. (1994). *Image and brain: The resolution of the imagery debate*. Cambridge, MA, US: The MIT Press.

Kosslyn, S. M., Ganis, G., & Thompson. (2006). Mental imagery and the human brain. In M. R. R. Q. Jing, G. D'Yewalle, H.-C. Zhang & K. Zhang (Ed.), *Progress in psychological science around the world: Neural, cognitive and developmental issues.* (Vol. 1, pp. 195-209). Hove, UK: Psychological Press.

Kosslyn, S. M., Heldmeyer, K. H., & Locklear, E. P. (1977). Children's drawings as data about internal representations. *Journal of Experimental Child Psychology, 23*(2), 191-211.

Kosslyn, S. M., Heldmeyer, K. H., & Locklear, E. P. (1980). Children's drawings as data about their internal representations. *Journal of Experimental Child Psychology*. doi:10.1016/0022-0965(77)90099-6

Kosslyn, S. M., Margolis, J. A., Barrett, A. M., Goldknopf, E. J., & Daly, P. F. (1990). Age differences in imagery abilities. *Child Development, 61*(4), 995-1010. doi:10.2307/1130871

Kozhevnikov, M., Kosslyn, S. M., & Shephard, J. (2005). Spatial versus object visualizers: a new characterization of visual cognitive style. *Memory and Cognition*, *33*(4), 710-726. doi:10.3758/BF03195337

Krüger, M., Amorim, M. A., & Ebersbach, M. (2014). Mental rotation and the motor system: embodiment head over heels. *Acta Psychologica, 145*, 104-110. doi:10.1016/j.actpsy.2013.11.004

Krüger, M., Kaiser, M., Mahler, K., Bartels, W., & Krist, H. (2013). Analogue mental transformations in 3-year-olds: introducing a new mental rotation paradigm suitable for young children. *Infant and Child Development*, *23*(2), 123-138. doi:10.1002/icd.1815

Krüger, M., & Krist, H. (2009). Imagery and motor processes - when are they connected? The mental rotation of body parts in development. *Journal of Cognition and Development, 10*(4), 239-261. doi:10.1080/15248370903389341

Lachmann, T., & Van Leeuwen, C. (2008). Goodness is central: task invariance of perceptual organization in a dual-task setting. *Japanese Psychological Research, 50*(4), 193-203. doi:10.1111/j.1468-5884.2008.00375.x

Lange-Küttner, C. (1994a). *Gestalt und Konstruktion*. Bern: Huber.

Lange-Küttner, C. (1994b). *Gestalt und Konstruktion [Gestalt and construction]*. Bern: Huber.

Lange-Küttner, C. (2004). More evidence on size modification in spatial axes systems of varying complexity. *Journal of Experimental Child Psychology, 88*, 171-192. doi:10.1016/s0022-0965(04)00028-1

Lange-Küttner, C. (2008a). Size and contour as crucial parameters in children drawing images. In C. Milbrath & H.-M. Trautner (Eds.), *Children's understanding and production of pictures, drawing, and art* (pp. 89-106). Göttingen: Hogrefe.

Lange-Küttner, C. (2008b). When one is really two: switching from object to place memory in the a-not-b search task. *European Journal of Developmental Science, 2*(4), 370-383. doi:10.3233/DEV-2008-2403
Lange-Küttner, C. (2009). Habitual size and projective size: the logic of spatial systems in children's drawings. *Developmental Psychology, 45*, 913-927. doi:10.1037/a0016133
Lange-Küttner, C. (2010a). Gender-specific developmental pathways for boys and girls: the wertheimer common-region-test can predict spatial memory. *European Journal of Developmental Science, 4*, 46-66. doi:10.3233/DEV-2010- 15817-004
Lange-Küttner, C. (2010b). Ready-made and self-made facilitation effects of arrays: priming and conceptualization in children's visual memory. *Swiss Journal of Psychology, 69*(4), 189-200. doi:10.1024/1421-0185/a000023
Lange-Küttner, C. (2012). The importance of reaction times for developmental science: what a difference milliseconds make. *International Journal of Developmental Science, 6*, 51-55. doi:10.3233/DEV-2012-11089
Lange-Küttner, C. (2013). Array effects, spatial concepts, or information processing speed. *Swiss Journal of Psychology, 72*(4), 197-217. doi:doi:10.1024/1421-0185/a000113
Lange-Küttner, C., & Ebersbach, M. (2012). Girls in detail, boys in shape: gender differences when drawing cubes in depth. *British Journal of Psychology*. doi:10.1111/bjop.12010
Lange-Küttner, C., & Green, H. (2007). What is the age of mental rotation ? *Proceedings of the 6th IEEE International Conference on Development and Learning, IEEE Press*, 259-263. doi:10.1109/DEVLRN.2007.4354043
Lange-Küttner, C., Kerzmann, A., & Heckhausen, J. (2002). The emergence of visually realistic contour in the drawing of the human figure. *British Journal of Developmental Psychology, 20*(3), 439-463. doi:10.1348/026151002320620415
Lange-Küttner, C., Korte, M. A., & Stamouli, C. (2019). Parents' autistic personality traits and sex-biased family ratio determine the amount of technical toy choice. *Frontiers in psychology, 10*, 2101. doi:10.3389/fpsyg.2019.02101
Lange-Küttner, C., & Küttner, E. (2015). How to learn places without spatial concepts: does the what-and-where reaction time system in children regulate learning during stimulus repetition? *Brain and Cognition, 97*, 59-73. doi:10.1016/j.bandc.2015.04.008
Lauer, J. E., Udelson, H. B., Jeon, S. O., & Lourenco, S. F. (2015). An early sex difference in the relation between mental rotation and object preference. *Frontiers in psychology, 6*, 558-558. doi:10.3389/fpsyg.2015.00558

Lauer, J. E., Yhang, E., & Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: a meta-analytic review. *Pscyhological Bulletin, 145*(6), 537-565. doi:10.1037/bul0000191

Lehmann, J., Quaiser-Pohl, C., & Jansen, P. (2014). Correlation of motor skill, mental rotation, and working memory in 3- to 6-year-old children. *European Journal of Developmental Psychology, 11*(5), 560-573. doi:10.1080/17405629.2014.888995

Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16*(11), 841-845. doi:10.1111/j.1467-9280.2005.01623.x

Liben, L. S., & Belknap, B. (1981). Intellectual realism: Implications ofor investigations of perceptual perspective taking in young children. *Child Development, 52*(3), 921-924. doi:10.2307/1129095

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 1479-1498. doi:10.2307/1130467

Luquet, G. H. (1927). *Children's drawings: le dessin enfantin.* (A. Costal, Trans.). Paris, France: Alcan.

Lütke, N. (2009). *The rotated colour cube task as a measure of spatial ability*. Unpublished dissertation. London Metropolitan University. London, United Kingdom.

Lütke, N., & Lange-Küttner, C. (2015). Keeping it in three dimensions: measuring the development of mental rotation in children with the rotated colour cube test (RCCT). *International Journal of Developmental Science, 9*(2), 95-114. doi:10.3233/DEV-14154

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: a meta-analysis. *Intelligence, 32*(5), 481-498. doi:10.1016/j.intell.2004.06.008

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences* (Vol. 1): Stanford University Press.

Machilsen, B., Pauwels, M., & Wagemans, J. (2009). The role of vertical mirror symmetry in visual shape detection. *Journal of Vision, 9*(12). doi:10.1167/9.12.11

Marmor, G. S. (1975). Development of kinetic images: when does the child first represent movement in mental images? *Cognitive Psychology, 7*(4), 548-559. doi:10.1016/0010-0285(75)90022-5

Marmor, G. S. (1977). Mental rotation and number conservation: are they related? *Developmental Psychology, 13*(4), 320. doi:10.1037//0012-1649.13.4.320

Metzler, J., & Shepard, R. N. (1974). Transformational studies of the internal representation of three-dimensional objects. In *Theories in cognitive psychology: The Loyola Symposium.* (pp. xi, 386-xi, 386). Oxford, England: Lawrence Erlbaum.

Mezzacappa, E. (2004). Alerting, orienting, and executive attention: developmental properties and sociodemographic correlates in an epidemiological sample of young, urban children. *Child Development, 75*(5), 1373-1386. doi:10.1111/j.1467-8624.2004.00746.x

Milbrath, C. (2005). Epistemology and creative invention. *Journal of Cultural and Evolutionary Psychology, 3*(2), 119-140. doi:10.1556/JCEP.3.2005.2.2

Milbrath, C. (2009). Comparative developmental and social perspectives on the mystery of upper paleolithic art. In C. M. C. Lightfoot (Ed.), *Arts and Human Development*. New York, NY: Psychology Press.

Mitchelmore, M. C. (1978). Developmental stages in children's representation of regular solid figures. *The Journal of Genetic Psychology, 133*(2), 229-239. doi:10.1080/00221325.1978.10533380

Möhring, W., & Frick, A. (2013). Touching up mental rotation: effects of manual experience on 6-month-old infants' mental object rotation. *Child Development, 84*(5), 1554-1565. doi:10.1111/cdev.12065

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: a sex difference. *Psychological Science, 19*(11), 1063-1066. doi:10.1111/j.1467-9280.2008.02200.x

Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old Infants. *Infancy, 16*(4), 435-445. doi:10.1111/j.1532-7078.2010.00058.x

Moore, V. (1986). The use of a colouring task to elucidate children's drawings of a solid cube. *British Journal of Developmental Psychology, 4*(4), 335-340. doi:10.1111/j.2044-835X.1986.tb01028.x

Moreau, D. (2012). Differentiating two- from three-dimensional mental rotation training effects. *The Quarterly Journal of Experimental Psychology*, 1-15. doi:10.1080/17470218.2012.744761

Morra, S. (2008). Spatial structures in children's drawings: how do they develop? In A. Vinter & C. Lange-Küttner (Eds.), *Drawing and the Non-Verbal Mind: A Life-Span Perspective* (pp. 159-194). Cambridge: Cambridge University Press.

Morra, S., Moizo, C., & Scopesi, A. (1988). Working memory (or the M operator) and the planning of children's drawings. *Journal of Experimental Child Psychology, 46*(1), 41-73. doi:10.1016/0022-0965(88)90022-7

Neuburger, S., Heuser, V., Jansen, P., & Quaiser-Pohl, C. (2012, 2012//). *Influence of rotational axis and gender-stereotypical nature of rotation stimuli on the mental-rotation performance of male and female fifth graders.* Paper presented at the Spatial Cognition VIII, Berlin, Heidelberg.

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2011). Gender differences in pre-adolescents' mental-rotation performance: do they depend on grade and stimulus type? *Personality and Individual Differences, 50*(8), 1238-1242. doi:10.1016/j.paid.2011.02.01

Neville, H. J., Stevens, C., Pakulak, E., Bell, T. A., Fanning, J., Klein, S., & Isbell, E. (2013). Family-based training program improves brain function, cognition, and behavior in lower socioeconomic status preschoolers. *Proceedings of the National Academy of Sciences, 110*(29), 12138-12143. doi:10.1073/pnas.1304437110

Nicholls, A. L., & Kennedy, J. M. (1995). Foreshortening in cube drawings by children and adults. *Perception, 24*(12), 1443-1456. doi:10.1068/p241443

Noble, K. G., Houston, S. M., Brito, N. H., Bartsch, H., Kan, E., Kuperman, J. M., . . . Sowell, E. R. (2015). Family income, parental education and brain structure in children and adolescents. *Nat Neurosci, 18*(5), 773-778. doi:10.1038/nn.3983

Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object. *Memory & Cognition, 41*(8), 1212-1227. doi:10.3758/s13421-013-0333-6

Okagaki, L., & Frensch, P. A. (1994). Effects of video game playing on measures of spatial performance: gender effects in late adolescence. *Journal of Applied Developmental Psychology, 15*(1), 33-58. doi:10.1016/0193-3973(94)90005-1

Palmer, S. E. (1991). Goodness, gestalt, groups, and garner: local symmetry subgroups as a theory of figural goodness. In G. R. Lockhead & J. R. Pomerantz (Eds.), *The Perception of Structure: Essays in honor of Wendell R. Garner*. (pp. 23-39). Washington, DC, US: American Psychological Association.

Palmer, S. E., Rosch, E., & Chase, P. (1981). Canonical perspective and the perception of objects. In *Attention and performance* (Vol. 9, pp. 135-151). Hilsdale, NJ: Erlbaum.

Pascual-Leone, J., & Smith, J. (1969). The encoding and decoding of symbols by children: A new experimental paradigm and a neo-Piagetian model. *Journal of Experimental Child Psychology, 8*(2), 328-355. doi:10.1016/0022-0965(69)90107-6

Permana, D. (2019). Using stereoscopic 3D images for effective learning in primary school. *International Journal for Educational and Vocational Studies, 1*. doi:10.29103/ijevs.v1i5.1369

Permana, D., Sarwanto, S., & Rintayati, P. (2018). Integration of stereoscopic 3D images in primary school textbooks. *International Journal of Multicultural and Multireligious Understanding, 5*, 45. doi:10.18415/ijmmu.v5i4.180

Perrucci, V., Agnoli, F., & Albiero, P. (2008). Children's performance in mental rotation tasks: orientation-free features flatten the slope. *Developmental Science, 11*(5), 732-742. doi:10.1111/j.1467-7687.2008.00723.x

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*(2), 176-184. doi:10.1016/j.bandc.2004.08.052

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*(1), 39-58. doi:10.1006/brcg.1995.1032

Peters, M., Lehmann, W., Takahira, S., Takeuchi, Y., & Jordan, K. (2006). Mental rotation test performance in four cross-cultural samples (N= 3367): overall sex differences and the role of academic program in performance. *Cortex, 42*(7), 1005-1014.

Peters, M., Manning, J. T., & Reimers, S. (2007). The effects of sex, sexual orientation, and digit ratio (2D:4D) on mental rotation performance. *Archives of Sexual Behaviour, 36*(2), 251-260. doi:10.1007/s10508-006-9166-8

Piaget, J. (1969). *The mechanisms of perception*. London: Routledge & Kegan Paul.

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

Piaget, J., & Inhelder, B. (1971). *Mental imagery in the child: a study of the development of imaginal representation*. London, United Kingdom: Routledge & Kegan Paul.

Pylyshyn, Z. W. (1977). What the mind's eye tells the mind's brain: A critique of mental imagery. In J. Nicholas (Ed.), *Images, perception, and knowledge* (pp. 1-36). Dordrecht: Springer.

Pylyshyn, Z. W. (1994). Some primitive mechanisms of spatial attention. *Cognition, 50*(1), 363-384. doi:10.1016/0010-0277(94)90036-1

Pylyshyn, Z. W. (2002). Mental imagery: In search of a theory. *Behavioral and Brain Sciences*.

Pylyshyn, Z. W. (2003). *Seeing and visualizing: It's not what you think*. Cambridge, MA: MIT Press.

Quaiser-Pohl, C. (2003). The mental cutting test" schnitte" and the picture rotation test-two new measures to assess spatial ability. *International Journal of Testing, 3*(3), 219-231. doi:10.1207/s15327574ijt0303_2

Quaiser-Pohl, C., Geiser, C., & Lehmann, W. (2006). The relationship between computer-game preference, gender, and mental-rotation ability. *Personality and Individual Differences, 40*(3), 609-619. doi:10.1016/j.paid.2005.07.015

Quaiser-Pohl, C., & Lehmann, W. (2010). Girls' spatial abilities: charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology, 72*(2), 245-260. doi:10.1348/000709902158874

Quaiser-Pohl, C., Neuburger, S., Heil, M., Jansen, P., & Schmelter, A. (2014). Is the male advantage in mental-rotation performance task independent? On the usability of chronometric tests and paper-and-pencil tests in children. *International Journal of Testing, 14*(2), 122-142. doi:10.1080/15305058.2013.860148

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19*(11), 1067-1070. doi: 10.1111/j.1467-9280.2008.02201.x

Quinn, P. C., & Liben, L. S. (2014). A sex difference in mental rotation in infants: convergent evidence. *Infancy, 19*(1), 103-116. doi:10.1111/infa.12033

Raven, J. C. (1939). *Progressive Matrices: A Perceptual Test of Intelligence.* London: H.K. Lewis.

Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*: Oxford Psychologists Press.

Rensink, R. A. (2001). Four-sight in hindsight: The existence of magical numbers in vision. *Behavioral and Brain Sciences, 24*(1), 141-142. doi:10.1017/S0140525X01483920

Richards, M., & Wadsworth, M. E. J. (2004). Long term effects of early adversity on cognitive function. *Archives of Disease in Childhood, 89*(10), 922-927. doi:10.1136/adc.2003.032490

Riggs, K. J., McTaggart, J., Simpson, A., & Freeman, R. P. J. (2006). Changes in the capacity of visual working memory in 5- to 10-year-olds. *Journal of Experimental Child Psychology, 95*(1), 18-26. doi:10.1016/j.jecp.2006.03.009

Rosander, K., & von Hofsten, C. (2002). Development of gaze tracking of small and large objects. *Experimental Brain Research, 146*(2), 257-264. doi:10.1007/s00221-002-1161-2

Rosser, R. A. (1980). *Acquisition of spatial concepts in relation to age and sex.* Arizona University, Washington, D.C.

Rosser, R. A., Ensing, S. S., & Mazzeo, J. (1985). The role of stimulus salience in young children's ability to discriminate two-dimensional rotations: Reflections on a paradigm. *Contemporary Educational Psychology, 10*(2), 95-103. doi:10.1016/0361-476x(85)90010-4

Ruthsatz, V., Neuburger, S., Rahe, M., Jansen, P., & Quaiser-Pohl, C. (2017). The gender effect in 3D-Mental-rotation performance with familiar and gender-stereotyped objects – a study with elementary school children. *Journal of Cognitive Psychology, 29*(6), 717-730. doi:10.1080/20445911.2017.1312689

Schwamborn, A., Mayer, R. E., Thillmann, H., Leopold, C., & Leutner, D. (2010). Drawing as a generative activity and drawing as a prognostic activity. *Journal of Educational Psychology, 102*, 872-879. doi:10.1037/a0019640

Schwarzer, G., Freitag, C., Buckel, R., & Lofruthe, A. (2012). Crawling is associated with mental rotation ability by 9-month-old infants. *Infancy, 18*(3), 432-441. doi:10.1111/j.1532-7078.2012.00132

Schweizer, K., Moosbrugger, H., & Goldhammer, F. (2005). The structure of the relationship between attention and intelligence. *Intelligence, 33*(6), 589-611. doi:10.1016/j.intell.2005.07.001

Serbin, L., Zelkowitz, P., Doyle, A., Gold, D., & Wheaton, B. (1990). The socialization of sex-differentiated skills and academic performance: A mediational model. *Sex Roles, 23*(11-12), 613-628. doi:10.1007/bf00289251

Shepard, R. N., & Cooper, L. A. (1982). *Mental images and their transformations*. Cambridge, Mass.: MIT Press.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701-703. doi:10.1126/science.171.3972.701

Singh-Manoux, A., Richards, M., & Marmot, M. (2005). Socioeconomic position across the lifecourse: how does it relate to cognitive function in mid-life? *Annals of Epidemiology, 15*(8), 572-578. doi:10.1016/j.annepidem.2004.10.007

Smith, J. R., Brooks-Gunn, J., & Klebanaov, P. (1997). The consequences of living in poverty for young children’s cognitive and verbal ability and early school achievement. In G. J. Duncan & J. Brooks-Gunn (Eds.), *Consequences of growing up poor* (pp. 132-189). New York, N.Y.: Russell Sage.

Stupica, B., & Cassidy, J. (2014). Priming as a way of understanding children’s mental representations of the social world. *Developmental Review, 34*(1), 77-91. doi:10.1016/j.dr.2013.12.003

Subrahmanyam, K., & Greenfield, P. M. (1994). Effect of video game practice on spatial skills in girls and boys. *Journal of Applied Developmental Psychology, 15*(1), 13-32. doi:10.1016/0193-3973(94)90004-3

Tapley, S. M., & Bryden, M. (1977). An investigation of sex differences in spatial ability: Mental rotation of three-dimensional objects. *Canadian Journal of Psychology/Revue canadienne de psychologie, 31*(3), 122.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study on the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin & Review, 2*(1), 55-82. doi:10.3758/BF03214412

Thompson, J., Nuerk, H., Moeller, K., & Kadosh, R. (2013). The link between mental rotation ability and basic numerical representations. *Acta Psychologica, 144*, 324-331. doi:10.1016/j.actpsy.2013.05.009

Thurstone, L. L. (1938). *Primary mental abilities* (Vol. 8). Chicago, IL: University of Chicago Press.

Titze, C., Heil, M., & Jansen, P. (2008). Gender differences in the Mental Rotations Test (MRT) are not due to task complexity. *Journal of Individual Differences, 29*(3), 130-133.

Titze, C., Jansen, P., & Heil, M. (2010a). Mental rotation performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology, 7*(4), 432-444. doi:10.1080/17405620802548214

Titze, C., Jansen, P., & Heil, M. (2010b). Mental rotation performance in fourth graders: no effects of gender beliefs (yet?). *Learning and Individual Differences, 20*(5), 459-463. doi:10.1016/j.lindif.2010.04.003

Toomela, A. (1999). Drawing development: stages in the representation of a cube and a cylinder. *Child Development, 70*(5), 1141-1150. doi:10.1111/1467-8624.00083

Toomela, A. (2003). Developmental stages in children's drawings of a cube and a doll. *Trames, 7*(3), 164-182.

Treisman, A. M. (1986). Properties, parts and objects. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, pp. 1-70). New York, NY: Wiley.

Treisman, A. M. (1988). Features and objects: the fourteenth barlett memorial lecture. *The Quarterly Journal of Experimental Pscyhology Section A, 40*(2), 201-237. doi:10.1080/02724988843000104

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*(1), 97-136. doi:10.1016/0010-0285(80)90005-5

Tzuriel, D., & Egozi, G. (2007). Dynamic assessment of spatial abilities of young children: effects of gender and task characteristics. *Journal of Cognitive Education and Psychology, 6*(2), 219-247. doi:10.1891/194589507787382160

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81*(5), 1417-1430. doi:10.1111/j.1467-8624.2010.01482.x

Ullstadius, E., Carlstedt, B., & Gustafsson, J.-E. (2004). Multidimensional item analysis of ability factors in spatial test items. *Personality and Individual Differences, 37*(5), 1003-1012. doi:10.1016/j.paid.2003.11.009

Valencia, R. R. (1984). Reliability of the raven coloured progressive matrices for anglo and for mexican-american children. *Psychology in the Schools, 21*(1), 49-52. doi:10.1002/1520-6807

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599-604. doi:10.2466/pms.1978.47.2.599

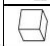
Vederhus, L., & Krekling, S. (1996). Sex differences in visual spatial ability in 9-year-old children*. *Intelligence, 23*(1), 33-43. doi:10.1016/s0160-2896(96)80004-3

Vijver, F. V. D. (1997). Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-Cultural Psychology, 28*(6), 678-709. doi:10.1177/0022022197286003
Voyer, D. (1995). Effect of practice on laterality in a mental rotation task. *Brain and Cognition, 29*(3), 326-335. doi:DOI: 10.1006/brcg.1995.1285
Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychonomic Bulletin and Review, 18*(2), 267-277. doi:10.3758/s13423-010-0042-0
Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the mental rotations test. *Canadian Journal of Experimental Psychology, 60*(2), 91-100. doi:10.1037/cjep2006010
Voyer, D., Rodgers, M., & McCormick, P. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory and Cognition, 32*(1), 72-82. doi:10.3758/BF03195821
Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: a factor analysis. *Acta Psychologica*. doi:10.1016/j.actpsy.2004.05.003
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. doi:10.1037/0033-2909.117.2.250
Waszak, F., Drewing, K., & Mausfeld, R. (2005). Viewer-external frames of reference in the mental transformation of 3-D objects. *Perception & Psychophysics, 67*(7), 1269-1279. doi:10.3758/BF03193558
Wechsler, D. (1981). *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
Wertheimer, M. (1923). Untersuchungen zur Lehre von der gestalt. II. *Psychologische Forschung, 4*(1), 301-350. doi:10.1007/BF00410640
Wilder, N. E., & Green, D. R. (1963). Expression of concepts through writing and drawing and effects of shifting medium. *Journal of Educational Psychology, 54*(4), 202-207. doi:10.1037/h0046691
Wimmer, M., Robinson, E., & Doherty, M. (2017). Are developments in mental scanning and mental rotation related? *PLoS ONE, 12*, e0171762. doi:10.1371/journal.pone.0171762
Wundt, W. (1899). Zur Kritik tachistoskopischer: Versuche I. *Philosophische Studien, 15*, 287-317.
Wundt, W. (1900). Zur Kritik tachistoskopischer: Versuche II. *Philosophische Studien, 16*, 61-81.
Yuille, J. C., & Steiger, J. H. (1982). Nonholistic processing in mental rotation: some suggestive evidence. *Attention, Perception, and Psychophysics, 31*(3), 201-209. doi:10.3758/BF03202524
Zhang, Z., Liu, J., Li, L., & Xu, H. (2018). The long arm of childhood in china: early-life conditions and cognitive function among middle-

aged and older adults. *Journal of Aging and Health*, *30*(8), 1319-1344. doi:10.1177/0898264317715975

Appendix Table A1

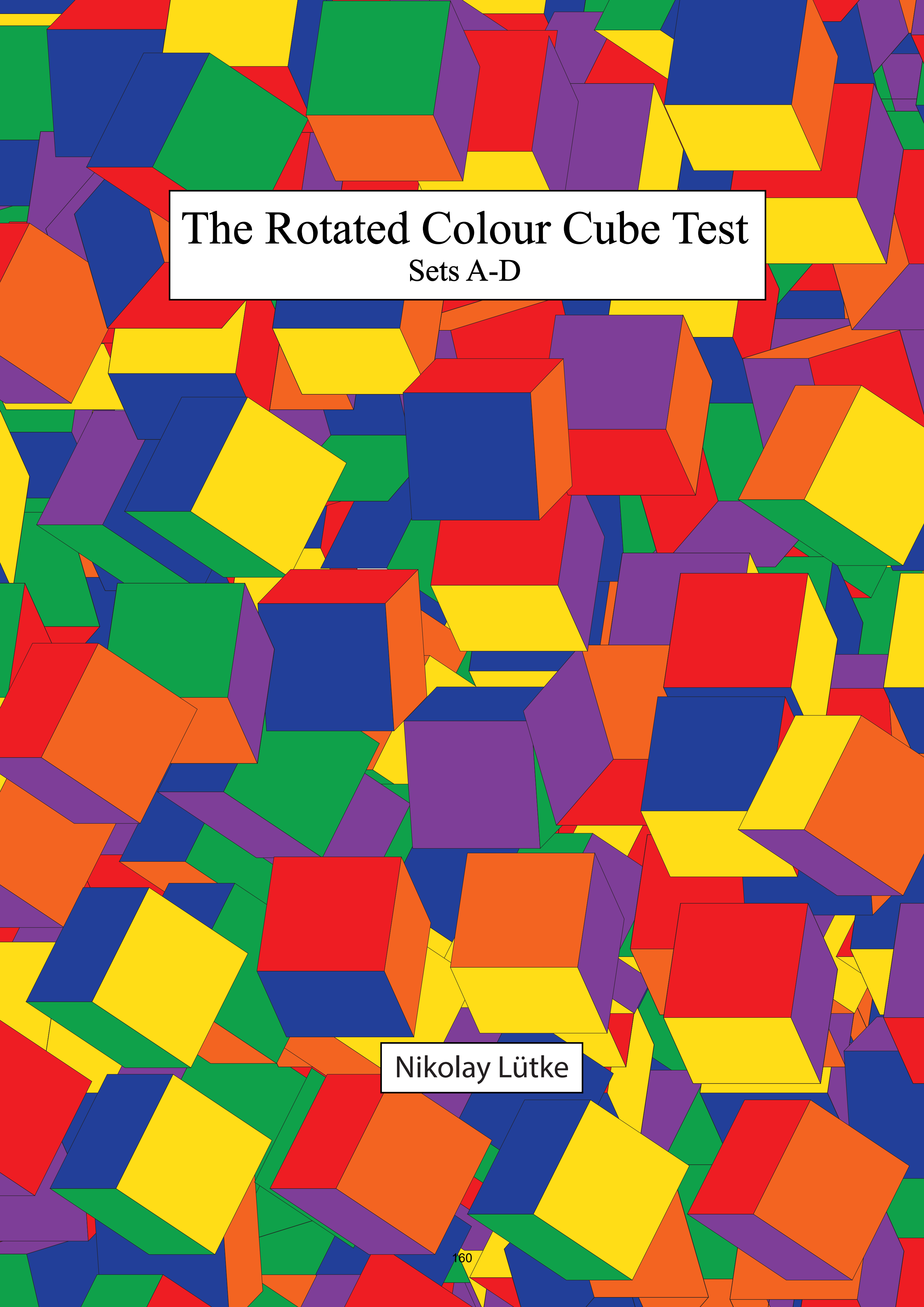
Properties of the Target Cube, the Test Cube and Distracters

Level	Page	Cubes	Section	Rotation
			RCCT A	
I	**A1**	6	100 % distracters with 1 colour identical to test cube	
	A2	8		
II	**A3**	6	100 % distracters with 2 colours identical to test cube	
	A4	8		
III	**A5**	6	50% distracters with 3 colours identical to test cube	
	A6	8		
IV	**A7**	6	100 % distracters with 3 colours identical to test cube	
	A8	8		
			RCCT B	
I	**B1**	6	100 % distracters with 1 colour identical to test cube	
	B2	8		
II	**B3**	6	100 % distracters with 2 colours identical to test cube	
	B4	8		
III	**B5**	6	50% distracters with 3 colours identical to test cube	
	B6	8		
IV	**B7**	6	100 % distracters with 3 colours identical to test cube	
	B8	8		
			RCCT C	
I	**C1**	6	100 % distracters with 1 colour identical to test cube	
	C2	8		
II	**C3**	6	100 % distracters with 2 colours identical to test cube	
	C4	8		
III*	**C5**	6	100% distracters with 2 colours identical to test cube	
	C6	8		
IV*	**C7**	6	50 % distracters with 3 colours identical to test cube	
	C8	8		
			RCCT D	
I	**D1**	6	100% distracters with 3 colours identical to test cube	
	D2	8		
II	**D3**	6	100% distracters with 3 colours identical to test cube	
	D4	8		
III*	**D5**	6	100% distracters with 3 colours identical to test cube	
	D6	8		
IV*	**D7**	6	100 % distracters with 3 colours identical to test cube	
	D8	8		

Note. * indicates mental rotation task

The Rotated Colour Cube Test

Sets A-D

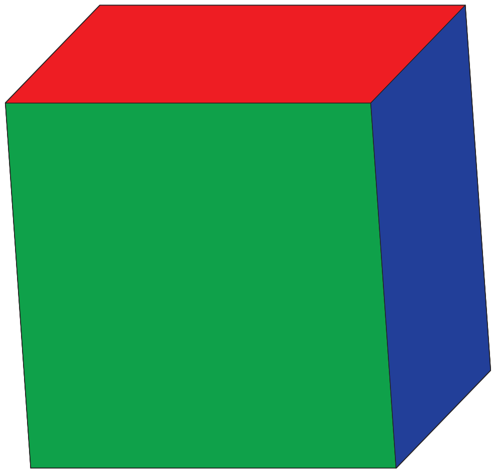


Nikolay Lütke

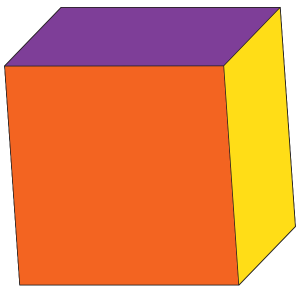
The Rotated Colour Cube Test

Practice Section

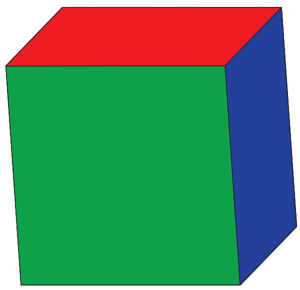
P 1



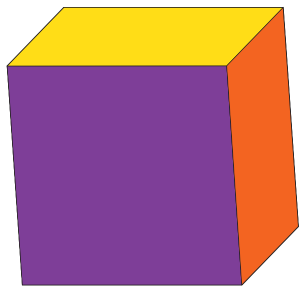
1



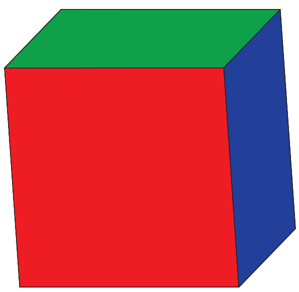
2



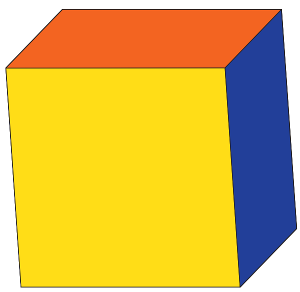
3



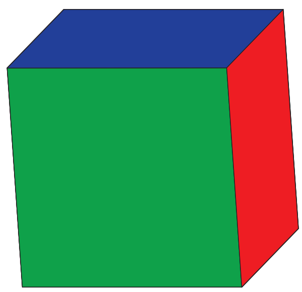
4



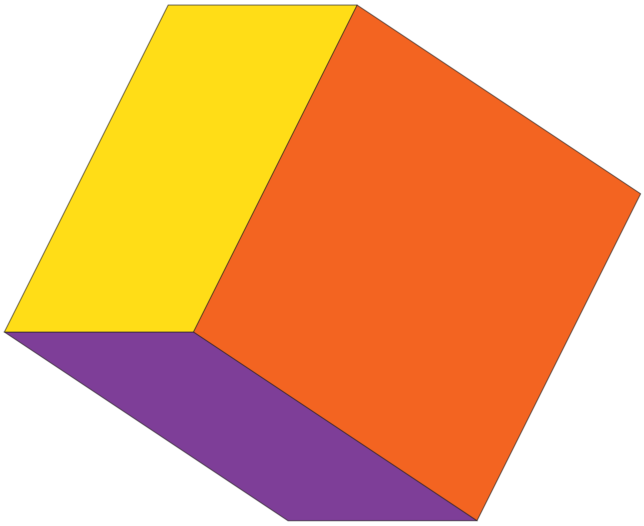
5



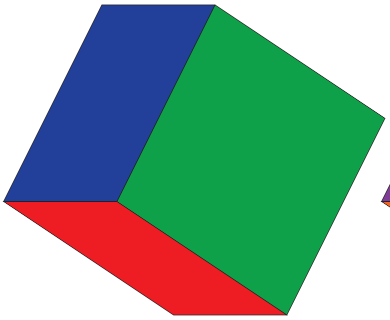
6



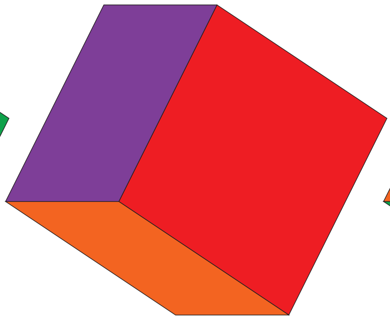
P 2



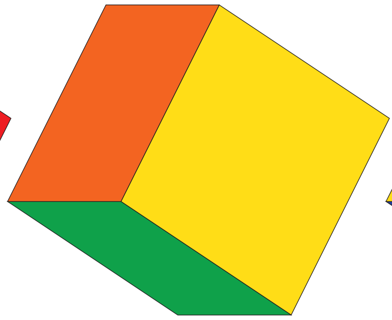
1



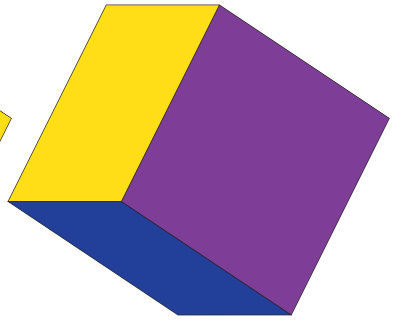
2



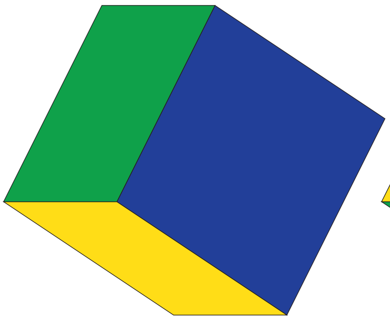
3



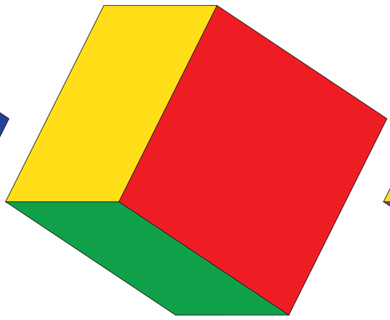
4



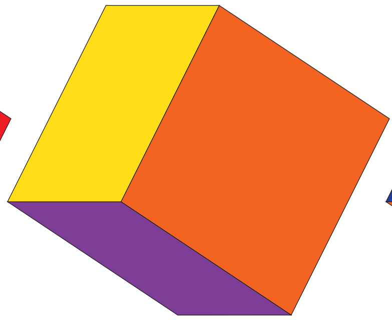
5



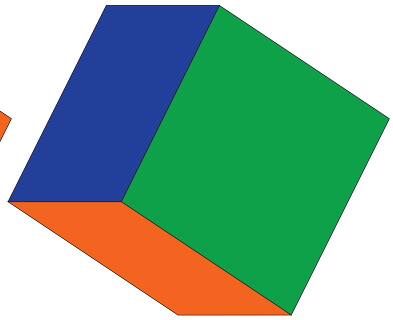
6



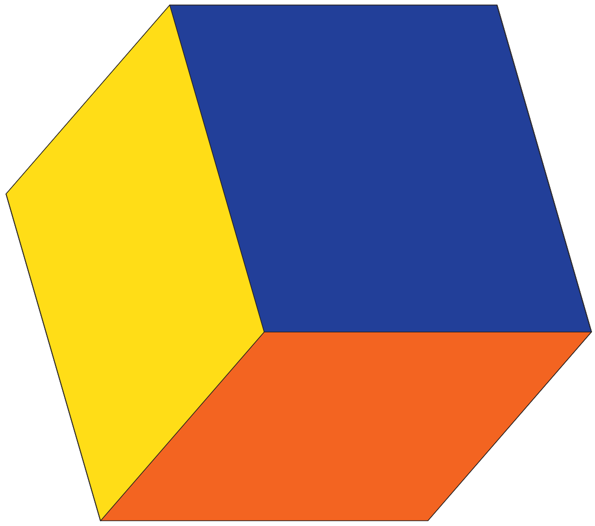
7



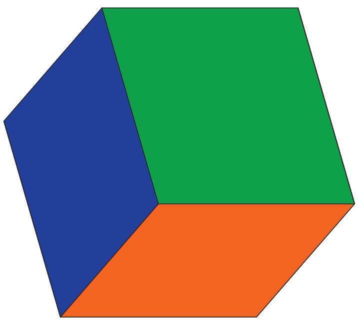
8



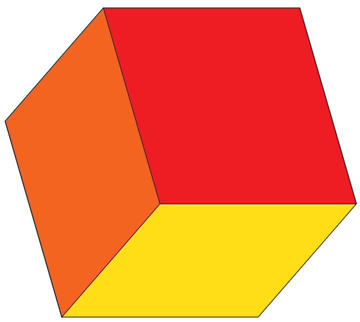
P 3



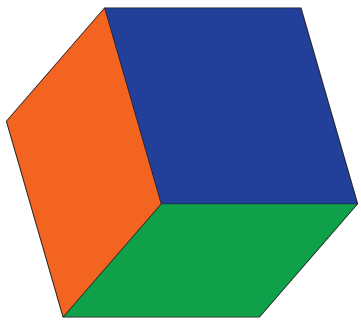
1



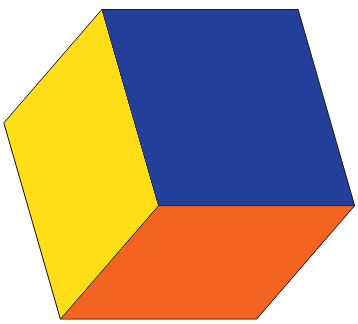
2



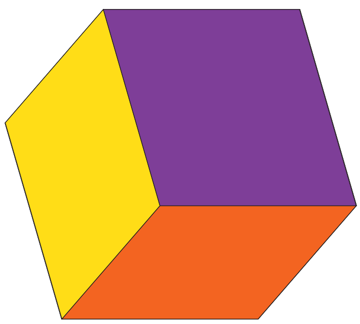
3



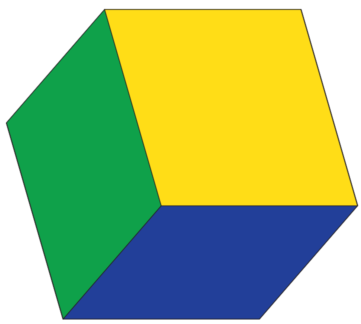
4



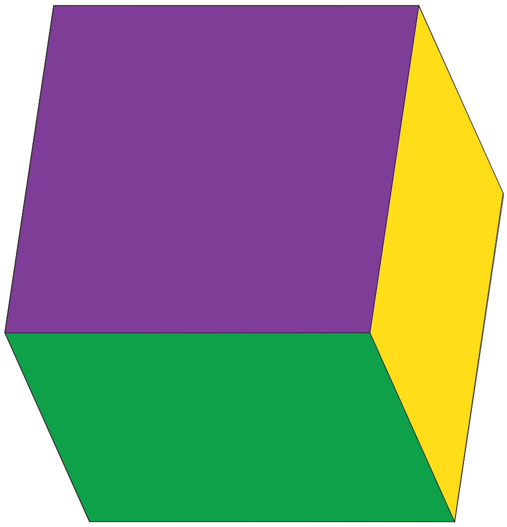
5



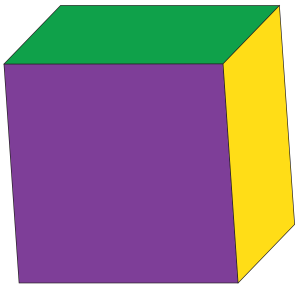
6



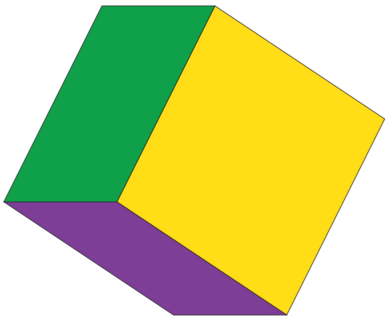
P 4



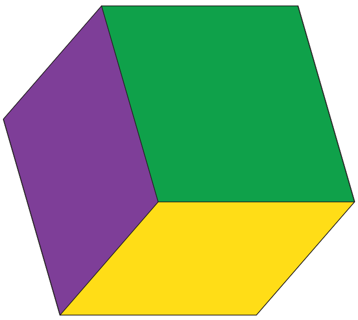
1



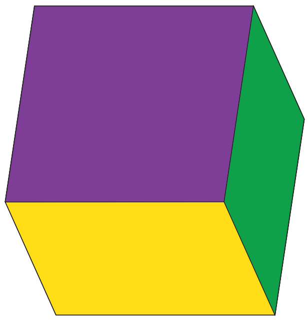
2



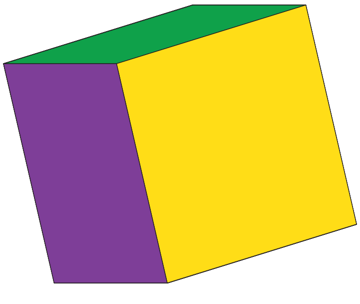
3



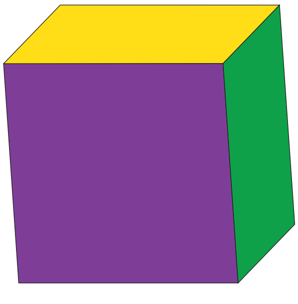
4



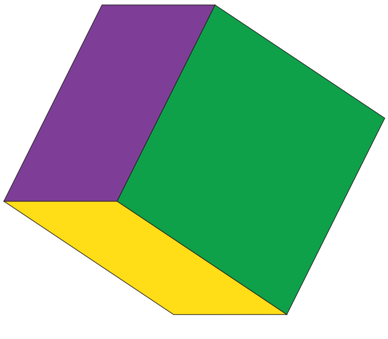
5



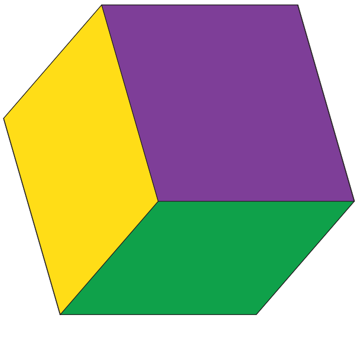
6



7



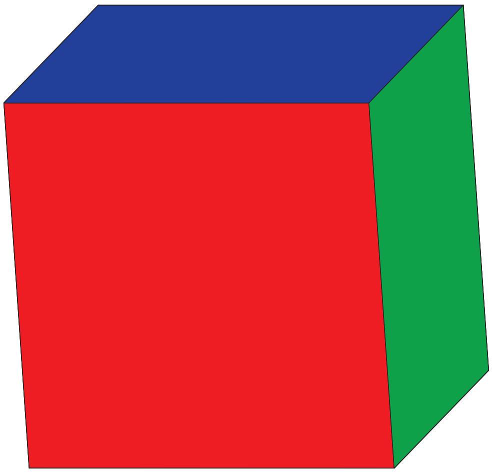
8



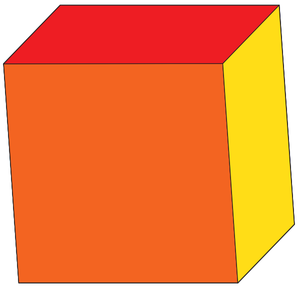
The Rotated Colour Cube Test

Set A

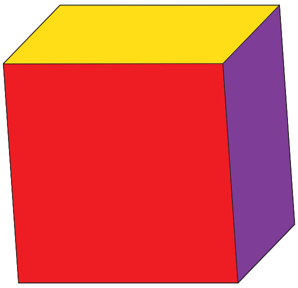
A 1



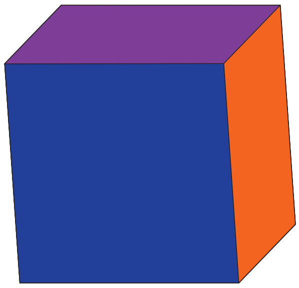
1



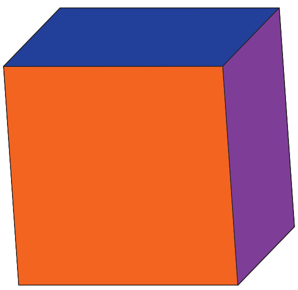
2



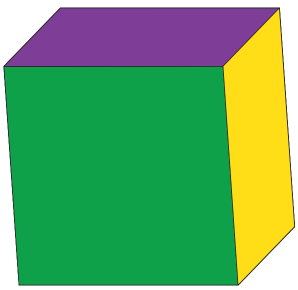
3



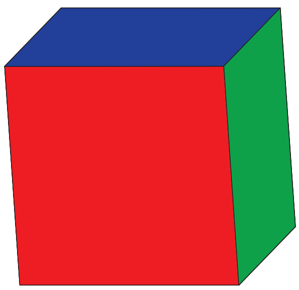
4



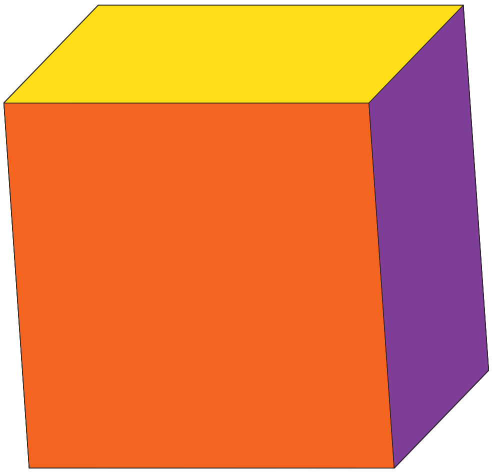
5



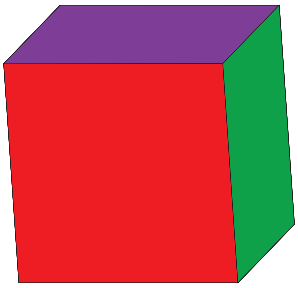
6



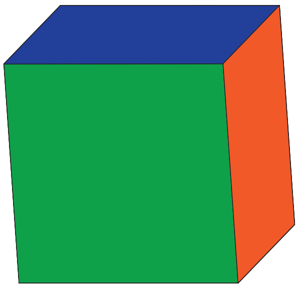
A 2



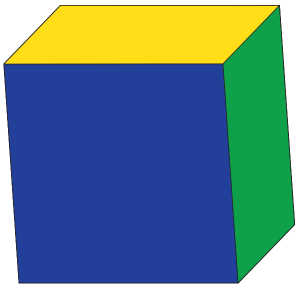
1



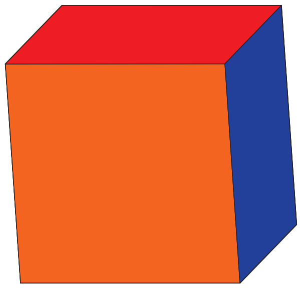
2



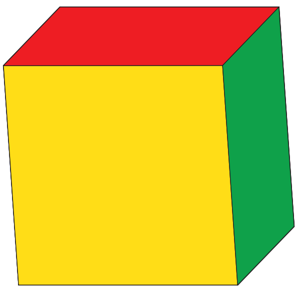
3



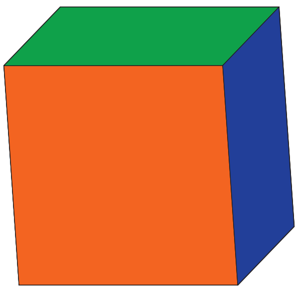
4



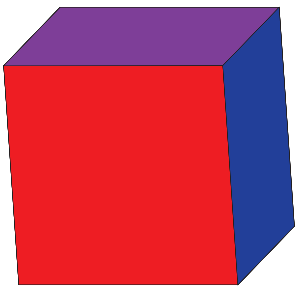
5



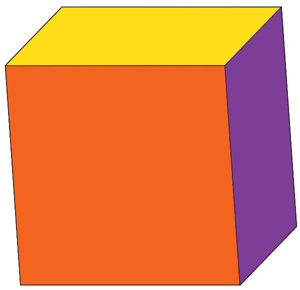
6



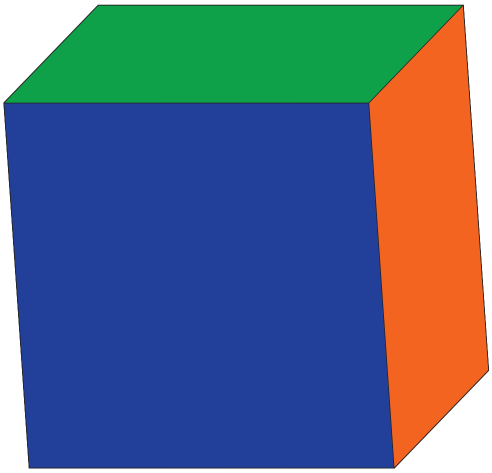
7



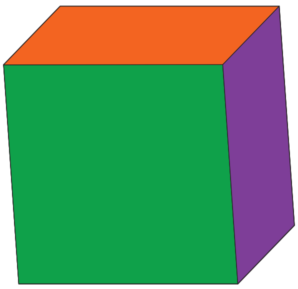
8



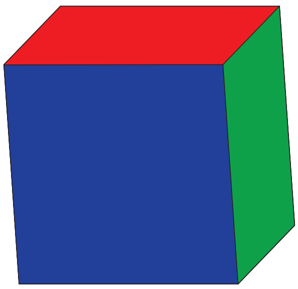
A 3



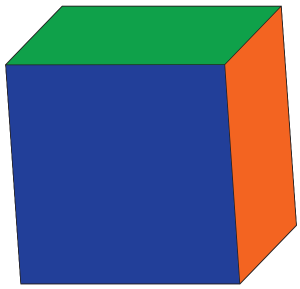
1



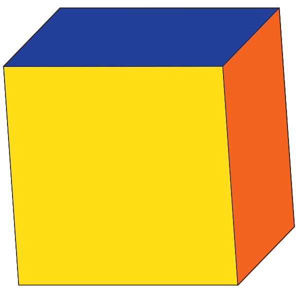
2



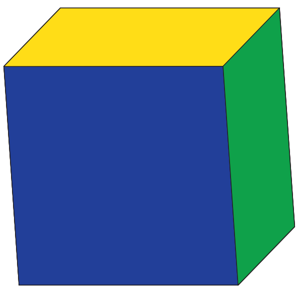
3



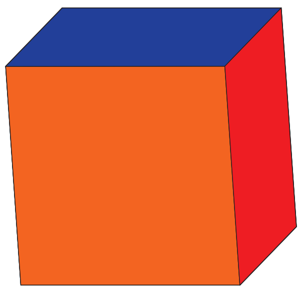
4



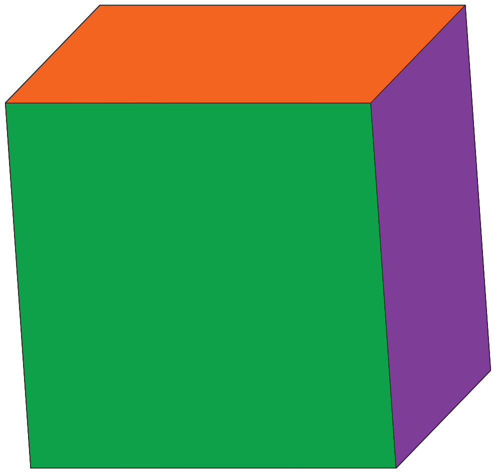
5



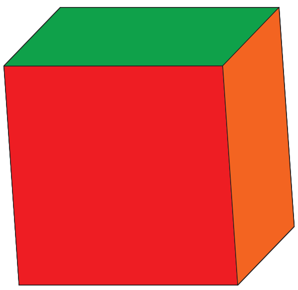
6



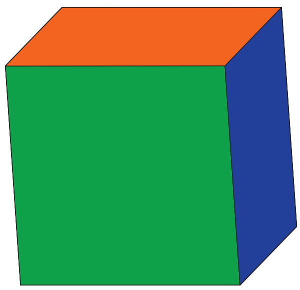
A 4



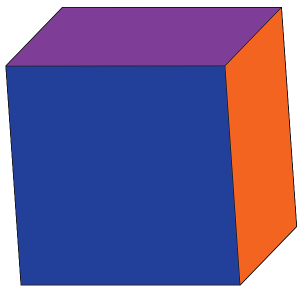
1



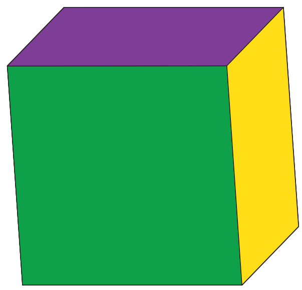
2



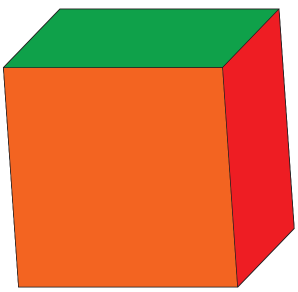
3



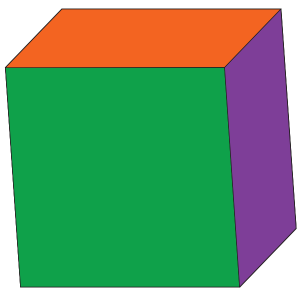
4



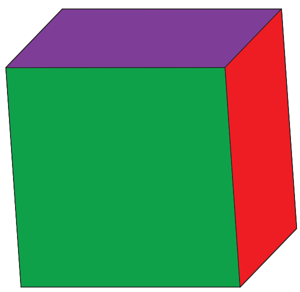
5



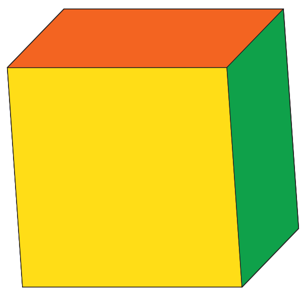
6



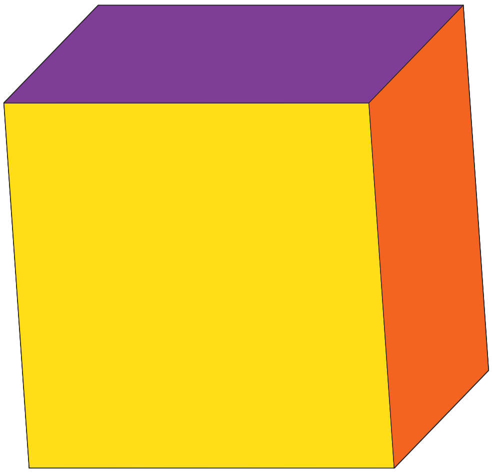
7



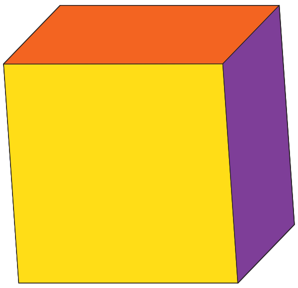
8



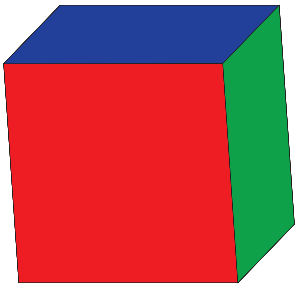
A 5



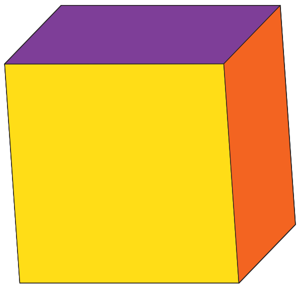
1



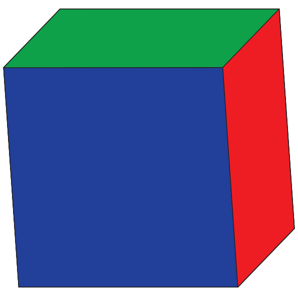
2



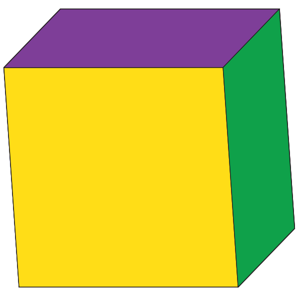
3



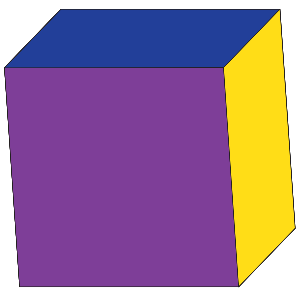
4



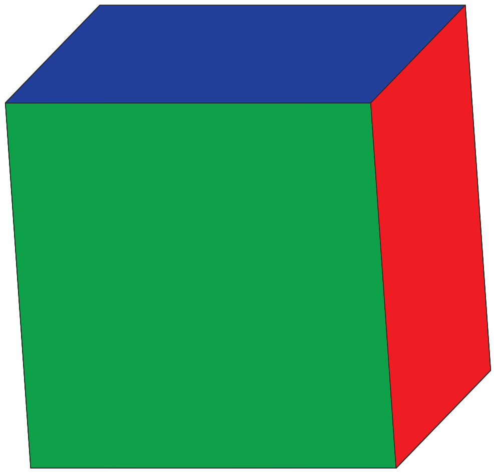
5



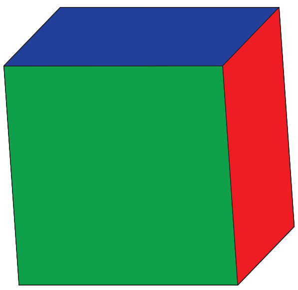
6



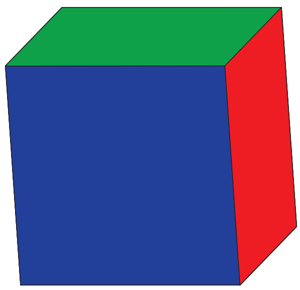
A 6



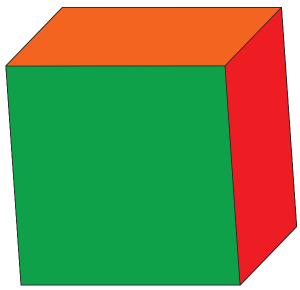
1



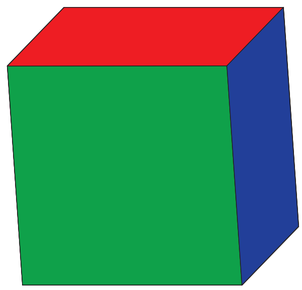
2



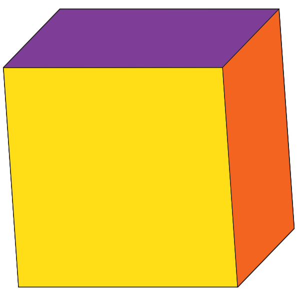
3



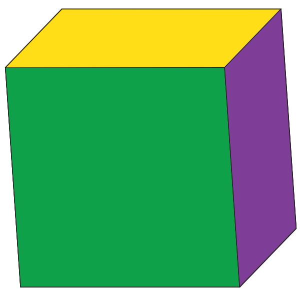
4



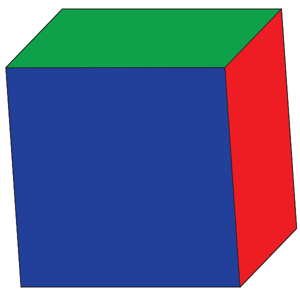
5



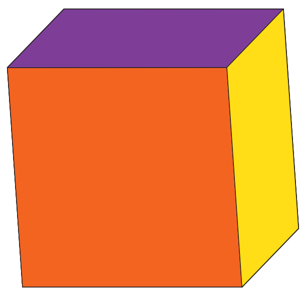
6



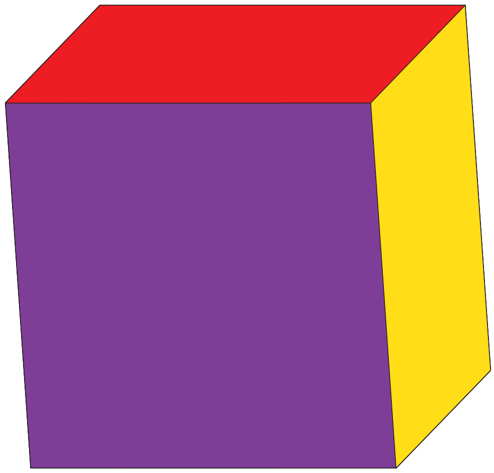
7



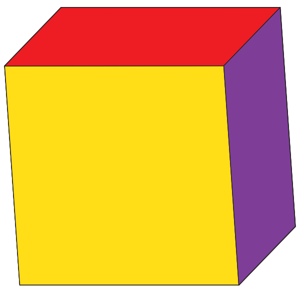
8



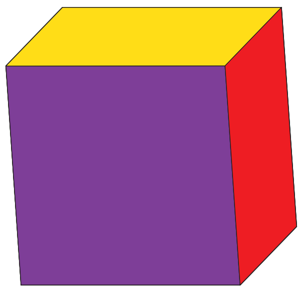
A 7



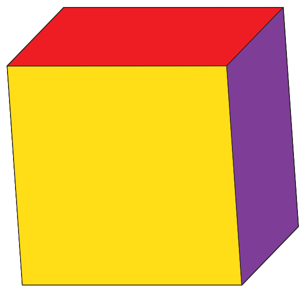
1



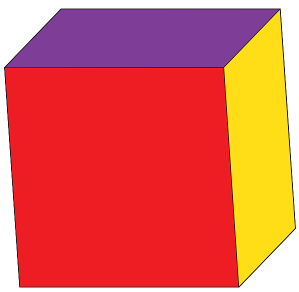
2



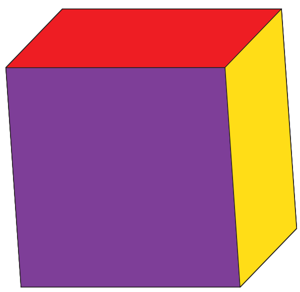
3



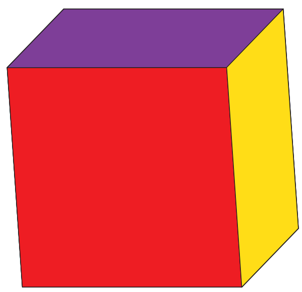
4



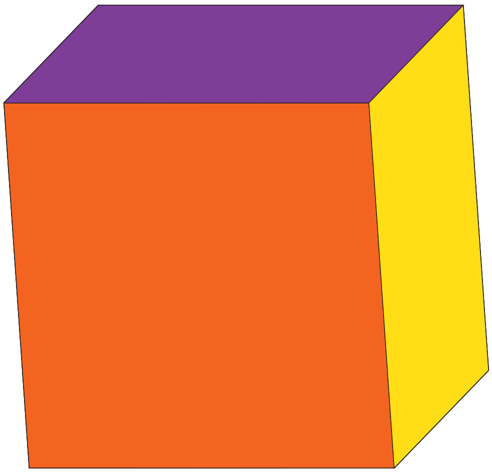
5



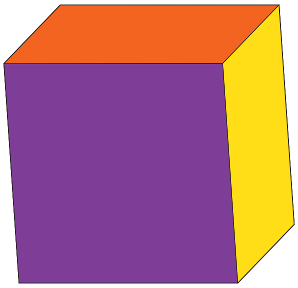
6



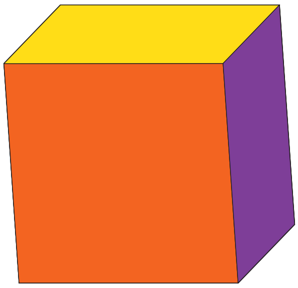
A 8



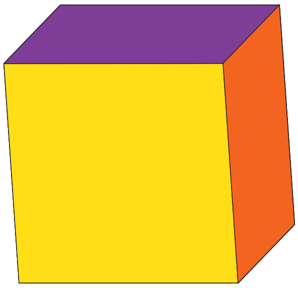
1



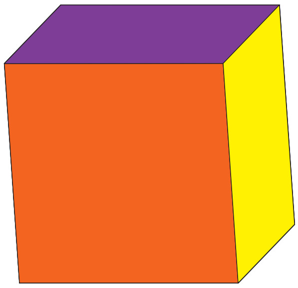
2



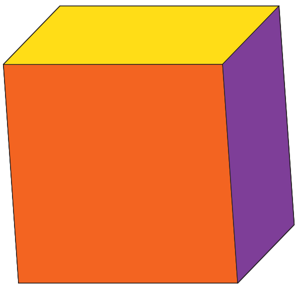
3



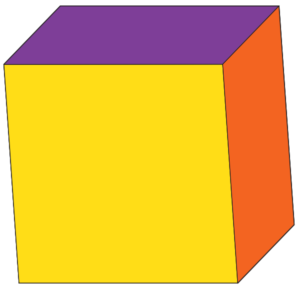
4



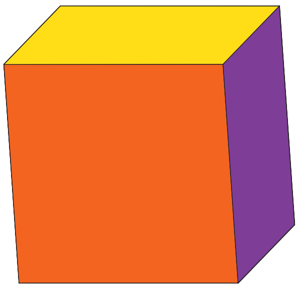
5



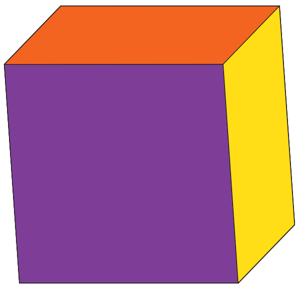
6



7



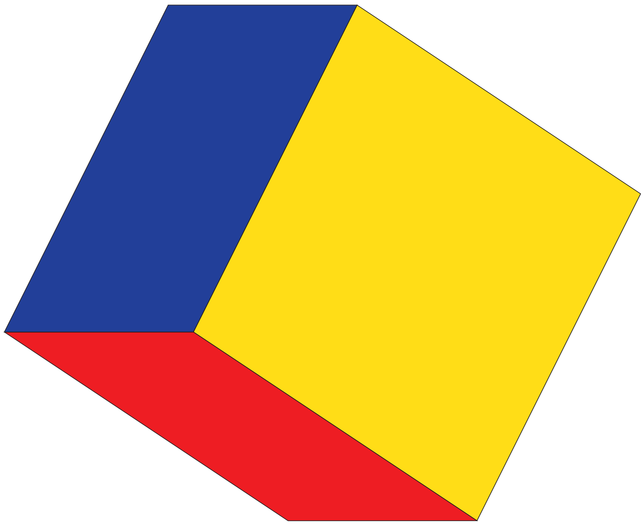
8



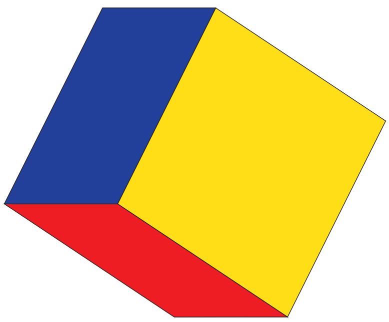
The Rotated Colour Cube Test

Set B

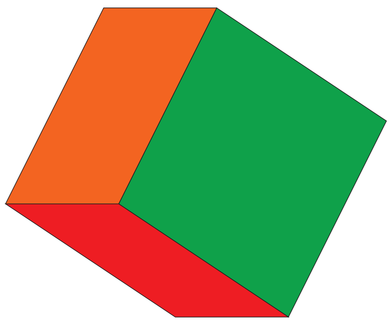
B 1



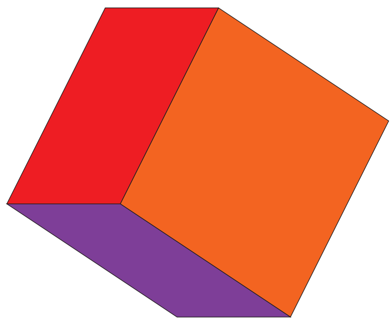
1



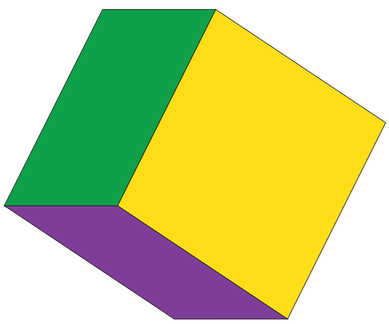
2



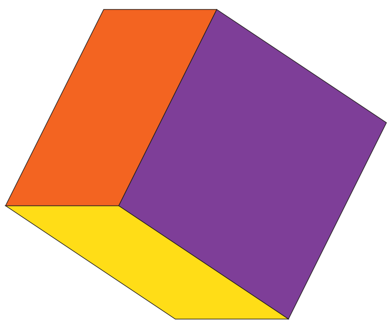
3



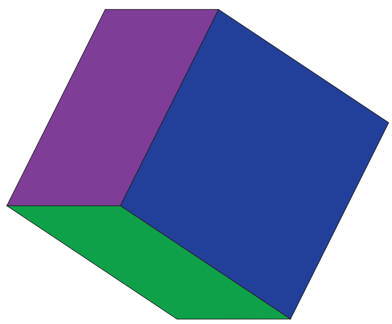
4



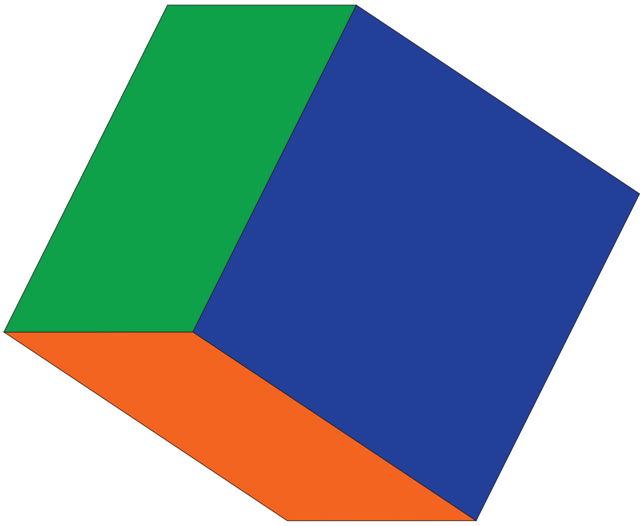
5



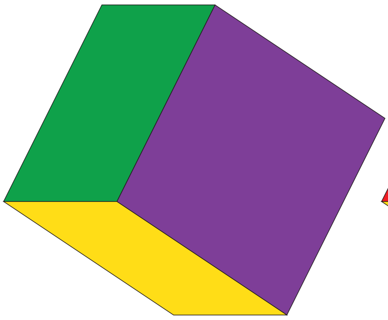
6



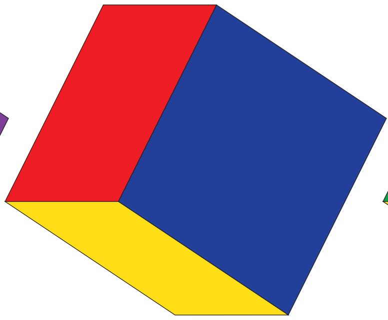
B 2



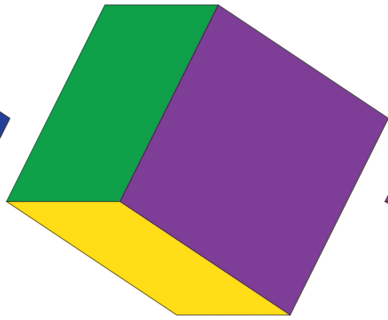
1



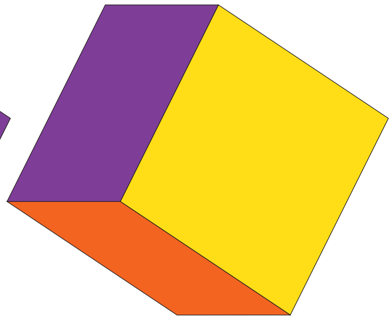
2



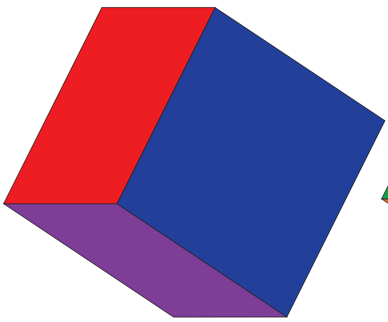
3



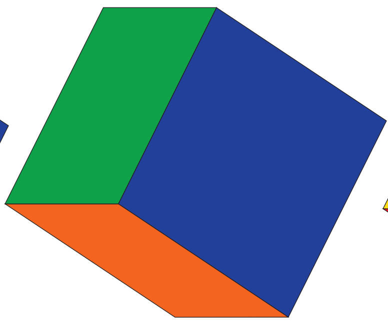
4



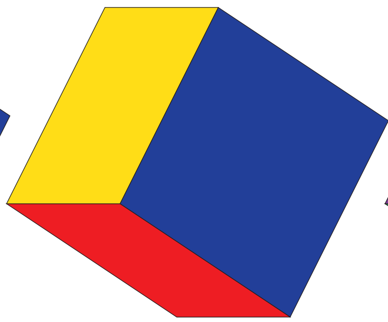
5



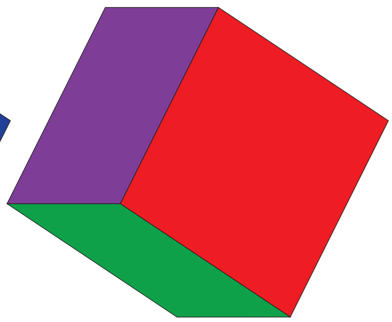
6



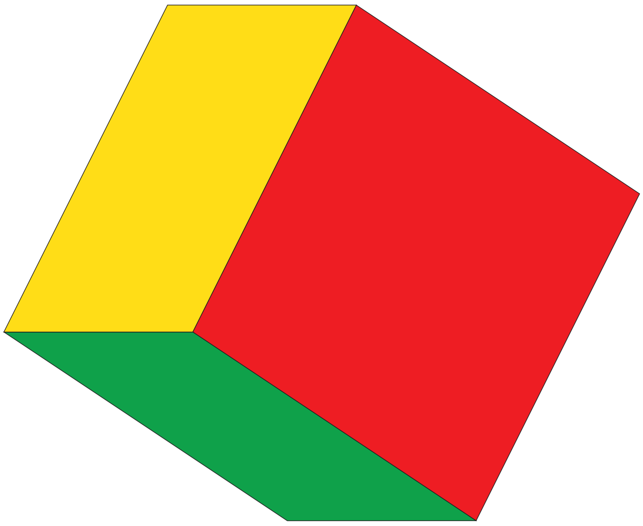
7



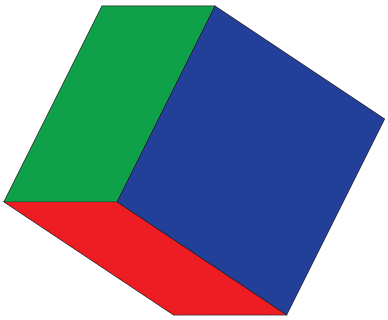
8



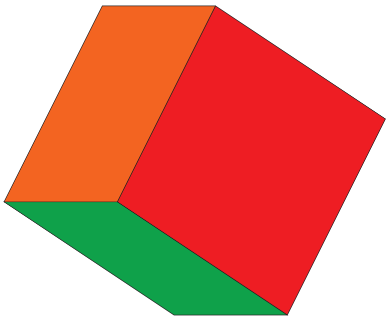
B 3



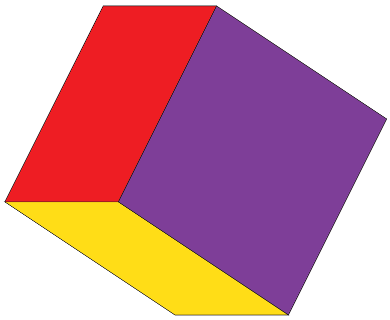
1



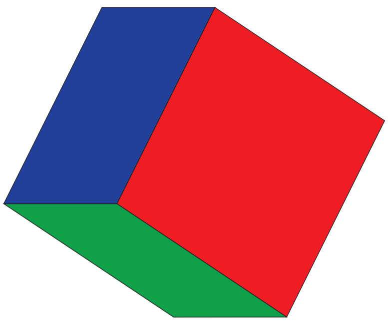
2



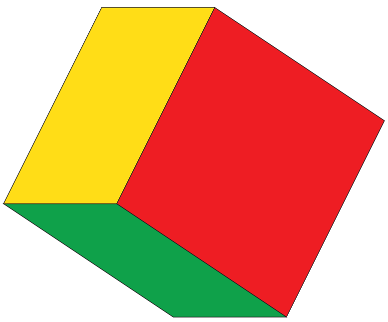
3



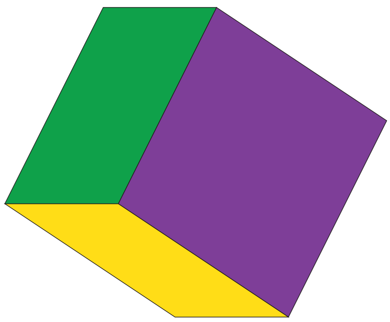
4



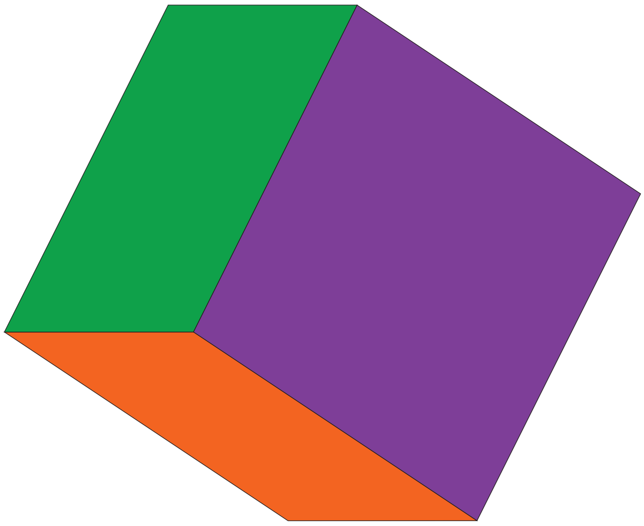
5



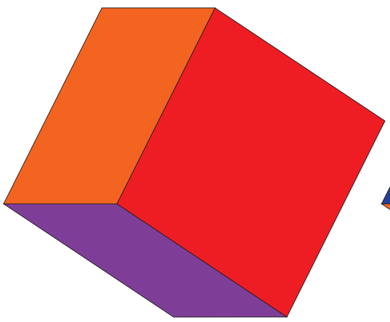
6



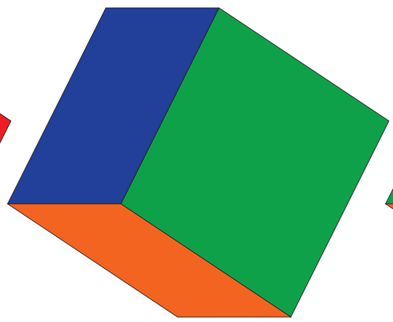
B 4



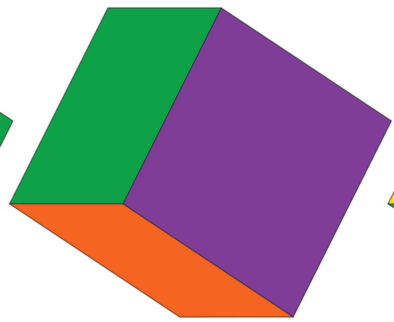
1



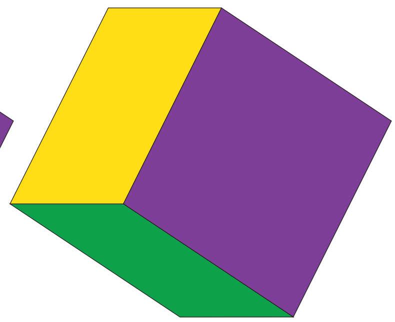
2



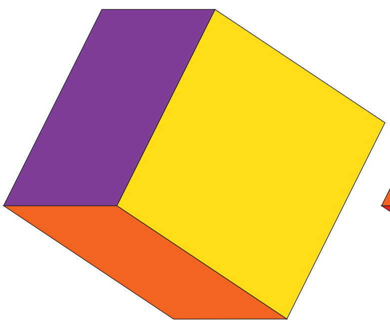
3



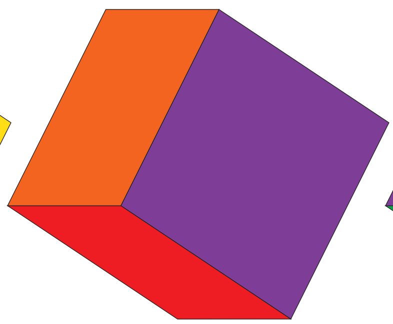
4



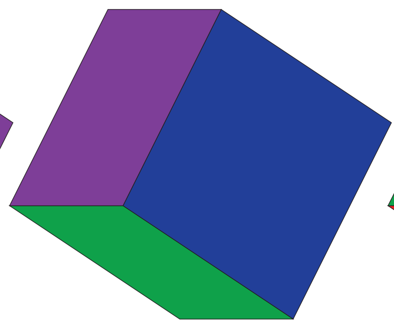
5



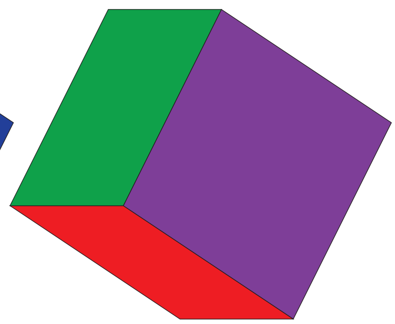
6



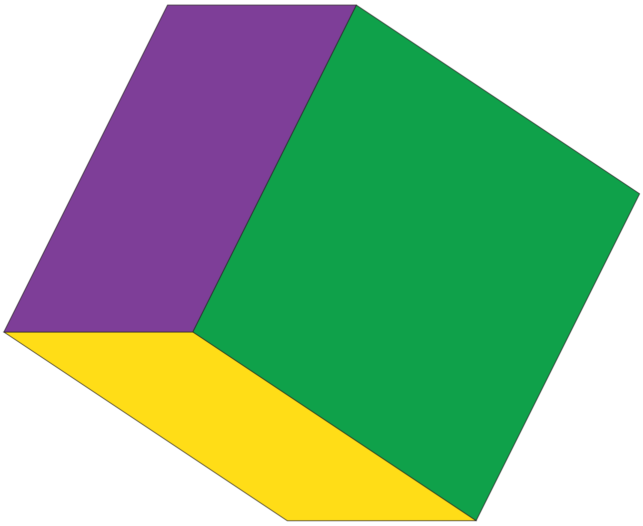
7



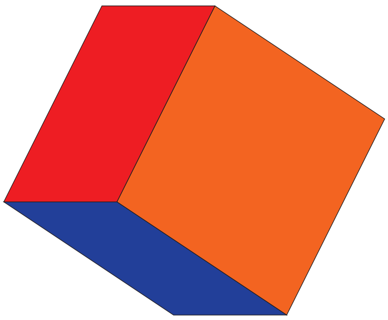
8



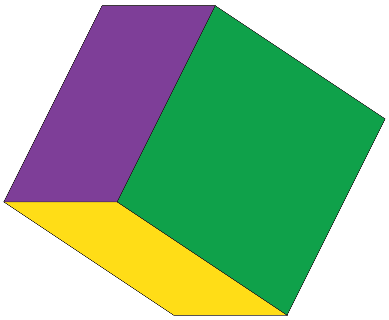
B 5



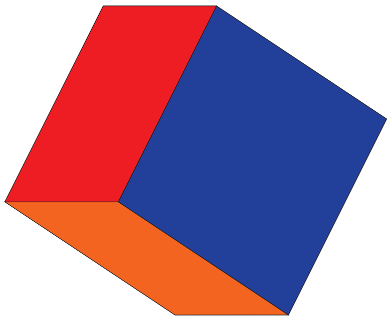
1



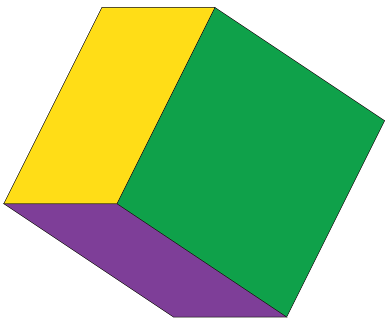
2



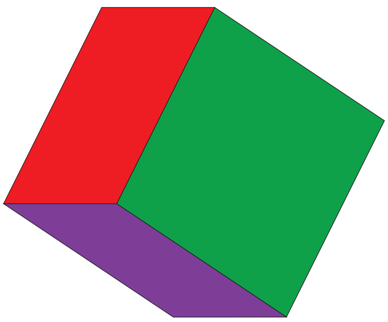
3



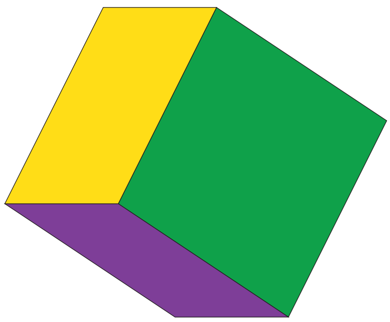
4



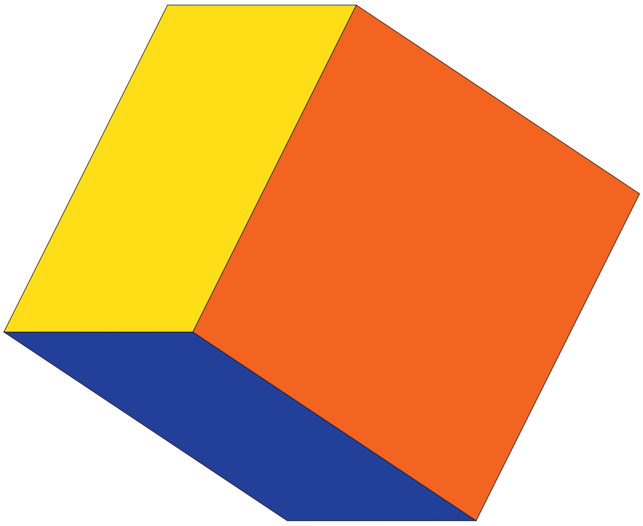
5



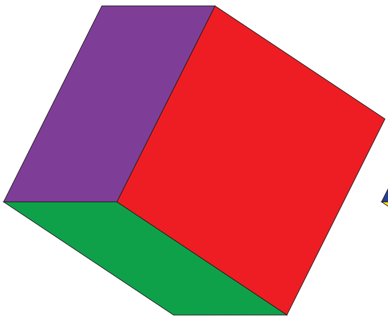
6



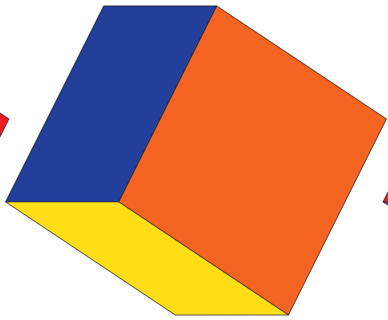
B 6



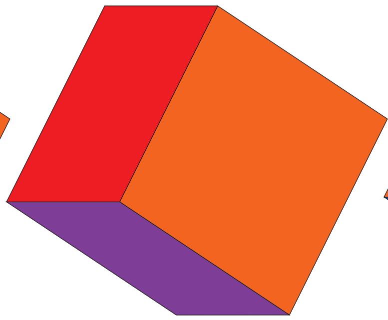
1



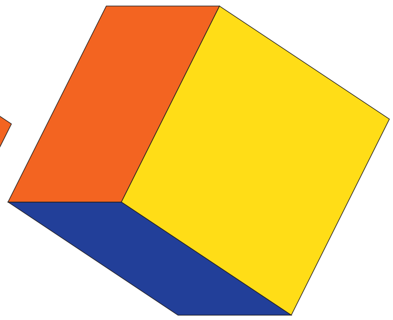
2



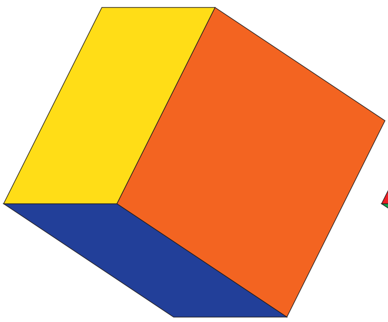
3



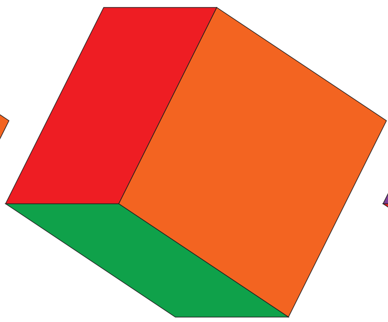
4



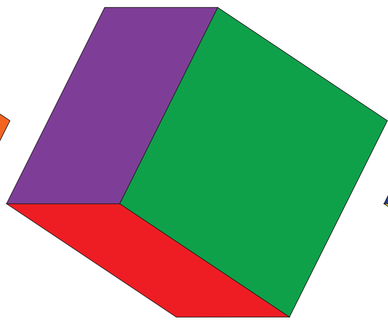
5



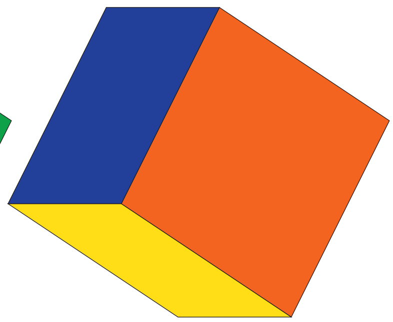
6



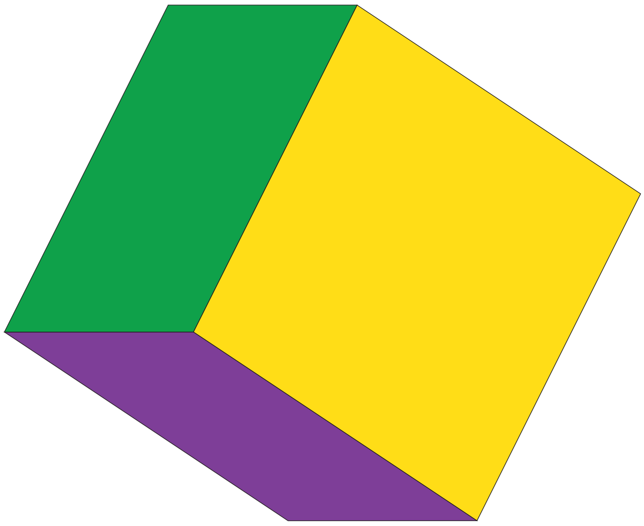
7



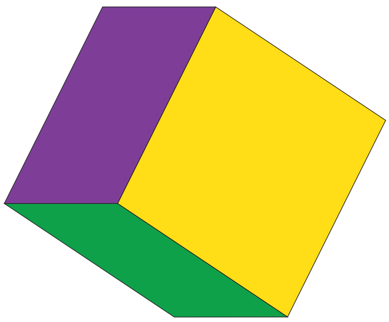
8



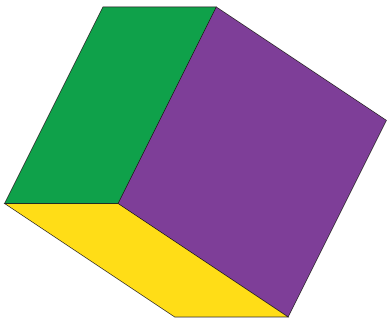
B 7



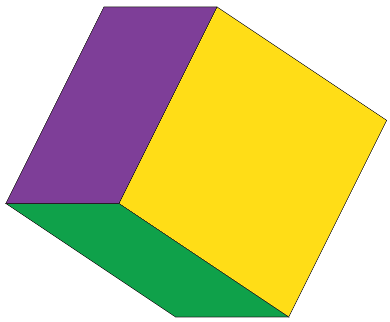
1



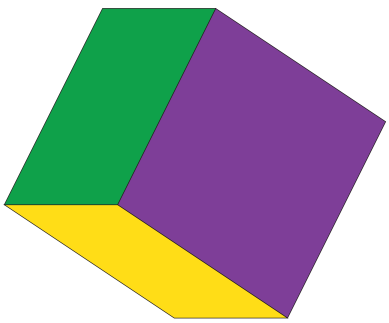
2



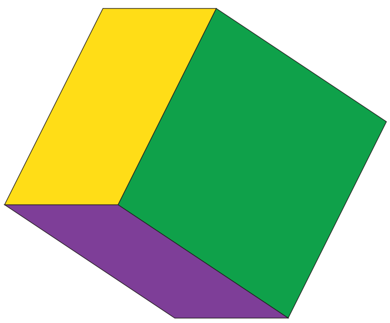
3



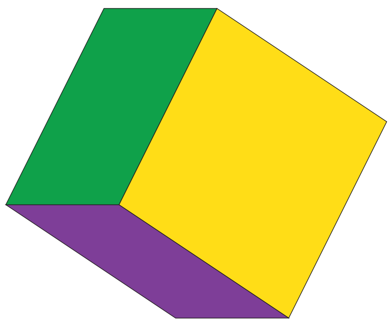
4



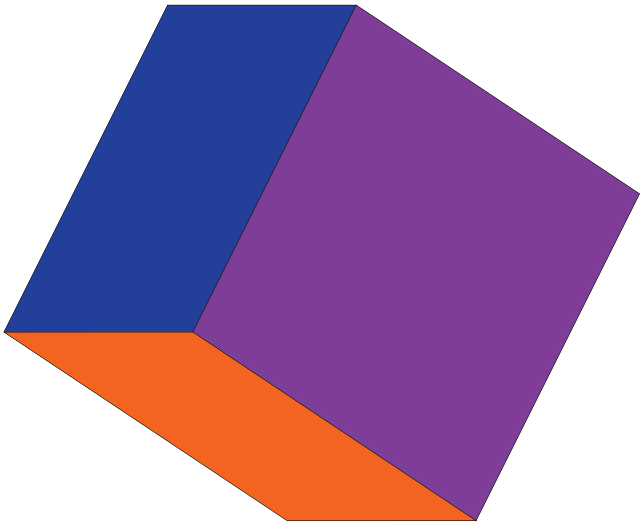
5



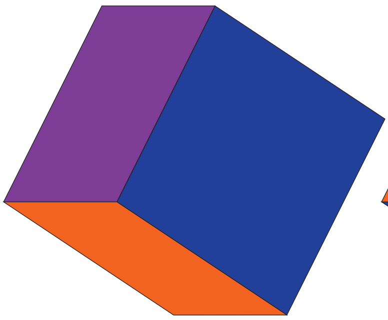
6



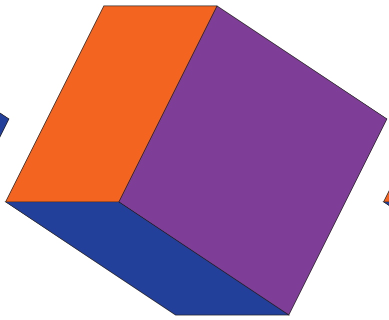
B 8



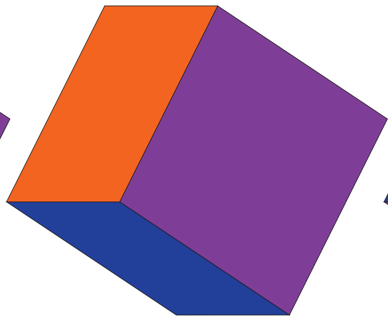
1



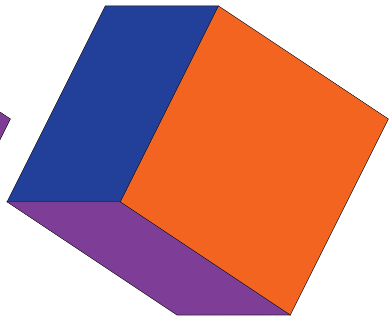
2



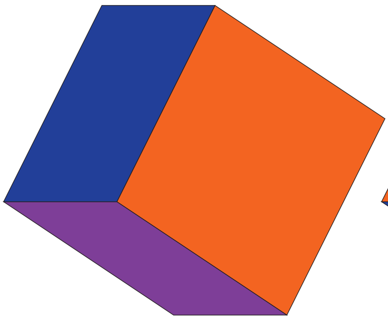
3



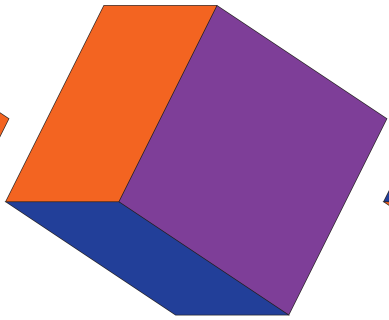
4



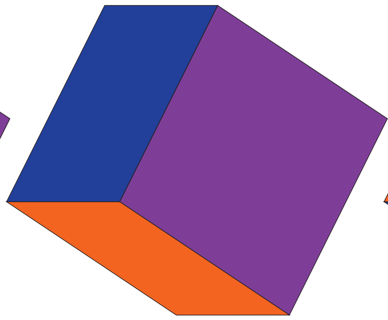
5



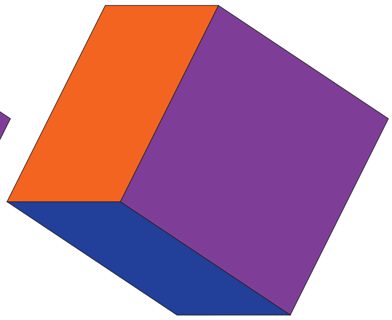
6



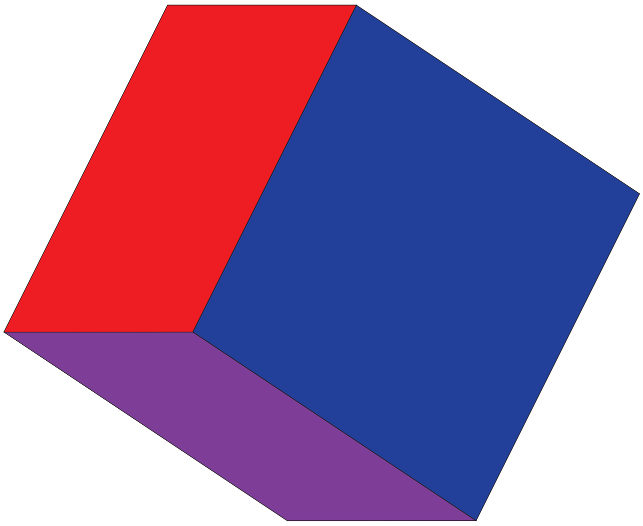
7



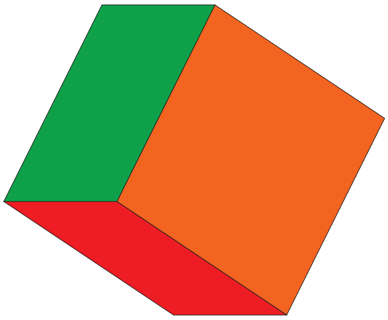
8



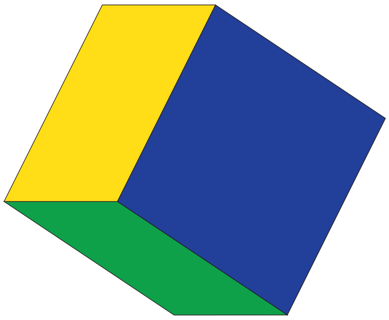
C 1



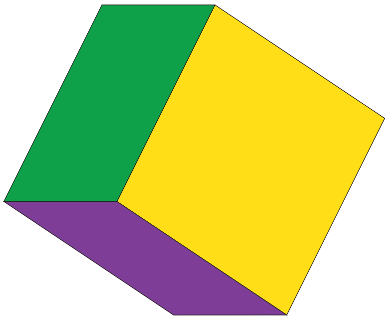
1



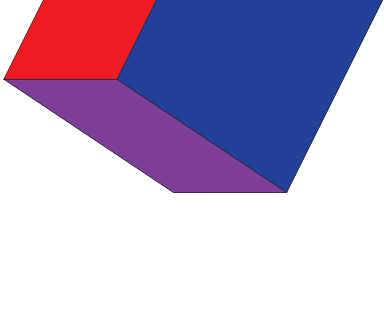
2



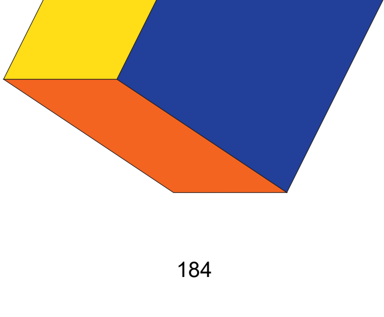
3



4



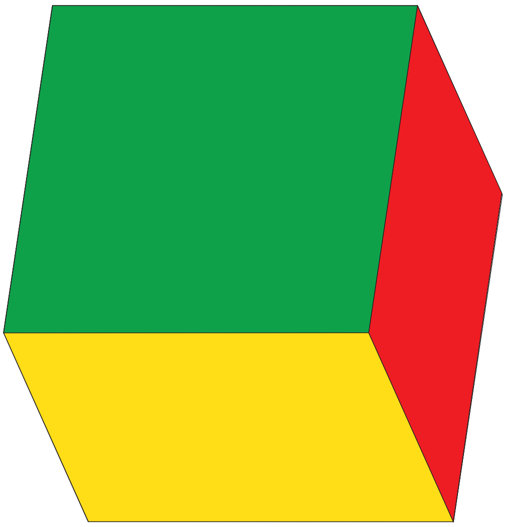
5



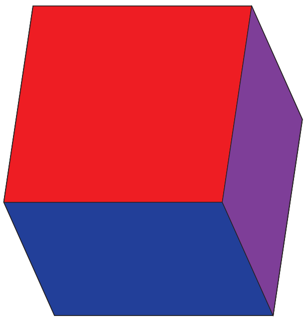
6

The Rotated Colour Cube Test
Set C

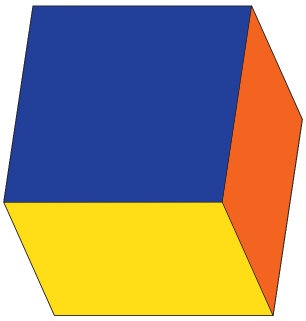
C 2



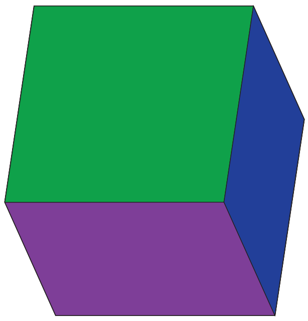
1



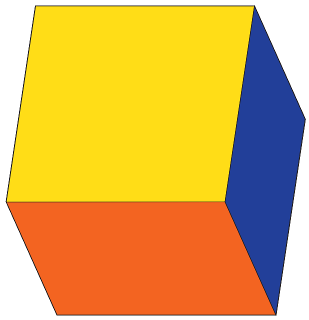
2



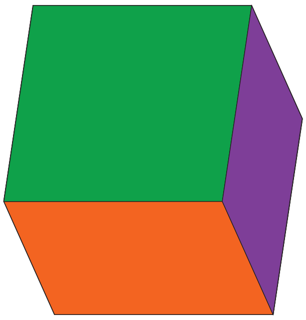
3



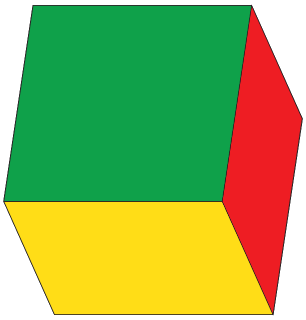
4



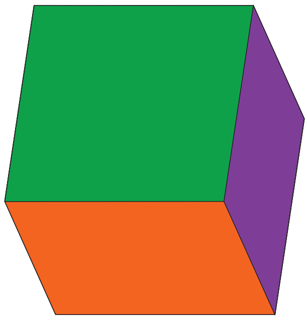
5



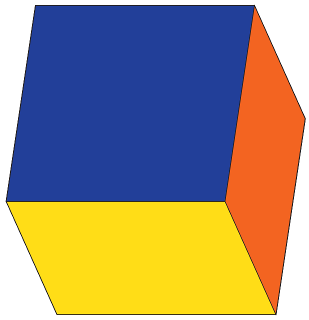
6



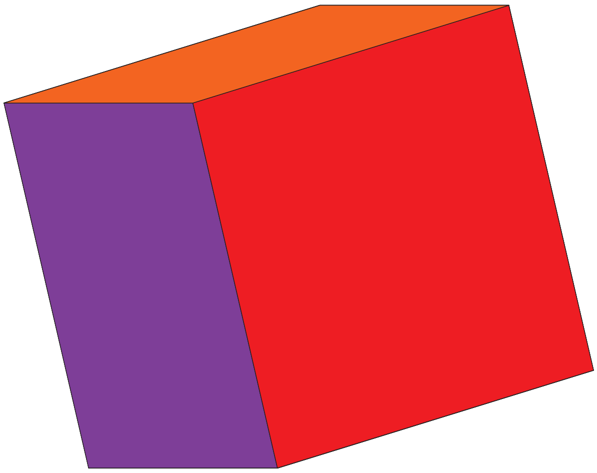
7



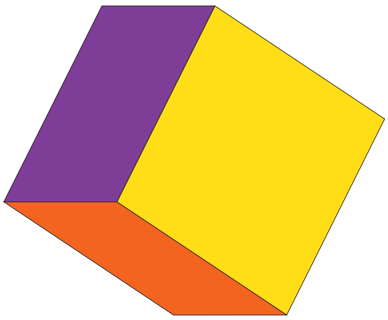
8



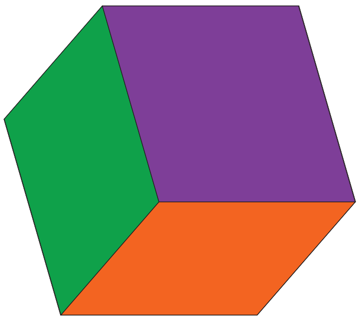
C 3



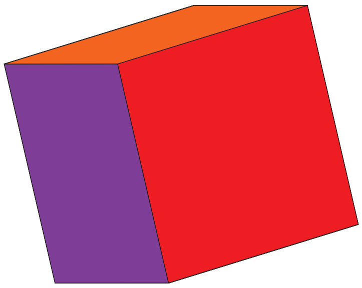
1



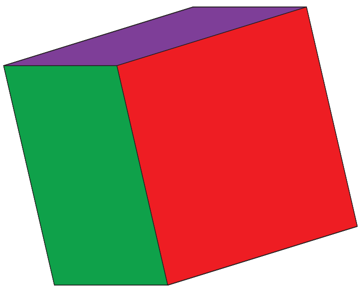
2



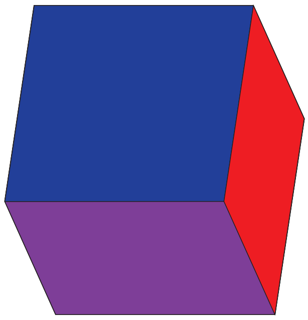
3



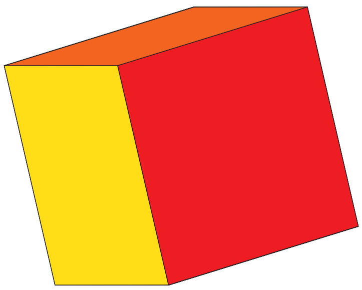
4



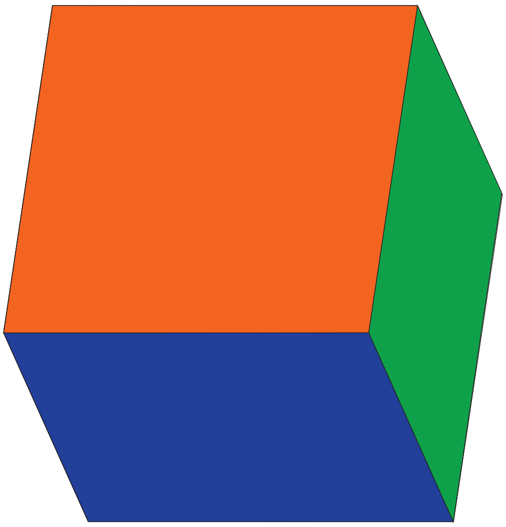
5



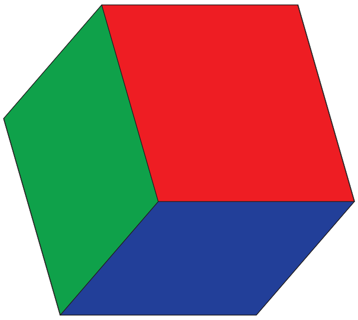
6



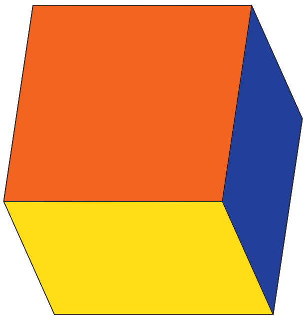
C 4



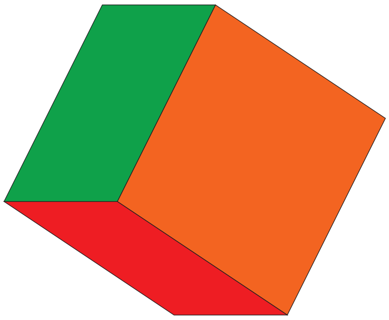
1



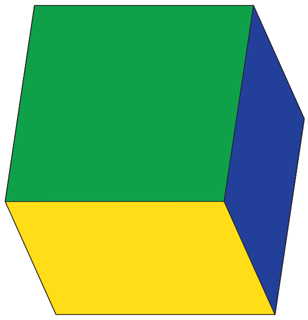
2



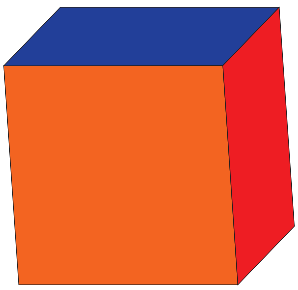
3



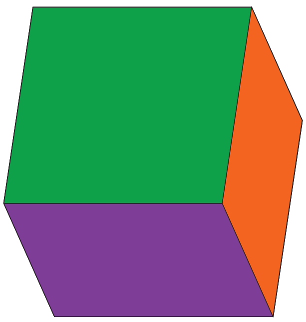
4



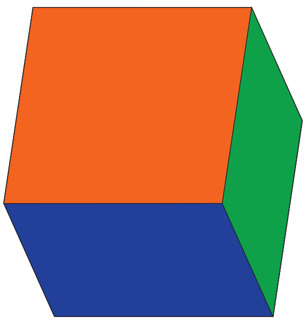
5



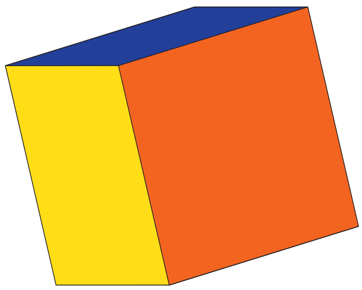
6



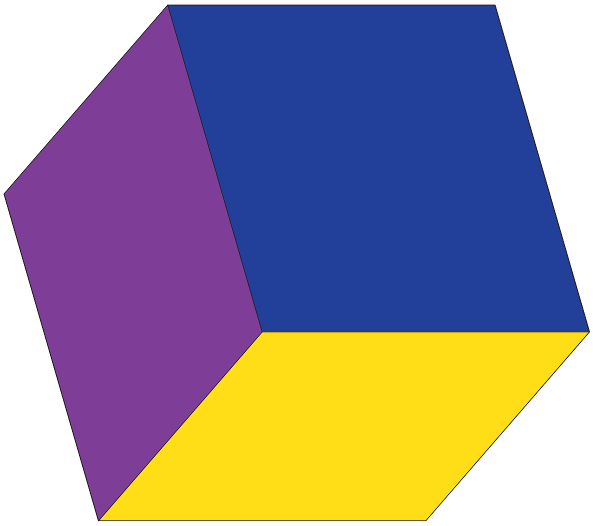
7



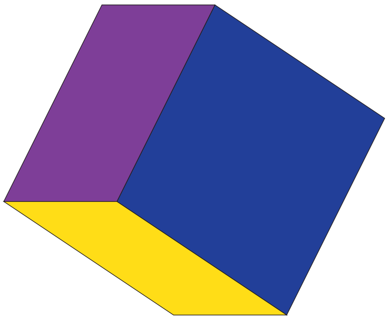
8



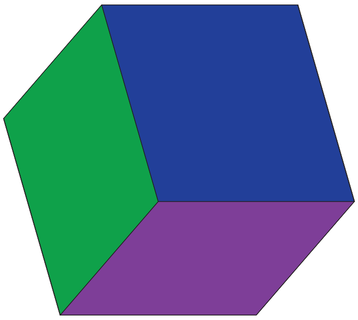
C 5



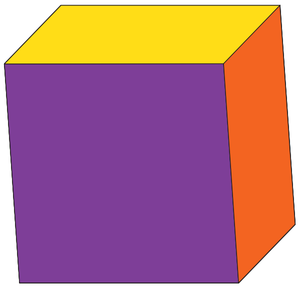
1



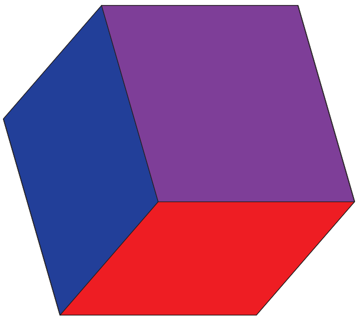
2



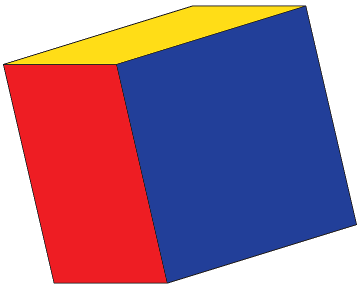
3



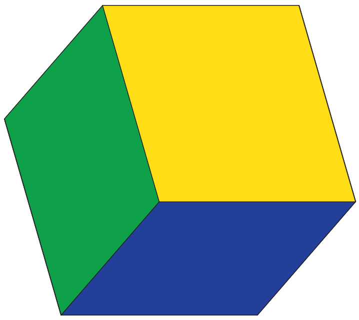
4



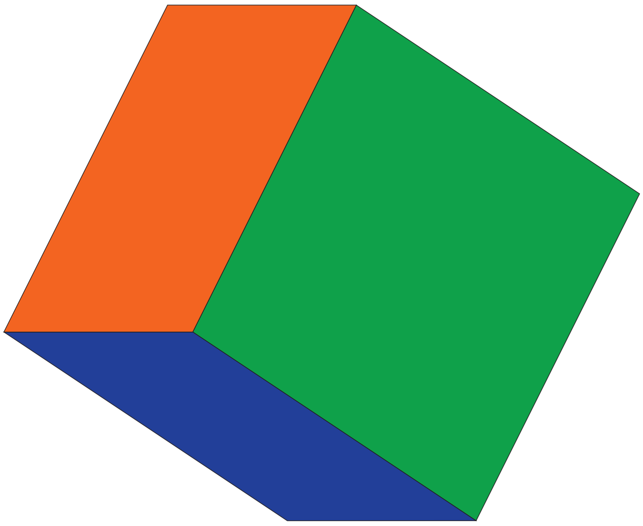
5



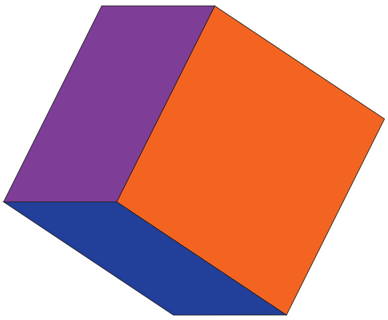
6



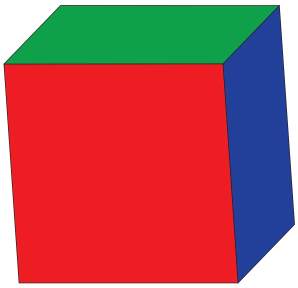
C 6



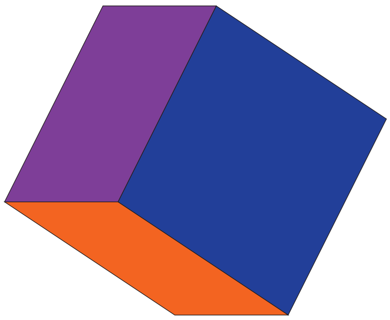
1



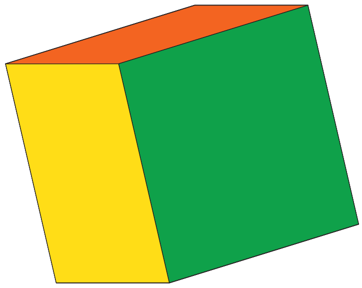
2



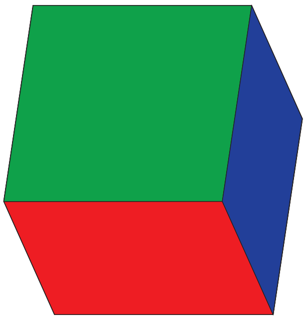
3



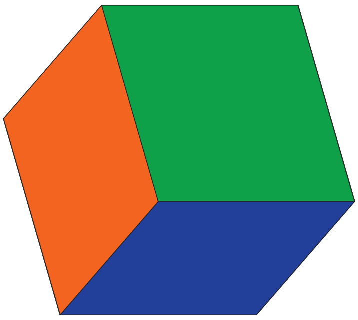
4



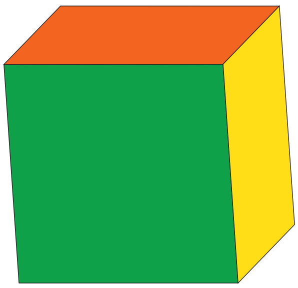
5



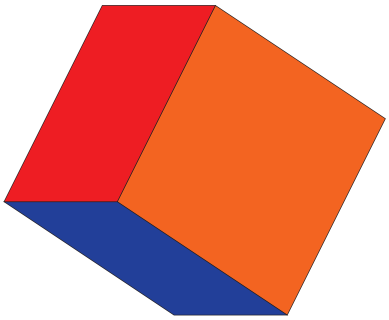
6



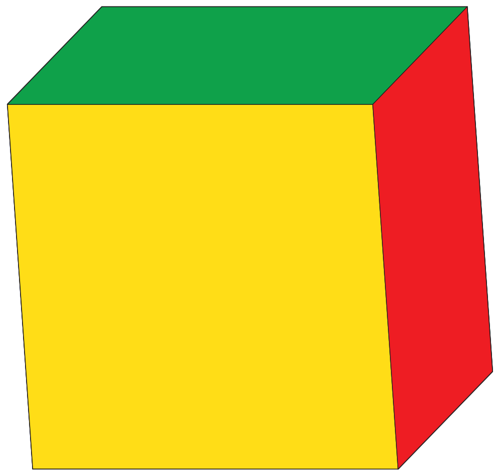
7



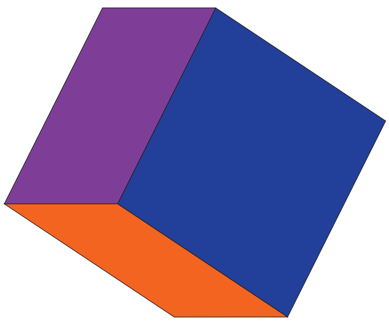
8



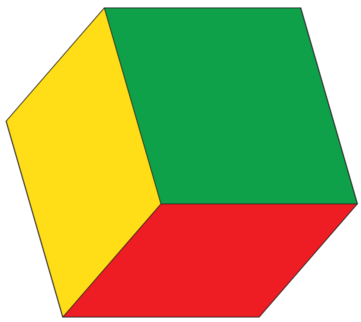
C 7



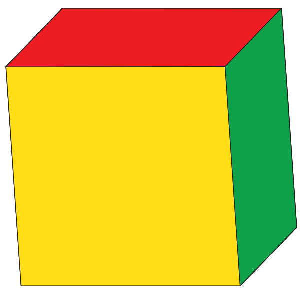
1



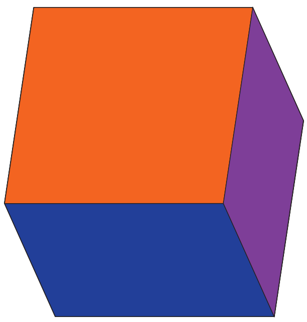
2



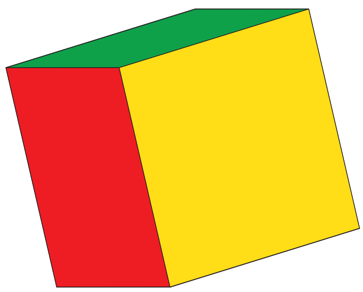
3



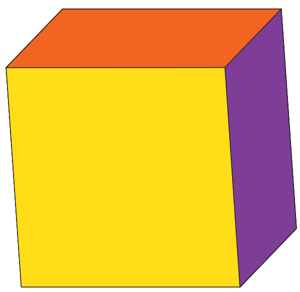
4



5

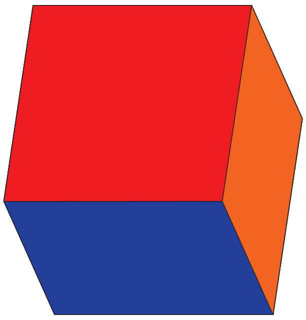


6

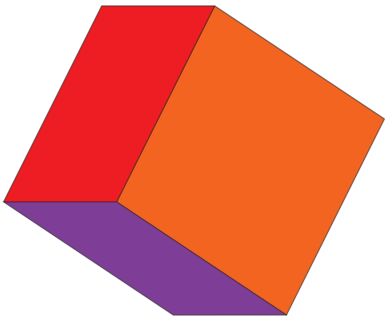


C 8

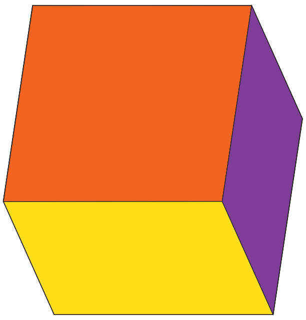
1



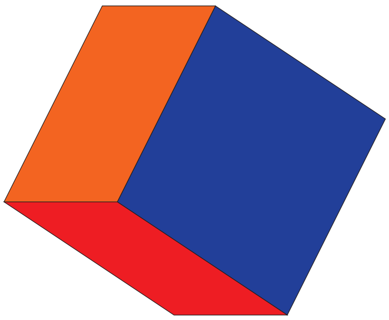
2



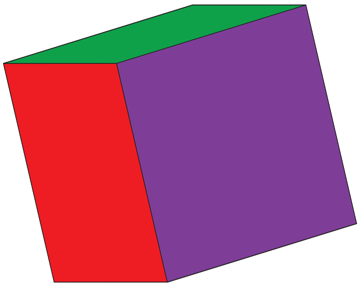
3



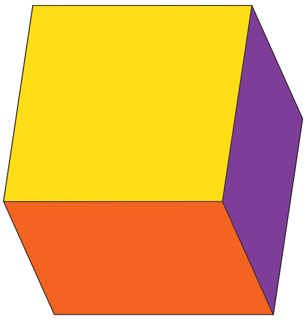
4



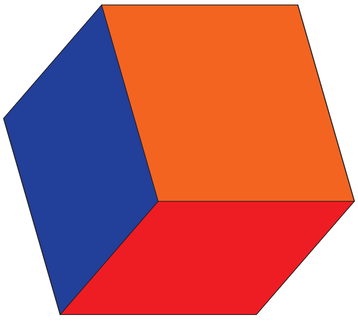
5



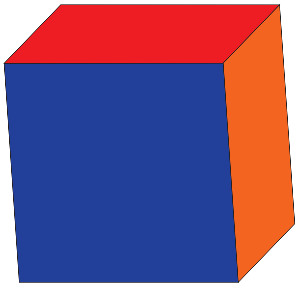
6



7

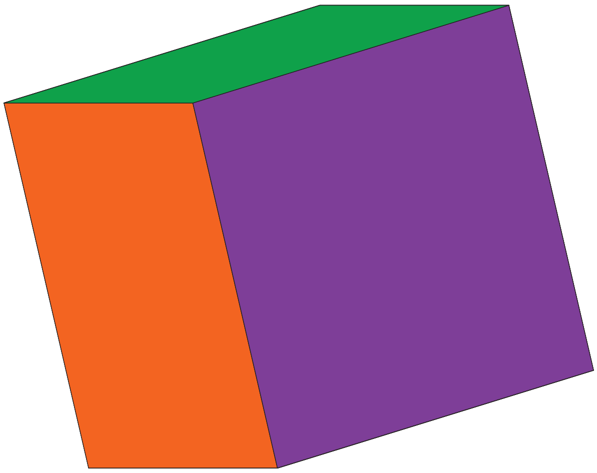


8

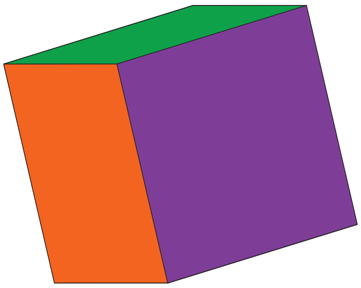


The Rotated Colour Cube Test
Set D

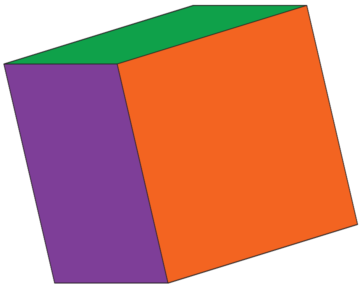
D 1



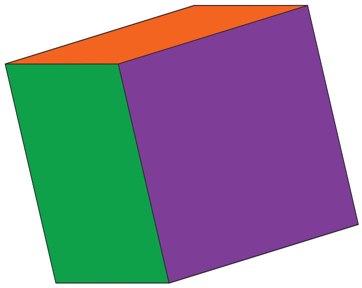
1



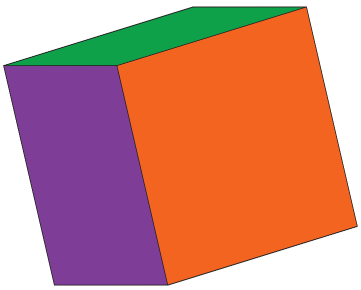
2



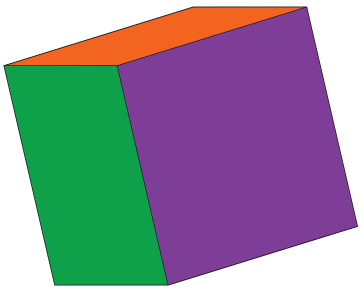
3



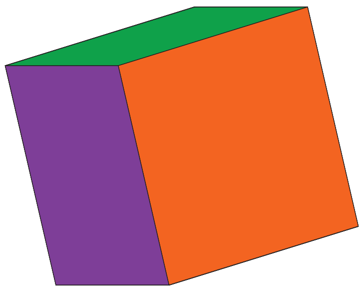
4



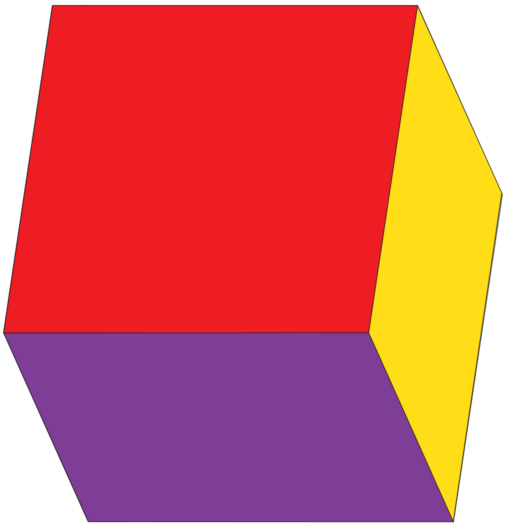
5



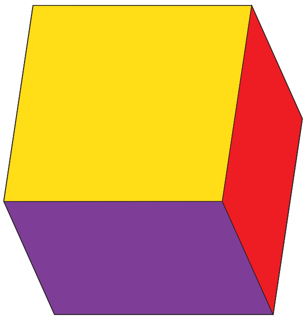
6



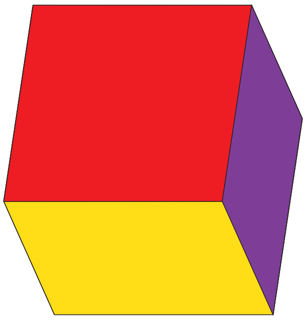
D 2



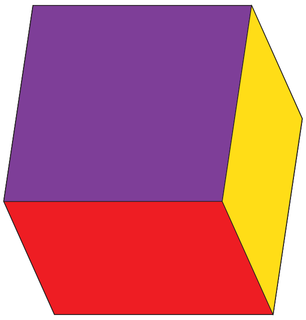
1



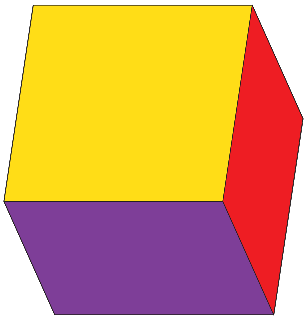
2



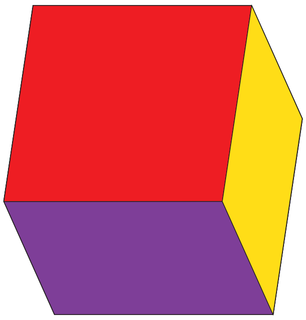
3



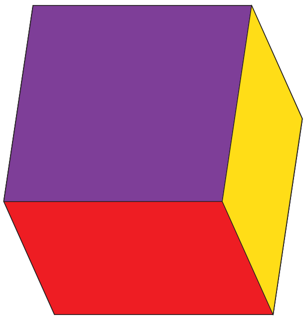
4



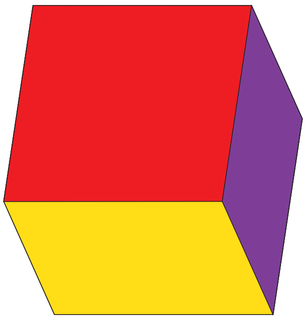
5



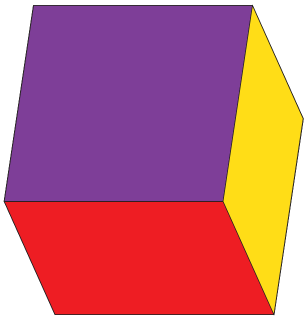
6



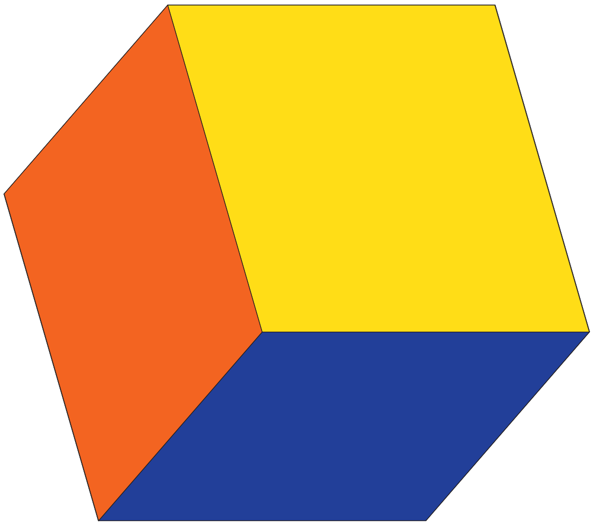
7



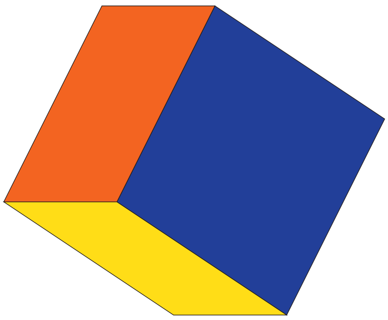
8



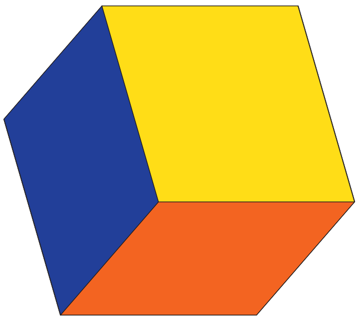
D 3



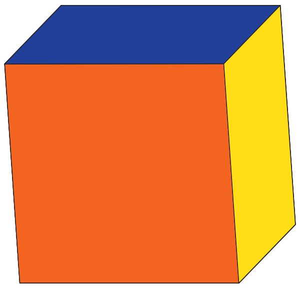
1



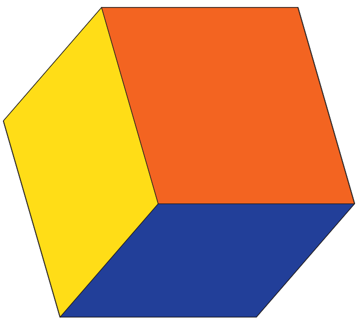
2



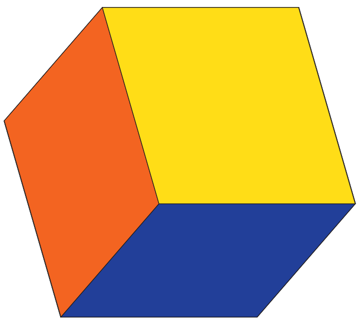
3



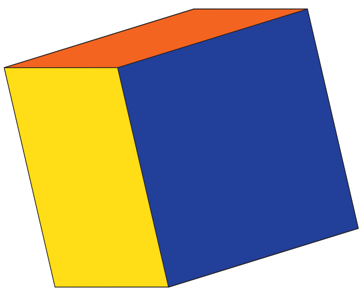
4



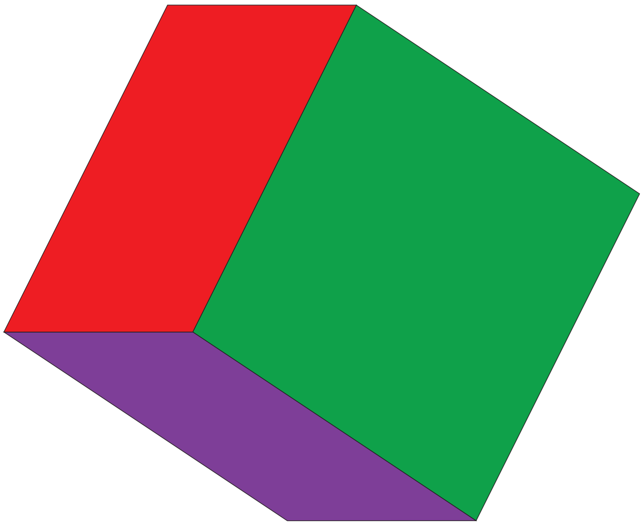
5



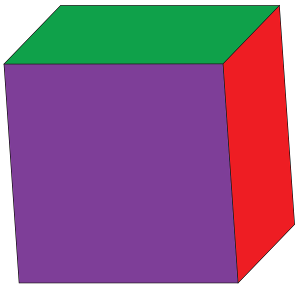
6



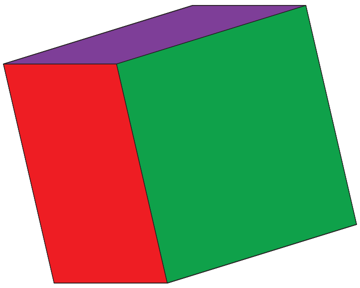
D 4



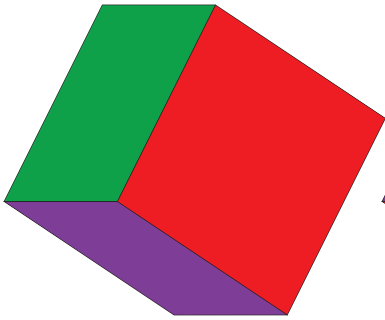
1



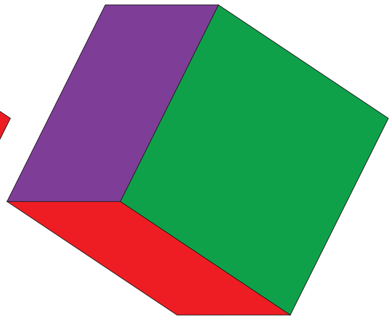
2



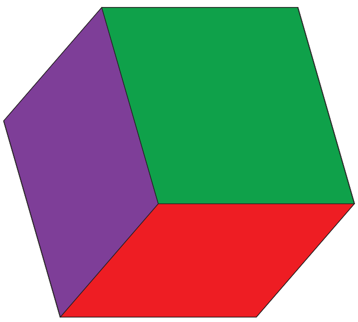
3



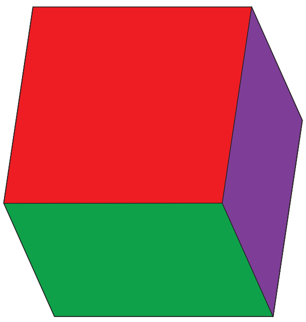
4



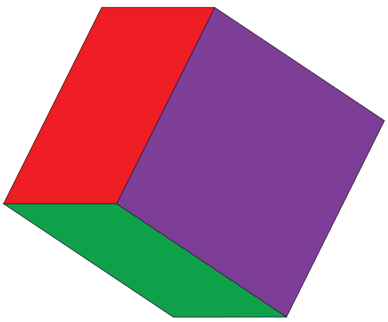
5



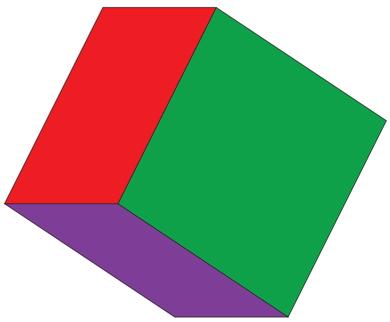
6



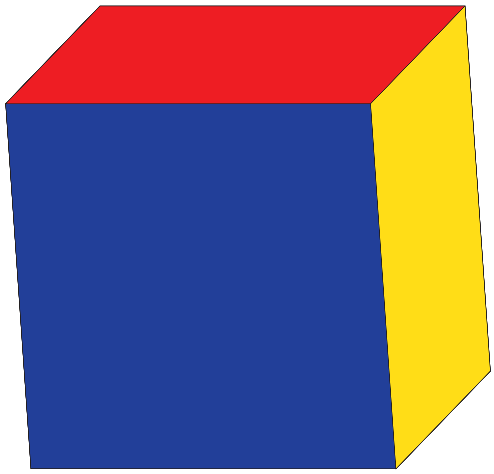
7



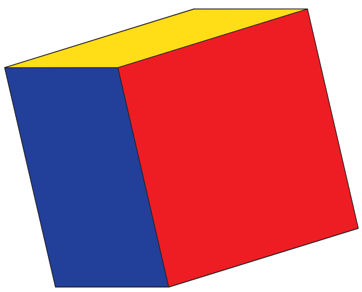
8



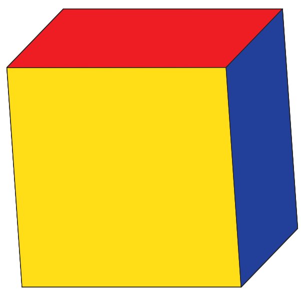
D 5



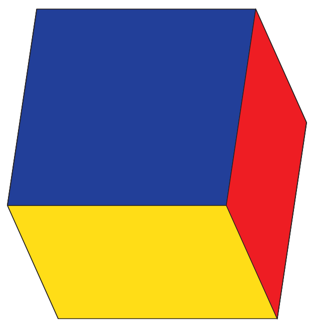
1



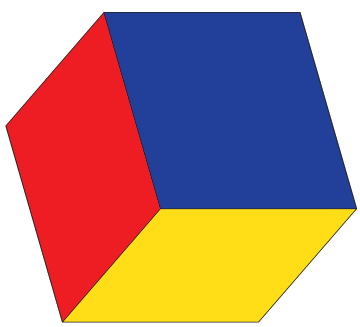
2



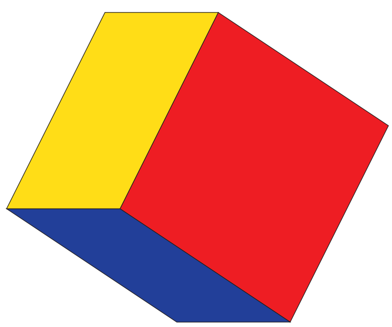
3



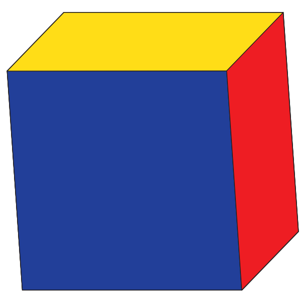
4



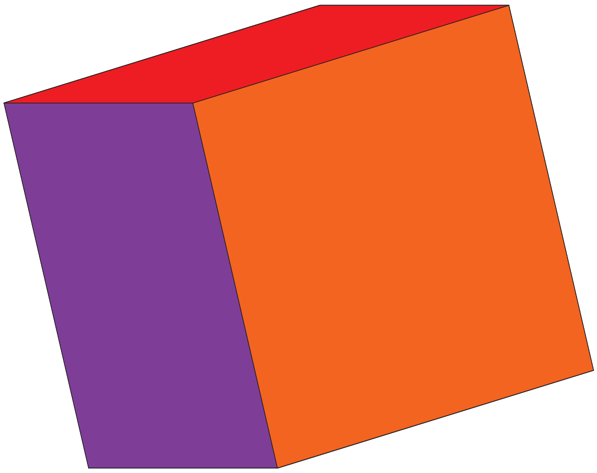
5



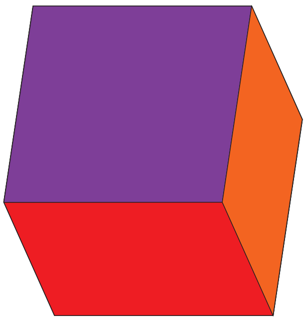
6



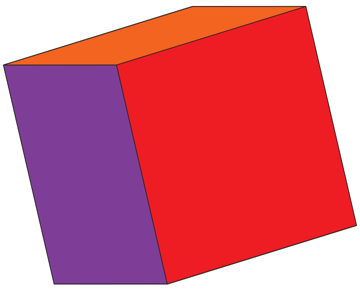
D 6



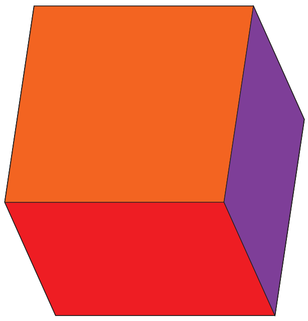
1



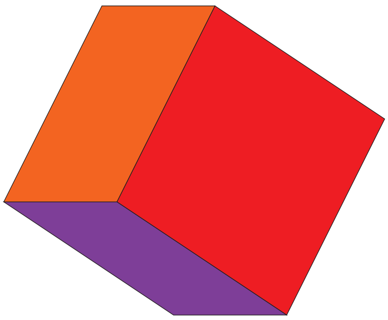
2



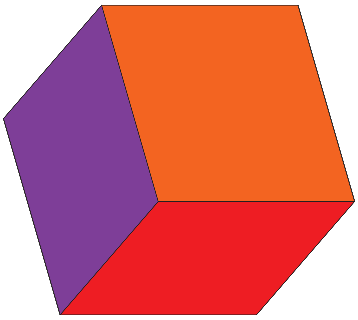
3



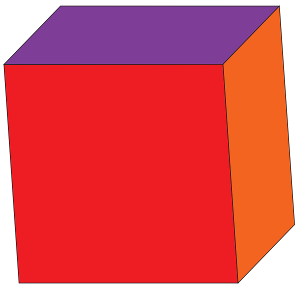
4



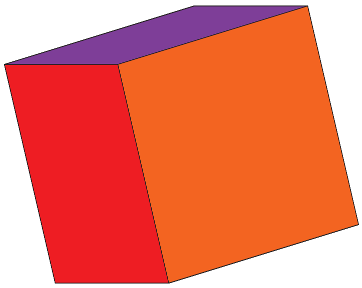
5



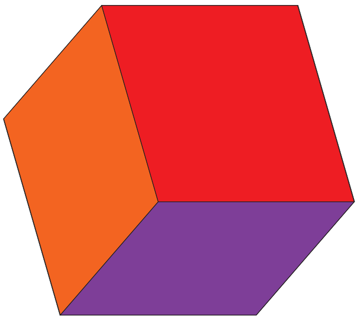
6



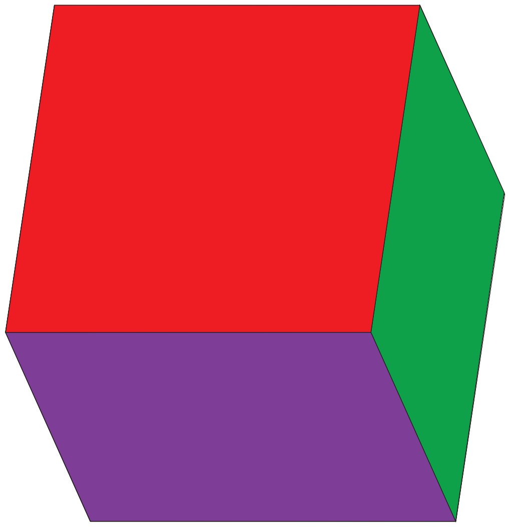
7



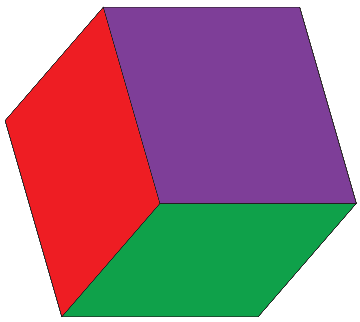
8



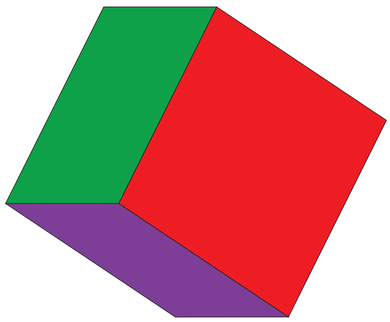
D 7



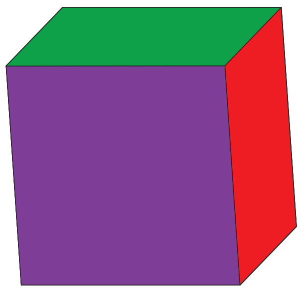
1



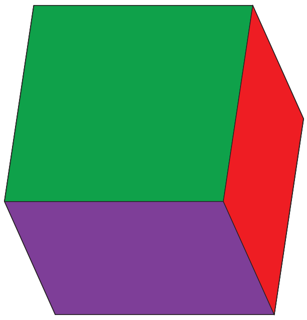
2



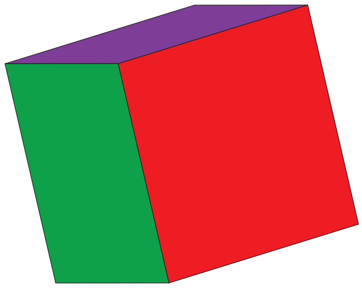
3



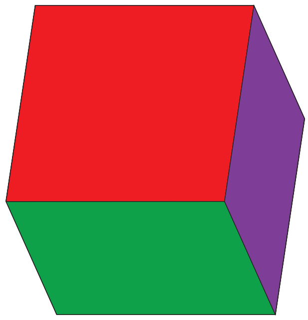
4



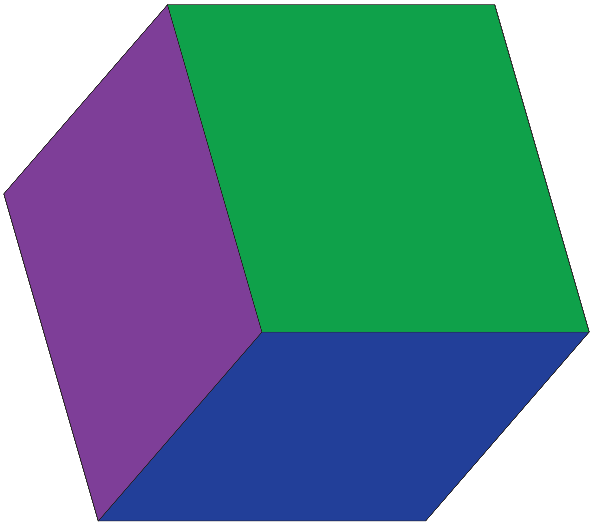
5



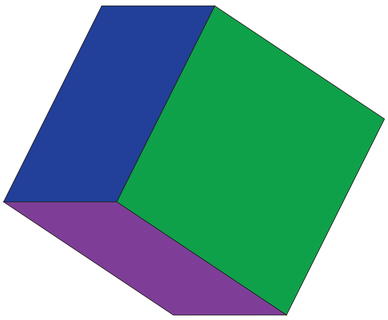
6



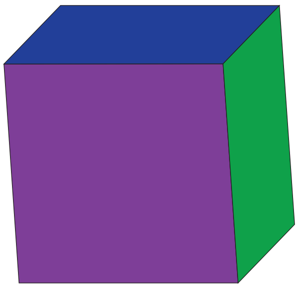
D 8



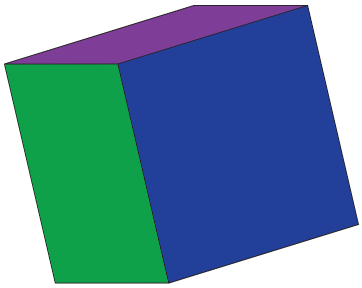
1



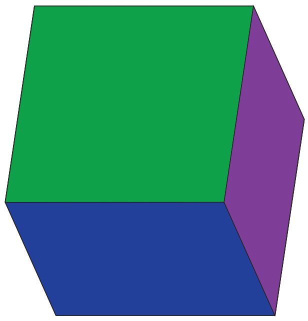
2



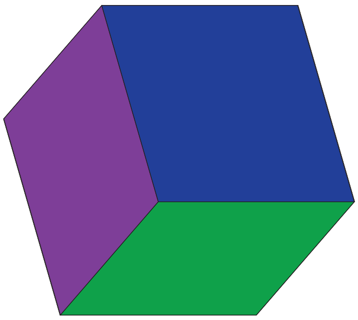
3



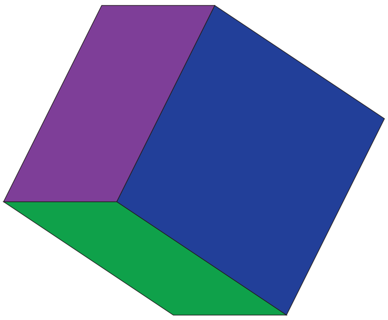
4



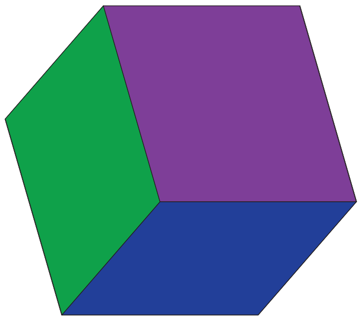
5



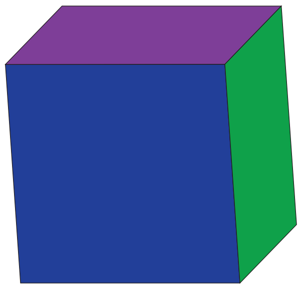
6

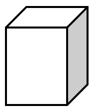


7



8





RECORD FORM FOR THE ROTATED COLOUR CUBE TEST V.1

Sets A, B, C, D

Name: **Date & time:**

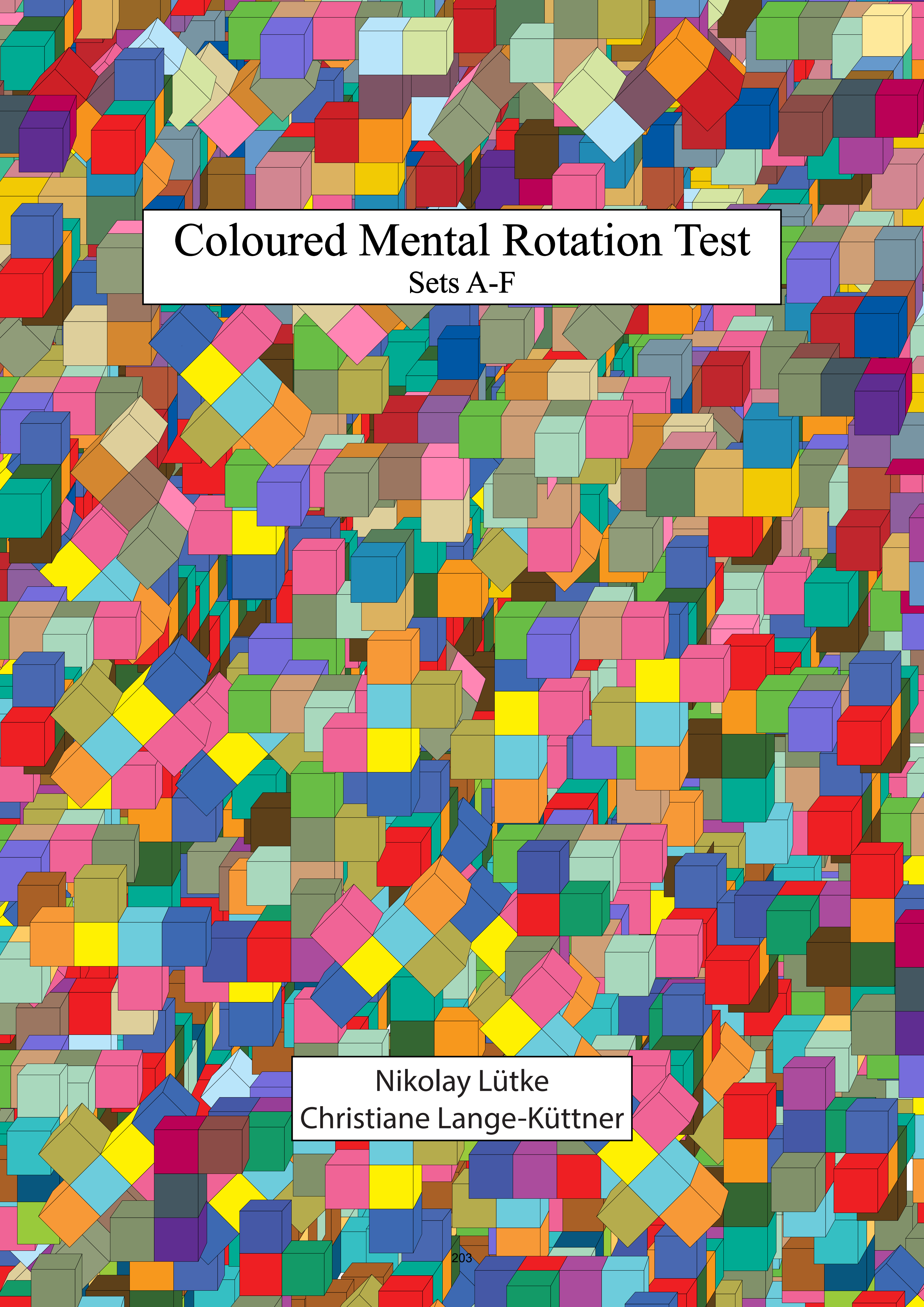
Age: **Place of Testing:**

A1	1	2	3	4	5	**6**		
A2	1	2	3	4	5	6	7	**8**
A3	1	2	**3**	4	5	6		
A4	1	2	3	4	5	**6**	7	8
A5	1	2	**3**	4	5	6		
A6	**1**	2	3	4	5	6	7	8
A7	1	2	3	4	**5**	6		
A8	1	2	3	**4**	5	6	7	8

B1	**1**	2	3	4	5	6		
B2	1	2	3	4	5	**6**	7	8
B3	1	2	3	4	**5**	6		
B4	1	2	**3**	4	5	6	7	8
B5	1	**2**	3	4	5	6		
B6	1	2	3	4	**5**	6	7	8
B7	1	2	3	4	5	**6**		
B8	1	2	3	4	5	6	**7**	8

C1	1	2	3	**4**	5	6		
C2	1	2	3	4	5	**6**	7	8
C3	1	2	**3**	4	5	6		
C4	1	2	3	4	5	6	**7**	8
C5	**1**	2	3	4	5	6		
C6	1	2	3	4	5	**6**	7	8
C7	1	**2**	3	4	5	6		
C8	1	2	3	4	5	6	**7**	8

D1	**1**	2	3	4	5	6		
D2	1	2	3	4	**5**	6	7	8
D3	1	2	3	4	**5**	6		
D4	1	2	3	4	5	6	7	**8**
D5	1	2	**3**	4	5	6		
D6	1	2	**3**	4	5	6	7	8
D7	1	2	3	4	**5**	6		
D8	1	2	3	4	5	6	7	**8**



Coloured Mental Rotation Test

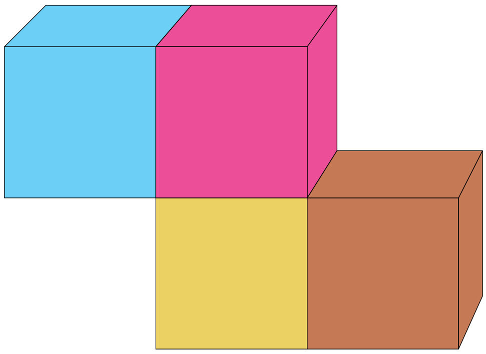
Sets A-F

Nikolay Lütke
Christiane Lange-Küttner

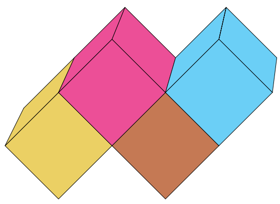
The Coloured Mental Rotation Test

Practise Section

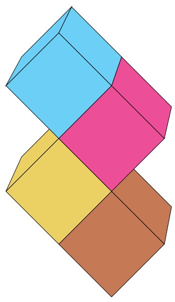
P 1



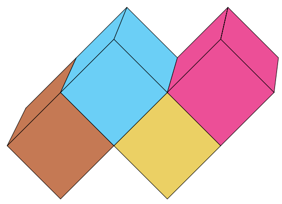
1



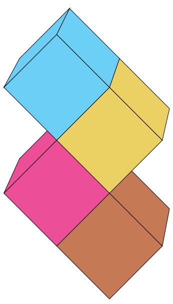
2



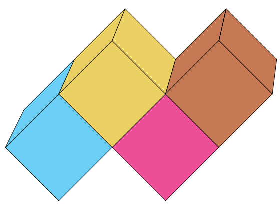
3



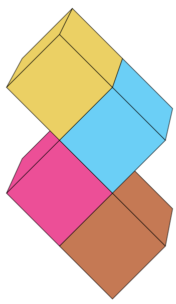
4



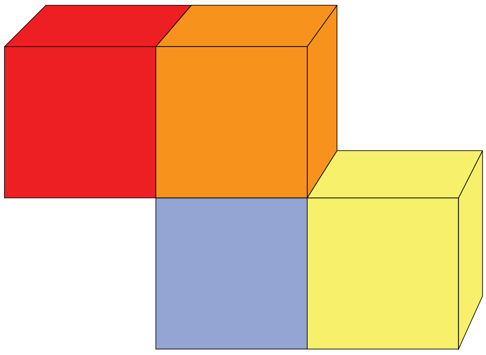
5



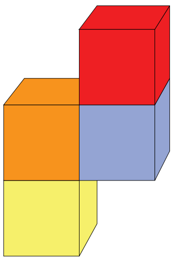
6



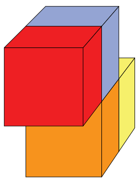
P 2



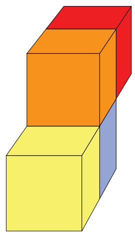
1



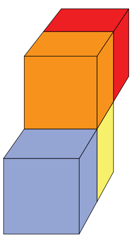
2



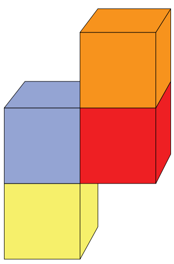
3



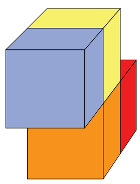
4



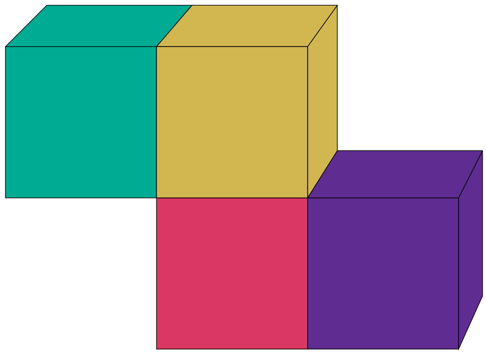
5



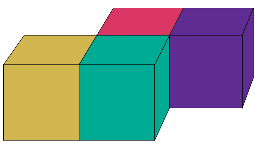
6



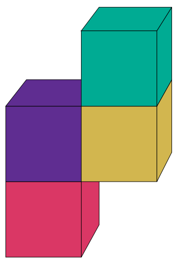
P 3



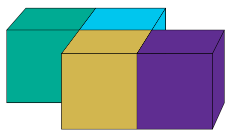
1



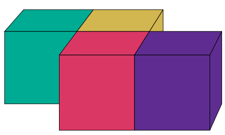
2



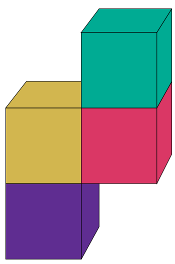
3



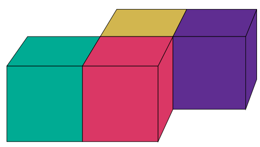
4



5



6

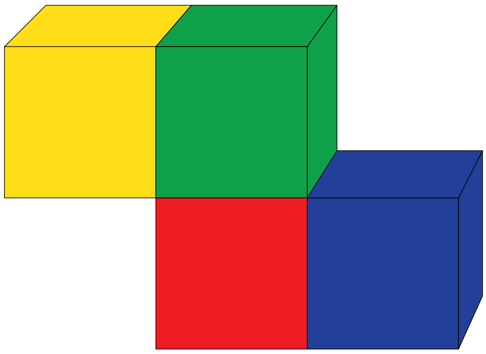


The Coloured Mental Rotation Test

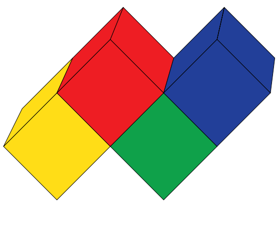
Set A

Four Cube Aggregates

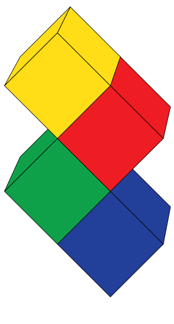
A 1



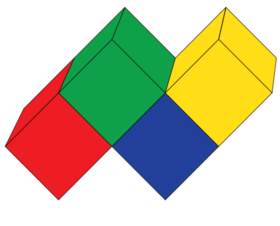
1



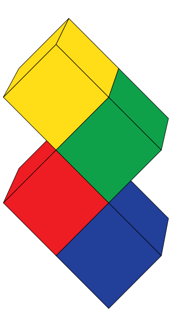
2



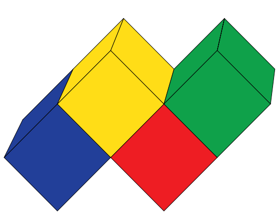
3



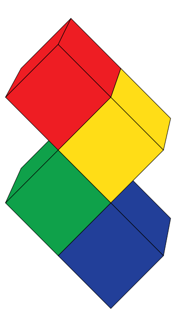
4



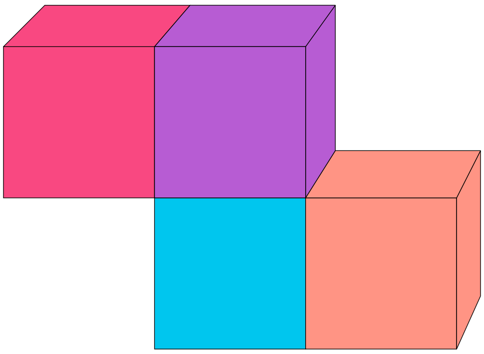
5



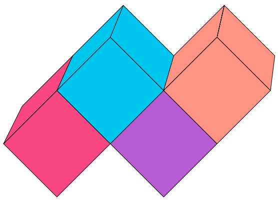
6



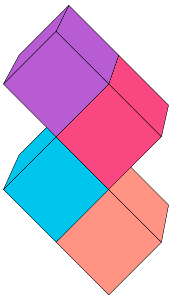
A 2



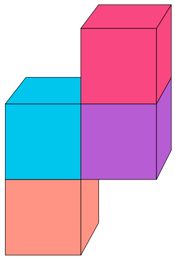
1



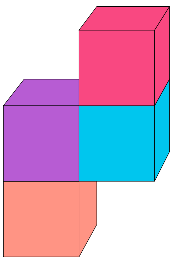
2



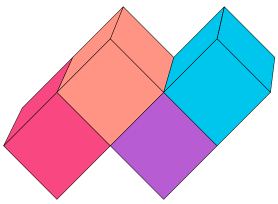
3



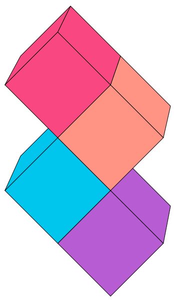
4



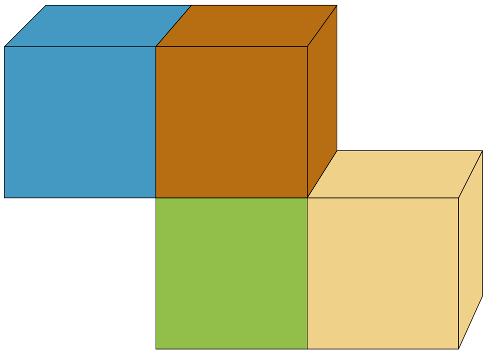
5



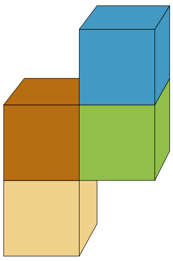
6



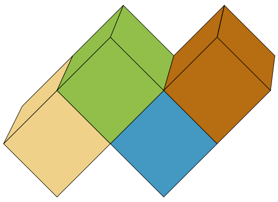
A 3



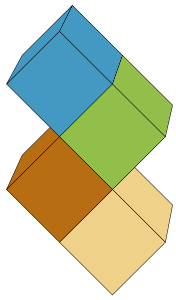
1



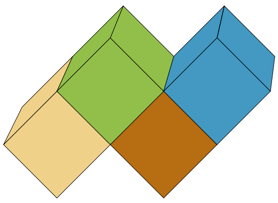
2



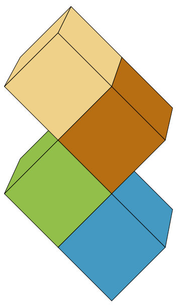
3



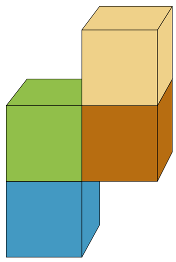
4



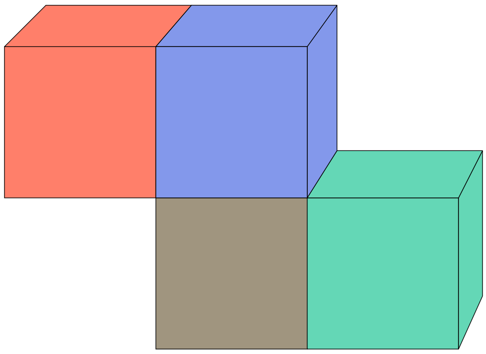
5



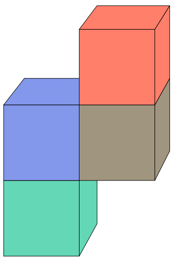
6



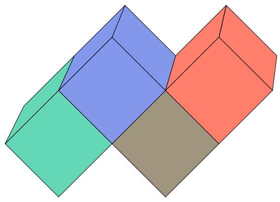
A4



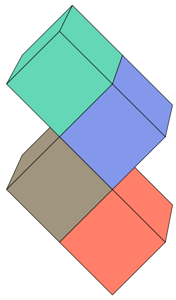
1



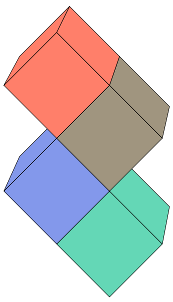
2



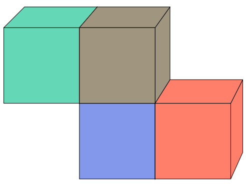
3



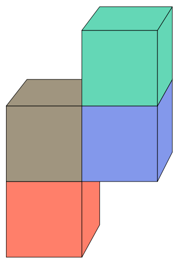
4



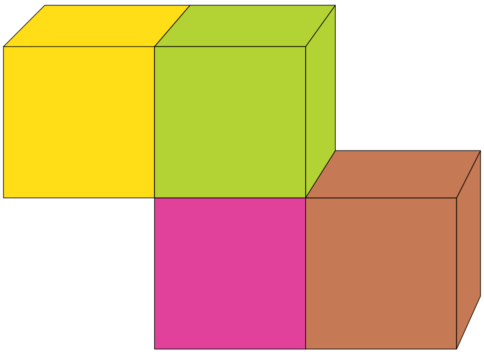
5



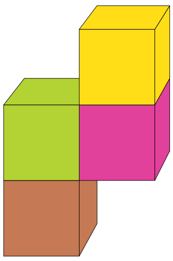
6



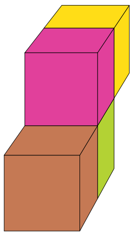
A 5



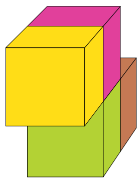
1



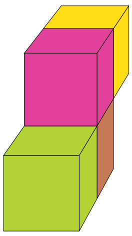
2



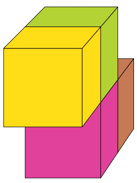
3



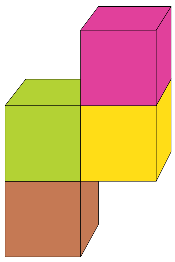
4



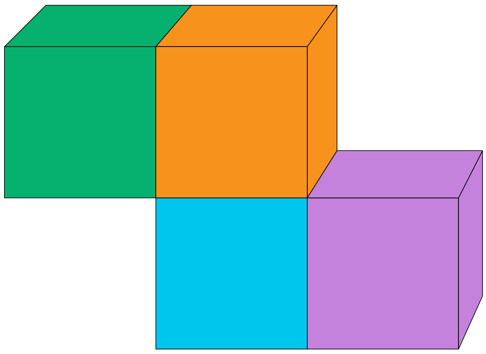
5



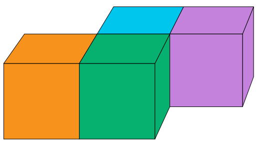
6



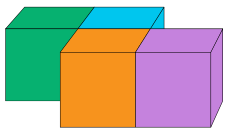
A 6



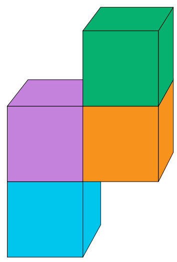
1



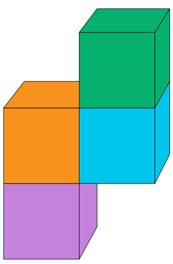
2



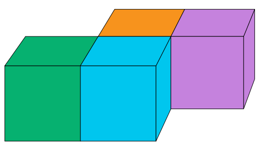
3



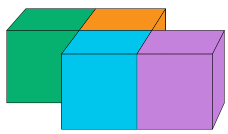
4



5



6

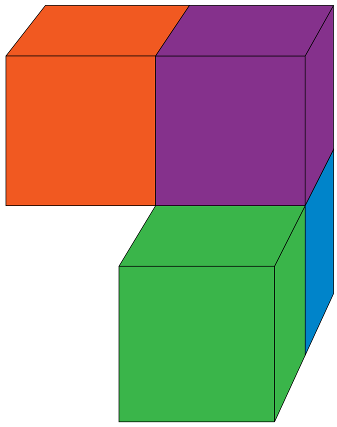


The Coloured Mental Rotation Test

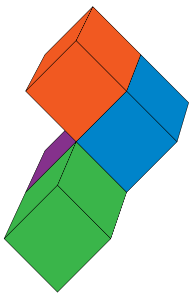
Set B

Four Cube Aggregates

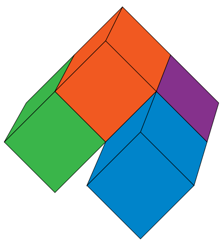
B 1



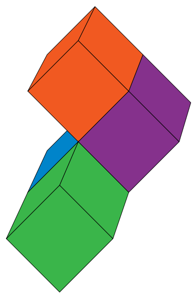
1



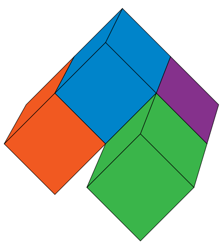
2



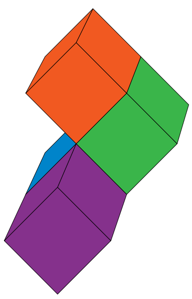
3



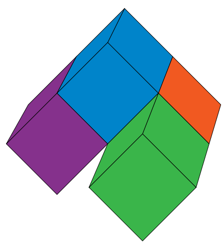
4



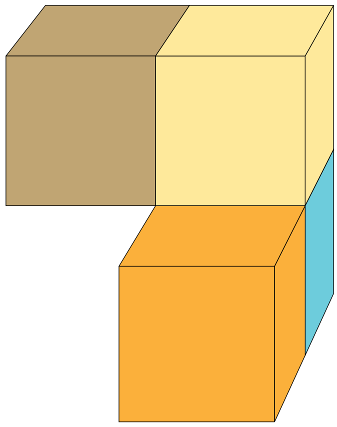
5



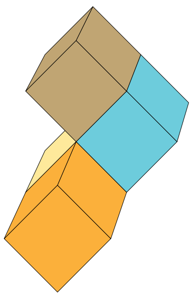
6



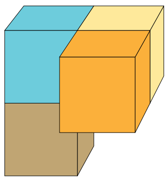
B 2



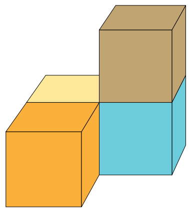
1



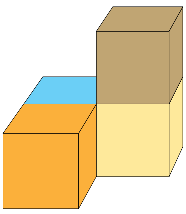
2



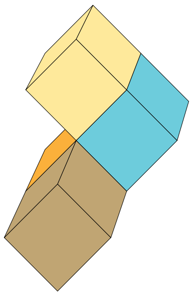
3



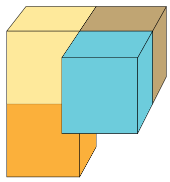
4



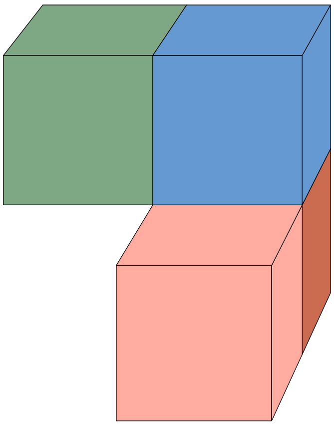
5



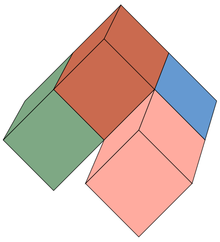
6



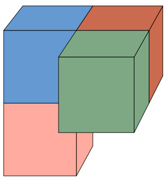
B 3



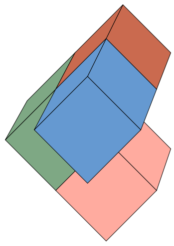
1



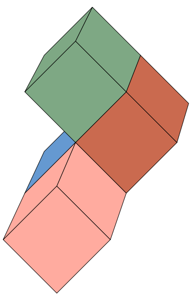
2



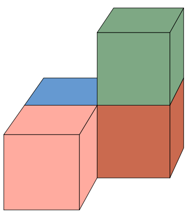
3



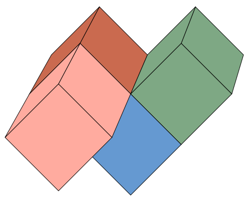
4



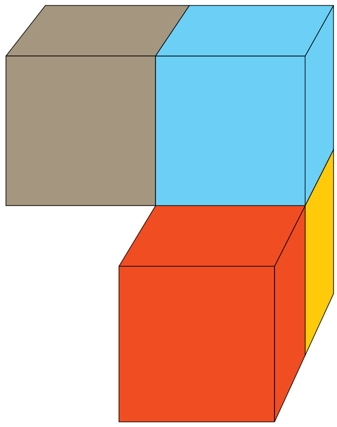
5



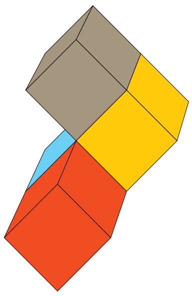
6



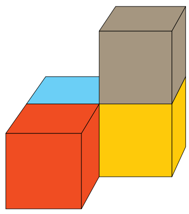
B4



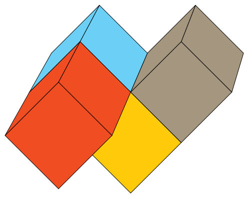
1



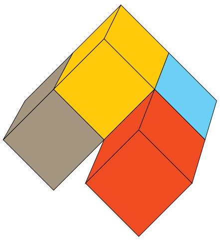
2



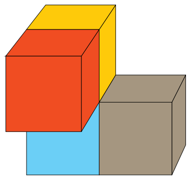
3



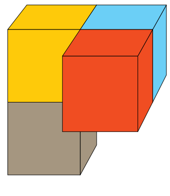
4



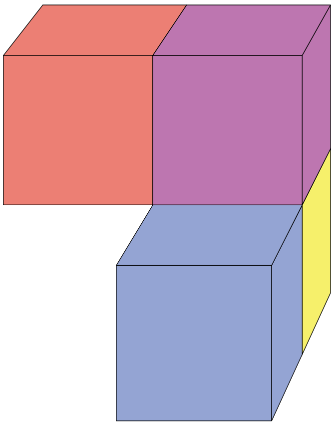
5



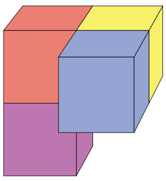
6



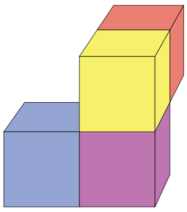
B 5



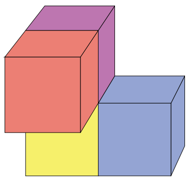
1



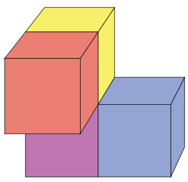
2



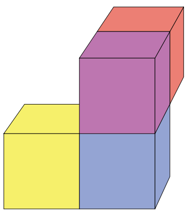
3



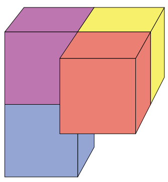
4



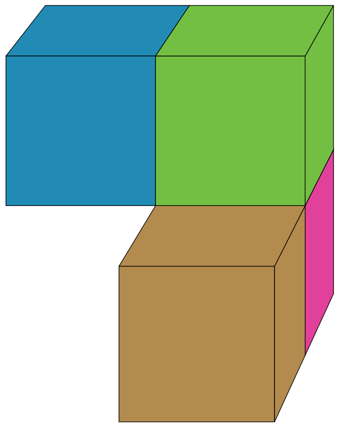
5



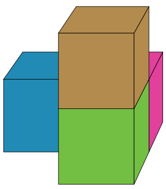
6



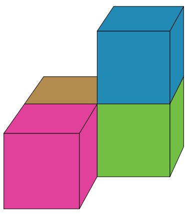
B6



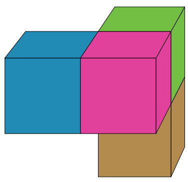
1



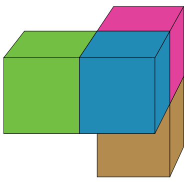
2



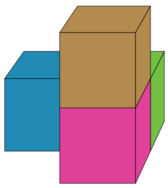
3



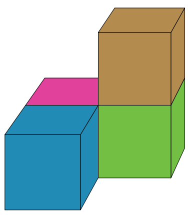
4



5



6

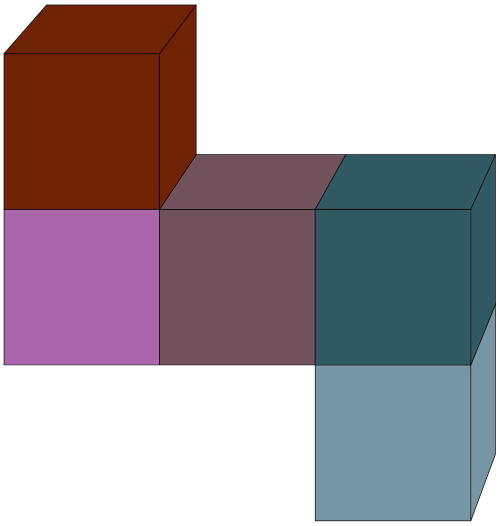


The Coloured Mental Rotation Test

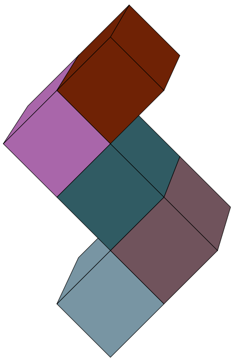
Set C

Five Cube Aggregates

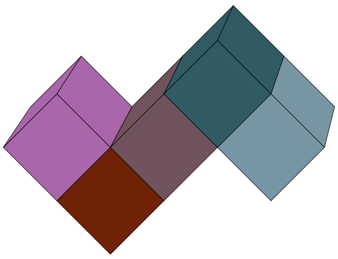
C 1



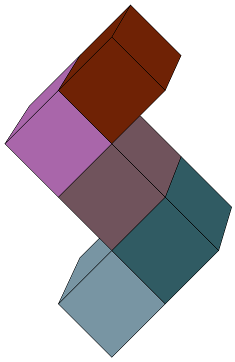
1



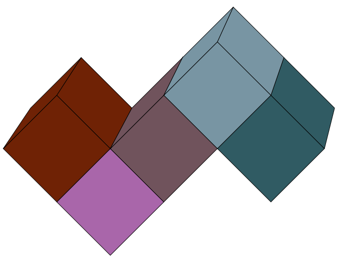
2



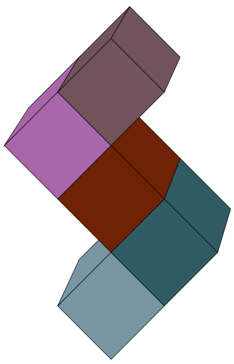
3



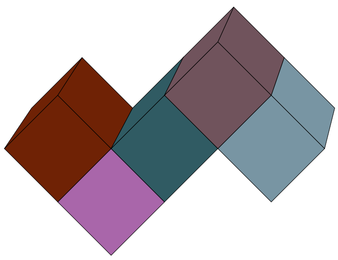
4



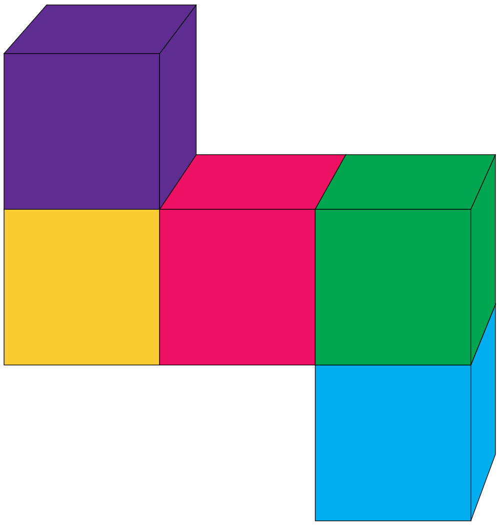
5



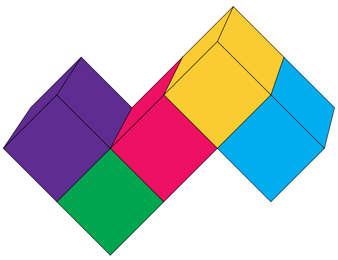
6



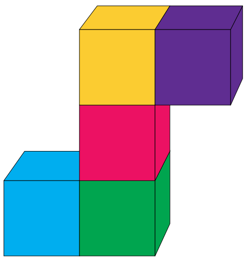
C 2



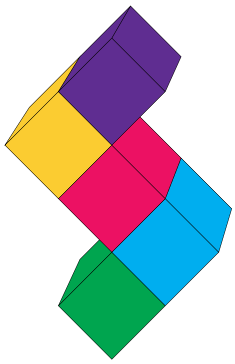
1



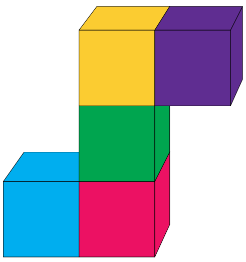
2



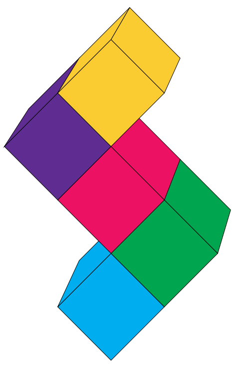
3



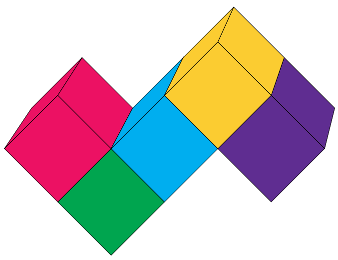
4



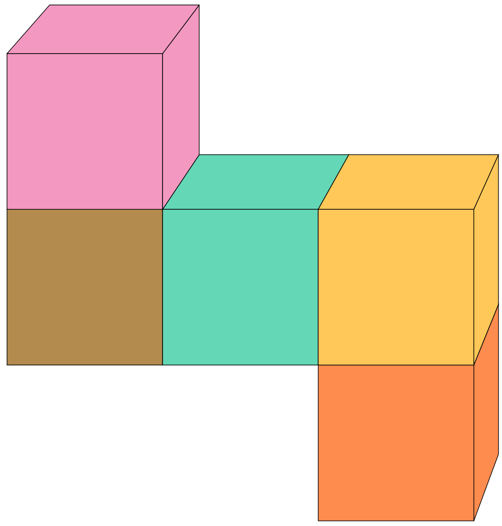
5



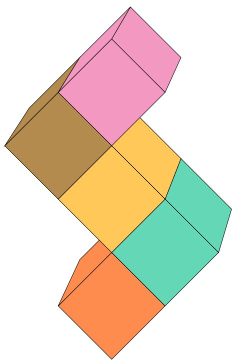
6



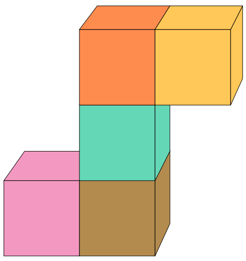
C 3



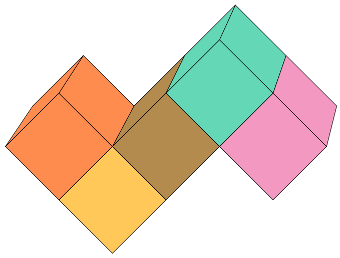
1



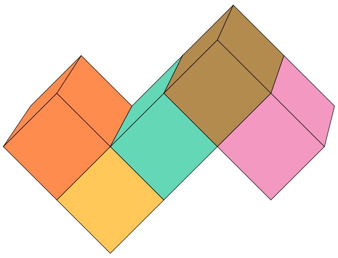
2



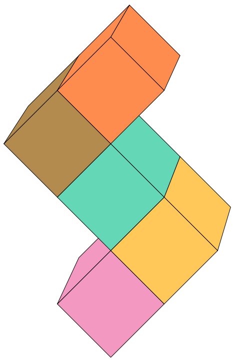
3



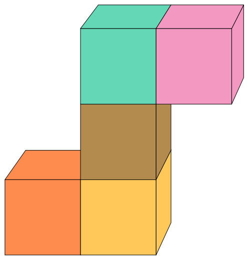
4



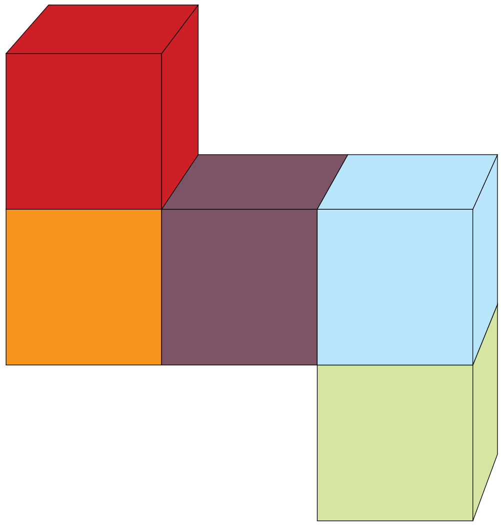
5



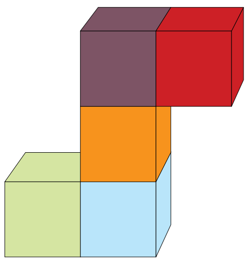
6



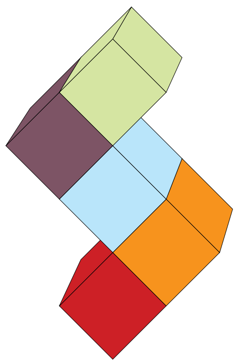
C 4



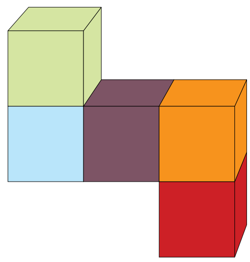
1



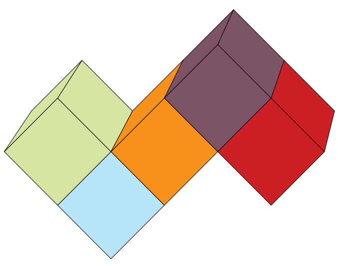
2



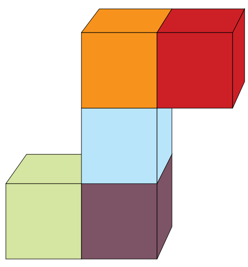
3



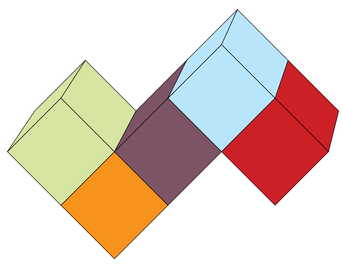
4



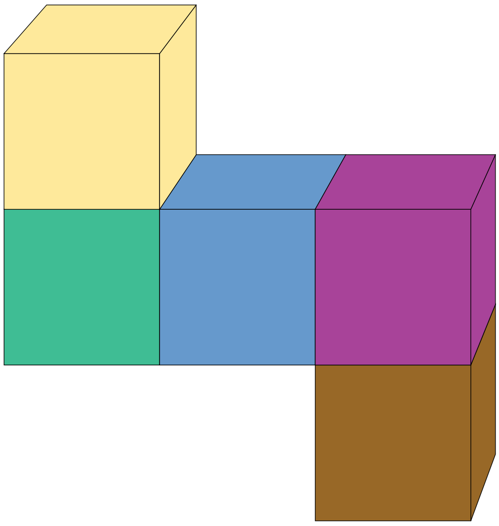
5



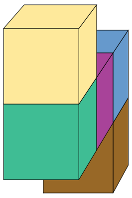
6



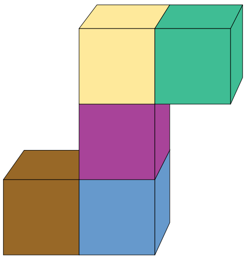
C 5



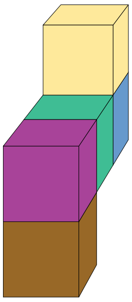
1



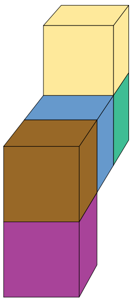
2



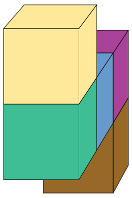
3



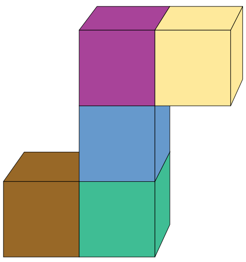
4



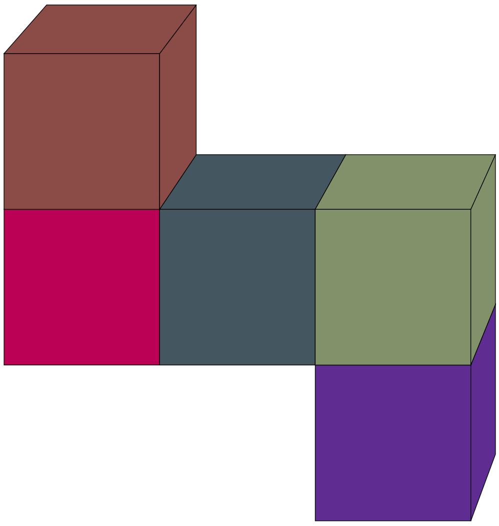
5



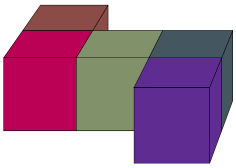
6



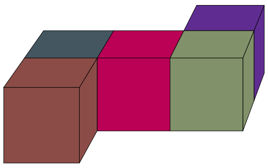
C 6



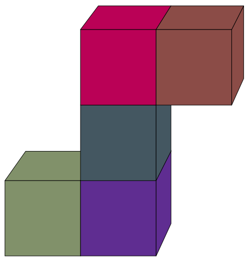
1



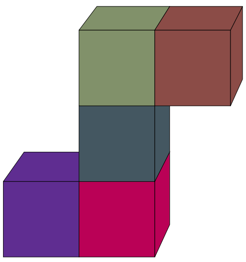
2



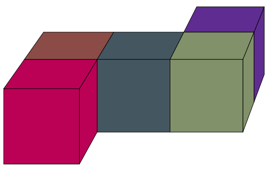
3



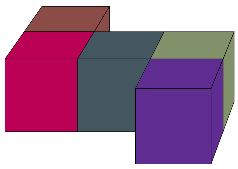
4



5



6

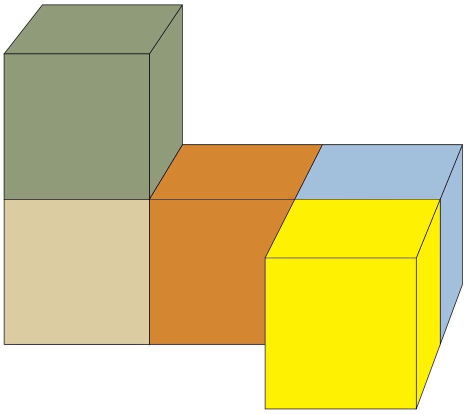


The Coloured Mental Rotation Test

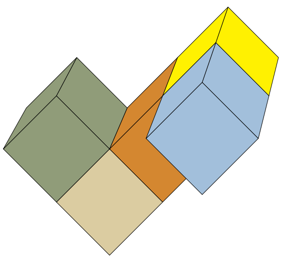
Set D

Five Cube Aggregates

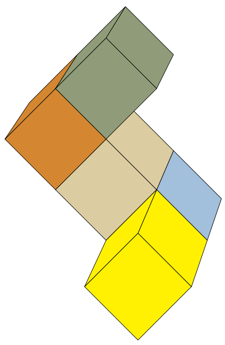
D 1



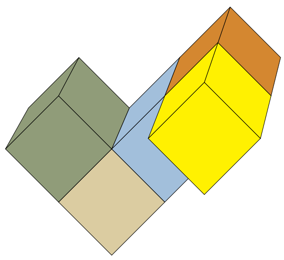
1



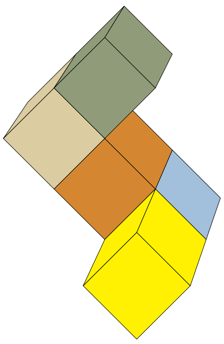
2



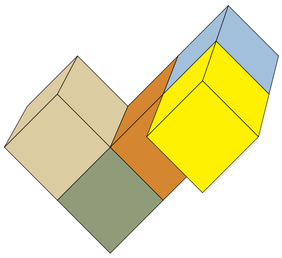
3



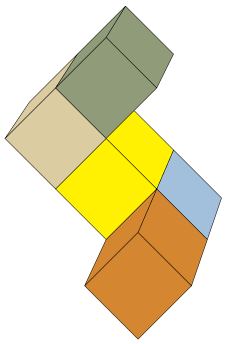
4



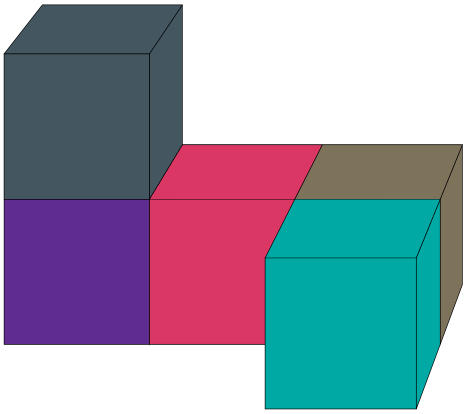
5



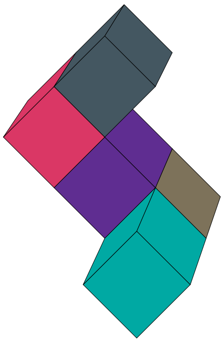
6



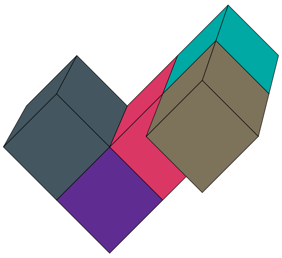
D 2



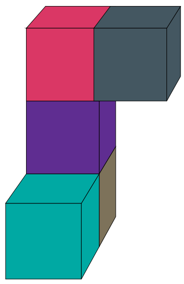
1



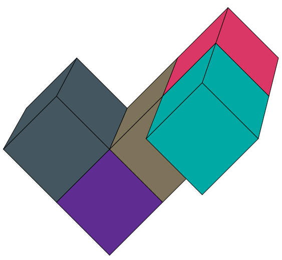
2



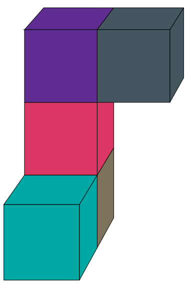
3



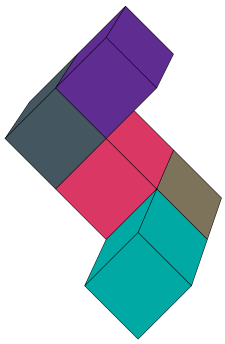
4



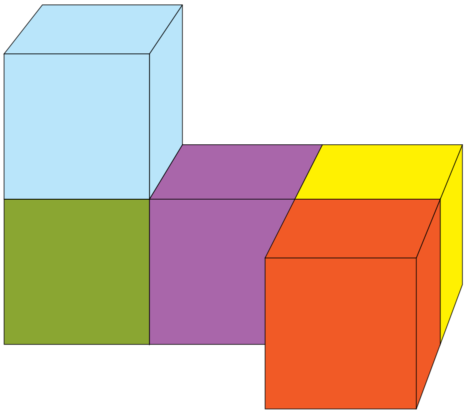
5



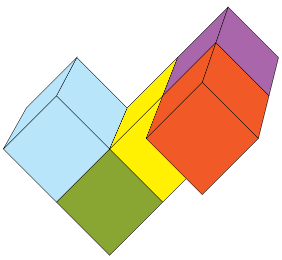
6



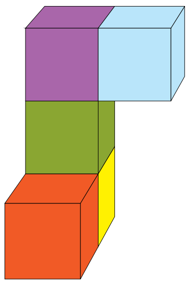
D 3



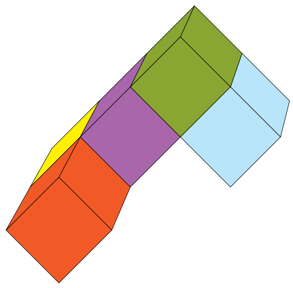
1



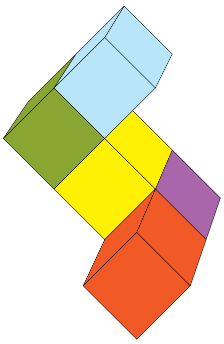
2



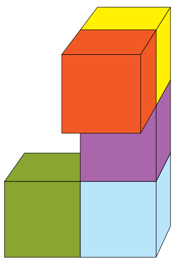
3



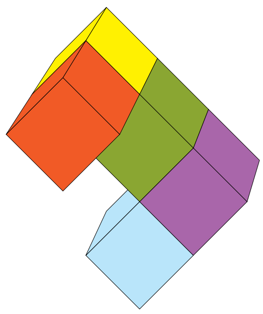
4



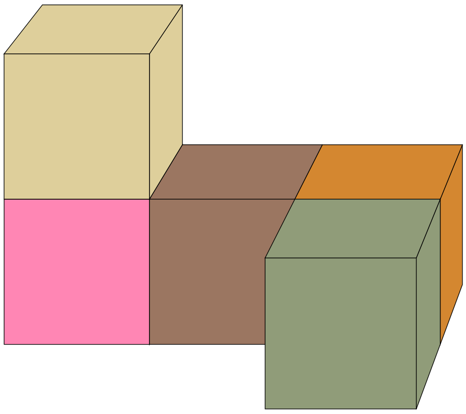
5



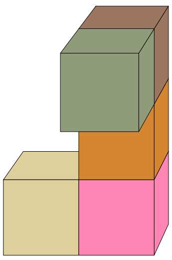
6



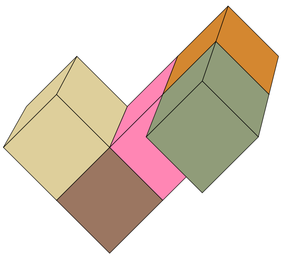
D 4



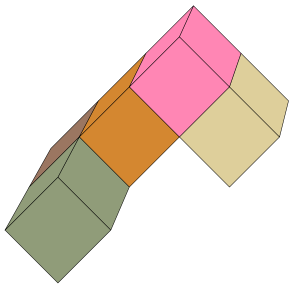
1



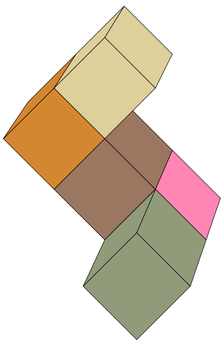
2



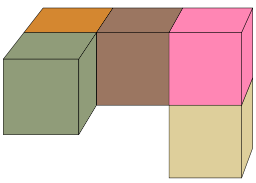
3



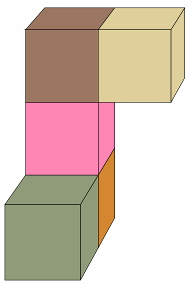
4



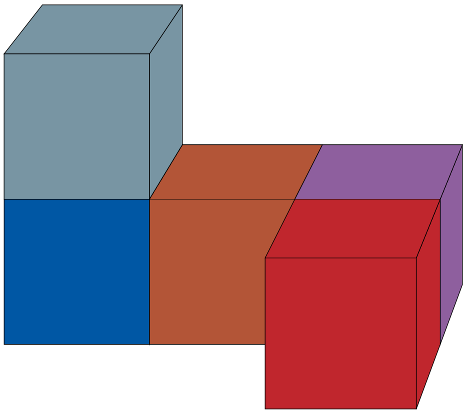
5



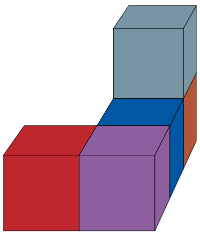
6



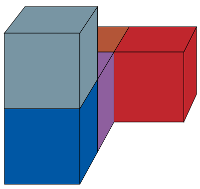
D 5



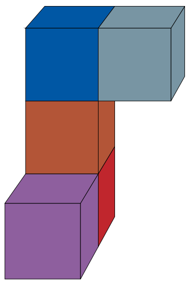
1



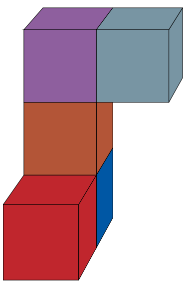
2



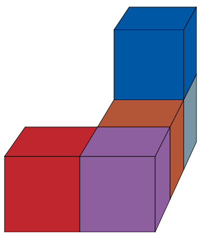
3



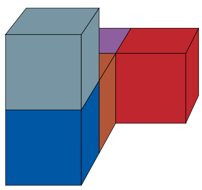
4



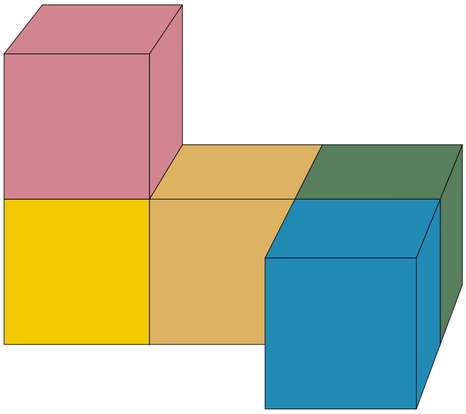
5



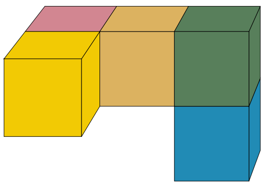
6



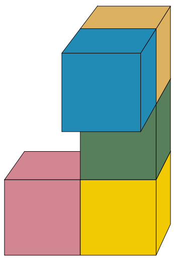
D 6



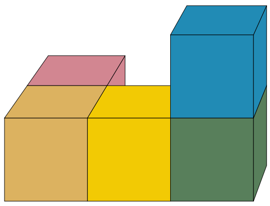
1



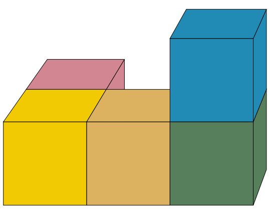
2



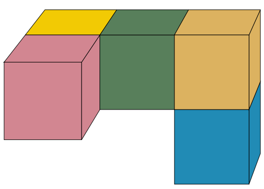
3



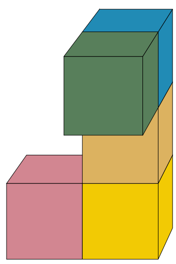
4



5



6

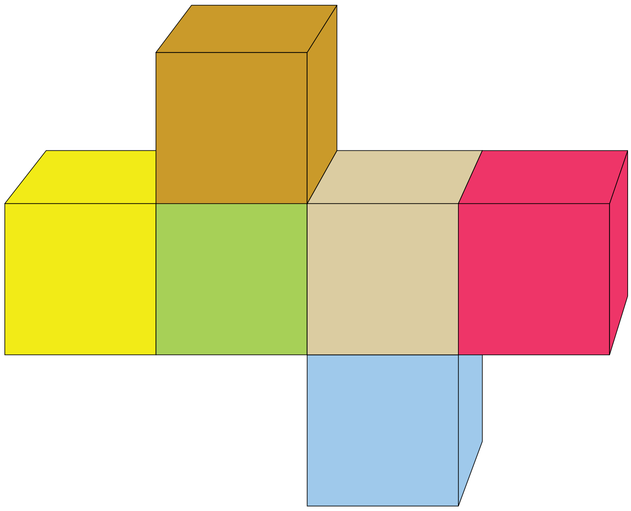


The Coloured Mental Rotation Test

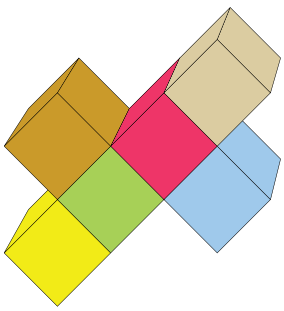
Set E

Six Cube Aggregates

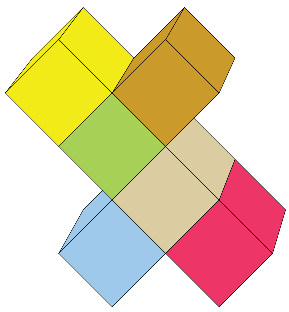
E 1



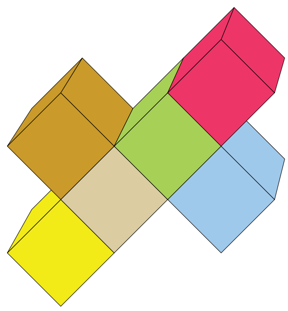
1



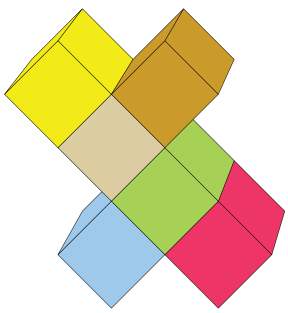
2



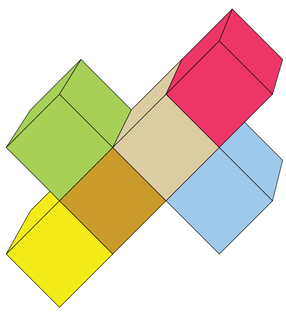
3



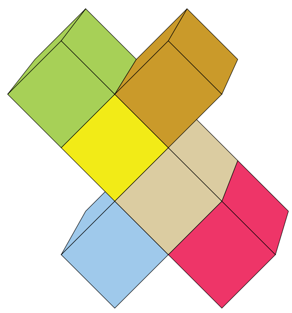
4



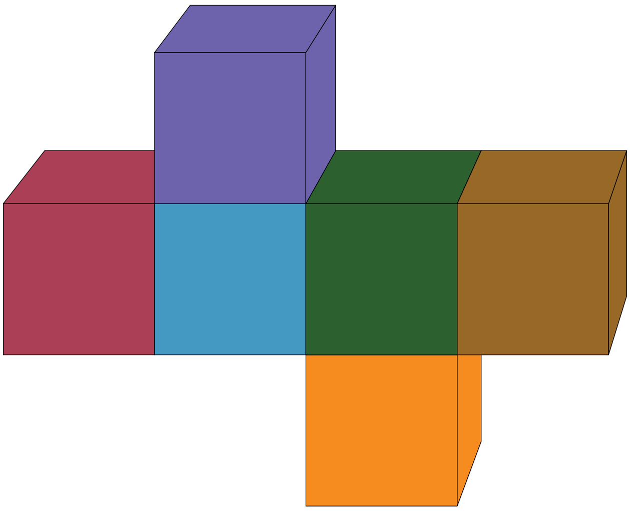
5



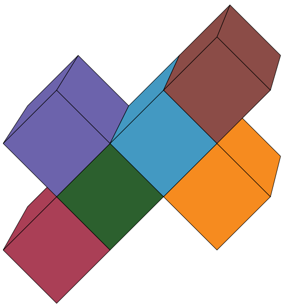
6



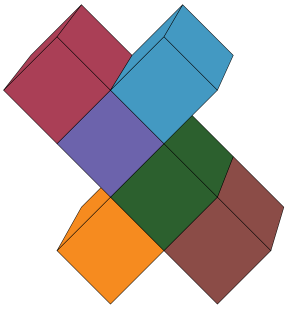
E 2



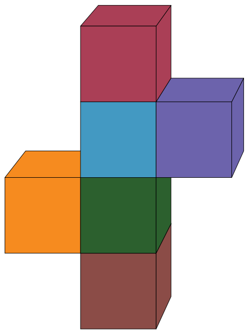
1



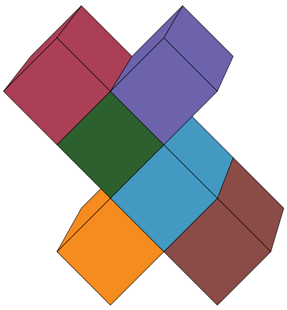
2



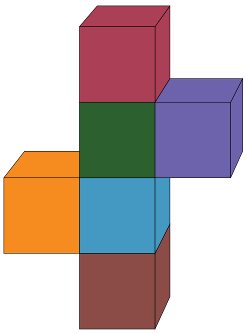
3



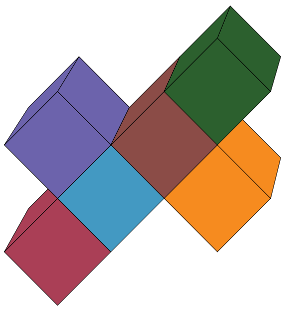
4



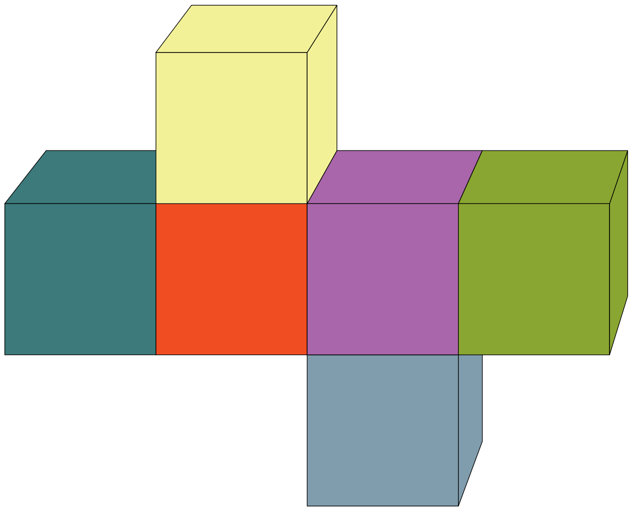
5



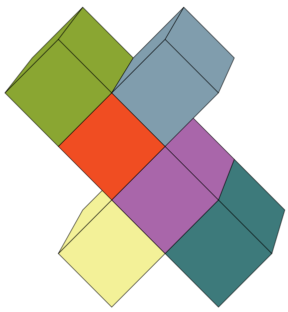
6



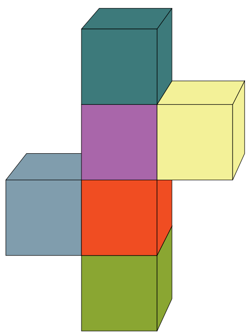
E 3



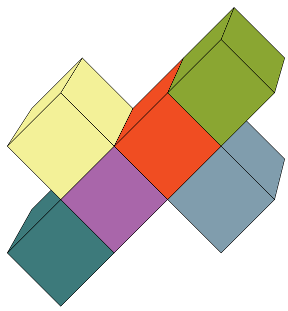
1



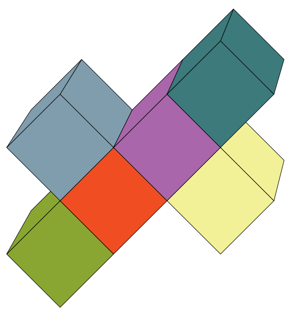
2



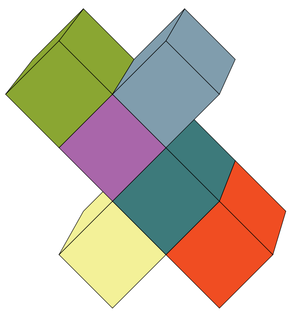
3



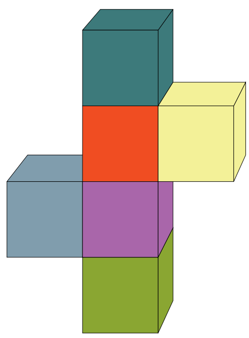
4



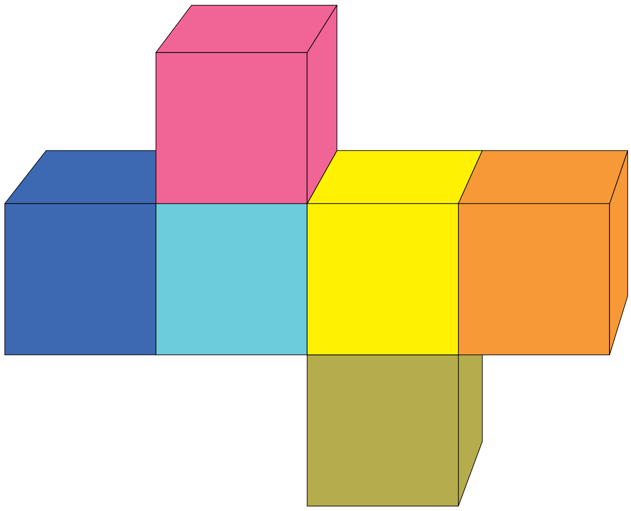
5



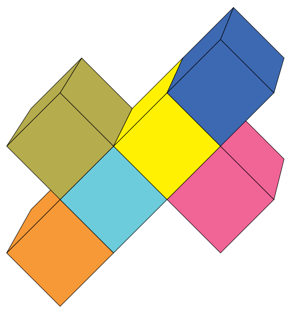
6



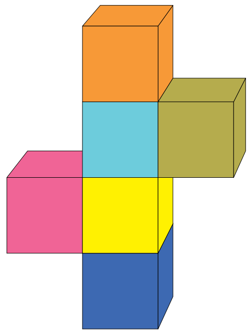
E 4



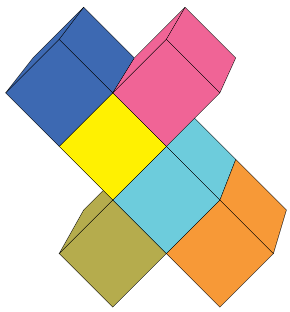
1



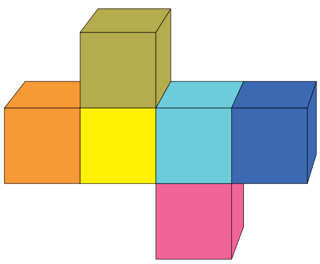
2



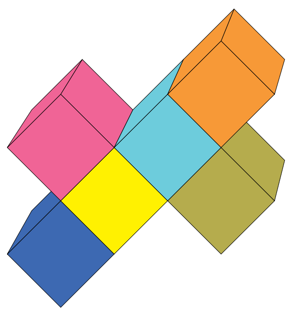
3



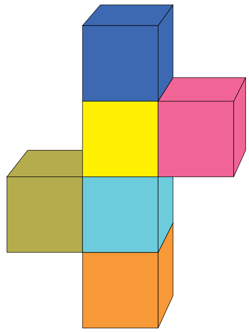
4



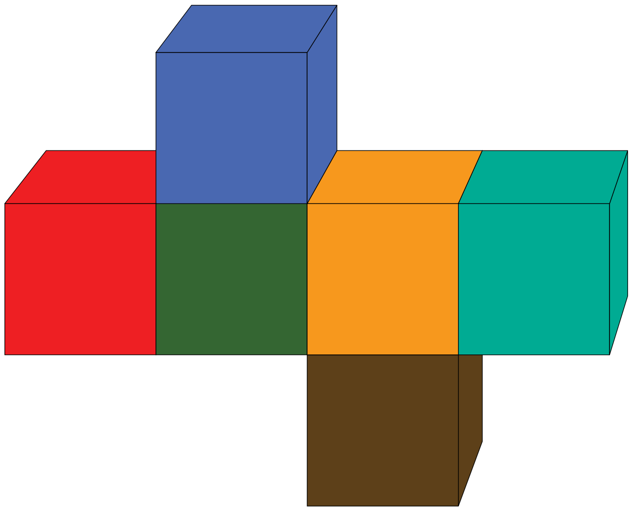
5



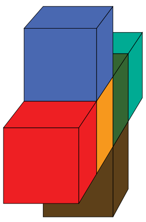
6



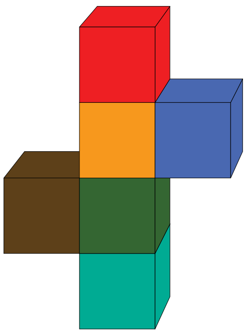
E 5



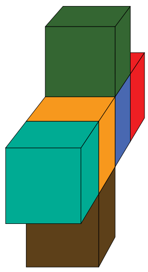
1



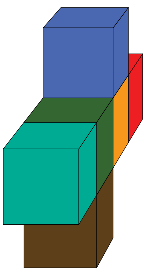
2



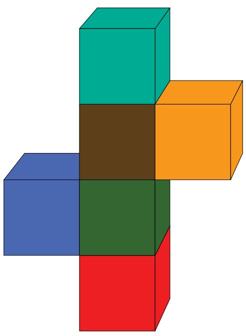
3



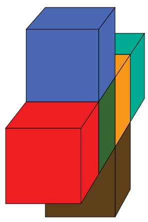
4



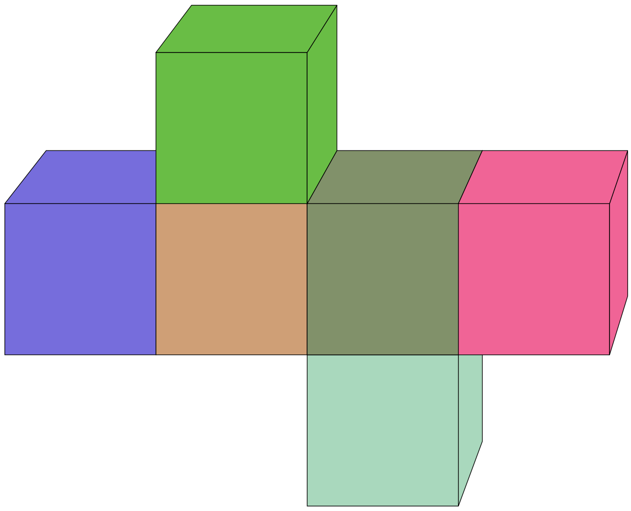
5



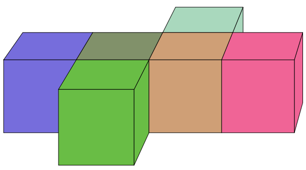
6



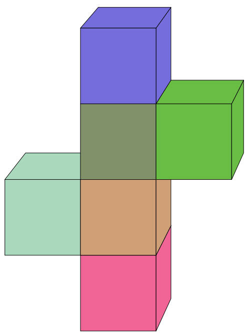
E 6



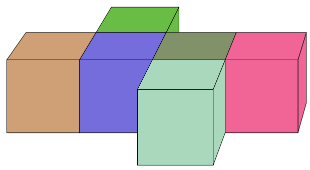
1



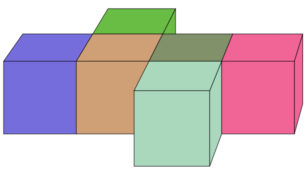
2



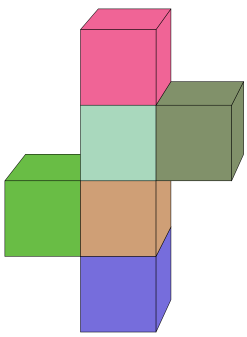
3



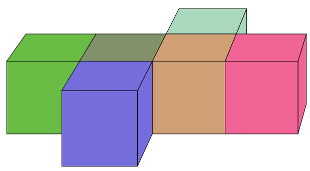
4



5



6

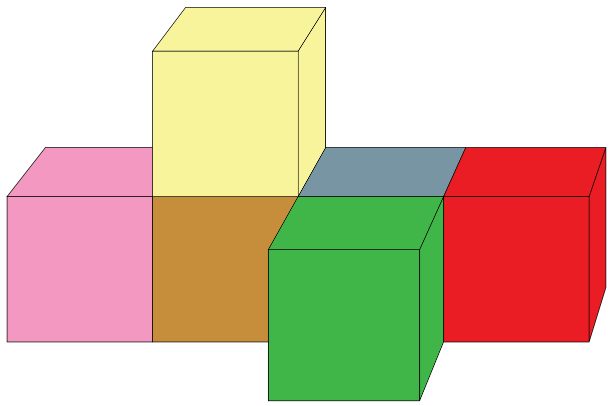


The Coloured Mental Rotation Test

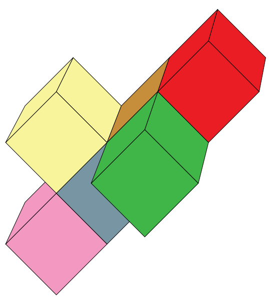
Set F

Six Cube Aggregates

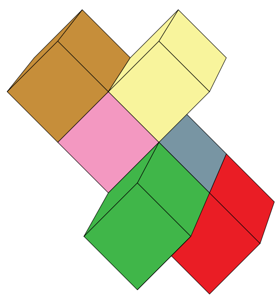
F 1



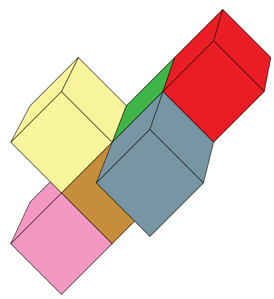
1



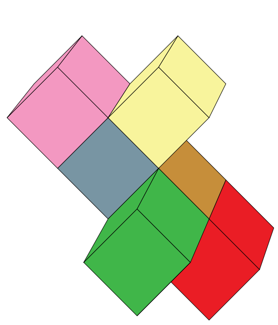
2



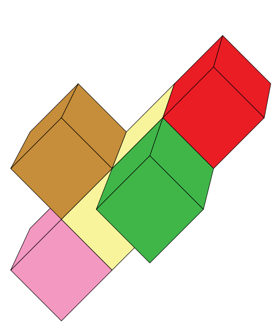
3



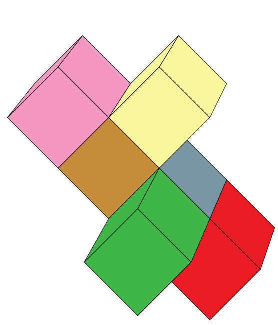
4



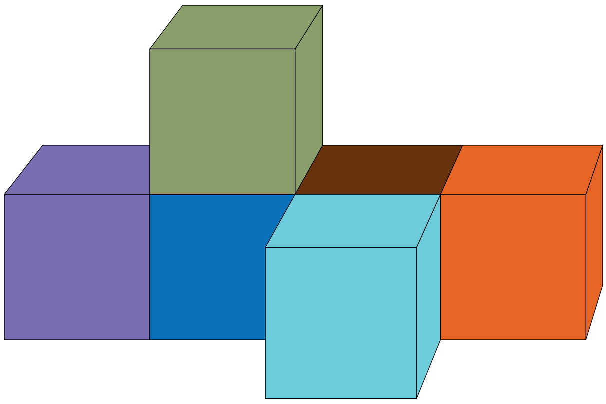
5



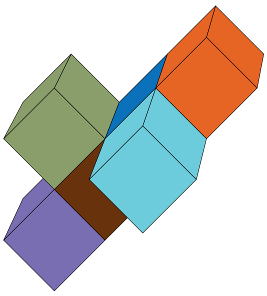
6



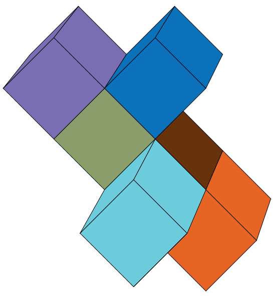
F 2



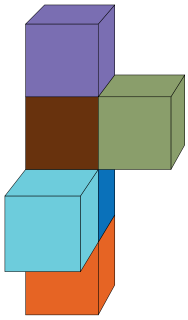
1



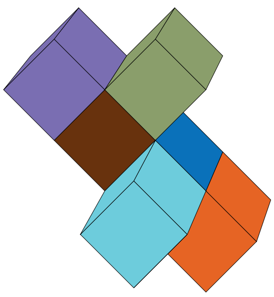
2



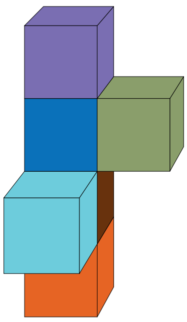
3



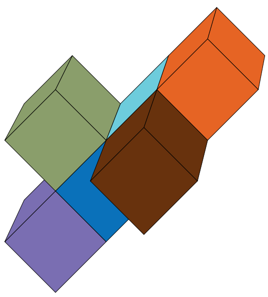
4



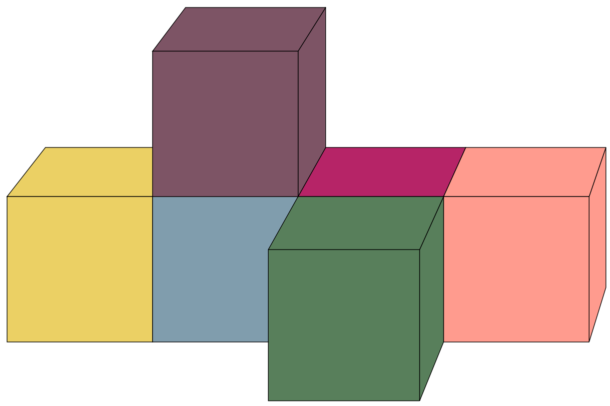
5



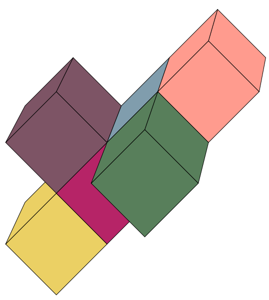
6



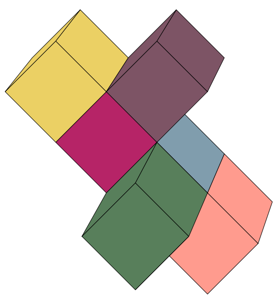
F 3



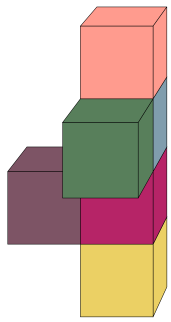
1



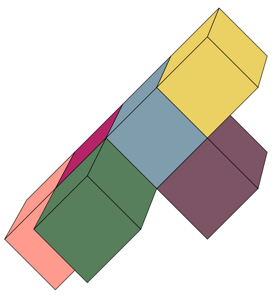
2



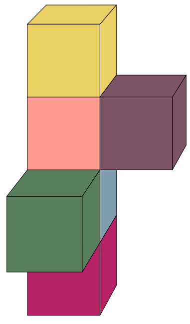
3



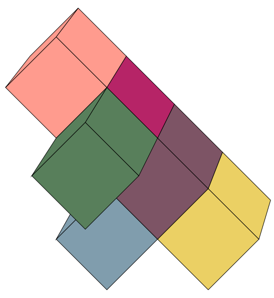
4



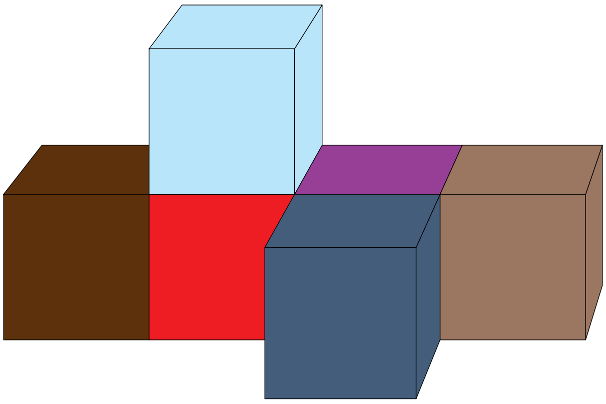
5



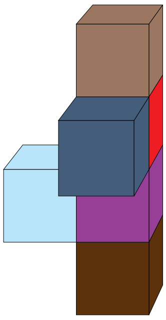
6



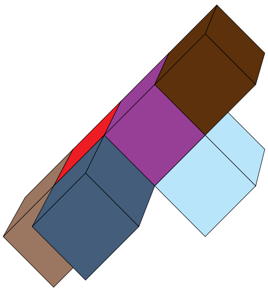
F 4



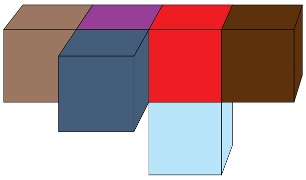
1



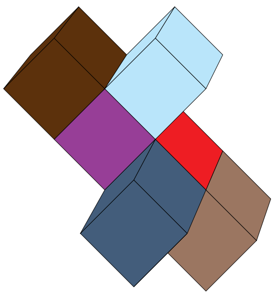
2



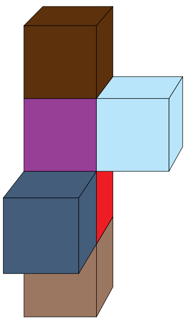
3



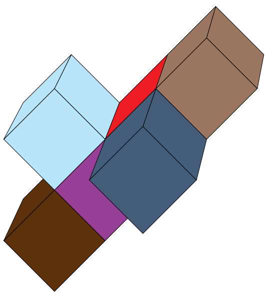
4



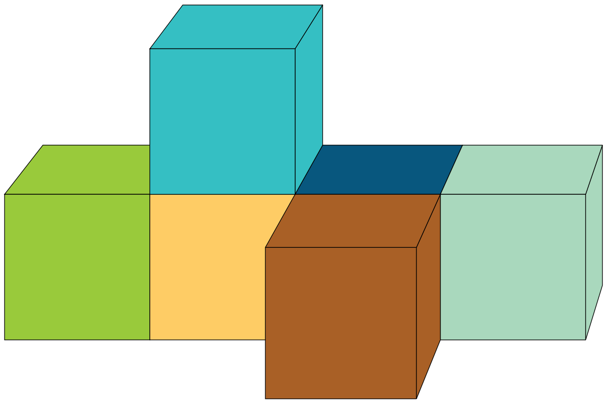
5



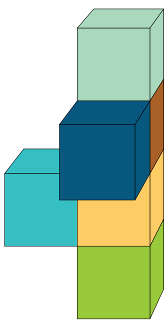
6



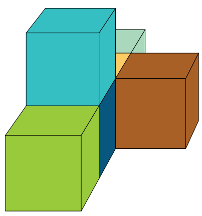
F 5



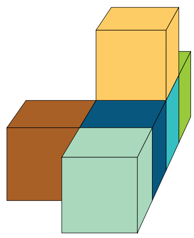
1



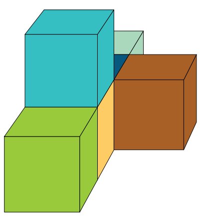
2



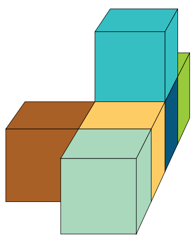
3



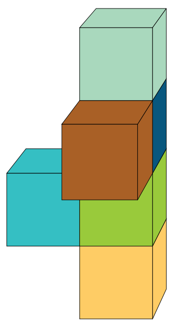
4



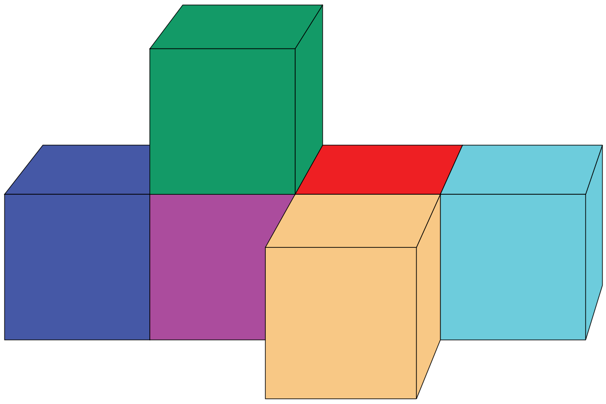
5



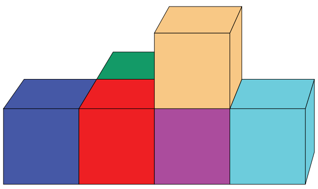
6



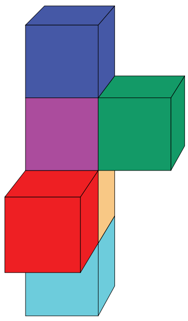
F 6



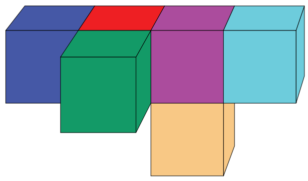
1



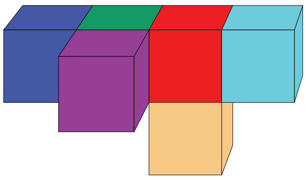
2



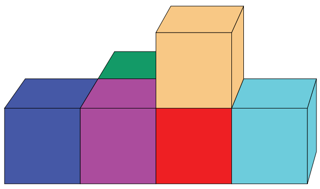
3



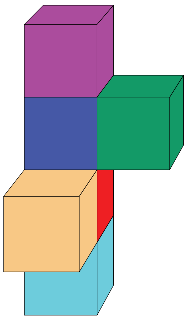
4



5



6



No:	
Gender:	
Age:	
Test sequence	
CMRT:	
CPMT:	
Name:	
Time limit:	

CMRT v.1	**1**	**2**	**3**	**4**	**5**	**6**
P1						
P2						
P3						
A1						
A2						
A3						
A4						
A5						
A6						
B1						
B2						
B3						
B4						
B5						
B6						
C1						
C2						
C3						
C4						
C5						
C6						
D1						
D2						
D3						
D4						
D5						
D6						
E1						
E2						
E3						
E4						
E5						
E6						
F1						
F2						
F3						
F4						
F5						
F6						
CALC I	0					

CPMT	**1**	**2**	**3**	**4**	**5**	**6**
A1						
A2						
A3						
A4						
A5						
A6						
A7						
A8						
A9						
A10						
A11						
A12						
AB1						
AB2						
AB3						
AB4						
AB5						
AB6						
AB7						
AB8						
AB9						
AB10						
AB11						
AB12						
B1						
B2						
B3						
B4						
B5						
B6						
B7						
B8						
B9						
B10						
B11						
B12						
CALC	0					