HEINRICH HEINE
UNIVERSITÄT
DÜSSELDORF

Psycho-soziale und testspezifische Einflüsse auf den Geschlechtseffekt beim Lösen des psychometrischen „Mental-Rotations-Test“

Inaugural-Dissertation

zur Erlangung des Doktorgrades

der Mathematisch-Naturwissenschaftlichen Fakultät

der Heinrich-Heine-Universität Düsseldorf

vorgelegt von

Corinna Titze

geboren in Duisburg

Düsseldorf, Mai 2010

Aus dem Institut für Experimentelle Psychologie

der Heinrich-Heine-Universität Düsseldorf

Gedruckt mit Genehmigung

der Mathematisch-Naturwissenschaftlichen Fakultät der

Heinrich-Heine-Universität Düsseldorf

Referentin: Prof. Dr. Petra Jansen

Koreferent: Prof. Dr. Martin Heil

Tag der mündlichen Prüfung:

Inhaltsverzeichnis

Zusammenfassung

Die Aufklärung des Geschlechtseffekts zugunsten der Männer bezogen auf die mentale Rotationsfähigkeit beschäftigt Kognitionsforscher seit Jahren, gilt dieser Effekt doch als der stabilste in der kognitiven Psychologie (Linn & Petersen, 1985).

Aus einer ersten chronometrischen Studie (Shepard & Metzler, 1971) wurde das psychometrische Testverfahren MRT entwickelt (Peters et al., 1995; Vandenberg & Kuse, 1978). Mit Hilfe dieses Tests lässt sich der Geschlechtseffekt reliabel nachweisen, die Leistungsdifferenz zwischen den Geschlechtern beträgt bis zu einer Standardabweichung (Voyer, Voyer & Bryden, 1995). Zur Aufklärung des Effekts werden zahlreiche Erklärungsansätze angeboten, welche sich überwiegend der Klasse der biologisch-neuronalen Ansätze oder der Klasse der psycho-sozialen Ansätze zuordnen lassen.

Die vorliegende Arbeit setzt sich speziell mit dem Geschlechtseffekt beim Lösen des psychometrischen MRT auseinander und untersuchte Möglichkeiten zur Reduktion des Geschlechtseffekts bei Grundschulkindern, Jugendlichen und Erwachsenen. Dazu wurden zunächst in drei Experimenten Hypothesen, welche auf psycho-sozialen Erklärungsansätzen basieren, überprüft. Es wurde zunächst untersucht, ob sich der Geschlechtseffekt bereits bei Viertklässlern nachweisen lässt. Nach erfolgreicher Bestätigung wurde überprüft, inwiefern dieser – wie bei Erwachsenen nachgewiesen (Moè & Pazzaglia, 2006; Moè 2009) – sich durch die Manipulation des Geschlechtsstereotyps bereits bei Viertklässlern beeinflussen lässt. Die Übertragung dieses konzeptuellen Forschungsansatzes von Erwachsenen auf Grundschulkinder bestätigte die Hypothese nicht. Basierend auf dem Befund, dass adoleszente Mädchen, welche monoedukativ unterrichtet werden, geringere Geschlechtsstereotype gegenüber als „männlich“ beurteilten Interessen aufweisen, wurde untersucht, ob sich in dieser Stichprobe der Geschlechtseffekt beim Lösen des MRT gegenüber koedukativ unterrichteten Schülern verringert (Kessels, 2007; Schoon, 2001). Eine Reduktion des Geschlechtseffekts konnte hypothesenkonform bei Schüler/innen der zwölften Jahrgangsstufe gezeigt werden.

Abschließend wurden in zwei weiteren Experimenten testspezifische Faktoren des MRT – die zeitliche Begrenzung sowie die Komplexität des Antwortformats – hinsichtlich ihres Einflusses auf die Größe des Geschlechtseffekts untersucht. Jedoch konnte kein Einfluss dieser Faktoren mit der hier vorgenommenen Operationalisierung gezeigt werden.

Basierend auf diesen Befunden zeigt sich, dass eine Reduktion des Geschlechtseffekts unter spezifischen Bedingungen um etwa die Hälfte möglich ist. Die Manipulation testspezifischer Faktoren bei Erwachsenen sowie die des Geschlechtsstereotyps bei Grundschulkindern zeigten hingegen keinen Einfluss auf die Größe des Geschlechtseffekts. Es muss festgestellt werden, dass die konkrete Beantwortung der Frage nach den verursachenden Einflussfaktoren des Geschlechtseffekts nach wie vor nicht hinreichend aufgeklärt werden konnte.

Abstract

The research on the gender effect on mental rotation ability in favor of men and on its causes have become increasingly popular in recent decades since this effect turned out to be the most stable gender effect in cognitive psychology (Linn & Petersen, 1985).

The psychometric MRT (Peters et al., 1995; Vandenberg & Kuse, 1978) has been developed from the first chronometric study presented on mental rotation (Shepard & Metzler, 1971). The gender effect in mental rotation is observed most consistently with this test where gender differences amount to one standard deviation. To understand the underlying causes numerous explanations have been offered with biological-neuronal ones on the one hand and psycho-social ones on the other hand.

The present work investigated possibilities to reduce the gender effect in primary school children, adolescents and adults. In three experiments hypotheses based on psycho-social explanations were tested. First, it was investigated, whether the gender effect is already present in fourth graders. After having confirmed this assumption, it was secondly tested, whether this gender effect – as already shown for adults (Moè & Pazzaglia, 2006; Moè 2009) – can be influenced by manipulating gender beliefs already in fourth graders. The transfer of this conceptual approach from adults to children did not confirm the hypothesis. Afterwards, it was examined – based on the finding that adolescent girls attending a single-sex school show less gender stereotyped beliefs against typically "male" interests – whether the gender effect solving the MRT will be reduced towards co-educative pupils. A reduction of the effect showed for pupils in grade twelve.

Finally, two more experiments investigated test-specific factors of the MRT – as it has been time limits on the one hand and complexity of answering format on the other hand - and their influence to effect size of the gender effect. Nevertheless, no influence showed.

Based on these findings, a reduction of the gender effect by half seems to be possible under specific conditions. The manipulation of test specific factors with adults as well as the manipulation of gender beliefs in primary school children did not seem to

have any influence to the size of the mental rotation gender effect. It has become apparent, that a concrete answer regarding the underlying factors for the occurrence of this gender effect still is far from being understood.

1 Einleitung

Die Mentale Rotationsfähigkeit zählt nach Linn und Petersen (1986) zu den drei Komponenten visuell-räumlicher Fähigkeiten (räumliche Wahrnehmung, räumliche Visualisierung, mentale Rotation). Sie beschreibt die Fähigkeit, zwei- oder dreidimensionale Objekte gedanklich zügig und akkurat hinsichtlich ihrer Orientierung zu manipulieren, d.h. sie beispielsweise zu drehen, zu spiegeln oder zu kippen (Linn und Petersen, 1985).

Die Untersuchung visuell-räumlicher Fähigkeiten und des wiederholt nachgewiesenen Geschlechtseffekts stellt nicht nur in der wissenschaftlichen Forschung ein zunehmend beliebteres Thema dar. Gerade auch die populärwissenschaftliche Literatur bedient sich gerne der Klischees hinsichtlich der (vermeintlich) unterschiedlichen Fähigkeiten von Männern und Frauen in Bezug auf ihre räumlichen Fähigkeiten (Pease & Pease, 2001). Doch trotz aller Übertreibungen und Vereinfachungen der Populärliteratur lässt sich ein Geschlechtseffekt zugunsten der Männer beim Lösen visuell-räumlicher Aufgaben in der Regel nicht von der Hand weisen (Halpern, 1989). Schon 1974 stellten Maccoby und Jacklin in einem klassischen Werk über psychologische Geschlechtsunterschiede fest, dass neben den mathematischen und den sprachlichen Fertigkeiten sowie der Aggressivität immer wieder auch hinsichtlich der räumlichen Fähigkeiten Unterschiede zwischen den Geschlechtern auftreten. Insbesondere die mentale Rotationsfähigkeit nimmt unter dem Sammelbegriff der visuell-räumlichen Fähigkeiten eine besondere Stellung ein.

Dieses Faktum mag zunächst wenig gravierend erscheinen, nimmt die mentale Rotation als ein Bestandteil der visuell-räumlichen Fähigkeiten im Alltag auf den ersten Blick scheinbar keine herausragende Rolle ein. Das Klischee, dass eine Frau gemeinhin schlecht einparken und keine Stadtpläne lesen könne, erscheint im Zeitalter von Einparkhilfen und Navigationsgeräten nebensächlich. Dennoch, der Einfluss visuell-räumlicher Fähigkeiten ist nicht zu unterschätzen: Cooper und Mumaw (1985) wiesen darauf hin, dass die Vorhersage des beruflichen Erfolgs in bestimmten Berufsfeldern durch räumliche Testverfahren valider erfolgt als durch allgemeine Intelligenztests. Radiologen, Chirurgen, Piloten, Fluglotsen oder Architekten beispielsweise benötigen eine besonders ausgeprägte visuell-räumliche Kompetenz, um den Anforderungen ihres Berufsalltags gerecht werden zu können (Brandt, 2005; Rozencwajg, 2005; Smoker, 1984; Wanzel, 2002). Gleichzeitig fällt auf, dass genau diese Berufszweige vornehmlich von Männern dominiert werden, obwohl diese Bereiche im Zuge der Gleichberechtigung und Emanzipation für Frauen seit mehreren Jahrzehnten zugänglich gemacht worden sind und Mädchen und junge Frauen nunmehr als „Gewinnerinnen“ des (deutschen) Schulsystems angesehen werden (Herwartz-Emden, 2007).

Die Relevanz der Aufklärung dieses Geschlechtseffekts bzgl. der visuell-räumlichen Fähigkeiten erscheint daher zentral, um dessen Beitrag zu dem bestehenden Ungleichgewicht erfassen und idealerweise reduzieren zu können. Fragen nach dem ersten Auftretensalter und der Manifestation des Effekts sowie Möglichkeiten zur Manipulation und damit Reduktion des Effekts erscheinen in diesem Zusammenhang interessant.

1.1 Das Paradigma der Mentalen Rotation

Shepard und Metzler waren die ersten, die 1971 das Paradigma der Mentalen Rotation im wissenschaftlichen Kontext vorstellten. Mittlerweile zählt das Paradigma zu einem der best untersuchtesten innerhalb der kognitiven Psychologie. Bis heute wurden allein 1371 Studien zur mentalen Rotation veröffentlicht (Stand Mai 2010, Datenbank PubMed).

In der Originalarbeit von Shepard und Metzler (1971) wurden den Probanden in einem chronometrischen Testverfahren paarweise Abbildungen dreidimensionaler identischer oder zueinander spiegelverkehrter Würfelfiguren, welche entweder in der Tiefen- oder der Bildebene um unterschiedliche Winkelgrade rotiert waren, präsentiert. Die Probanden mussten entscheiden, ob sich die Objekte durch Rotation ineinander überführen ließen oder nicht (siehe Abbildung 1). Es zeigte sich, dass die Rotationsgeschwindigkeit mit zunehmender Winkeldisparität linear anstieg und dass die Rotation in der Bildebene nicht mehr Zeit in Anspruch nahm, als die Rotation in der Tiefenebene. Dieser Befund wird dahingehend interpretiert, dass der Prozess der mentalen Rotation analog zu einer real durchgeführten Rotation zu verstehen ist.

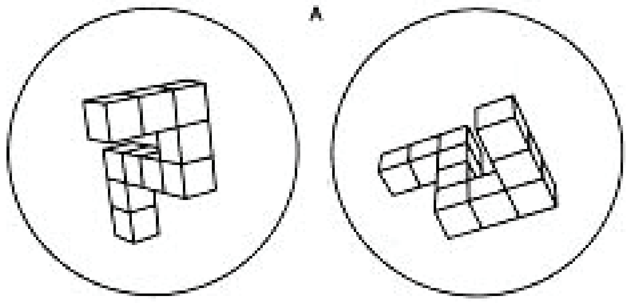


Abbildung 1: Reizpaar aus dem chronometrischen Originalverfahren von Shepard und Metzler (1971). Es muss geprüft werden, ob sich die Reize durch Rotation ineinander überführen lassen oder nicht.

Aus diesem chronometrischen Testverfahren, bei welchem neben der Erfassung der Antwortgenauigkeit der Probanden auch die Messung der Reaktionszeiten mittels Computer möglich war, entwickelten sich in der folgenden Zeit psychometrische sog. Papier-Bleistift-Tests. Diese Testverfahren sind in der Regel zeitlich begrenzt und erfassen die Antwortgenauigkeit, d.h. die Anzahl korrekt gelöster Items, die genaue Messung der Reaktionszeit entfällt jedoch. Chronometrische und psychometrische Testverfahren unterscheiden sich aber nicht nur hinsichtlich der Darbietungsform und der Datenerfassung, sondern auch in Bezug auf ihre Komplexität. Während in chronometrischen Verfahren in der Regel mit Paarvergleichen gearbeitet wird, werden in psychometrischen Verfahren oftmals komplexere Auswahlentscheidungen von den Probanden verlangt. Nachfolgend wird der bekannteste psychometrische Test zur Erfassung der mentalen Rotationsfähigkeit vorgestellt.

1.2 Der psychometrische „Mental-Rotations-Test (MRT)”

Der “Mental-Rotations-Test” ist eine von Peters, Laeng, Latham, Jackson, Zaiyouna und Richardson (1995) neu entworfene und erweiterte Version des ersten psychometrischen mentalen Rotationstests von Vandenberg und Kuse (1978), welcher auf den Originalitems von Shepard und Metzler’s (1971) chronometrischem Test basierte. Der Test von Vandenberg und Kuse erwies sich aufgrund qualitativ schlechter werdender Kopiervorlagen zunehmend als ungeeignet zum Einsatz in der Forschung.

Der MRT besteht aus insgesamt 24 Items, welche in zwei Sets mit je zwölf Items dargeboten werden. Als Items werden zweidimensionale Abbildungen dreidimensional dargestellter Würfelfiguren genutzt. Eine Würfelfigur ist dabei aus zehn einzelnen Quadern zusammengesetzt, ein Item besteht aus einer Zielfigur auf der linken Seite und vier Vergleichsfiguren auf der rechten Seite. Zwei dieser vier Vergleichsfiguren lassen sich durch Rotation um die vertikale Achse in die jeweilige Zielfigur überführen. Die verbleibenden zwei Vergleichsfiguren stellen Distraktoren dar, bei denen es sich entweder um Spiegelbilder der Zielfigur oder um vollständige andere Würfelfiguren handelt (siehe Abbildung 2).

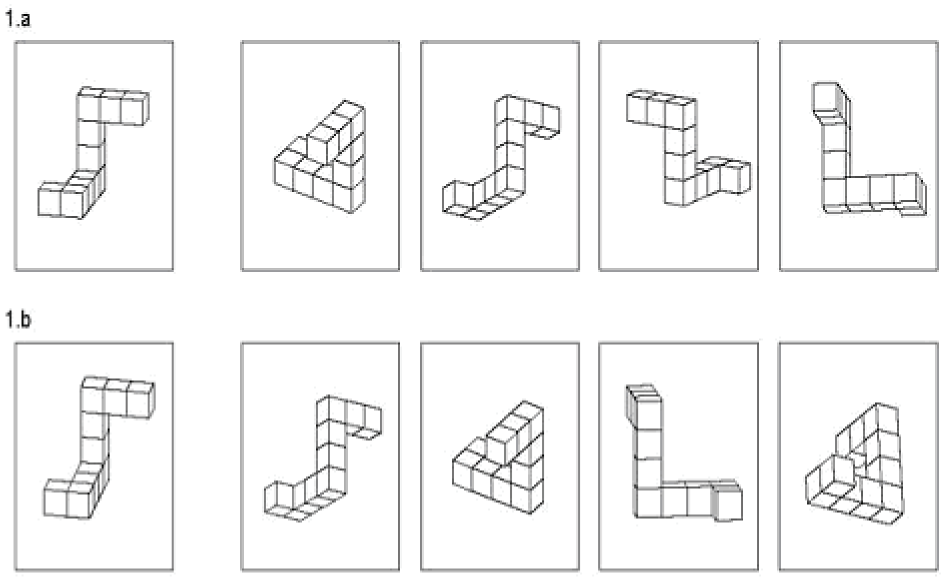


Abbildung 2: Diese Abbildung zeigt jeweils das erste Item aus den psychometrischen Testverfahren MRT A und MRT B. Zwei der vier Vergleichsfiguren lassen sich durch Rotation in die Zielfigur überführen (Lösung 1a: Vergleichsfigur 1 und 3, Lösung 1b: Vergleichsfigur 2 und 3).

Die Aufgabe des Probanden besteht darin, innerhalb einer pro Set auf drei Minuten begrenzten Zeitspanne, die beiden jeweils korrekten Vergleichsfiguren zu identifizieren und zu markieren. Zwischen den beiden Sets wird eine vierminütige Pause eingehalten. Ein Punkt wird immer nur dann vergeben, wenn tatsächlich beide Figuren richtig erkannt worden sind, so dass ein Maximum von 24 Punkten in diesem Testverfahren erzielt werden kann. Durch dieses konservative Bewertungskriterium liegt die Ratewahrscheinlichkeit bei 3.96 Items pro Test.

Der MRT wurde von Peters et al. (1995) in drei verschiedenen Versionen entwickelt. Der MRT A beinhaltet die Originalitems, während der MRT B aus einer Neuzusammensetzung der Items der ersten Version besteht. Es wurde gezeigt, dass sich die Schwierigkeit beider Versionen durch dieses Vorgehen nicht veränderte und der Einsatz des Verfahrens in Gruppentestungen möglich ist. Der MRT C ist der schwierigste Test und verlangt die Rotation der Items sowohl um die horizontale als auch die vertikale Achse.

Nachfolgend präsentierte Studien greifen im Wesentlichen auf den psychometrischen MRT im Original zurück oder benutzen leichte Modifizierungen, da mit diesem Testverfahren die Überlegenheit der Männer in der mentalen Rotation am konsistentesten beobachtet werden kann.

1.3 Geschlechtseffekte in der mentalen Rotationsfähigkeit

Obwohl Shepard und Metzler (1971) in ihrer klassischen chronometrischen Studie keine Geschlechtseffekte untersuchten, stellte sich später heraus, dass genau das von ihnen verwendete Reizmaterial den größten, stabilsten und reliabelsten Geschlechtseffekt zugunsten der Männer in der kognitiven Psychologie erzeugt (Halpern, 1989; Voyer, Voyer & Bryden, 1995) – insbesondere dann, wenn psychometrische Testverfahren eingesetzt werden.

Nachfolgend dargestellte Befunde zu Geschlechtseffekten beziehen sich auf Leistungsdifferenzen aus Verhaltensdaten, wobei psychometrische und chronometrische Studien vorgestellt werden. Auf die Darstellung der biologisch-neuronalen Datenlage (z.B. Unterschiede in der Hirnaktivität) wird hingegen verzichtet, wenngleich deren Einfluss zur Aufklärung des Effekts nicht zu vernachlässigen ist.

1.3.1 Geschlechtseffekte hinsichtlich der mentalen Rotationsfähigkeit bei Erwachsenen

Vandenberg und Kuse (1978) berichteten bei der Vorstellung ihres psychometrischen Testverfahrens bereits von deutlichen Hinweisen auf einen Geschlechtseffekt zugunsten der Männer, welcher sich bis heute zeigen lässt (Linn & Petersen, 1985, Voyer et al., 1995; Karadi, Csatho, Kovacs & Kosztolanyi, 2003). Verglichen mit Geschlechtseffekten aus anderen kognitiven Bereichen (z.B. der Überlegenheit von Frauen im bestimmten verbalen und dem Vorteil der Männer in mathematischen Bereichen), hat sich dieser Geschlechtseffekt bei Erwachsenen als durchgehend robust und reliabel erwiesen, entsprechende Effektgrößen wurden gefunden (Halpern, 1989; Linn & Hyde, 1989; Halpern, 1992; Masters & Sanders, 1993; Halpern et al., 2007).

Aber nicht jedes Testverfahren zur mentalen Rotationsfähigkeit ruft Geschlechtseffekte mit ähnlich hohen Effektgrößen hervor. Wird der psychometrische MRT (Peters et al., 1995) verwendet, sind die gefundenen Effektstärken am größten, wie in zwei Metaanalysen hervorgehoben wird (Linn & Petersen, 1985; Voyer et al., 1995). Die mittlere Leistung von Männern und Frauen weicht um etwa eine Standardabweichung voneinander ab ($d = 0.9$; siehe Cohen, 1977). Andere Testverfahren zur mentalen Rotationsfähigkeit (z.B. Cards Rotation Test; Ekstrom, French, Harman, 1976) weisen den Geschlechtseffekt zugunsten der Männer zwar gleichermaßen nach, die Effektstärken fallen mit $d = 0.31$ bis $d = 0.49$ jedoch wesentlich geringer aus.

Werden chronometrische Testverfahren, welche die Reize von Shepard und Metzler (1971) nutzen, verwendet, verändert sich das bisher homogene Bild der Geschlechtseffekte. Jansen-Osmann und Heil (2007) fanden keinen Geschlechtseffekt, wenn die Würfelfiguren von Shepard und Metzler (1971) chronometrisch im Paarvergleich präsentiert wurden. Peters (2005) fand in einer messwiederholten Studie, dass keine Geschlechtseffekte auftraten, wenn die Reize des MRT chronometrisch als Paarvergleiche dargeboten wurden. Sie traten jedoch auf (sowohl vor als auch nach der chronometrischen Präsentation), wenn der MRT in der klassischen psychometrischen Version verwendet wurde und die korrekte Identifikation von zwei aus vier Alternativen gefordert wurde. Auch Desrocher, Smith und Taylor (1995) sowie Jordan, Wuestenberg, Heinze, Peters und Jaencke (2002) konnten keinen Geschlechtseffekt in chronometrischen Studien nachweisen. Andererseits fanden Monahan, Harke und Shelley (2008) bei der Präsentation des MRT in einer chronometrischen Version einen reduzierten aber dennoch signifikanten Geschlechtseffekt (d = 0.59). Des Weiteren fanden Parsons et al. (2004) in ihrer chronometrischen Studie einen Geschlechtseffekt.

1.3.2 Geschlechtseffekte hinsichtlich der mentalen Rotationsfähigkeit bei Kindern

Während das Auftreten des Geschlechtseffekts bei Erwachsenen als reliabel angesehen werden kann (zumindest unter Verwendung psychometrischer Verfahren wie dem MRT), tritt der entsprechende Effekt bei Kindern nicht in dieser Einheitlichkeit auf. In Abhängigkeit von Alter, Reizmaterial, Darbietungsform und weiteren experimentellen Faktoren variieren die Befunde stark.

Zunächst vermuteten die Entwicklungspsychologen Piaget und Inhelder (1956), dass Kinder vor dem siebten/achten Lebensjahr nicht in der Lage sind, kinetische Abbildungen mental zu repräsentieren, so dass die Fähigkeit zur mentalen Rotation grundsätzlich noch nicht gegeben ist. Marmor (1975) war daraufhin die erste, welche die mentale Rotationsfähigkeit von Kindern systematisch untersuchte und zeigen konnte, dass bereits Fünf-jährige 2-D Rotationsaufgaben lösen können. Zehn-jährige waren in der Lage auch komplexes, abstraktes Reizmaterial ähnlich schnell und korrekt wie Erwachsene mental zu rotieren (Hardwick, McIntyre & Pick, 1976).

Hinsichtlich des Geschlechtseffekts vermuteten Maccoby und Jacklin (1974), dass die Adoleszenz als kritische Phase zum erstmaligen Auftreten dieses Geschlechtseffekts angesehen werden kann. Die Vermutung, dass es sich hier um ein entwicklungsbedingtes Phänomen handelt, wird auch durch sehr aktuelle Studien gestützt, bei denen keine Geschlechtseffekte in der Untersuchung präadoleszenter Kinder gefunden werden konnten: Hahn, Jansen und Heil (in press) konnten in einer chronometrischen Studie mit Vorschulkindern, in welcher Abbildungen von Buchstaben im Paarvergleich als Reizmaterial dargeboten worden sind, keinen Geschlechtseffekt – weder hinsichtlich der Reaktionszeit noch bzgl. der Anzahl korrekt gelöster Items – finden. Karadi, Szabo, Szepesi, Kallai und Kovacs (2001) fanden in einem psychometrischen Verfahren keinen Geschlechtseffekt bei Kindern im Alter von acht bis elf Jahren. Als Reizmaterial wurden Abbildungen von Händen benutzt.

Abweichend zu Maccoby und Jacklin’s Hypothese jedoch konnten Pezaris und Casey (1991) keinen Geschlechtseffekt bei 13 bis 15-Jährigen finden, obwohl ein psychometrisches Testverfahren genutzt wurde, welches die Reize von Shepard und Metzler (1971) verwandte und ähnlich dem MRT konstruiert war.

Dennoch weist die Mehrzahl der Studien auf das Auftreten von Geschlechtseffekten in verschiedenen experimentellen Settings bereits vor dem Eintritt in die Adoleszenz hin. Erste Habituationsstudien lieferten sogar Evidenz dafür, dass bereits männliche Säuglinge im Alter von fünf Monaten eine bessere Rotationsleistung zeigen als gleichaltrige Mädchen (Moore & Johnson, 2008; Quinn & Lyben, 2008). Hahn, Jansen und Heil (2010) fanden in einer weiteren chronometrischen Untersuchung von Vorschulkindern einen Geschlechtseffekt zugunsten der Jungen hinsichtlich der Anzahl korrekt gelöster Items, wenn Abbildungen von Tieren als Reizmaterial verwendet wurden.

Psychometrische Studien hingegen weisen auf das Auftreten des Geschlechtseffekts erst im späten Grundschulalter hin (z.B. Johnson & Meade, 1987). Untersucht wurden Kinder und Jugendliche im Alter von sechs bis 18 Jahren mit Hilfe sieben verschiedener räumlicher psychometrischer Testverfahren, welche sich hinsichtlich der Dimension der Objekte (2D vs. 3D), der Art der Objekte (abstrakt vs. konkret) und des Antwortformats (gleich-ungleich Entscheidungen vs. Auswahl aus bis zu fünf Alternativen) unterschieden. Unter anderem wurden auch die Items von Shepard und Metzler (1971) verwendet. Entsprechend dem Alter der Kinder wurde die

Testbatterie in Bezug auf die Testdauer und die Anzahl der Testtage modifiziert. Geschlechtseffekte zugunsten von Jungen wurden ab dem zehnten Lebensjahr reliabel für Rotationsaufgaben in der Bildebene gefunden. Die dreidimensionalen Würfelfiguren von Shepard und Metzler wurden hingegen als wenig aussagekräftig vor dem Erreichen des fünften Schuljahres bewertet.

Des Weiteren fanden Kerns und Berenbaum (1991) Geschlechtseffekte bei neun bis 13-jährigen Kindern in vier verschiedenen psychometrischen visuell-räumlichen Testverfahren (dabei wurden in einem Testverfahren Items aus dem MRT in Form von realen Objektnachbildungen als Paarvergleiche verwendet), wobei die Effektstärken im mittleren Bereich (d = 0.30 bis 0.65) lagen. Aus diesem Befund wurde abgeleitet, dass bereits bei präadoleszenten Kindern der klassische Geschlechtseffekt zugunsten der Jungen gefunden werden kann, wenn kindgerechte, adäquate Testverfahren genutzt werden.

Hinsichtlich des MRT stellten Linn und Petersen (1985) fest, dass dieses psychometrische Testverfahren – welches bei Erwachsenen für die stabilsten und reliabelsten Effekte sorgt – bisher nicht bei Kindern angewandt wurde, welche jünger als 13 Jahre waren, da die kognitive Beanspruchung als zu hoch angesehen wurde. Stattdessen wurden alternative Testverfahren eingesetzt, die – angewandt bei Erwachsenen – deutlich geringere Effektstärken zeigten als der MRT selbst. Dennoch wurde vermutet: „Sex differences are detected as soon as mental rotation could be measured.” (Linn & Petersen, 1985, S. 1488). Gleichzeitig wiesen Vandenberg und Kuse (1978) darauf hin, dass mit ihrem psychometrischen Testverfahren, welches die Grundlage für den MRT bildete, auch Grundschulkinder untersucht worden sind und das Testverfahren meist innerhalb von zehn Minuten bearbeiten konnten. Es wurden jedoch keine Angaben über die Leistungen oder gar Geschlechtseffekte präsentiert. Es erfolgte lediglich der Hinweis, dass das Testverfahren als geeignet angesehen wird, die Entwicklung der visuell-räumlichen Fähigkeiten und des Geschlechtseffekts zu untersuchen.

Aufgrund der vorgefundenen Methodenvielfalt, dem Fehlen aussagekräftiger Längsschnittstudien und den untersuchten unterschiedlichen Altersstufen der Kinder ist eine einheitliche Interpretation vorliegender Geschlechtseffekte nicht möglich.

1.4 Erklärungsansätze für Geschlechtsunterschiede

Die berichteten Effektstärken des Geschlechtseffekts zugunsten der Männer wirken beeindruckend und erwiesen sich in der Vergangenheit als sehr stabil. Während Geschlechtseffekte aus anderen kognitiven Bereichen im Verlauf der Jahre abnahmen, blieb der Geschlechtseffekt hinsichtlich der mentalen Rotation unverändert (Feingold, 1988; Masters & Sanders, 1993).

Zahlreiche Erklärungsansätze aus unterschiedlichen Forschungsbereichen werden angeboten, um das Auftreten dieses Geschlechtseffekts zu erklären. Insbesondere biologisch-neuronale Faktoren nehmen eine zentrale Rolle bei der Erklärung der Entstehung des Geschlechtseffekts ein. Faktoren wie zerebrale Lateralisierung und Asymmetrien (Hugdahl, Thomes & Ersland, 2006; Schöning et al., 2007), genetische Komponenten (McGee, 1979), Reifungsprozesse (Newcombe, Bandura, & Taylor, 1983; Sanders & Soares, 1986) und hormonelle Veränderungen (z.B. Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis & Güntürkün, 2000; Hines et al., 2003; Imperato-McGinley, Pichardo, Gautier, Voyer, & Bryden, 1991) beeinflussen die Performanz von Männern und Frauen in der mentalen Rotation. Zusätzlich liefern auch Einflüsse wie Übung/ Erfahrung oder das Alter (Alington, Leaf, & Monahagan, 1992; Geiser, Lehmann & Eid, 2008; Lizzaraga & Ganuza, 2003; Terlecki & Newcombe, 2005; Voyer, Nolan, & Voyer, 2000) und psychosoziale Faktoren (z.B. Shih, Pittinsky, & Ambady, 1999; Quaiser-Pohl & Lehmann, 2002; Moè & Pazzaglia, 2006) jeweils ein gewisses Maß an empirischer Evidenz, so dass von einem multifaktoriellen „bio-psycho-sozialen“ Erklärungsansatz ausgegangen werden kann. Asendorpf (2003) erklärte: „Psychologische Geschlechtsunterschiede beruhen auf einer durch Geschlechtsstereotypisierung bedingten kulturellen Verstärkung genetisch und ökologisch bedingter Geschlechtsunterschiede auf hormoneller, neuronaler und Verhaltensebene.“ (Asendorpf, 2003, S. 411).

Ein weiterer Ansatz zur Erklärung der Geschlechtseffekte, welche speziell bei der Verwendung des MRT auftreten, fokussiert auf diejenigen testspezifischen Eigenschaften des MRT, welche einen modulierenden Effekt auf den Geschlechtseffekt haben könnten. So geht Voyer (1997) davon aus, dass testimmanente Faktoren einen wesentlichen Einfluss auf die Größe des auftretenden Geschlechtseffekts haben. Stumpf (1993) weist darauf hin, dass testspezifische Faktoren in der Lage sind, Geschlechtseffekte zu minimieren ohne sie vollständig aufzuheben.

Trotz zahlreicher Erklärungsansätze ist die Forschung dennoch weit davon entfernt, die Ursachen für den Geschlechtseffekt vollständig aufzuklären. Uneinheitliche Befunde v.a durch eine unsystematische Methodenvielfalt erschweren die Aufklärung zusätzlich.

Im Folgenden wird auf die psycho-sozialen Faktoren sowie die testspezifischen Faktoren eingegangen. Die Bedeutung der biologischen Aspekte soll damit ausdrücklich nicht in Abrede gestellt werden.

1.4.1 Psycho-soziale Einflussfaktoren

Auf den möglichen Einfluss von Geschlechterrollenerwartungen beim Lösen visuell-räumlicher Faktoren wiesen bereits Vandenberg und Kuse (1978) hin. Nach Ruble und Martin (1998) ähneln die Inhalte von Geschlechtsstereotypien bereits bei zehn- bis elfjährigen Kindern denen von Erwachsenen. Kinder sind dann auch bereit, ihre eigenen Erwartungen und Stereotype zugunsten allgemein gängiger Stereotypien aufzugeben (Eccles, Wigfield & Schiefele, 1998). Nash (1979) stellte außerdem fest, dass bereits bei Kindern ab dem zweiten Schuljahr davon ausgegangen werden kann, dass sozial-verbale sowie künstlerische Fähigkeiten als weibliche und räumliche, mechanische sowie athletische Fähigkeiten als männliche angesehen werden. Diese geschlechtsbezogenen Vorurteile ebnen den Weg für kognitive Leistungseinbußen in unterschiedlichsten Bereichen. So geht die sog. „Stereotype Threat Theory“ von Steele und Aronson (1995) davon aus, dass Probanden einen Leistungsabfall beim Lösen kognitiver Aufgaben allein aufgrund der Erwartung zeigen, ein bestehendes negatives Vorurteil (z.B. „Frauen sind schlecht in visuell-räumlichen Aufgaben") über eine bestimmte Gruppe (z.B. Frauen), zu welcher sie gehören, zu bestätigen. Gleichzeitig wurde gefunden, dass Frauen empfänglich sind, Defizite bei Lösen schwieriger Mathematikaufgaben zu zeigen, in Abhängigkeit von der Anzahl der Männer, welche in der Testsituation anwesend waren.

Speziell im Bereich der Mentalen Rotationsleistung zeigen aktuelle Befunde (Moè & Pazzaglia, 2006, Moè, 2009), dass Frauen ihre Leistung im MRT signifikant steigern können, wenn ihre Einstellung gegenüber den eigenen Fähigkeiten beim Lösen eines visuell-räumlichen Tests durch eine Manipulation der Testinstruktionen positiv beeinflusst wird. Frauen, denen gesagt wurde, dass sie in dieser Aufgabe besser abschneiden würden als Männer, erbrachten eine signifikant bessere Leistung als Frauen, die davon ausgingen, dass Männer diesen Aufgabentyp besser bewältigen würden (dieser Effekt wurde umgekehrt auch bei Männern gefunden). Ähnliche Befunde konnten auch Quaiser-Pohl und Lehmann (2002) zeigen: Frauen, welche ein starkes leistungsbezogenes Selbstkonzept hinsichtlich der eigenen Fähigkeiten zur Bewältigung visuell-räumlicher Aufgaben zeigten, erbrachten signifikant bessere Leistungen, als Frauen mit einem schwachen Selbstkonzept.

Bisher liegen keine Daten zum Einfluss psychosozialer Faktoren auf die mentale Rotationsleistung von Kindern und Jugendlichen vor. Die Relevanz psychosozialer Faktoren kann in Bezug auf Kinder und Jugendliche jedoch aus Befunden zu Leistungsunterschieden im schulischen Bereich geschlossen werden. Hinsichtlich der mathematischen Leistungen von Kindern und Erwachsenen (Ambady, Shih, Kim & Pittinski, 2001; Muzzatti und Agnoli, 2007; Neuville & Croizet, 2007; Shih, Pittinski & Ambady, 1999) besteht nach wie vor ein Ungleichgewicht zugunsten der Jungen/ Männer. Einflüsse z.B. durch die unterschiedlichen Erwartungen an die Leistung von Kindern und Jugendlichen durch Eltern oder Lehrer (Jacobs & Eccles, 1992) sowie übergeordnete kulturelle Differenzen zwischen den Nationen hinsichtlich der (Un-)

Gleichbehandlung der Geschlechter tragen zur Aufrechterhaltung oder Reduktion dieses Geschlechtseffekts bei (Guiso, Monte, Sapienza & Zingales, 2008). Lippa, Collaer und Peters (2009) konnten für Erwachsene zeigen, dass auch in Bezug auf die mentale Rotation eine Assoziation von Gleichberechtigung und ökonomischer Entwicklung mit einer besseren Leistung einhergeht.

Neben diesen „äußeren“ Einflüssen spielen aber auch bei Kindern und Jugendlichen kognitive Strategien eine Rolle. Befunde der sog. „self-to-prototype-matching“ Hypothese (Niedenthal, Cantor & Kihlstrom, 1985), zeigen, dass ein Proband, welcher zwischen mehreren Optionen auswählen soll, sich zunächst den jeweiligen sog. (klischeehaften) „Prototypen“ vorstellen würde, welcher sich für die eine oder die andere Option entscheiden würde. Die wahrgenommene Ähnlichkeit zwischen einem bestimmten Prototypen und dem Selbstbild des Probanden trägt dann entscheidend dazu bei, welche Entscheidung er treffen wird. Hannover und Kessels (2004) konnten zeigen, dass Schüler diese kognitive Strategie benutzen, um z.B. Schulfächer als „beliebt“ oder „unbeliebt“ einzustufen. So wird nachvollziehbar, warum z.B. Mädchen das Fach Physik insbesondere während der Pubertät ablehnen: Der Prototyp eines Mädchens, welches Physik mag, wird zunächst als weniger attraktiv für Jungen eingestuft. Typischerweise männliche Interessen oder Verhaltensweisen – visuell-räumliche Aufgabentypen eingeschlossen – werden daher abgelehnt. Leistungseinbußen aufgrund fehlender Motivation (siehe auch Dweck, 1986) sind die Folge.

Insgesamt legen die Befunde nahe, dass der Einfluss psycho-sozialer Faktoren nicht unterschätzt werden darf und zeigt die Brisanz um die Frage nach der Relation von tatsächlich vorhandenen visuell-räumlichen Fähigkeiten und einer möglichen Verzerrung der Befundlage durch übergeordnete Stereotype.

1.4.2 Testspezifische Einflussfaktoren (bezogen auf den MRT)

Die testspezifischen Einflussfaktoren können vermutlich als Untergruppe der psycho-sozialen Faktoren betrachtet werden. In einem ersten Schritt erscheint zunächst eine spezifische Charakteristik des Testverfahrens MRT ausschlaggebend für die

Entstehung des Geschlechtseffekts und damit für diesen ursächlich zu sein. In einem zweiten Schritt stellt sich jedoch die Frage, weshalb genau diese Faktoren zwischen Männern und Frauen diskrimieren sollten, die Frage nach dem „warum“ ist noch nicht geklärt. Zum Verständnis bieten sich erneut die psycho-sozialen Erklärungsansätze an.

Als erster spezifischer Einflussfaktor auf die Leistungen von Männern und Frauen im MRT wurde die *Zeitbegrenzung* des Verfahrens diskutiert. Goldstein, Haldane und Mitchell (1990) berichteten, dass der Geschlechtseffekt zugunsten der Männer aufgehoben wird, wenn die Zeitbegrenzung von 3 Minuten pro 12 Items nicht eingesetzt wird (siehe auch Voyer, 1997). Einerseits besteht die Vermutung, dass Frauen sorgfältiger und dadurch langsamer arbeiten (Maccoby & Jacklin, 1974). Andererseits wird angenommen, dass eine – im Vergleich zu Männern – erhöhte Unsicherheit zu einem zögerlicheren Antwortverhalten führen kann (Goldstein et al., 1990). Im Gegensatz zu diesem Befund gibt es jedoch weitaus mehr Daten, die dafür sprechen, dass eine Aufhebung der Zeitbegrenzung keinen Einfluss auf das Auftreten des Geschlechtseffekts hat (Resnick, 1993; Delgado & Prieto, 1996, Masters, 1998). Peters (2005) zeigte, dass – obwohl Frauen grundsätzlich weniger Items bearbeiten als Männer, wenn eine Zeitbegrenzung vorgegeben wird – eine Verdoppelung des Zeitlimits keine Reduktion des Geschlechtseffektes mit sich brachte. Es kann davon ausgegangen werden, dass der Faktor „Zeitbegrenzung“ nicht als die kritische Einflussgröße für den Geschlechtseffekt im MRT angesehen werden kann.

Des Weiteren wurde postuliert, dass der Geschlechtseffekt gleichermaßen nicht auftritt bzw. reduziert wird, wenn bei der *Auswertung* die tatsächlich korrekt gelösten Items zur Anzahl der bearbeiteten Items ins Verhältnis gesetzt werden (Goldstein et al., 1990; Voyer, 1997). Diese Annahme folgt aus dem Befund, dass Frauen signifikant weniger Items bearbeiten als Männer, was ihnen bei der Standardauswertung als Nachteil reicht. Damit könnte der Geschlechteseffekt als künstlich induziert erachtet werden, da nicht unbedingt die Fähigkeiten der Frauen reduziert sind, sondern vielmehr deren Arbeitsgeschwindigkeit. Voyer et al. (1995) bestätigten, dass der Geschlechtseffekt dann größer ist, wenn die konservative Auswertungsmethode für den

MRT benutzt wird (d.h. ein Punkt, wenn beide korrekten Alternativen identifiziert worden sind). Masters (1998) hingegen zeigte, dass der Geschlechtseffekt nicht durch die Methode der Auswertung beeinflusst werden kann.

Der Faktor *Dimensionalität* der Aufgabe (2D vs. 3D und Rotation in der Bildtiefe vs. Rotation in der Bildebene) scheint hingegen nicht ausschlaggebend für die Größe des Geschlechtseffekts zu sein (Collins & Kimura, 1997), wohl aber scheint die *Aufgabenschwierigkeit* modulierend zu wirken. Sie fanden bereits Geschlechtseffekte in der klassischen Größenordnung unter Verwendung von 2D-Reizmaterial in verschiedenen Schwierigkeitsgraden. Weiterhin verglichen Peters et al. (1995) die Leistung von Männern und Frauen, die entweder den MRT in seiner Standardversion bearbeiteten oder aber eine schwierigere Form des MRT lösten. Die Größe des Geschlechtseffekts bei der schwierigen Version reduzierte sich um die Hälfte verglichen mit der Standardversion des MRT, plausible Erklärungen für diesen Befund stehen noch aus.

Als weiterer Faktor kann die *Komplexität der Entscheidung* beim MRT diskutiert werden. Die Aufgabenstellung erfordert von den Probanden den Vergleich von vier Antwortalternativen mit einer Zielfigur. Zwei der vier Alternativen müssen dabei als korrekt identifiziert werden. Bereits Kerkman, Wise und Hardwood (2000) beobachteten jedoch, dass bei Aufgaben, welche einen einfacheren Paarvergleich verlangten (hier wurden allerdings keine Items des MRT verwendet), keine Geschlechtseffekte bei den Durchgängen auftraten, in denen die Reize übereinstimmten. Sie schlossen, dass Frauen grundsätzlich genau so schnell und korrekt rotieren können wie Männer, dass sie jedoch bei Durchgängen, in denen die Reize nicht übereinstimmen, mehr Zeit darauf verwenden diesen Mismatch wirklich abzusichern. Weitere Evidenz für den Einfluss der Komplexität auf den Geschlechtseffekt ergibt sich aus der Diskrepanz zwischen chronometrischen und psychometrischen Testverfahren. Während in psychometrischen Testverfahren der Geschlechtseffekt reliabel nachgewiesen werden konnte (siehe Voyer, Voyer & Bryden, 1995), ist die Befundlange in chronometrische Studien heterogen (siehe Jansen-Osmann und Heil, 2007; Monahan et al., 2008; Peters, 2005).

Zuletzt sind die Arten der im MRT verwendeten *Items* untersucht worden. In der Originalversion werden zum Teil Abbildungen von Items benutzt, bei welchen Überlappungen auftreten. Voyer und Hou (2006) fanden, dass bei diesem Itemtyp der Geschlechtseffekt am größten ist und führten dies auf die Dreidimensionalität des Verfahrens zurück.

Grundsätzlich muss festgestellt werden, dass der tatsächliche Einfluss oben beschriebener Faktoren bisher nur unzureichend aufgeklärt werden konnte (siehe z.B. Voyer & Saunders, 2004), was größtenteils auf eine unübersichtliche Methodenvielfalt zurückzuführen ist. Bisher liegen zudem keine Daten zur Befundlage bei Kindern und/ oder Jugendlichen vor.

1.5 Zentrale Fragestellungen

Die vorliegende Arbeit widmet sich in fünf Einzelexperimenten der Beeinflussbarkeit des Geschlechtseffekts beim Lösen des psychometrischen Testverfahrens MRT durch Manipulation psycho-sozialer und testspezifischer Einflussfaktoren. In der Einleitung wurde gezeigt, dass diese beiden Erklärungsansätze neben den biologisch-neuronalen Erklärungsmodellen eine nicht zu vernachlässigende Rolle einzunehmen scheinen, wenngleich die Datenlage nach wie vor heterogen ist. Insbesondere in Bezug auf die Untersuchung von Kindern erschwert eine große Methodenvielfalt den Vergleich und die Interpretation der Ergebnisse.

Die ersten drei Experimente beschäftigten sich mit den psycho-sozialen Einflussfaktoren auf die mentale Rotationsleistung von Grundschulkindern, Jugendlichen und Erwachsenen.

Dabei befassten sich die ersten beiden Experimente mit der Untersuchung des Geschlechtseffekts bei Viertklässlern erstmalig unter Verwendung des MRT, welcher dafür bekannt ist, bei Erwachsenen große Geschlechtseffekte hervorzubringen (siehe z.B. Voyer et al., 1995). Es wurde einerseits der Frage nachgegangen, inwiefern der MRT bei Kindern, welche jünger als 13 Jahre alt sind, überhaupt eingesetzt werden kann

(Linn & Petersen, 1985). Andererseits sollte geprüft werden, ob es ein bestimmtes Alter gibt (Johnson & Meade, 1987; Maccoby & Jacklin, 1974), ab welchem der Geschlechtseffekt erstmals mit dem MRT nachgewiesen werden kann (Experiment 1).

Weiterhin wurde im zweiten Experiment untersucht, ob Viertklässler ähnlich wie Erwachse auch auf die Manipulation von geschlechtsbezogenen Stereotypien mit Leistungsverbesserungen oder Leistungsverschlechterungen reagieren (Moè & Pazzaglia, 2006).

Da gezeigt werden konnte, dass Schülerinnen einer Mädchenschule im Vergleich zu Schülerinnen einer koedukativen Schule sich weniger vorurteilsbehaftet gegenüber „typisch männlichen“ Schulfächern zeigen und weniger geschlechtsbezogene Interessen aufweisen (Kessels, 2007; Schoon, 2001), wurde der Einfluss einer monoedukativen Schulform auf die Größe des Geschlechtseffekts bei Jugendlichen der achten und 12. Klasse untersucht. Es wurde erwartet, dass Mädchenschülerinnen die Schülerinnen einer koedukativen Schule hinsichtlich der mentalen Rotationsleistung übertreffen würden, selbst wenn die koedukativen Schülerinnen in Abwesenheit der männlichen Klassenkameraden getestet würden. Es sollte zudem überprüft werden, ob sich der klassische Geschlechtseffekt zwischen Jungen und Mädchenschülerinnen zumindest reduzieren lässt.

Die letzten beiden Experimente beschäftigten sich mit den testspezifischen Einflussfaktoren auf das Auftreten von Geschlechtseffekten bei der mentalen Rotation, im Speziellen mit der Komplexität des Antwortformats im MRT. Der Vergleich von Studien, welche chronometrische Testverfahren einsetzten und mit Paarvergleichen der Reize arbeiteten, konnten den klassischen Geschlechtseffekt nicht nachweisen (Jansen-Osmann & Heil, 2007; Peters, 2005). Zunächst wurde daher die Komplexität des MRT mit Hilfe einer Schablone reduziert (Experiment 4). Im Weiteren wurde die Komplexität mit Hilfe von Paarvergleichen in einem neu entworfenen Papier-Bleistift-Test reduziert (Experiment 5).

2 Experimente

Die Auswirkungen verschiedener Manipulationen (testspezifisch vs. psychosozial) auf das Auftreten des Geschlechtseffekts beim Lösen eines klassischen mentalen Rotationstests werden nachfolgend in fünf Einzelexperimenten beschrieben. In dieser Arbeit werden der MRT A und B (Peters et al., 1995) als Grundlage verwendet, welche sich in der Forschung mittlerweile gegenüber seiner Vorgängerversion (Vandenberg & Kuse, 1978) durchgesetzt hat.

2.1 Experiment 1: Der Geschlechtseffekt bei Viertklässlern und Erwachsenen

Im ersten Experiment wurde zunächst überprüft, ob der typischerweise auftretende Geschlechtseffekt bei der Mentalen Rotation zugunsten der Männer unter Verwendung des MRT sich bereits bei Kindern im Alter von 9 und 10 Jahren zeigen lässt.

Viertklässler und Erwachsene bearbeiteten den MRT (Version A oder B) unter regulären Versuchsbedingungen im Rahmen einer Gruppentestung. Zuvor erhielten alle Probanden den „Tiere-MRT“, der speziell für diese Studie in Anlehnung an den MRT entwickelt wurde (siehe Abbildung 3). Dieser Test entspricht vom Aufbau her dem MRT, verwendet jedoch einfachere, zweidimensionale Tierfiguren als Reizmaterial, welche in der Bildebene rotiert wurden. Dieser Test diente dazu, insbesondere die Kinder mit dem Konzept der mentalen Rotation sowie dem Testformat vertraut zu machen.

Abbildung 3 : Diese Abbildung zeigt das erste Item des Tiere-MRT. Zwei der vier Vergleichsfiguren lassen sich durch Rotation im oder gegen den Uhrzeigersinn in die Zielfigur überführen.

Zur Auswertung wurden die Viertklässler hinsichtlich ihres Alters entweder der Gruppe der neunjährigen oder der Gruppe der zehnjährigen Kinder zugeordnet (mittleres Alter der Kindergruppen: 9.3 vs. 10.3 Jahre).

Es zeigte sich, dass sowohl die älteren Kinder als auch die Erwachsenen den klassischen großen Geschlechtseffekt zugunsten des männlichen Geschlechts zeigten, wohingegen bei den jüngeren Kindern kein signifikanter Unterschied gefunden wurde. Allerdings deutete die Effektstärke auf das Vorliegen zumindest eines kleinen bis mittleren Effekts hin (siehe Tabelle 1).

Tabelle 1 :
Diese Tabelle zeigt die mittlere Anzahl korrekt beantworteter Items und die entsprechenden Standardfehler für den MRT als Funktion von Alter und Geschlecht. Zusätzlich zeigt Cohen’s *d* die Effektstärke des jeweiligen Geschlechtseffekts innerhalb der Gruppen und insgesamt an.

Anzahl korrekter Items

Gruppe	M	SE	Cohen's *d*
jüngere Kinder	6.5	.53	0.45
Mädchen	5.67	.62	
Jungen	7.33	.85	
ältere Kinder	8.27	.61	1.45
Mädchen	5.75	.52	
Jungen	10.79	.85	
Erwachsene	10.85	.75	1.24
Frauen	8.13	.70	
Männer	13.58	1.06	
Gesamt	8.54	.39	0.94
weiblich	6.51	.38	
männlich	10.57	.61	

Zudem konnte gezeigt werden, dass zwischen neunjährigen Jungen und erwachsenen Männern ein signifikanter Unterschied hinsichtlich der Leistung besteht. Dies legt die Vermutung nahe, dass sich die mentale Rotationsleistung mit zunehmendem Alter verbessert. Für das weibliche Geschlecht konnte eine derartige Feststellung nicht getroffen werden: Zwischen Mädchen und Frauen zeigten sich keinerlei signifikante Leistungsunterschiede – Gründe für dieses Ungleichgewicht sind bisher nicht bekannt.

Entgegen der Vermutung von Linn und Petersen (1985), die erforderliche Aufmerksamkeitsspanne zur Bewältigung des MRT sei für Kinder unter 13 Jahren zu hoch, ergaben sich in dieser Studie diesbezüglich keine Hinweise. Zwischen der Anzahl korrekt beantworteter Items beider Sets des MRT ergab sich ein signifikanter Unterschied. Dieser erwies sich jedoch als unabhängig von Alter und Geschlecht der Probanden, d.h. auch die erwachsenen Teilnehmer beantworteten im zweiten Set signifikant weniger Items richtig als im ersten Set.

Dieses Ergebnis unterstützt die These von Johnson und Meade (1987), dass sich der Geschlechtseffekt zugunsten der Jungen bereits ab einem Alter von zehn Jahren manifestiert und widerspricht der Annahme von Maccoby und Jacklin (1974), dass derartige Effekte erstmalig in der Pubertät auftreten. Das Auftreten des Effekts kann dennoch bisher nicht hinreichend erklärt werden: Biologische und/oder hormonelle Aspekte allein scheinen nicht ausschlaggebend zu sein, wenngleich hier der Einfluss präpubertärer Hormone nicht ausgeschlossen werden kann. Insbesondere sozial-kognitive Erklärungsansätze – z.B. in Form von Geschlechtsstereotypen - könnten eine ebenso wichtige Rolle spielen.

2.2 Experiment 2: Einfluss von Geschlechtsstereotypen bei Kindern

Im zweiten Experiment wurde der Einfluss von Geschlechtsstereotypien auf die Leistung von Viertklässlern beim Lösen des MRT untersucht. Wohingegen bei Erwachsenen die Wirkung einer Manipulation dieser Stereotype bereits im Hinblick auf die visuell-räumliche Leistung nachgewiesen werden konnte (siehe Moè & Pazzaglia, 2006; Moè, 2009), steht ein entsprechender Nachweis für Kinder noch aus.

Viertklässler lösten im Rahmen einer Gruppentestung zunächst unvoreingenommen 12 Items einer speziell für diese Untersuchung zusammengesetzten

Parallelversion des MRT. Anschließend wurden zwei unterschiedliche Geschlechtsstereotype induziert. Einem Teil der Kinder wurde erklärt, dass bei diesem Test stets die Jungen besser abschneiden würden als die Mädchen. Dem anderen Teil wurde das Gegenteil berichtet. Eine Kontrollgruppe erhielt geschlechtsneutrale Informationen zum Test. Daraufhin bearbeiteten die Kinder 12 weitere Items des MRT.

Entgegen der Erwartung zeigte sich, dass die Leistung der Kinder nicht durch die Manipulation beeinflusst worden war (siehe Tabelle 2 und 3).

Tabelle 2:
Die Tabelle zeigt die Anzahl korrekt gelöster Items (Mittelwerte und Standardabweichungen) für die Jungen vor und nach der experimentellen Manipulation in Abhängigkeit von der jeweiligen Instruktion. Cohen’s *d* – als Maß für die Effektstärke – wird angegeben.

	Anzahl korrekt gelöster Items		
Instruktion	vor der Instruktion	nach der Instruktion	**Cohen's *d***
Jungen besser			
Mittel	3.83	4.76	0.34
S.D.	2.41	3.00	
Mädchen besser			
Mittel	2.86	4.12	0.29
S.D.	1.85	2.51	
neutral			
Mittel	3.19	4.00	0.36
S.D.	2.13	2.32	

Hinweis: Maximale Punktzahl = 12.

Es zeigte sich weder eine Annäherung der Leistung von Jungen und Mädchen noch zeigten die Kinder, welche eine positive Manipulation erhalten hatten, eine bessere Leistung als die Kinder gleichen Geschlechts, welche eine negative Manipulation erhalten hatten. Der klassische Geschlechtseffekt zugunsten der Jungen hingegen trat auf – wenngleich die Effektstärke d = 0.39 verhältnismäßig niedrig ausfiel. Gleichzeitig zeigte sich, dass eine simple Testwiederholung (auch mit unterschiedlichen Items) zu einer Verbesserung der Leistung, sowohl bei Jungen als auch bei Mädchen, führte.

Tabelle 3:
Die Tabelle zeigt die Anzahl korrekt gelöster Items (Mittelwerte und Standardabweichungen) für die Mädchen vor und nach der experimentellen Manipulation in Abhängigkeit von der jeweiligen Instruktion. Cohen's *d* – als Maß für die Effektstärke – wird angegeben.

	Anzahl korrekt gelöster Items		
Instruktion	vor der Instruktion	nach der Instruktion	**Cohen's *d***
Jungen besser			
Mittel	2.36	3.93	0.78
S.D.	1.51	2.41	
Mädchen besser			
Mittel	2.05	3.26	0.56
S.D.	1.79	2.45	
neutral			
Mittel	2.71	3.17	0.23
S.D.	1.68	2.24	

Hinweis: Maximale Punktzahl = 12.

Erklärungen für diesen Befund sind weitgehend unklar. Methodische Probleme oder aber eine Überlagerung des Effekts durch andere Prozesse wie beispielsweise Reaktanz (siehe Brehm, 1966) oder „stereotype stratification“ (Steele, 2003) sind denkbar.

2.3 Experiment 3: Einfluss der Monoedukation

Da sowohl eine Reduktion des Geschlechtseffekts zwischen den Jungen und Mädchen als auch eine erwartete Verbesserung der Leistung der Mädchen, welche eine positive Manipulation erhalten hatten, im Vergleich zu den Mädchen, welche eine negative Manipulation erhalten hatten, scheiterte, stellte sich die Frage inwiefern andere Faktoren modulierend dazu beitragen können, die Leistung der Mädchen zu verbessern.

Da Inzicht und Ben-Zeev (2000) zeigten, dass Frauen in der Gegenwart von Männern dazu neigen, Defizite in ihrer Problemlösefähigkeit zu entwickeln, wurde in einem dritten Experiment der Frage nachgegangen, ob Mädchen, welche eine Mädchenschule besuchen, im Vergleich zu Mädchen einer koedukativen Schule eine bessere Leistung beim Lösen des MRT zeigen. Es wurden Acht- und Zwölftklässler aus beiden Schulformen untersucht. Die Jugendlichen der koedukativen Schule wurden dabei getrennt nach Geschlecht getestet.

Die erwarteten Unterschiede zeigten sich in der 12. Klasse (siehe Tabelle 4).

Tabelle 4:
Die Tabelle zeigt die Mittelwerte und Standardabweichungen der Anzahl korrekt beantworteter Items im MRT in Abhängigkeit von Alter, Geschlecht und Schulform.

			MRT	
		N	*M*	*SD*
Mädchenschule	8. Klässlerinnen	42	8.45	4.27
Koedukative Schule	8. Klässler	42	10.38	4.24
	8. Klässlerinnen	42	9.40	4.66
Mädchenschule	12. Klässlerinnen	42	11.00	4.54
Koedukative Schule	12.Klässler	42	12.90	4.62
	12. Klässlerinnen	42	8.64	3.53

Schülerinnen der Mädchenschule erbrachten signifikant bessere Leistungen als ihre Kolleginnen der koedukativen Schule (d = 0.58). Zwischen den Jungen und den Mädchen der 12. Klasse zeigte sich der klassische große Geschlechtseffekt (d = 1.04) zugunsten der Jungen. Zwischen den Jungen und den Schülerinnen der Mädchenschule zeigte sich kein signifikanter Unterschied, wenngleich die Effektstärke auf einen mittleren Effekt zugunsten der Jungen hindeutet (d = 0.41). Bezüglich der Achtklässler zeigte sich ein signifikanter Unterschied nur zwischen der Leistung der Jungen und der Mädchen der Mädchenschule (d = 0.45). Zwischen den Mädchengruppen beider Schulformen zeigten sich keine Unterschiede (d = 0.21).

Die reine Abwesenheit der Jungen während der Testphase zeigte hier- im Gegensatz zur Befundlage bei den Erwachsenen – nicht dieselbe leistungsfördernde Wirkung. Die langjährige Monoedukation könnte hingegen spätestens in der Adoleszenz zu einer Leistungssteigerung beitragen.

Diese Ergebnisse unterstützen die Vermutung, dass psycho-soziale Faktoren bei der Entstehung des Geschlechtseffekts eine nicht zu unterschätzende Rolle spielen und eine Modulierung des Effekts bewirken können.

2.4 Experiment 4: Komplexitätsreduktion mit Hilfe einer Schablone

Dieses Experiment untersucht den Einfluss des Faktors *Komplexität* auf die Leistung Erwachsener beim Lösen des MRT in zwei verschiedenen Versionen. Auf die Untersuchung von Kindern wurde an dieser Stelle verzichtet, um den Einfluss weiterer moderierender Variablen zunächst auszuschließen.

Das Experiment basiert auf dem Befund, dass das Lösen einer chronometrischen mentalen Rotationsaufgabe, bei welcher paarweise die klassischen Shepard und Metzler (1971) Reize am Computer präsentiert wurden, keine Geschlechtseffekte zugunsten der Männer gefunden wurden (Peters, 2005; Jansen-Osmann & Heil, 2007). Diese Geschlechtsunterschiede treten in der psychometrischen Version des MRT hingegen sehr zuverlässig auf (Voyer, Voyer & Bryden, 1995).

Die eine Hälfte der Probanden löste den MRT in seiner Originalversion. Auf die im Originaltest vorgesehene sechsminütige Zeitbegrenzung wurde verzichtet. Die andere Hälfte der Probanden löste den MRT mit Hilfe einer speziellen Schablone, welche lediglich den Zielreiz sichtbar machte und je einen der vier Auswahlreize. Pro Item musste der Proband also vier einzelne Entscheidungen treffen. Zu jedem Zeitpunkt waren stets nur zwei Reize – wie in den chronometrischen Aufgaben üblich – sichtbar.

Es zeigte sich, dass die durchgeführte Komplexitätsreduktion (Vereinfachung des Antwortformats und Wegfall der Zeitbegrenzung) nicht zu einem Ausbleiben des Geschlechtseffekts führte. Männer zeigten in beiden Experimentalgruppen eine signifikant bessere Leistung als Frauen – sowohl in Bezug auf die Antwortgenauigkeit als auch auf die Reaktionszeit (siehe Tabelle 5).

Tabelle 5:
Die Tabelle zeigt die Mittelwerte und Standardabweichungen der Anzahl korrekt beantworteter Items im MRT in Abhängigkeit von Geschlecht und Experimentalgruppe.

Geschlecht	Komplexität	Mittelwert	S.D.
Männer (N = 150)	Mit Schablone (N = 75)	21.21	.54
	Ohne Schablone (N = 75)	20.76	.50
	Gesamt	20.99	.37
Frauen (N = 150)	Mit Schablone (N = 75)	19.01	.56
	Ohne Schablone (N = 75)	18.91	.59
	Gesamt	18.99	.40
Gesamt (N = 300)	Mit Schablone (N = 150)	20.14	.40
	Ohne Schablone (N= 149)	19.83	.39
	Gesamt	19.98	.28

Zusammenfassend lässt sich der Faktor *Komplexität* – in der hier realisierten Form – als einem wesentlichen testspezifischen Faktor, welcher an der Entstehung des klassischen Geschlechtseffekts im MRT beteiligt sein könnte, ausschließen.

2.5 Experiment 5: Komplexitätsreduktion durch direkte Paarvergleiche

Das letzte Experiment beschäftigt sich erneut mit dem testspezifischen Faktor *Komplexität*. Anstatt der klassischen psychometrischen Version, in welcher zwei von vier Auswahlreize als korrekt identifiziert werden müssen, mussten die Probanden hier einfachere Paarvergleiche auf dem Papier lösen (gleich/ ungleich). Die Originalitems des MRT wurden dazu in vier Paarvergleiche verwandelt und zufällig angeordnet. Es wurde kein Zeitlimit vorgegeben, die jeweils benötigte Gesamtzeit wurde jedoch notiert. Zur Auswertung wurden zwei Methoden verwandt. Einerseits wurde ein Punkt für jede korrekte „Gleich-Ungleich-Entscheidung“ vergeben. Andererseits wurde die klassische, konservative Methode verwandt: Ein Punkt wurde nur dann vergeben, wenn beide zu einem Items gehörenden Auswahlreize korrekt identifiziert worden sind.

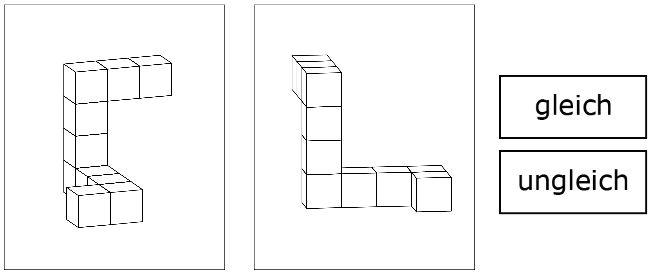


Abbildung 4: Zeigt Item 1 der neu entworfenen Version der Paarvergleiche des MRT. Von den Probanden wird eine „gleich-ungleich“ Entscheidung verlangt.

Zwischen beiden Auswertungsmethoden zeigte sich jedoch kein statistischer Unterschied, Männer zeigten eine signifikant bessere Leistung als Frauen und arbeiteten zudem signifikant schneller. Erneut konnte kein Einfluss des testspezifischen Faktors *Komplexität* zur Reduktion des Geschlechtseffekts nachgewiesen werden, wie dies in chronometrischen Studien der Fall ist.

3 Allgemeine Diskussion

Die vorliegenden Einzelexperimente geben Hinweise auf die Einflussmöglichkeiten und die Einschränkungen psycho-sozialer und testspezifischer Faktoren zur Aufklärung des Geschlechtseffekts, welcher reliabel durch den psychometrischen „Mental-Rotations Test“ MRT hervorgerufen wird.

So konnte in zwei Experimenten gezeigt werden, dass der testspezifische Faktor *Komplexität* (Vereinfachung der Entscheidung und Wegfall der Zeitbegrenzung) keinen Einfluss auf die Größe des auftretenden Geschlechtseffekts zugunsten der Männer hat. Diese Annahme wurde durch Befunde aus chronometrischen Studien abgeleitet, in denen gezeigt wurde, dass der Geschlechtseffekt nicht auftritt, wenn die Reize paarweise am Computer präsentiert wurden (siehe z.B. Peters, 2005). Zeitgleich konnten kein Geschlechtseffekt in den Reaktionszeiten gefunden werden. Eine Reduktion der zu treffenden Entscheidung von einer Auswahl zweier Reize aus vier Alternativen zu einer two-alternative-forced choice Aufgabe brachte weder mit Hilfe einer Schablone (Experiment 1) noch mittels direkten Paarvergleichs (Experiment 2) die erwartete Minimierung des Geschlechteseffekts. Die Diskrepanz der Befunde aus chronometrischen und psychometrischen Studien ist somit nicht auf die Entscheidungskomplexität in der hier dargestellten Weise zurückzuführen. Die Vermutung, dass Frauen sich einerseits durch die enge Zeitvorgabe und andererseits durch die Komplexität der zu treffenden Entscheidung in ihrem Antwortverhalten von den Männern unterscheiden, ist so nicht haltbar.

Drei weitere Experimente befassten sich mit der Untersuchung der mentalen Rotationsleistung von Kindern, Jugendlichen und Erwachsenen. Die gewählten experimentellen Manipulationen bezogen sich dabei auf psycho-soziale Faktoren der Beeinflussbarkeit des Geschlechtseffekts durch Alter, Instruktionen und Schulform.

Das erste der drei Experimente zeigte, dass bereits Kinder im Alter von zehn Jahren den Geschlechtseffekt zugunsten der Jungen auf demselben hohen Niveau wie Erwachsene zeigen, wenn der MRT benutzt wird – was meines Wissen bis zu diesem

Zeitpunkt in der Form noch nicht untersucht worden ist. Neun-jährige Kinder hingegen zeigten keinen signifikanten Geschlechtseffekt, wenngleich die Effektstärke auf einen mittleren Effekt hinwies. Weiterhin zeigte sich, dass die Leistung der weiblichen Teilnehmer in den verschiedenen Altersklassen stagnierte, während bei den männlichen Teilnehmern ein kontinuierlicher Zuwachs mit zunehmendem Alter beobachtet werden konnte. Dieser Befund macht auf zweierlei Aspekte aufmerksam. Zum einen kann der Geschlechtseffekt zugunsten des männlichen Geschlechts bereits bei Kindern vor dem Erreichen der Adoleszenz (Maccoby & Jacklin, 1987) nachgewiesen werden. Die frühe Annahme von Linn und Petersen (1985), dass Geschlechtseffekte sich nachweisen lassen, sobald die Fähigkeit der mentalen Rotation gemessen werden kann, scheint bestätigt. Das Fehlen des Geschlechtseffekts bei den neun-Jährigen ist dabei möglicherweise auf die hohe Aufgabenschwierigkeit zurückzuführen, die zu einer Reduktion des Geschlechtseffekts beigetragen haben könnte (siehe Peters et al., 1995). Zum anderen zeigte sich, dass weibliche Teilnehmerinnen sich nicht in demselben Maße mit zunehmendem Alter verbessern wie männliche Teilnehmer. Es wird einschränkend jedoch darauf hingewiesen, dass es sich hier nicht um eine Längsschnitt-, sondern um eine Querschnittstudie handelte.

Orientiert man sich bei der Interpretation der Befunde zum Geschlechtseffekt bei Grundschulkindern an psychosozialen Erklärungsansätzen aus dem Forschungsbereich der Erwachsenen, scheinen Konzepte wie das *Selbstkonzept* oder *Geschlechtsstereotype* nahe liegend (Steele, 1997; Quaiser-Pohl & Lehmann, 2002), denn es konnte gezeigt werden, dass beispielsweise bei der kindlichen Bewertung schulischer Anforderungen diese Konzepte bereits eine Rolle spielen (siehe z.B. Eccles, Wigfield & Schiefele, 1998; Nash, 1979; Ruble & Martin, 1998). Entgegen der entsprechenden Befunde bei Erwachsenen jedoch, konnte bei Kindern im zweiten Experiment kein Einfluss der Manipulation des Geschlechtsstereotypes auf die Leistung im MRT gezeigt werden, der bekannte Geschlechtseffekt blieb bestehen. Im Widerspruch dazu stehen Befunde aus dem mathematischen Bereich (siehe Ambady et al., 2001): Hier beeinflusste die Manipulation die Leistung der Kinder, ohne dass gleichzeitig Geschlechtseffekte aufgetreten sind. Möglicherweise handelt es sich bei diesem Befund um einen Effekt der Vertrautheit des Testmaterials und damit gleichzeitig um ein Problem der Akzeptanz der Manipulation. Grundschüler haben bereits mehrjährige Erfahrungen mit mathematischen Tests und können hier auf einen eigenen Erfahrungsschatz zurückgreifen: sie können einschätzen, inwiefern sie in der Vergangenheit Tests besser oder aber schlechter als das Gegengeschlecht gelöst haben. Bei der mentalen Rotation handelt es sich um ein überwiegend unvertrautes Konzept. Die eigene Erfahrung im Umgang mit dem Testverfahren fehlte, die vorgegebene Manipulation des Geschlechtsstereotyps mag bei vielen Kindern nicht Zustimmung sondern vielmehr Reaktanz (Brehm, 1966) oder aber „Wettbewerbsgedanken“ ausgelöst haben. Tendenziell schienen Mädchen sogar eher von der Manipulation, dass Jungen besser abschneiden würden, zu profitieren, als von der Bekräftigung der Leistung des eigenen Geschlechts. Wenngleich der erwartete Effekt nicht gezeigt werden konnte, wird der Einfluss sozialer Manipulationen deutlich und sollte nicht unterschätzt werden.

Im Gegensatz zu der relativ kurzfristigen, experimentellen konkreten Manipulation des Geschlechtsstereotyps in Bezug auf die mentale Rotationsleistung – welche hier keine Reduktion des Geschlechtseffekts bewirkte – stellte sich die Frage, ob es bestehende soziale Settings gibt, in welchen sich Jungen und Mädchen bzgl. der zuvor diskutierten psycho-sozialen Faktoren (z.B. Selbstkonzept, Geschlechtsstereotype) bereits unabhängig von einer experimentellen Manipulation unterscheiden. Befunde aus monoedukativen (Mädchen-)Schulen weisen darauf hin, dass Mädchen sich weniger geschlechtsstereotypisiert verhalten und typisch männlichen Schulfächern offener gegenüberstehen. Die Untersuchung der mentalen Rotationsleistung im Vergleich koedukativer Schüler/innen und monoedukativer Schülerinnen zeigte einerseits eine signifikant bessere Leistung der Mädchenschülerinnen in der zwölften Klasse (nicht jedoch in der achten Klasse), verglichen mit den Schülerinnen der koedukativen Schule und deutet auf den möglichen Einfluss der Schulform und dem daraus resultierenden positiven Einfluss auf das Selbstkonzept und/oder die vorherrschenden Geschlechtsstereotype hin. Weiterhin weist die Reduktion des Geschlechtseffekts zwischen den koeduaktiven Schülern und den monoedukativen Schülerinnen auf eine halbe Standardabweichung (im Vergleich zu der sonst üblichen ganzen Standardabweichung) auf eine Modifizierbarkeit des Geschlechtseffekts hin. Zusätzlich konnte gefunden werden, – entgegen der Ergebnisse des ersten Experiments – dass monoedukative Schülerinnen und koedukative Schüler sich im querschnittlichen Vergleich von der achten zur zwölften Klasse gleichmäßig verbesserten. Dieses Experiment betont erneut den nicht zu vernachlässigenden Einfluss psycho-sozialer Faktoren. Dabei wurde zudem gezeigt, dass kurzfristige Interventionen – hier die Trennung der Geschlechter in der Testsituation an der koedukativen Schule – nicht denselben Effekt brachte wie die (langfristige) Monoedukation (siehe dazu Inzlicht & Ben-Zeev, 2000).

Abschließend wird darauf hingewiesen, dass – trotz zahlreicher Forschungsarbeiten – zum jetzigen Zeitpunkt weder der Einfluss der testspezifischen noch der psycho-sozialen Faktoren hinreichend verstanden und nachvollzogen werden konnten, um eindeutige und verallgemeinerbare Aussagen zu den Möglichkeiten zur Reduktion des Geschlechtseffekts geben zu können.

3.1 Ausblick

Die Diskussion um den Einfluss testspezifischer Faktoren auf den Geschlechtseffekt des MRT ist nach wie vor nicht abgeschlossen – systematische, vergleichbare Studien fehlen bisher. Die Komponenten Zeitbegrenzung, Auswertung, Dimensionalität sowie Komplexität der Entscheidung scheinen in den bisher untersuchten Formen keinen Beitrag zur Aufklärung des Effekts zu leisten, Frauen scheinen in ihrer Leistungsfähigkeit hierdurch nicht beeinträchtigt zu sein (Collins & Kimura, 1997; Goldstein, Haldane und Mitchell, 1990; Voyer, 1997). Dennoch weist die Diskrepanz zwischen chronometrischen und psychometrischen Befunden auf weiteren Forschungsbedarf hin. Während in dieser Arbeit versucht wurde, die Diskrepanz ansetzend an dem psychometrischen Testverfahren MRT aufzuklären, könnten künftige Arbeiten untersuchen, welche Modifikationen notwendig sind, um den

Geschlechtseffekt in der entsprechenden chronometrischen Testversion zu induzieren. Zusätzlich sollte der Auswahl der verwendeten Reize in den zu vergleichenden chronometrischen und psychometrischen Testversion selbst mehr Beachtung geschenkt werden (siehe Jansen-Osmann, 2007; Voyer & Hou, 2006), um die bestehenden Unterschiede zwischen beiden Verfahren zu reduzieren und so den entscheidenden Faktor, welcher die Diskrepanz des Geschlechtseffekts verursacht, aufzudecken.

Schwieriger als die Untersuchung der testspezifischen Faktoren gestaltet sich die Aufklärung des Einflusses der psycho-sozialen Faktoren, verallgemeinerbare Aussagen sind bisher schwer zu treffen. So konnte noch nicht abschließend geklärt werden, aus welchem Grund Frauen ihre Leistung mit zunehmendem Alter nicht in demselben Umfang wie Männer zu verbessern scheinen. Insbesondere das Fehlen von Längsschnittstudien erscheint an dieser Stelle problematisch, denn diese wären insbesondere interessant um den entwicklungspsychologischen Verlauf der Geschlechtseffekte zu dokumentieren. Dabei wäre gleichzeitig die Gefahr verfälschter Ergebnisse durch die notwendigen Testwiederholungen und den damit einhergehenden Trainingseffekten hoch und würde die Sinnhaftigkeit von aufwendigen Längsschnittstudien in Frage stellen (siehe z.B. Peters et al., 1995; Spence, Yu, Feng & Marshman, 2009). Weiter ist unklar, welcher entwicklungsbedingte Aspekt für den großen Geschlechtseffekt verantwortlich ist, welcher bei Zehnjährigen beobachtet wurde, (noch) nicht hingegen bei Neunjährigen. Während der Einfluss hormoneller Veränderungen zumindest bei Erwachsenen als gesichert angesehen werden kann (Hausmann et al., 2000), sollte auch bei künftigen Untersuchungen mit Kindern, welche sich an der Schwelle zur Adoleszenz befinden (siehe Kluge & Jansen, 1996) ein Hormonstatus erfasst werden, um mögliche Einflüsse berücksichtigen zu können.

Weiterhin zeigte die konzeptuelle Übertragung eines psycho-sozialen Forschungsansatzes, welcher erfolgreich bei Erwachsenen getestet wurde, auf Kinder im Grundschulalter weitere Unterschiede im Verhalten von Kindern und Erwachsenen. Zudem stehen die Befunde des zweiten Experiments im Widerspruch zu

Forschungsergebnissen aus Studien zur Untersuchung der mathematischen Leistungsfähigkeit von Kindern. Grundsätzlich ist auffallend, dass in den letzten Jahrzehnten der Geschlechtseffekt im mathematischen Bereich zugunsten der Jungen kontinuierlich abnahm, während der Effekt hinsichtlich der mentalen Rotation stabil blieb (Hyde, Fennema & Lamon, 1990). Möglicherweise lassen sich mathematische Leistungen stärker durch psycho-soziale Faktoren beeinflussen, wohingegen die mentale Rotationsleistung relativ stärker von biologisch-neuronalen Komponenten geprägt ist und damit verhältnismäßig robust gegen psycho-soziale Faktoren scheint. Um diese Diskrepanz aufzuklären, wäre es notwendig, in nachfolgenden Studien sowohl die mathematische Leistung als auch die mentale Rotationsleistung intraindividuell zu testen.

Zusammenfassend wird festgestellt, dass die beiden Ansätze, welche in dieser Arbeit aufgegriffen worden sind weiter als zentral erachtet werden, um insbesondere die Forschung um aufrechterhaltende Faktoren des Geschlechtseffekts in der mentalen Rotationsleistung entwicklungspsychologisch nachvollziehen zu können. Die zahlreichen und methodisch vielfältigen bisher geleisteten Forschungsarbeiten zeigen dabei deutlich, dass vor allem kleinschrittige und systematische Studien notwendig sein werden, um das Puzzle zur Aufklärung des Geschlechtseffekts zusammensetzen zu können.

3.2 Zusammenfassende Thesen

1. Die Reduktion der Testkomplexität des MRT in Form des Wegfalls der Zeitbegrenzung einerseits sowie der Vereinfachung des Antwortformats andererseits trägt nicht zur Reduktion der Größe des Geschlechtseffekts bei.

2. Der Einsatz des MRT erscheint ab einem mittleren Alter von zehn Jahren möglich. Der klassische Geschlechteffekt zugunsten des männlichen Geschlechts lässt sich auf demselben Effektstärkeniveau wie das der Erwachsenen nachweisen, wenngleich Kinder noch eine insgesamt niedrigere Performanz hinsichtlich der Antwortgenauigkeit zeigen als Erwachsene.

3. Die positive/ negative Manipulation des Geschlechtsstereotyps bei Grundschulkindern durch die Testinstruktion des MRT führt nicht wie bei Erwachsenen zu einer signifikant verbesserten/ verschlechterten Rotationsleistung im Vergleich zu ihren gleichgeschlechtlichen Gegenspielern.

4. Der Einfluss psycho-sozialer Faktoren auf den Geschlechtseffekt in der mentalen Rotation ist nicht zu unterschätzen, wobei deren Wirksamkeit nicht kurzfristig induzierbar zu sein scheint. Eine Reduktion des Geschlechtseffekts um etwa die Hälfte ist möglich und verdeutlicht die Modifizierbarkeit des Effekts unter spezifischen Bedingungen.

4 Literatur

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*, 539-553.

Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: Effects of identity activation on quantitative performance. *Psychological Science, 12*, 385-390.

Asendorpf, J. (2003). Geschlechtsunterschiede. In J. Asendorpf (Ed.), *Psychologie der Persönlichkeit* (Vol. 3, pp. 375-414). Heidelberg: Springer.

Brandt, M. G., & Wright, E. D. (2005). Medical student career choice and mental rotations ability. *Clinical and Investigative Medicine, 28*, 112-117.

Brehm, J. W. (1966). *A theory of psychological reactance*. New York - London: Academic Press.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience, 111*, 845-849.

Cooper, L. A., & Mumaw, R. J. (1985). Spatial Aptitude. In R. F. Dillon (Ed.), *Individual Differences in Cognition* (Vol. 2, pp. 67-94). New York: Academic Press.

Delgado, A. R., & Prieto, G. (1996). Sex differences in visuospatial ability: do performance factors play such an important role? *Memory & Cognition, 24*, 504-510.

Desrocher, M. E., Smith, M. L., & Taylor, M. J. (1995). Stimulus and sex differences in performance of mental rotation: Evidence from event-related potentials. *Brain and Cognition, 28*, 14-38.

Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist, 41*, 1040-1048.

Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology* (5 ed., Vol. 3: Social, emotional, and personality development. , pp. 1017-1095). Hoboken, NJ: John Wiley & Sons Inc.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for the kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist, 43*, 95-103.

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence, 36*, 556-563.

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

Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Diversity. Culture, gender, and math. *Science, 320*, 1164-1165.

Hahn, N., Jansen, P., & Heil, M. (2010). Preschoolers' mental rotation: Sex differences in hemispheric asymmetry. *Journal of Cognitive Neuroscience, 22*, 1244-1250.

Hahn, N., Jansen, P., & Heil, M. (in press). Preschoolers' mental rotation of letters: Sex differences in hemispheric asymmetry. *Cognitive Neuroscience*.

Halpern, D. F. (1989). The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist, 44*, 1156-1158.

Halpern, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest, 8*, 1-51.

Hannover, B., & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction, 14*, 51-67.

Hardwick, D. A., McIntyre, C. W., & Pick, H. L. (1976). The content and manipulation of cognitive maps in children and adults. 1976. *Monographs of the Society for Research in Child Development, 41*, 1-55.

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience, 114*, 1245-1250.

Herwartz-Emden, L. (2007). "Schulkultur, Geschlechtersegregation und Mädchensozialisation" - eine Studie über Mädchenschulen ["Schooling culture, gender segregation and socialisation of girls" - a study about single-sex girls schools]. In L. Herwartz-Emden (Ed.), *Neues aus alten Schulen - Empirische Studien in Mädchenschule [News from old schools - empirical studies in single-sex girls schools]* Opladen/ Farmington Hills: Barbara Budrich Verlag.

Hines, M., Fane, B. A., Pasterski, V. L., Matthews, G. A., Conway, G. S., & Brook, C. (2003). Spatial abilities following prenatal androgen abnormality: Targeting and mental rotations performance in individuals with congenital adrenal hyperplasia. *Psychoneuroendocrinology 28*, 1010-1026.

Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing: an fMRI study of mental rotation. *Neuropsychologia, 44*, 1575-1583.

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

Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., & Bryden, M. P. (1991). Cognitive abilities in androgen insensitive subjects - Comparison with control males and females from the same kindred. *Clinical Endocrinology, 34*, 341-347.

Inzlicht, M., & Ben Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365-371.

Jansen Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition, 64*, 217-227.

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

Jordan, K., Wustenberg, T., Heinze, H.-J., Peters, M., & Jancke, L. (2002). Women and men exhibit different cortical activation patterns during mental rotation tasks. *Neuropsychologia, 40*, 2397-2408.

Karadi, K., Csatho, A., Kovacs, B., & Kosztolanyi, P. (2003). Subgroup analysis of sex difference on the Vandenburg-Kuse mental rotation test. *Perceptual and Motor Skills, 96*, 197-200.

Karadi, K., Kovacs, B., Szepesi, T., Szabo, I., & Kallai, J. (2001). Egocentric mental rotation in Hungarian dyslexic children. *Dyslexia, 7*, 3-11.

Kerkman, D. D., Wise, J. C., & Hardwood, E. A. (2000). Impossible "mental rotation" problems. A mismeasure of women's spatial abilities? *Learning and Individual Differences, 12*, 253-269.

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

Kessels, U. (2007). Identifikation mit naturwissenschaftlichen Fächern: Ein Vergleich von Schülerinnen einer monoedukativen und einer koedukativen Schule [Identification with sciences: a comparison of girls attending a single-sex and a co-educative school]. In L. Herwartz-Emden (Ed.), *Neues aus alten Schulen - Empirische Studien in Mädchenschule [News from old schools - empirical studies in single-sex girls schools]* Oplanden/ Farmington Hills: Barbara Budrich Verlag.

Kluge, N., & Jansen, G. (1996). *Körperentwicklung in der Pubertät*. Frankfurt/ Bern: Lang.

Linn, M. C., & Hyde, J. S. (1989). Gender, Mathematics, and Science. *Educational Researcher, 18*, 17-27.

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

Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore: The Johns Hopkins University Press.

Lippa, R. A., Collaer, M. L., & Peters, M. (2009). Sex Differences in Mental Rotation and Line Angle Judgments Are Positively Associated with Gender Equality and Economic Development Across 53 Nations. *Archives of Sexual Behavior,* DOI 10.1007/s10508-008-9460-8.

Lizarraga, M. L. S. d. A., & Ganuza, J. M. G. (2003). Improvement of Mental Rotation in Girls and Boys. *Sex Roles, 49*, 277-286.

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

Marmor, G. S. (1975). Development of kinetic images: When does the child first represent movement in mental images? *Cognitive Psychology, 7*, 548-559.

Masters, M. S. (1998). The gender difference on the Mental Rotations test is not due to performance factors. *Memory & Cognition, 26*, 444-448.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*, 337-341.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin, 86*, 889-918.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*, 21-27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369-377.

Monahan, J. S., Harke, M. A., & Shelley, J. R. (2008). Computerizing the mental rotations test: Are gender differences maintained? *Behavior Research Methods, 40*, 422-427.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science, 19*, 1063-1066.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology, 43*, 747-759.

Nash, S. C. (1979). Sex role as mediator of intellectual functioning. In M. A. Wittig & A. C. Peterson (Eds.), *Sex-related differences in cognitive functioning* (pp. 263-302). New York: Academic Press.

Neuville, E., & Croizet, J.-C. (2007). Can salience of gender identity impair math performance among 7-8 year old girls? The moderating role of task difficulty. *European Journal of Psychology of Education, 22*, 307-316.

Newcombe, N., Bandura, M. M., & Taylor, D. G. (1983). Sex differences in spatial ability and spatial activities. *Sex Roles, 9*, 377-386.

Niedenthal, P. M., Cantor, N., & Kihlstrom, J. F. (1985). Prototype matching: A strategy for social decision making. *Journal of Personality and Social Psychology, 48*, 575-584.

Parsons, T. D., Larson, P., Kratz, K., Thiebaux, M., Bluestein, B., Buckwalter, J. G., et al. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia, 42*, 555-562.

Pease, A., & Pease, B. (2001). *Why Men Don't Listen & Women Can't Read Maps*. London: Orion.

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*, 176-184.

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

Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition, 17*, 1-22.

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

Quaiser Pohl, C., & Lehmann, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology, 72*, 245-260.

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19*, 1067-1070.

Resnick, S. M. (1993). Sex differences in mental rotations: an effect of time limits? *Brain and Cognition, 21*, 71-79.

Roberts, A. S. (2007). Predictors of future performance in architectural design education. *Educational Psychology, 27*, 447-463.

Rozencwajg, P., Lemoine, C., Rolland Grot, M., & Bompard, A. (2005). Les effets conjugues de l'age et de l'experience professionnelle sur les competences spatiales: Le cas des controleurs aeriens Combined Effects of Age and Job Experience on spatial abilities: The case of Air-Traffic Controllers. *Psychologie du Travail et des Organisations, 11*, 47-57.

Ruble, D. N., & Martin, C. L. (1998). Gender development. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology* (5 ed., Vol. 3: Social, emotional, and personality development, pp. 933-1016). Hoboken, NJ: John Wiley & Sons Inc.

Sanders, B., & Soares, M. P. (1986). Sexual maturation and spatial ability in college students. *Developmental Psychology, 22*, 199-203.

Schoning, S., Engelien, A., Kugel, H., Schafer, S., Schiffbauer, H., Zwitserlood, P., et al. (2007). Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia, 45*, 3203-3214.

Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: A 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *International Journal of Behavioral Development, 25*, 124-132.

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

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80-83.

Smoker, W. R., Berbaum, K. S., Luebke, N. H., & Jacoby, C. G. (1984). Spatial perception testing in diagnostic radiology. *American Journal of Roentgenology, 143*, 1105-1109.

Spence, I., Yu, J. J., Feng, J., & Marshman, J. (2009). Women match men when learning a spatial skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 35*, 1097-1103.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist, 52*, 613-629.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Steele, J. (2003). Children's Gender Stereotypes About Math: The Role of Stereotype Stratification. *Journal of Applied Social Psychology, 33*, 2587-2606.

Stumpf, H. (1993). Performance factors and gender-related differences in spatial ability: Another assessment. *Memory and Cognition, 21*, 826-836.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles, 53*, 433-441.

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

Voyer, D. (1997). Scoring procedure, performance factors, and magnitude of sex differences in spatial performance. *American Journal of Psychology, 110*, 259-276.

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale, 60*, 91-100.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles, 43*, 891-915.

Voyer, D., & Saunders, K. A. (2004). Gender differences on the mental rotations test: A factor analysis. *Acta Psychologica, 117*, 79-94.

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*, 250-270.

Wanzel, K. R., Hamstra, S. J., Anastakis, D. J., Matsumoto, E. D., & Cusimano, M. D. (2002). Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The Lancet, 359*, 230-231.

5 Einzelarbeiten

5.1 Experiment 1

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

5.2 Experiment 2

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

5.3 Experiment 3

Titze, C., Jansen, P., & Heil, M. (under review). Single-sex school girls outperform girls attending a co-educative school in mental rotation performance.

5.4 Experiment 4

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*, 130-133.

5.5 Experiment 5

Titze, C., Heil, M., & Jansen, P. (2010). Pairwise Presentation of Cube Figures Does Not Reduce Gender Differences in Mental Rotation Performance. *Journal of Individual Differences, 31, 101-105*.

EUROPEAN JOURNAL OF DEVELOPMENTAL PSYCHOLOGY
0000, 00 (00), 000–000, i*First* article

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

Mental rotation performance and the effect of gender in fourth graders and adults

Corinna Titze, Petra Jansen, and Martin Heil
Heinrich-Heine-University, Düsseldorf, Germany

Whereas a gender effect favouring males in adults' mental rotation performance can be reliably found, the respective gender effect in children's performance does not appear consistently in the literature. Therefore, this study investigated whether there is a "crucial" time slot on infantile development when a gender effect is detectable. Following Johnson and Meade's (1987) hypothesis that age ten might be crucial, we investigated 96 fourth graders—split into two age groups (mean: 9.3 vs. 10.3 years)—and 48 adults that performed the MRT (Peters et al., 1995). We observed large gender effects in older children as well as in adults. However, the younger children did not reveal a significant gender effect, although effect sizes indicated a small-to-medium effect. Surprisingly, the performance of women, older girls, and younger girls did not differ significantly, whereas older boys and adults clearly outperformed younger boys. Possible reasons for these effects are discussed.

Keywords: Developmental aspects; Fourth graders; Gender effect; Mental rotation.

It is a popular, commonly held, assumption that men outperform women on visual-spatial tasks (Pease & Pease, 2001), even though there is evidence that the magnitude of the gender effect can be manipulated by different variables as for example the instructions (Moè & Pazzaglia, 2006). Furthermore, the implications of this effect are often over interpreted (Halpern, 1996).

Correspondence should be addressed to Corinna Titze, Institute of Experimental Psychology, Heinrich-Heine-University, Universitätsstr. 1, D-40225 Düsseldorf, Germany. E-mail: corinna.titze@uni-duesseldorf.de

This study was supported by the German Research Foundation.

We thank Michael Peters for his friendly permission to use the Mental Rotations Test (MRT) in our study. We thank all primary school children, their teachers and all the students who participated in this study.

© 2009 Psychology Press, an imprint of the Taylor & Francis Group, an Informa business

http://www.psypress.com/edp DOI: 10.1080/17405620802548214

Visual-spatial abilities contain three discrete aspects: spatial perception, spatial visualization, and mental rotation (Linn & Petersen, 1986). According to several meta-analyses (Linn & Petersen, 1986; Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995), the apparent visual-spatial advantage favouring men does not apply to all aspects of visual-spatial abilities. Instead, the male advantage shows most obviously if mental rotation is considered (i.e., ''. . . the ability to rotate a two or three dimensional figure rapidly and accurately''; Linn & Petersen, 1985, p. 1483).

This gender effect in mental rotation performance is the largest documented cognitive gender effect (Halpern, 1989). Compared to other cognitive gender effects (e.g., females tend to excel in verbal tasks whereas males tend to outperform females in math) the gender effect in mental rotation performance is reliable and stable over the life span, beginning at 18 years (at least). The large effect size of the mental rotation gender effect $d = 0.9$, however, is only observed on the basis of the three-dimensional Mental Rotations Test (MRT; Peters et al., 1995; originally developed by Vandenberg & Kuse, 1978). Other mental rotation tests—such as the Primary Mental Abilities Test (PMA; Thurstone, 1958) or the Cards Rotation Test (Ekstrom, French, & Harman, 1976)—achieve effect sizes ranging from $d = 0.31$ to $d = 0.49$ (which indicate small-to-medium effects as defined by Cohen, 1977), see, e.g., Voyer et al. (1995). Although hypothesized reasons for this effect are numerous (for example ''age'', e.g., Lizarraga & Ganuza, 2003; Willis & Schaie, 1988; ''experience and practice'', e.g., Alington, Leaf, & Monaghan, 1992; Terlecki & Newcombe, 2005; Voyer, Nolan, & Voyer 2000; ''social and cognitive factors'', e.g., Feingold, 1993, 1996; Quaiser-Pohl & Lehmann, 2002; ''specific characteristics of the stimuli used'', e.g., Jansen-Osmann & Heil, 2007; ''hormonal influences'', Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis, & Güntürkün, 2000), its sheer existence is widely accepted as long as adults' performance is measured with the MRT.

Whether or not a gender effects in children's mental rotation performance exists, however, is still unclear. Maccoby and Jacklin (1974) hypothesized adolescence to be the critical date of the first occurrence of a gender effect in mental rotation. Supporting this assumption, Karadi, Szabo, Szepesi, Kallai, and Kovacs (1999) did not observe any gender effect in 8- to 11-year-old children when egocentric mental hand rotation was measured. Adults (20–23 years), however, showed the gender effect leading to the assumption of a developmental phenomenon. Nevertheless, Linn and Petersen (1985) pointed out that the MRT itself has never been used with children younger than 13 years. Instead, alternative measures were used that, when applied to adults, produced substantially lower effect sizes than the MRT itself. As a consequence, Linn and Petersen postulated that: ''Sex differences are detected as soon as mental rotation could be measured.'' (Linn & Petersen,

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

1985, p. 1488). A number of later studies supported Linn and Petersen's (1985) assumption and were inconsistent with Maccoby and Jacklin's (1974) hypothesis, although according to our knowledge no study at all used the MRT with children. Johnson and Meade (1987) showed a male advantage in picture plane mental rotation that is reliably found by age 10 but, interestingly, not before. They tested a large sample of 1800 public-school students aged 6 to 18 years with a set of spatial tests. They chose seven adult spatial tests that differed regarding the dimensionality of the objects (2D vs. 3D), the kind of objects (abstract vs. concrete) and the kind of answer (same–different judgements vs. selection from up to five alternatives). Furthermore, Johnson and Meade (1987) modified the test battery for different ages. Children at kindergarten level and younger children (7–10 years; grades 1–4) completed the tests on 7 consecutive days with each testing session lasting 20 minutes. The oldest children (13–18 years; grades 7–12) completed the tests on 2 consecutive days. One session lasted about 1 hour. Although Johnson and Meade (1987) used the three-dimensional Shepard and Metzler (1971) block figures for all children other than the children at kindergarten level, they reported that: "... the test is of questionable value below grade 5." (Johnson & Meade, 1987, p. 735).

Similarly, Kerns and Berenbaum (1991) used four spatial tests similar to those that are well established to induce large gender effects in adults. First, they chose two tests measuring spatial visualization developed by Tuddenham (1970): the Geometric Forms and the House Plans test. Furthermore, they constructed two new mental rotation tests: the Mirror Images Test and the Three-Dimensional (3D) Mental Rotations Test. This last test used the same shapes as the MRT but as real-life three-dimensional objects. Kerns and Berenbaum (1991) observed that 9- to 13-year-old boys scored higher on all tests than girls of the same age. Additionally, they reported moderate effect sizes ranging from $d = 0.30$ to 0.65. Kerns and Berenbaum (1991) concluded that the gender effect can be found in pre-adolescents if appropriate tests are used. It has to be remarked critically, however, that Kerns and Berenbaum (1991) used three-dimensional objects rather than their pictorial presentations. Additionally, the decision complexity differed (same–different judgments vs. selecting two correct answers out of four alternatives). Pezaris and Casey (1991), on the other hand, did not find any gender effect in juveniles aged from 13 to 15 years using a mental-rotation task that was built with the Shepard and Metzler (1971) figures and realized a decision mode similar to the MRT itself.

In summary, it is still not clear when gender effects in mental rotation emerge. Different findings might be due to the use of modified tests on different age groups—within different groups of children (Johnson & Meade, 1987) and between adults and children (see Linn & Petersen, 1985)—or to the use of real-life, non-abstract or two-dimensional objects (Alyman & Peters, 1993; Foulkes & Hollifield, 1989). As a consequence, the results are not fully comparable between the studies cited above—precluding an unequivocal interpretation. In our study, therefore, we considered this problem by using the classical MRT itself—which is well known to induce a gender effect in adults—for children and adults without adapting the testing conditions for age. As far as we know, no previous study has used the MRT itself with children younger than 13 years. Furthermore, we wanted to test the hypotheses of Maccoby and Jacklin (1974) and of Johnson and Meade (1987) more thoroughly by testing 9-year-old children, 10-year-old children and adults: if gender effects arise primarily in adolescence as proposed by Maccoby and Jacklin (1974), no appreciable gender effects should be present in either group of children. If, however, a gender effect in children's mental rotation performance indeed is found only by age 10 (as proposed by Johnson & Meade, 1987), then 9-year-old children should show this effect in a much reduced magnitude.

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

METHODS

Participants

Ninety-six fourth graders (48 boys and 48 girls) and 48 adults (24 men and 24 women) participated in this study. For statistical analysis the children's data were split into two age groups: one group with a mean age of 9.3 years (range: 8.3 to 9.6 years, $SD = 0.24$; 24 boys and girls each) and the other group with a mean age of 10.3 years (range: 9.9 to 11.0 years, $SD = 0.28$; 24 boys and girls each). Both groups of children consisted of fourth graders to ensure that years of formal education was not confounded with age. To ensure that the findings were not affected by a difference on general intelligence between both sexes, we used the CFT-20 (Weiß, 1998) as co-variable. The CFT-20 is a non-verbal and culture-fair intelligence test, which measures reasoning with four subtests on the basis of concrete problems (classifications, matrices, . . .). An ANOVA revealed no main effects of "gender", $F(1, 92) = 1.88$; $p = .17$; $\eta^2 = .020$, and "age", $F(1, 92) = 0.21$; $p = .65$; $\eta^2 = .002$. There was no interaction effect, $F(1, 92) = 0.002$; $p = .96$; $\eta^2 = .000$. Children were recruited from primary schools near Düsseldorf and were tested during normal class time. A maximum of 20 children was tested at once and each class was paid €100 for their participation. Adults were students aged from 19 to 43 years ($M = 23.1$ years; $SD = 3.94$). They were tested during normal course time in the University of Düsseldorf. Group sizes varied from 10 up to 15 participants per session. An ANOVA revealed that the averaged age of male and female participants did not differ between the sexes, $F(1, 138) = 0.19$; $p = .657$; $\eta^2 = .001$.

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

Material and procedure

Participants had to solve the "Mental Rotations Test (MRT)" in its versions A or B (Peters et al., 1995), which is a redrawn version of the Vandenberg and Kuse (1978) "Mental Rotations Test" with figures provided by Shepard and Metzler (1971). The MRT is made of three-dimensional cube figures in two sets with 12 items each (see Figure 1 for an example). Six items were presented per DIN-A4-sized sheet of paper. Each item contains a target item on the left side and four sample stimuli on the right. Two items of the sample stimuli were identical but in-depth rotated versions of the target item. The two remaining sample stimuli did not match regardless of rotation.

The participants were asked to cross out both correct sample stimuli. Participants were given three minutes for the first 12 items and another three minutes for the remaining 12. We used the MRT in its versions A and B within the group testing sessions to avoid participants copying their answers. Stimulus figures in both versions were identical but rearranged to form different items. On this account, the level of difficulty did not change (Peters et al., 1995).

One problem of using the MRT with children is that they may produce poor results just because (a) they did not really understand this abstract kind of task and/or (b) the level of attention required was too high. This problem is important to consider because instructions were given in written form in the original version of this test. To avoid the children not understanding the task, we gave them another mental rotations test (MRT-Animal; see Figure 2), which was specially designed for this study. This test was designed similarly to the MRT but used two-dimensional familiar stimuli

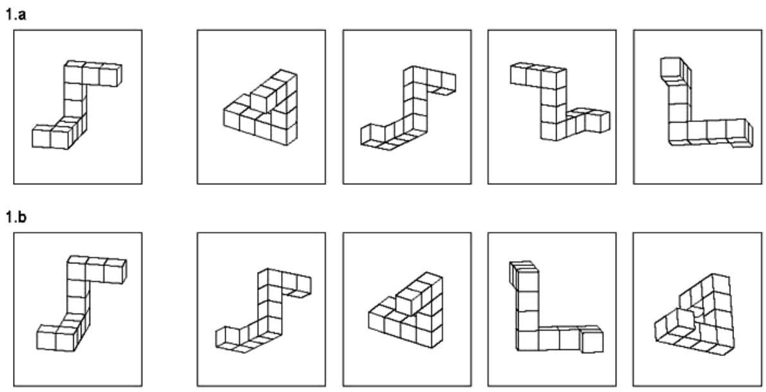


Figure 1. This figure shows the first item of the MRT in its versions A and B. The subject has to mentally rotate the four sample stimuli on the right to decide whether they match the target item. Two out of the four alternatives match the target item.

(Snodgrass & Vanderwart, 1980) rotated in the picture plane instead of three-dimensional abstract cube figures rotated in depth. In our test we used black and white pictures to slightly increase difficulty (Alington et al., 1992) and to enhance the similarity to the MRT.

As in the MRT, participants had to cross out both animals that matched the target animal after rotation (i.e., children were asked to imagine the rotated animals standing on their feet in an upright position and afterwards to decide which two were looking in the same direction as the target animal). This test is more concrete than the MRT and is thought to make children familiar with the kind of test and its answer format as well as with the concept of mental rotation. Children were given 60 seconds to complete 16 items presented on four DIN-A4-sized sheets of paper (horizontal format). Afterwards, the MRT was given. Children were told that they had to solve another test that worked exactly as the animal test but now used cube figures instead of animals. Verbal instructions were followed by three practice items, the correct solutions of which were given after all participants had finished them.

An ANOVA revealed that children's "gender" and "age" had no effect on the number of items answered correctly in the animal-MRT, $F(1, 92) = 1.08$; $p = .301$; $\eta^2 = .012$; $F(1, 92) = 0.12$; $p = .729$; $\eta^2 = .001$. Neither was an interaction found, $F(1, 92) = 0.09$; $p = .925$; $\eta^2 = .000$. Hence, it is plausible to assume that boys and girls understood the task equally well. Adults completed the MRT-Animal in the same way, to keep the testing sessions constant between age groups (see Table 1 for descriptive results). A Pearson correlation of the performance on the MRT-Animal and the MRT revealed a significant positive correlation $r = .27$ and $p < .001$ (two-tailed).The MRT-Animal was not analysed further as it was just used to introduce the concept of mental rotation to the children. The completion of both tests required about 25 minutes.[1]

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

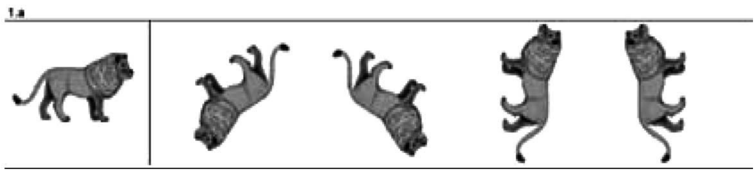


Figure 2. This figure shows the first item from the MRT-Animal. This test was given to make children more familiar with the concept of mental rotation tasks and to prevent poor results resulting from poor task comprehension.

[1]A specimen test is available from the authors upon request.

Design and statistical analysis

Peters et al. (1995) proposed the following standard scoring method that was used in the present study as well: one point was given if both and only both correct sample stimuli of a target item were marked correctly. A maximum of 24 points could thus be obtained.

The design of this study involved two factors: ''age'' (adults vs. younger children, 9 years, vs. older children, 10 years) and ''gender'' (male vs. female). The number of correctly answered items in the MRT was the dependent variable.

Given a total sample size of $N = 144$ and a desired level of $\alpha = .05$, $\beta = .05$, an effect of size $d = 0.76$ (i.e., a medium-to-large effect as defined by Cohen, 1977) could be detected with a probability of $1 - \beta = .95$.[2]

RESULTS

Analysis of variance (ANOVA) revealed a main effect of ''age'', $F(2, 138) = 15.52$; $p < .001$; $\eta^2 = .184$, and of ''gender'', $F(1, 138) = 39.95$; $p < .001$; $\eta^2 = .224$. Furthermore, an interaction was found, $F(2, 138) = 3.50$; $p = .033$; $\eta^2 = .048$. This interaction remained significant, too, only if both groups of children were included and did not change if ''intelligence'' was used as co-variable, $F(1, 92) = 5.47/6.47$; both ps $< .021$. The post hoc tests described in the following paragraph were conducted using conservative and α-error-revised Scheffé tests. The means and standard errors are shown in Table 1. Adults significantly outperformed the older children, as well as the younger children. However, the older children did not outperform the younger children. Collapsing across age groups, males outperformed females. There is no statistically significant difference on the performance of boys and girls in the group of younger children. There was, however, a significant difference between boys and girls of the older group, as well as between men and women of the adult group. Furthermore, younger boys showed a significantly lower performance than men. The performance of older boys and adults and the performance of older and younger boys did not differ statistically. There were, however, no statistically significant differences in performance of younger girls, older girls and women.

The effect size of the gender effect within the older group was $d = 1.47$ that is—following Cohen (1977)—a large gender effect. The effect size of the gender effect within the adults was $d = 1.24$. The gender effect thus seemed to be stable with increasing age. As reported above, there was no significant

[2]All effect size and power calculations reported in this article were conducted using the G*Power 3 program (Faul, Erdfelder, Lang, & Buchner, 2007).

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

gender effect within the group of younger children but an effect size of $d = 0.45$ indicated that there was a small-to-medium gender effect.

Linn and Petersen (1985) claimed that the MRT was too difficult for children younger than 13 years because the level of attention required was too high. To test this hypothesis, we decided to analyse the MRT data more closely: the MRT consists of two sets and three minutes were given to the participants to solve each set. There was no break, implying that participants had to pay attention for six minutes—which might be too long for children. Therefore, we analysed the number of items solved correctly on set one and on set two of the MRT (see Table 2). A repeated-measures MANOVA with "set" as within-subject factor and "age" and "gender" as between-subject factors revealed a significant main effect "set",

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

TABLE 1
Mean number of correct answers and standard-errors for the MRT-Animal as well as the MRT as a function of gender and age

	MRT-Animal		*MRT*	
	Mean correct	*Standard error*	*Mean correct*	*Standard error*
Younger group	6.48	0.47	6.5	0.53
Girls	6.17	0.68	5.67	0.62
Boys	6.79	0.65	7.33	0.85
Older group	6.25	0.45	8.27	0.61
Girls	5.88	0.59	5.75	0.52
Boys	6.63	0.70	10.79	0.85
Adult group	11.73	0.47	10.85	0.75
Women	10.67	0.67	8.13	0.70
Men	12.79	0.48	13.58	1.06
Over all	8.15	0.33	8.54	0.39
Female	7.57	0.45	6.51	0.38
Male	8.74	0.49	10.57	0.61

TABLE 2
Mean number of items answered correctly separately for the first and the second half of the MRT as a function of age

	MRT Set 1		*MRT Set 2*	
	Mean correct	*Standard error*	*Mean correct*	*Standard error*
Children	4.1	0.24	3.4	0.23
Younger group	3.5	0.31	3.1	0.28
Older group	4.7	0.36	3.7	0.37
Adults	5.6	0.45	5.2	0.36

$F(1, 138) = 9.56$, $p = .002$, $\eta^2 = .065$. Since no interaction with "gender", $F(1, 138) = 0.40$, $p = .528$, $\eta^2 = .003$, and since especially no interaction with "age", $F(2, 138) = 1.05$, $p = .352$, $\eta^2 = .015$, was observed, we obtained no evidence to suggest that the MRT requires a level of attention too high for children at the age of 10.

DISCUSSION

We found that adults and older children showed the typical large gender effect in mental rotation in favour of males. Younger children did not show this large gender effect although time of formal education had been kept constant for both age groups. But the effect size $d = 0.45$ found in the group of younger children pointed out that there is a small difference that cannot be ignored. The non-significant results might be due to sample size. However, given an estimated effect size of $d = 0.45$ for the group of younger children and a level of $\alpha = \beta = .05$, a considerable sample size of $N = 260$ would be needed to obtain this effect, a size rarely met in developmental cognitive psychology. Taken together, effect sizes were substantially smaller for younger children, reached a maximum in the older group and seemed to still maintain a high level for adults—indicating the manifestation of a gender effect at about 10 years as proposed by Johnson and Meade (1987) but contradicting the assumption of Maccoby and Jacklin (1974) that gender effects arise primarily in adolescence.

Furthermore, the results showed that adults outperformed the older as well as the younger children. Performance of the older and the younger children did not differ. A significant and positive correlation of the MRT and the MRT-Animal showed that using the MRT-Animal prior to the MRT introduced the concept of mental rotations test very well to primary-school children. There was no gender effect in the performance on the MRT-Animal suggesting the conclusion that girls understood the concept of two-dimensional picture plane mental rotation as well as boys. Additionally, we showed that children's performance did not decline more than adults' on set two compared to set one of the MRT. This indicates that children—assuming that they understood the concept of the mental rotation task—were able to solve the mental rotation task despite the concerns of Linn and Petersen (1985).

A further important result is that there was an interaction of age and gender: with increasing age, males continuously improved their mental-rotation performance as shown in a significant difference between the performance of adult men and the younger boys. But for females, we did not find any kind of significant improvement of performance with increasing age. As stated above, effect sizes of the mental rotation gender effect increased with increasing age because women—contrary to men—did not

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

improve their performance with increasing age significantly. Reasons for that disproportion are not yet known.

The findings displayed above raise at least two questions requiring detailed discussion: (1) Why do girls/women not improve their performance with increasing age as men do? and (2) Which developmental aspect is responsible for the large gender effect observed in 10-year-old boys and girls that 9-year-olds do not (yet) show?

In general, pubertal hormonal changes have been discussed as a critical factor on visual-spatial performance differences (see Linn & Petersen, 1986). At least for adults, there is strong evidence that hormonal differences contribute to the gender effect on mental rotation. Hausmann et al. (2000) showed that testosterone had a positive influence on mental rotation performance whereas estradiol did not. Women improved their performance on a mental rotation test during their menses—where the estradiol level is lowest. Focusing on this biological aspect, it is difficult to determine the age when children should show the gender effect for the first time as a function of beginning hormonal changes. Kluge and Jansen (1996) pointed out that within the last 150 years the initiation of puberty has started to develop earlier. Whereas the mean age of the first menstruation was 16.6 years in 1860, today it is 12.2 years. For boys there is a similar development. Furthermore, Kluge and Jansen (1996) noticed that the beginning of puberty today is 8 to 14 years for girls and 10 to 16 years for boys. This wide range indicates a rather individualized development and therefore we cannot exclude the possibility that the children tested in our study were already pubertal. Perhaps, the youngest girls performed at the same level as women because hormonal changes have already increased the level of estradiol. This would explain why we did not find an improvement in females' performance in the different age groups. For the males' performance, the range given by Kluge and Jansen (1996) might also explain the differences found: the older boys (mean age: 10.3 years) did not differ from men's performance because hormonal changes (i.e., increasing testosterone) had already improved their performance whereas the younger boys (mean age: 9.3 years) still showed a significantly worse performance compared to men. This interpretation, of course, remains very speculative. We conclude that, even testing primary-school children, further studies have to control the hormonal status of every participant.

However, as Casey (1996) pointed out, there is evidence that biological and social factors interact in determining the extent of this gender effect. We want to focus on the discussion of the "self-concept" as a possible social factor that influences the mental rotation performance. Nash (1979) argued that individuals perform better on cognitive tasks if their gender stereotype of the task and their self-concept match. Regarding visual-spatial tasks men were expected to perform better than women. This hypothesis was confirmed by Signorella, Jamison, and Krupa (1989), who predicted the

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

spatial performance on the basis of sex, self-concept and spatial activities. Furthermore, Quaiser-Pohl and Lehmann (2002) studied the ''achievement-related self-concept'' of men and women regarding their attitudes towards maths and physics. They found that women with a high achievement-related self-concept scored significantly higher on the MRT than those with a lower concept. For men no correlation between self-concept and mental rotation performance was obtained. Apparently, women are more sensitive to their own expectations concerning their abilities than men. This might be crucial if female participants were asked to solve a test on which they expect themselves to have low abilities—as on the MRT. Whether this relation is also valid for children has been largely studied as a function of performance on different school subjects but not in the context of visual-spatial abilities. Tiedemann and Faber (1995), for example, provided evidence that primary-school children (third and fourth graders) differed concerning their math self-concept. Girls judged their abilities lower than boys did and expected a lower test result even if no actual gender differences in test performance were found. Furthermore, the self-concept in children is more difficult to assess because it only develops during childhood, especially during primary-school education. Oerter (2002) stated that self-concept in children becomes more realistic and differentiated in late childhood (9 to 12 years). The feedback on their performance in school and the comparison with other children helps them to construct their own complex and hierarchically structured self-concept. This developmental progress might be an explanation why we found gender differences in 10-year-olds but not in 9-year-olds. Future research in our lab will test this hypothesis with fourth graders.

In summary, we found large gender effects in children aged 10.3 years and in adults using a classic mental rotation test that is well known to produce large gender effects in adults (MRT; Peters et al., 1995). With children aged 9.3 years, however, the effect size of the (non-significant) gender effect turned out to be substantially smaller. Furthermore, girls' performance did not differ from women's so it seems that females did not improve their performance with increasing age as males did. Hormonal as well as social-cognitive explanations may account for the pattern of findings.

Manuscript received 12 December 2007
Revised manuscript accepted 30 September 2008
First published online day/month/year

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

REFERENCES

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*, 539–553.

Alyman, C., & Peters, M. (1993). Performance of male and female children, adolescents and adults on spatial tasks that involve everyday objects and settings. *Canadian Journal of Experimental Psychology*, *47*, 730–747.

Casey, M. B. (1996). Gender, sex, and cognition: Considering the interrelationship between biological and environmental factors. *Learning and Individual Differences*, *8*, 39–53.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Ekstrom, R. B., French, J. W., & Harman, H. H. (1976). *Manual for the kit of factor-referenced cognitive tests*. Princeton, NJ: Educational Testing Service.

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

Feingold, A. (1993). Cognitive gender differences: A developmental perspective. *Sex Roles*, *29*, 91–112.

Feingold, A. (1996). Cognitive gender differences: Where are they, and why are they there? *Learning and Individual Differences*, *8*, 25–32.

Foulkes, D., & Hollifield, M. (1989). Responses to picture-plane and depth mental-rotation stimuli in children and adults. *Bulletin of the Psychonomic Society*, *27*, 327–330.

Halpern, D. F. (1989). The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist*, *44*, 1156–1158.

Halpern, D. F. (1996). Changing data, changing minds: What the data on cognitive sex differences tell use and what we hear. *Learning and Individual Differences*, *8*, 73–82.

Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, *114*, 1245–1250.

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition*, *64*, 217–227.

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

Karadi, K., Szabo, I., Szepesi, T., Kallai, J., & Kovacs, B. (1999). Sex differences on the hand mental rotation task for 9-yr.-old children and young adults. *Perceptual and Motor Skills*, *89*, 969–972.

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

Kluge, N., & Jansen, G. (1996). *Körperentwicklung in der Pubertät* [*Corporal development in puberty*]. Frankfurt, Germany/Bern, Switzerland: Lang.

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

Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67–101). Baltimore: Johns Hopkins University Press.

Lizarraga, M. L. S. d. A., & Ganuza, J. M. G. (2003). Improvement of mental rotation in girls and boys. *Sex Roles*, *49*, 277–286.

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

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics*, *23*, 337–341.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*, 369–377.

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

Nash, S. C. (1979). Sex role as mediator of intellectual functioning. In M. A. Wittig & A. C. Peterson (Eds.), *Sex-related differences in cognitive functioning* (pp. 263–302). New York: Academic Press.

Oerter, R. (2002). Kindheit [Childhood]. In R. Oerter & L. Montada (Eds.), *Entwicklungspsychologie* [*Developmental psychology*] (5th ed., pp. 209–255). Weinheim, Germany: Beltz.

Pease, A., & Pease, B. (2001). *Why men don't listen & women can't read maps*. London: Orion.

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

Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition*, *17*, 1–22.

Quaiser-Pohl, C., & Lehmann, W. (2002). Girls' spatial abilities: Charting the contributions of experiences and attitudes in different academic groups. *British Journal of Educational Psychology*, *72*, 245–260.

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

Signorella, M. L., Jamison, W., & Krupa, M. H. (1989). Predicting spatial performance from gender stereotyping in activity preferences and in self-concept. *Developmental Psychology*, *25*, 89–95.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Terlecki, M. S., & Newcombe, N. S. (2005). How important is the digital divide? The relation of computer and videogame usage to gender differences in mental rotation ability. *Sex Roles*, *53*, 433–441.

Thurstone, T. G. (1958). *Manual for the SRA Primary Mental Abilities*. Chicago: Science Research Associates.

Tiedemann, J., & Faber, G. (1995). Mädchen im Mathematikunterricht: Selbstkonzept und Kausalattributionen im Grundschulalter [Gender differences in elementary school children's self-concept and attributions in mathematics]. *Zeitschrift fur Entwicklungspsychologie und Pädagogische Psychologie*, *27*(1), 61–71.

Tuddenham, R. D. (1970). A Piagetian test of cognitive development. In W. B. Dockrell (Ed.), *On intelligence* (pp. 49–70). London: Methuen.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599–604.

Voyer, D., Nolan, C., & Voyer, S. (2000). The relation between experience and spatial performance in men and women. *Sex Roles*, *43*, 891–915.

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*, 250–270.

Weiß, R. H. (1998). *Grundintelligenztest Skala 2 (CFT 20) mit Wortschatztest (WS) und Zahlenfolgentest (ZF). Handanweisung* [*Culture Fair Intelligence Test Manual*] (4th ed.). Göttingen, Germany: Westermann Test.

Willis, S. L., & Schaie, K. W. (1988). Gender differences in spatial ability in old age: Longitudinal and intervention findings. *Sex Roles*, *18*, 189–203.

Downloaded By: [Titze, Corinna] At: 08:50 24 June 2009

Learning and Individual Differences xxx (2010) xxx–xxx

Contents lists available at ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

Mental rotation performance in fourth graders: No effects of gender beliefs (yet?)

Corinna Titze [a,b,*], Petra Jansen [b], Martin Heil [a]

[a] Institute of Experimental Psychology, Heinrich-Heine-University, Universitätsstr. 1, 40223 Düsseldorf, Germany
[b] Institute of Sports Science, University of Regensburg, Universitätsstr. 31, 93053 Regensburg, Germany

A R T I C L E I N F O

Article history:
Received 22 September 2009
Received in revised form 29 March 2010
Accepted 1 April 2010
Available online xxxx

Keywords:
Mental rotation
Gender differences
Fourth graders
Gender beliefs

A B S T R A C T

The influence of gender beliefs on cognitive task performance has been demonstrated repeatedly for adults. For children, there is evidence that gender beliefs can substantially impede or boost math performance — a task where gender differences in favour of boys declined over past decades. Therefore, we examined this phenomenon using the Mental Rotations Test (MRT), a task where gender differences still occur reliably favouring males — for adults as well as for children. A sample of 252 fourth graders, whose beliefs about spatial ability were manipulated experimentally (instructions given: boys are better, girls are better or independent of gender) had to complete the MRT. In contrast to adult's literature, children's performance did not decrease or increase as a function of instruction: boys always outperformed girls; girls not even outperformed their same-sex counterparts given the “girls better” instruction. The transfer of the conceptual approach failed — possible reasons are discussed.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Research on gender differences has become increasingly popular in recent decades. In cognitive psychology, there are at least three areas where gender differences have been found repeatedly: in verbal and mathematical abilities as well as in visual-spatial abilities (see e.g. Halpern et al., 2007; Linn & Hyde, 1989). Whereas in general, females outperform males on verbal tasks, males excel (probably) on mathematical performance and (reliably) on mental rotation, i.e. the cognitive process of imagining an object turning around (Shepard & Metzler, 1971). The male advantage in mental rotation is observed most consistently with the Mental Rotations Test, MRT (Peters et al., 1995; Vandenberg & Kuse, 1978), where the gender difference amounts to one standard deviation (Voyer, Voyer, & Bryden, 1995). The MRT consists of 24 items, each presenting a target figure and four sample stimuli. Two sample stimuli are rotated-in-depth versions of the target figure (see Fig. 1) and have to be identified.

The underlying causes of gender differences are still far from being understood. Two broad classes of explanations are the “psycho-social” variety (consisting of, e.g., stereotype threat, sex-role identification, or differential experience and socialization) and the biological-neuronal variety (e.g., rate of maturation, genetic complement, sex hormone level, or cerebral lateralization; for details and references of both, see e.g., Jansen-Osmann & Heil, 2007; Voyer et al., 1995). Each of these explanations can quote empirical support, and they are by no means mutually exclusive. Therefore, research efforts should be directed towards the question of the relative contribution of different explanations and especially towards the question of whether the relative contributions might differ between different cognitive areas. Although such systematic research efforts are not published yet (and indeed are difficult), unsystematic comparisons suggest that math performance and mental rotation might differ with respect to the relative contribution of psycho-social (probably larger for math performance) versus biological-neuronal (probably larger for mental rotation) explanations.

This impression is based on the following observations: whereas the gender differences in the MRT have remained stable over the last decades, those in math performance have declined considerably over time (e.g., Feingold, 1988; Hyde, Fennema, & Lamon, 1990). Whereas gender differences in mental rotation were reported for preschool children (Hahn, Jansen, & Heil, 2010) already (and probably even for infants, see Moore & Johnson, 2008; Quinn & Liben, 2008), those in math performance emerge only around age ten (Hyde et al., 1990). Mental rotation is accompanied by gender-specific lateralization differences of cortical activity (e.g., Hugdahl, Thomsen, & Ersland, 2006). Moreover, the continuous steroid fluctuations in normally cycling women leads to concomitant changes of gender differences both in mental rotation performance (Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis, & Güntürkün, 2000) and in the accompanying cerebral asymmetries (Schöning et al., 2007). By contrast, with respect to math performance, the effects of, inter alia, parent and teacher expectations (Jacobs & Eccles, 1992), family structure (Kao, 1995), motivation (Dweck, 1986), and cultural differences in gender-(in)equality (Guiso, Monte, Sapienza, & Zingales, 2008) have been observed. Moreover, a growing body of research indicates the effects of gender beliefs on math performance both in adults (e.g., Shih, Pittinsky, & Ambady, 1999) and in children

* Corresponding author. Institute of Experimental Psychology, Heinrich-Heine-University, Universitätsstr. 1, 40225 Düsseldorf, Germany. Tel.: +49 211 8112134; fax: +49 211 8115037.
E-mail address: Corinna.Titze@uni-duesseldorf.de (C. Titze).

1041-6080/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.lindif.2010.04.003

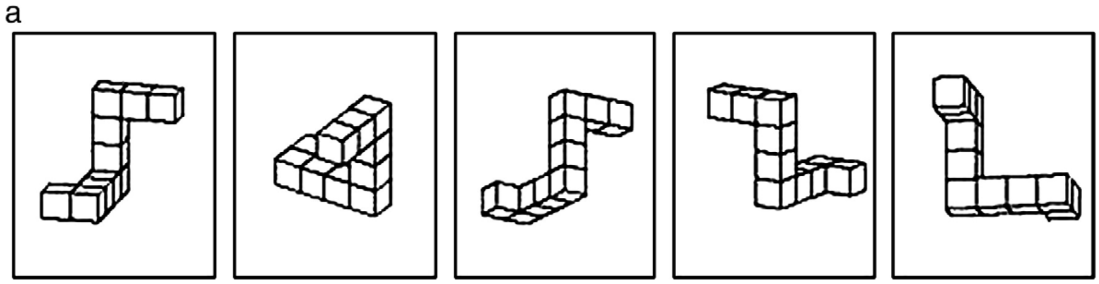


Fig. 1. This figure shows the first item of the MRT. Two of four sample stimuli match the target figure after rotation in depth.

(Ambady, Shih, Kim, & Pittinsky, 2001; Muzzatti & Agnoli, 2007; Neuville & Croizet, 2007). The activation of negative or positive gender beliefs can substantially impede or boost math performance (see also: Aronson, 2004; Nosek, Banaji, & Greenwald, 2002; Steele, 1997).

This comparison, of course, is a (over-)simplification and suffers from the fact that the corresponding counterparts in research are often missing. We simply do not know whether math performance depends upon steroid fluctuations across the menstrual cycle. Until recently, no research existed as to whether mental rotation performance is affected by gender beliefs at all. In the meantime, Moè (Moè, 2009; Moè & Pazzaglia, 2006) reported evidence that manipulating gender beliefs of mental rotation abilities in men and women – by telling them that researchers had found men/women to perform better on visual-spatial tasks – led adult participants to outperform their same-sex counterparts in the MRT. No evidence exists as to whether or not this gender belief effect on mental rotation is observed with children as was the respective effect on math performance (Ambady et al., 2001). This would be especially interesting since "there is considerably agreement that, beginning in the second grade and persisting through the twelfth grade, children perceive social-verbal and artistic skills as feminine, whereas spatial, mechanical, and athletic skills are viewed as masculine." (Nash, 1979, p.265) whereas perceived gender differences for math are less homogenous (e.g., Heyman & Legare, 2004).

Therefore, we investigated the influence of gender beliefs in children following the common stereotype that men are more able than women to solve tasks requiring spatial abilities (see e.g. Nash, 1979; Pease & Pease, 2001). We chose to investigate fourth graders, ten or eleven years old on average, because Ruble and Martin (1998) found that, at this age, the content of gender beliefs begins to resemble the content of adults' stereotypes. This fact might contribute to an expected poorer performance in cognitive tasks if a sex-role stereotype predicts this. As Eccles, Wigfield, and Schiefele (1998) pointed out, there is evidence that children often distort their own expectations towards their stereotypes about which gender is most talented – regarding a particular domain – so as to be consistent with gender beliefs. Moreover, Titze, Jansen and Heil (in press) demonstrated that gender differences in the MRT performance can reliably be found (at least) at age 10. Finally, since the stereotype threat literature describes test difficulty as a crucial feature in gender believe effects (e.g., Neuville & Croizet, 2007), the MRT can be regarded as well suited for children at this age because it is a rather difficult test.

2. Methods

2.1. Participants

252 fourth graders (126 boys and girls each) volunteered (age: $M = 10.47$ years, $SD = .52$). An ANOVA revealed no age difference between both sexes (F (1, 250) = 0.001; $p = .976$; $\eta^2 = .000$). Prior to testing, all parents gave their written consent. Children were divided (class by class) into three groups (all of the same size, i.e. $N = 84$ with 42 boys and girls each). Different experimental manipulations were given to the groups (for a more detailed description see Section 2.3). Children were recruited from primary schools near Düsseldorf. A maximum of 25 children was tested at once and each class was paid 100 € for participation.

2.2. Materials and procedure

All children were asked to complete two mental rotation tasks — the first one prior to and the second one after the experimental manipulation. Usually, the Mental Rotations Test (MRT, Peters et al., 1995) is not used with children younger than 13 years (Linn & Petersen, 1985). In a previous study with 4th graders we had good experience using a simple rotation test (MRT-animal) prior to the MRT that introduced the answer format as well as the concept of mental rotation (see Titze et al., in press). The layout of this test strictly follows the design of the MRT but uses two-dimensional familiar stimuli (Snodgrass & Vanderwart, 1980) rotated in picture plane. Furthermore, we used black and white pictures to slightly increase difficulty (Alington, Leaf, & Monaghan, 1992) and to enhance similarity to the MRT (see Fig. 2). A target animal on the left and four sample stimuli on the right were presented. Children had to identify these two sample stimuli that match the target animal after rotation. Sixteen items had to be solved within 60 s. The MRT-animal was not analyzed since it was only meant to introduce the concept of mental rotation. A specimen test is available from the authors.

Additionally, we used the MRT of Peters et al. (1995) in a shortened version. This test uses three-dimensional abstract cube figures rotated in depth (see Fig. 1). Two sample stimuli match the target figure on the left after rotation. We shortened the MRT for two reasons: first, solving 24 items (as in the original version) before and after the experimental manipulation might have been too demanding for children. Second, we needed to create two halves of comparable difficulty to present one before and one after experimental manipulation. Therefore, in a prior study, we

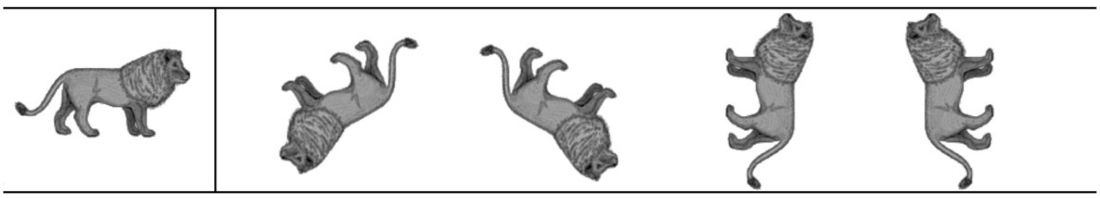


Fig. 2. This figure shows the first item of the two-dimensional MRT-Animal. After rotation in picture plane, two animals match the target animal on the left.

had administered the original version of the MRT to 120 4th graders who completed the MRT under standard conditions either in its version A or B. We then matched the items that had been answered by all children for level of difficulty (i.e. analyzing the percentage of children that answered each item correctly) and built up two versions containing 12 items each. Half of the children started completing the new short version A in the pre-test and version B in the post-test, the other half vice versa. Three minutes were given (following Peters et al., 1995) to solve the items as quickly and accurately as possible. We found that there was no difference regarding the mean number of correctly answered items in the pre-test depending on version (F (1, 250) = 0.311; p = .800).

2.3. Procedure

Children were tested during regular class time and all instructions were given verbally. They were informed that all questionnaires were anonymous, that they will not get any marks and that their teachers will not be allowed to have a look on their data. First, children had to complete the MRT-animal. Afterwards the first half of the MRT was given. Children were then given a five minutes break to move around and to drink something. Thereafter, we presented the experimental manipulation: the first group was told: "There is something we have to tell you about the MRT. Many primary school children have already completed the MRT and we always found that boys scored higher than girls — as you have surly imagined. Now, we want to check whether in this class the boys are better, too"; the second group was told: "There is something we have to tell you about the MRT. Many primary school children have already completed the MRT and we always found that girls scored higher than boys — as you have surly imagined. Now, we want to check whether in this class the girls are better, too". The third group was told: "There is something we have to tell you about the MRT. Many primary school children have already completed the MRT and this questionnaire measures your visual-spatial abilities. You need, for example, these abilities if you want to explain a friend the way to your home". For ethic reason, we decided not to give children a manipulated feedback of their performance in the first half of the MRT. After the instruction has been read out aloud, children had to complete the second half of the MRT. Finally, we informed them about the real intention of this study and answered all questions in detail.

2.4. Data scoring

Given a total sample size of $N = 252$ with 2×3 groups and a desired level of $\alpha = .05$, $\beta = .05$, an effect of size $d = .5$ (i.e. medium effect as defined by Cohen, 1977) could be detected with a probability of $1 - \beta = .95$.[1] The number of correctly answered items in the MRT was the dependent variable.

Peters et al. (1995) proposed the following standard scoring method that was used in the present study as well: one point was given if both and only both correct sample stimuli of a target item were marked correctly. A maximum of 12 points could thus be obtained for each half of the MRT respectively. The correlation between both parts of the MRT was $r = .556$ with $p < .001$.

3. Results

A repeated measures design was applied with the within-subject factor *time* (two levels: immediately before and after experimental manipulation) and two between-subject factors: *instruction* (three levels: boys better, girls better, neutral) and *gender* (two levels: boys, girls).

A repeated measures ANOVA revealed main effects of Gender (F (1, 246) = 12.90; $p < .000$; $\eta^2 = .050$) and Time (F (1, 246) = 57.41; $p < .000$; $\eta^2 = .189$) but did not reveal an interaction of Time × Instruction (F (2, 246) = 2.21; $p = .111$; $\eta^2 = .018$) or Time × Gender (F (1, 246) = .084; $p = .773$; $\eta^2 = .000$) and neither an interaction of Time × Instruction × Gender (F (2, 246) = 1.16; $p = .315$; $\eta^2 = .009$).

In general, boys outperformed girls (boys: $M = 3.79$, $SD = 2.39$; girls: $M = 2.91$, $SD = 2.03$.) with Cohen's $d = .39$ — as an estimate of gender effect size. Independent of this overall gender effect both, girls and boys, improved their performance from the pre-test — immediately before the experimental manipulation — to the post-test — immediately after the manipulation — about one point independent of instruction given (see Fig. 3). Cohen's d value of this repeated measures time effect was $d = .45$ (pre-test: $M = 2.83$, $SD = 1.98$; post-test: $M = 3.87$, $SD = 2.53$).

Additionally, we analyzed our data following the analysis of Moè and Pazzaglia (2006) whose paper built the basis of our study. Again, a repeated measures design was applied with the within-subject factor *time* (two levels: immediately before and after experimental manipulation) and two between-subject factors: *instruction* (three levels: boys better, girls better, neutral) and *gender* (two levels: boys, girls).

We ran the repeated measures ANOVA separately for girls and boys to examine whether the instructions had an influence to the performance of the three groups. For boys a main effect of Time (F (1, 123) = 27.91; $p < .000$; $\eta^2 = .185$) showed but no interaction of Time × Instruction (F (2, 123) = .51; $p = .601$; $\eta^2 = .008$) was found. Although the interactions have not become significant, we conducted post-hoc Duncan test – as Moè and Pazzaglia did – to verify whether our results obtained with children are tendentially in line with the results from their study with adults. But post-hoc tests did not show any differences between the three boys' groups (see Table 1 for accuracy scores).

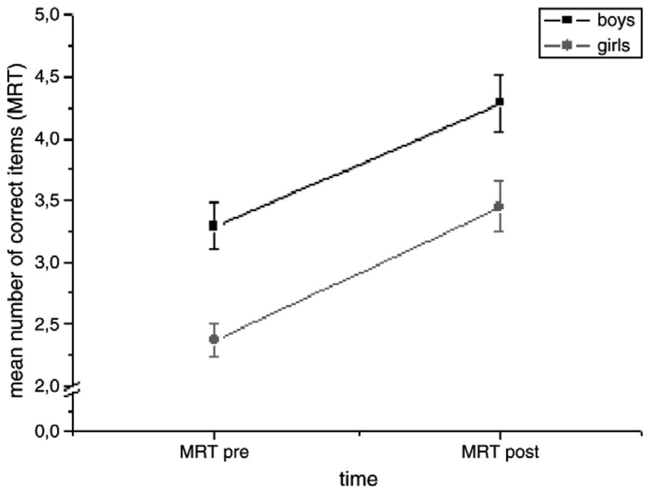


Fig. 3. This figure shows the mental rotation performance in the pre- and post-test as a function of sex.

Table 1
This table presents the accuracy score for the boys: the mean number and its standard deviations of correct answered items in the MRT before and after the instructions are shown. Additionally, Cohen's d – as an estimate of effect size – is given.

Instructions	Accuracy score		
	Before instruction	After instruction	Cohen's d
Boys are better			
Mean	3.83	4.76	.34
S.D.	2.41	3.00	
Girls are better			
Mean	2.86	4.12	.29
S.D.	1.85	2.51	
Gender neutral			
Mean	3.19	4.00	.36
S.D.	2.13	2.32	

Maximum score = 12.

[1] All effect size and power calculations reported in this article were conducted using the G•Power 3 program (Faul, Erdfelder, Lang & Buchner, 2007).

ARTICLE IN PRESS

Table 2
This table presents the accuracy score for the girls: the mean number and its standard deviations of correct answered items in the MRT before and after the instructions are shown. Additionally, Cohen's *d* – as an estimate of effect size – is given.

Instructions	Accuracy score		
	Before instruction	After instruction	Cohen's *d*
Boys are better			
Mean	2.36	3.93	.78
S.D.	1.51	2.41	
Girls are better			
Mean	2.05	3.26	.56
S.D.	1.79	2.45	
Gender neutral			
Mean	2.71	3.17	.23
S.D.	1.68	2.24	

Maximum score = 12.

For the girls, the repeated measures ANOVA revealed a main effect of Time ($F\,(1,\,123) = 29.51$; $p < .000$; $\eta^2 = .194$), too. An interaction of Time × Instruction ($F\ (2,\ 123) = 2.76$; $p = .067$; $\eta^2 = .43$) barely missed to reach significance. Nevertheless, for the girls, post-hoc Duncan test showed that the "boys better" instruction differed significantly from the "neutral" one (see Table 2 for accuracy scores). The "girls better" condition did not differ from "boys better" and "neutral".

4. Discussion

Whereas a substantial body of research demonstrated the effects of gender beliefs on math performance both in adults (e.g., Shih et al., 1999) and in children (e.g., Ambady et al., 2001), only two studies with adults from the same lab (Moè, 2009; Moè & Pazzaglia, 2006) indicated corresponding effects on mental rotation. Since no study existed yet that examined the relation between mental rotation performance and gender beliefs in children, we transferred the conceptual approach used with adults (see e.g. Moè, 2009; Moè & Pazzaglia, 2006) to 10- to 11-years old children.

Children's performance on the MRT, however, was not affected by a gender belief manipulation. Boys' and girls' performance turned out to be independent of the manipulation given. This is an interesting finding, all the more since at the same time we observed significant and medium-sized gender effects on mental rotation performance independent of the gender beliefs manipulation. By contrast, gender belief manipulations affected children's math performance (e.g., Ambady et al., 2001) when at the same time no gender effects on math performance were found. This double-dissociation underlines the social power of stereotype threats.

A possible explanation for the present results is offered by Steele (2003). Although the girls in the "boys better" condition of our study were confronted with negative, stereotypical information about their gender group, they might not have referred themselves as being a member of this group. This phenomenon is called "stereotype stratification" and is proposed to be used especially by girls. The question remains, however, why stereotype stratification should prevent the effects of gender beliefs with respect to mental rotation but should not do so with respect to math performance.

Thus, in contrast to Moè and Pazzaglia (2006), who found that both, women and men, outperformed their same-sex counterparts if they were told that it will be likely that their gender performs better, children did not show this pattern of behaviour. For girls, quite the reverse might be true: post-hoc test indicated that girls may tend to perform better, if they expect boys to outperform them. Nevertheless, this result has to be interpreted with caution because the interaction of Time × Instruction barely failed to reach significance.

Thus, this result should be kept in mind for further investigations without over-estimating the effect incidentally.

Taken together, it seems that with respect to mental rotation adults but not children are affected by gender beliefs (Moè, 2009; Moè & Pazzaglia, 2006). However, one has to keep in mind some subtle differences between the present study and those of Moè: first, we did not ask – prior to the test – whether children would judge mental rotation tasks as typically "boys or girls tasks" before we gave the manipulative instruction since children did not have a distinctive gender stereotype concerning the completely unfamiliar mental rotation task. As a consequence, our approach might have caused children not to believe the gender stereotype given. Second, we did not test boys and girls separately but decided to test children "class-by-class" in the normal setting.

What is definitely needed is more research directly comparing gender belief effects on mental rotation and on math performance in the same children and at different ages. If indeed gender beliefs affect math performance substantially earlier than mental rotation performance, then research has to identify if this discrepancy is exclusively based on differences in familiarity of the tests (high with respect to math and low with respect to mental rotation performance). When this explanation could successfully be excluded, one might start to think more closely about whether the developmental differences in gender belief effects for math versus mental rotation performance might probably reflect differences in the relative contribution of psycho-social versus biological-neuronal classes of explanations for the respective gender effects.

Finally, we found that the simple replication of the MRT lead to a uniform improvement in the performance of both boys and girls. Both groups improved about one point in the second test, although they had to solve different items. This finding indicates that practice can improve girls' performance in the same way as boys' performance (Spence, Yu, Feng, & Marshman, 2009).

Acknowledgements

This study was supported by the German Research Foundation. We thank Michael Peters for his friendly permission to use the Mental Rotations Test (MRT) in our study. We thank all primary school children and their teachers for participating in this study.

References

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*, 539—553.
Ambady, N., Shih, M., Kim, A., & Pittinsky, T. L. (2001). Stereotype susceptibility in children: effects of identity activation on quantitative performance. *Psychological Science*, *12*, 385—390.
Aronson, J. (2004). The threat of stereotype. *Educational Leadership*, *62*, 14—19.
Cohen, J. (1977). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.
Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist*, *41*, 1040—1048.
Eccles, J. S., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology, 5 ed. Social, emotional, and personality development*, *Vol. 3*. (pp. 1017—1095) Hoboken, NJ: John Wiley & Sons Inc.
Faul, F., Erdfelder, E., Lang, A. -G., & Buchner, A. (2007). G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175—191.
Feingold, A. (1988). Cognitive gender differences are disappearing. *American Psychologist*, *43*, 95—103.
Guiso, L., Monte, F., Sapienza, P., & Zingales, L. (2008). Diversity. Culture, gender, and math. *Science*, *320*, 1164—1165.
Hahn, N., Heil, M., & Jansen, P. (2010). Preschoolers' mental rotation: Sex differences in hemispheric asymmetry. *Journal of Cognitive Neuroscience*, *22*, 1244—1250.
Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*, 1—51.
Hausmann, M., Slabbekoorn, D., Van Goozen, S. H., Cohen-Kettenis, P. T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience*, *114*, 1245—1250.
Heyman, G. D., & Legare, C. H. (2004). Children's beliefs about gender differences in the academic and social domains. *Sex Roles*, *50*, 227—236.

Please cite this article as: Titze, C., et al., Mental rotation performance in fourth graders: No effects of gender beliefs (yet?), *Learning and Individual Differences* (2010), doi:10.1016/j.lindif.2010.04.003

ARTICLE IN PRESS

Hugdahl, K., Thomsen, T., & Ersland, L. (2006). Sex differences in visuo-spatial processing: an fMRI study of mental rotation. *Neuropsychologia*, *44*, 1575−1583.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: a meta-analysis. *Psychological Bulletin*, *107*, 139−155.

Jacobs, J. E., & Eccles, J. S. (1992). The impact of mothers' gender-role stereotypic beliefs on mothers' and children's ability perceptions. *Journal of Personality and Social Psychology*, *63*, 932−944.

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition*, *64*, 217−227.

Kao, G. (1995). Asian Americans as model minorities? A look at their academic performance. Feb 1995. *American Journal of Education*, *103*, 121.

Linn, M. C., & Hyde, J. S. (1989). Gender, mathematics, and science. *Educational Researcher*, *18*, 17−27.

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

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences*, *19*, 21−27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences*, *16*, 369−377.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: a sex difference. *Psychological Science*, *19*, 1063−1066.

Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: attitudes and stereotype threat susceptibility in Italian children. *Developmental Psychology*, *43*, 747−759.

Nash, S. C. (1979). Sex role as mediator of intellectual functioning. In M. A. Wittig & A. C. Peterson (Eds.), *Sex-related differences in cognitive functioning* (pp. 263−302). New York: Academic Press.

Neuville, E., & Croizet, J. -C. (2007). Can salience of gender identity impair math performance among 7–8 year old girls? The moderating role of task difficulty. *European Journal of Psychology of Education*, *22*, 307−316.

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Math = male, me = female, therefore math not-equal-to me. *Journal of Personality and Social Psychology*, *83*, 44−59.

Pease, A., & Pease, B. (2001). *Why Men Don't Listen & Women Can't Read Maps.* London: Orion.

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

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science*, *19*, 1067−1070.

Ruble, D. N., & Martin, C. L. (1998). Gender development. In N. Eisenberg & W. Damon (Eds.), *Handbook of child psychology, 5 ed. Social, emotional, and personality development*, *Vol. 3*. (pp. 933−1016) Hoboken, NJ: John Wiley & Sons Inc.

Schöning, S., Engelien, A., Kugel, H., Schafer, S., Schiffbauer, H., Zwitserlood, P., et al. (2007). Functional anatomy of visuo-spatial working memory during mental rotation is influenced by sex, menstrual cycle, and sex steroid hormones. *Neuropsychologia*, *45*, 3203−3214.

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

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype susceptibility: identity salience and shifts in quantitative performance. *Psychological Science*, *10*, 80−83.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174−215.

Spence, I., Yu, J. J., Feng, J., & Marshman, J. (2009). Women match men when learning a spatial skill. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*, 1097−1103.

Steele, C. M. (1997). A threat in the air: how stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613−629.

Steele, J. (2003). Children's gender stereotypes about math: the role of stereotype stratification. *Journal of Applied Social Psychology*, *33*, 2587−2606.

Titze, C., Jansen, P., & Heil, M. (in press). Mental Rotation Performance and the Effect of Gender in Fourth Graders and Adults. *European Journal of Developmental Psychology*. doi:10.1080/17405620802548214.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, *47*, 599−604.

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*, 250−270.

Please cite this article as: Titze, C., et al., Mental rotation performance in fourth graders: No effects of gender beliefs (yet?), *Learning and Individual Differences* (2010), doi:10.1016/j.lindif.2010.04.003

Single-sex school girls outperform girls attending a co-educative school in mental rotation performance

Corinna Titze[1, 2], Petra Jansen[2], and Martin Heil[1]

Correspondence concerning this article should be addressed to:

Corinna Titze, Institute of Experimental Psychology, Heinrich-Heine-University,

Universitätsstr. 1, 40225 Düsseldorf, Germany.

Phone: 0049-211-8112134

Fax : 0049-211-8115037

E-Mail: corinna.titze@uni-duesseldorf.de

Manuscript version: December 2009

[1] Institute of Experimental Psychology, Heinrich-Heine-University, Düsseldorf, Germany

[2] Institute of Sports Science, University of Regensburg, Regensburg, Germany

Abstract

Do girls attending a single-sex school are able to outperform their same-sex counterparts attending a co-educative school solving a mental rotation task? To explore these questions, 252 pupils of two different age groups and school types participated: 126 pupils were 8^{th} and 12^{th} graders each (mean age: 13.63 vs. 17.99 years). Thereof 84 pupils attended a single-sex school, 164 pupils attended a regular co-educative school (84 boys and girls each). All pupils first had to complete the Mental Rotations Test by Peters et al. (1995). Testing of co-educative pupils was separated by gender.

We found, that girls attending a single-sex school were able to outperform their same-sex counterparts at least in grade twelve. In grade eight no differences between both groups were obtained. As expected, the large, well known gender effect between 12^{th} grade boys and co-educative girls was found. Furthermore, we found that boys and single-sex school girls improved their performance from grade eight to twelve significantly – with nearly identical effect sizes, indicating a uniform improvement. Girls attending a co-educative school, however, did not improve their performance. Expectations as well as possible consequences for educators and policy-makers are discussed.

The question, whether boys and girls should be educated together or separated by gender is regarded as "one of the most researched topics in education" (Smithers & Robinson, 2006, p.1). Especially in the fields of mathematics, engineering and science, females still are underrepresented. This is astonishing because level of education assimilated between men and women over past decades (e.g. Herwartz-Emden, 2007) and girls are looked upon as 'winners' of the educational system, not only in Germany:

> "On average, girls enter the school systems earlier, they achieve better at the key competence of reading, they are less likely to leave school without qualification, they are quicker and more successful as regards the transition from school to vocational education and training, they qualify in the more demanding segments of profession, their rates for obtaining a higher education qualification are significantly higher and they drop out of study courses less often..." (see Bundesministerium für Bildung und Forschung, 2008, p.20).

Generally speaking, there still are at least three areas where cognitive gender differences between boys and girls can be found repeatedly: in verbal and mathematical abilities as well as in visual-spatial abilities (see e.g. Linn & Hyde, 1989; Halpern et al., 2007). Whereas in general, females outperform males on verbal tasks, males possibly excel on mathematical performance and definitely do so on visual-spatial abilities – at least if mental rotation performance is required, i.e. the cognitive process of imagining an object turning around (Shepard & Metzler, 1971). Nuttall, Casey and Pezaris (1992; p. 120) assumed "…that gender differences in spatial skills are the key to understanding gender differences in math achievement." Disadvantages of girls in the area of spatial abilities have to attract interest of research all the more since Cooper and Mumaw (1985) pointed out that the predictive validity regarding (especially industrial) job performance of spatial tests is higher than that of measures of general intelligence. Therefore, in this study we focused on the investigation of modulating factors on gender differences in mental rotation performance of school-aged adolescents.

For mental rotation performance differences, two broad classes of explanations are discussed: the "psycho-social" variety (consisting of, e.g., stereotype threat, sex-role identification, or differential experience and socialization) and the biological-neuronal variety (e.g., rate of maturation, genetic complement, sex hormone level, or cerebral lateralization; for details and references of both, see e.g., Voyer, Voyer & Bryden, 1995; Jansen Osmann & Heil, 2007). Each of these explanations can quote empirical support, and they are by no means mutually exclusive.

It has been shown repeatedly, that mental rotation performance increases with increasing age. But males' age-related improvement turned out to be larger than females', leading to the assumption that gender differences increase as a function of age (Voyer et al., 1995; Geiser, Lehmann, & Eid, 2008). Titze, Jansen, and Heil (in press) recently provided evidence that whereas males' mental rotation performance substantially increased between age 9 and age 25, females' performance did not increase at all.

Additionally, research suggests a relation between gender, stereotypes and their interference with intellectual performance. On the one hand, studies showed that women's mental rotation performance increased if their self-concept was pushed via instructions (see Moè & Pazzaglia, 2006; Moè, 2009). On the other hand, Inzlicht and Ben-Zeev (2000) showed that women are susceptible to experiencing problem-solving deficits in the presence of males. Women had to complete a difficult math task in groups of same-sex participants or in mixed groups. Women's performance decreased – following the stereotype-threat theory (Steele & Aronson, 1995) – as the relative number of males present increased. Stereotype-threat is defined as "being at risk of confirming, as self-characteristic, a negative stereotype about one's group" (Steele & Aronson, 1995, p. 797).

Moreover, "there is considerably agreement that, beginning in the second grade and persisting through the twelfth grade, children perceive social-verbal and artistic skills as feminine, whereas spatial, mechanical, and athletic skills are viewed as masculine." (Nash, 1979, p.265). From this point of view, it is obvious that the stereotype threat might contribute to gender differences in stereotyped domains in school: the expectation of girls to confirm a negative stereotype about their group (e.g. "Girls perform worse than boys in a math test."), might be responsible for decrements in cognitive task performance.

How do girls attending a single-sex school, who are therefore not (compared to co-educative schools) confronted with gender related stereotypes in their scholastic daily routine, perform on a "typically" male task? There is evidence that girls attending single-sex schools show decreased bias towards typically male (e.g. math and sciences, especially physics) and female (e.g. verbal) subjects (see e.g. Kessels, 2007). Following Schoon (2001), these girls develop less sex stereotyped interests and self-concepts and were more likely to aspire a career as a scientist compared to girls attending co-educative schools.

To investigate whether a performance difference depending on type of school between girls solving a mental rotation task (that normally is solved best by male participants) can be found, the Mental Rotations Test, MRT (Vandenberg & Kuse, 1978; Peters et al., 1995) was used. With this test, a male advantage can be reliably found and the gender difference amounts to about one standard deviation (Masters & Sanders, 1993; Voyer, Voyer, & Bryden, 1995). The MRT consists of 24 items, each presenting a target figure and four sample stimuli. Two sample stimuli are rotated-in-depth versions of the target figure (see Figure 1) and have to be identified.

We tested the following hypotheses:

(1) We expected girls attending a single-sex school to outperform their same sex counterparts attending a co-educative school significantly.

(2) We hypothesized that gender differences between boys and girls will be reduced (in terms of effect size) if girls attend a single-sex school.

(3) We hypothesized that mental rotation performance increases with increasing age: boys and single-sex school girls should improve equally well whereas girls

attending a co-educative school were not expected to improve significantly (see, Titze, Jansen, & Heil, in press).

Methods

Participants

In this study 252 pupils of two different age groups participated: 126 pupils were 8th and 12th graders each. 84 pupils attended a single-sex school, 164 pupils attended an regular co-educative school (84 boys and girls each). The mean age was M = 13.63 years (SD = 0.51) for the 8th graders and M = 17.99 years (SD = 0.74) for the 12th graders.

Pupils were recruited from academic high schools near Düsseldorf and were tested during normal class time in a group testing session within gender homogenous groups. A maximum of 25 pupils was tested at once. Participation in the testing session was optional, pupils were allowed to break off their participation at any time.

Material and Procedure

All pupils first had to complete the Mental Rotations Test by Peters et al. (1995). The completion of the test took about 15 minutes. The MRT is a redrawn version of the Vandenberg and Kuse (1978) ‘Mental Rotations Test’ with figures provided by Shepard and Metzler (1971). It is made of three-dimensional cube figures in two sets with 12 items each (see, for an example, Figure 1). Six items are presented per sheet of paper. Each item contains a target figure on the left side and four sample stimuli on the right. Two sample stimuli are identical but rotated versions of the target figure. The remaining two sample stimuli do not match regardless of any rotation made. Instructions were given verbally. Pupils were told that they had to mentally rotate the sample stimuli on the right to find out those two sample stimuli matching the target. Instructions were followed by three practice items whose correct solution was given after all pupils finished these items. Participants were given three minutes to solve the first 12 items and another three minutes for the remaining 12. We used the MRT in its versions A and B within the group testing sessions to prevent our results being tampered by participants who might just copy their answers from a neighbor. Stimulus figures in both versions were identical but rearranged to form different tasks. On this account, the level of difficulty does not change (Peters et al., 1995). Testing of co-educative pupils was done separated by gender.

Design and Statistical analysis

Altogether, a 3 (group: co-educative school boys vs. co-educative school girls vs. single-sex school girls) x 2 (age: 8^{th} and 12^{th} graders) design was used. Performance in the MRT (number of correctly answered items) was set as dependent variable whereas group and age were defined as independent variables.

Peters et al. (1995) proposed the following standard scoring method that was used in the present study as well: one point was given if both and only both correct sample stimuli of a target item were marked correctly. A maximum of 24 points could thus be obtained.

Given a total sample size of N = 252 and a desired level of $\alpha = .05$, an effect of size d = .56 (i.e. a medium to large effect as defined by Cohen, 1977) could be detected with a probability of $1-\beta = .95$[1].

Results

An ANOVA revealed a significant main effect of *age* ($F(1, 246) = 6.95$; $p < .05$ $\eta = .027$), indicating that 12^{th} graders performed better on the MRT than 8^{th} graders. Furthermore, a main effect of *group* was found, indicating that boys generally outperformed girls irrespective of attended school ($F(2, 246) = 8.25$; $p < .001$ $\eta = .063$). The interaction of *age*

* *group* turned out to be statistically relevant, too ($F(2, 246) = 4.07$; $p < .05$ $\eta = .032$). Post-hoc tests showed that girls attending an all girls-school outperformed their same-sex counterparts attending a co-educative school in grade 12 ($F(1,82) = 7.05$; $p < .05$ $\eta = .079$; effect size $d = .580$) but not in grade 8 ($F(1,82) = .95$; $p = .331$ $\eta = .012$; effect size $d = .212$). Moreover, grade 12 boys outperformed co-educative girls significantly ($F(1,82) = 22.58$; $p < .001$ $\eta = .216$; effect size $d = 1.036$) but tightly failed to significantly outperform single-sex school girls ($F(1,82) = 3.63$; $p = .06$ $\eta = .042$¸ effect size $d = .414$). However, in the 8^{th} grade boys significantly outperformed girls attending an all-girls school ($F(1,82) = 4.32$; $p < .05$ $\eta = .05$; effect size $d = .452$), whereas no difference between boys and girls attending a co-educative school was found ($F(1,82) = 1.01$; $p = .318$ $\eta = .012$; effect size $d = .221$). Boys improved their performance significantly from grade 8 to grade 12 ($F(1,82) = 6.82$; $p < .05$ $\eta = .077$; effect size $d = .568$), and so did the single-sex school girls ($F(1,82) = 7.02$; $p < .05$ $\eta = .079$; effect size $d = .578$). For girls attending a co-educative school, however, no performance increase between grades 8 and 12 was obtained ($F(1,82) = .71$; $p = .401$ $\eta = .009$; effect size $d = -0.183$), see Figure 2.

Discussion

The findings of the present study are widely in line with the literature and the hypotheses suggested above. Results support the contribution of psycho-social factors to explain gender differences in mental rotation performance.

First, we showed that girls attending a single-sex school outperformed their same-sex counterparts at least in grade twelve whereas in grade eight no difference between both groups was present. In fact, twelfth grade single-sex school girls performed more than half a standard deviation better than girls attending a co-educative school.

Second, as expected, the large, well known gender effect between (12^{th} grade) boys and co-educative girls was obtained. Effect sizes indicate that 12^{th} grade boys performed less than a half a standard deviation (i.e. medium effect as defined by Cohen, 1977) better than single-sex school girls but about a full standard deviation (i.e. a large effect) better than co-educative girls. Regarding performance in grade eight, no gender effect between boys and girls attending the co-educative school was obtained (with effect sizes indicating a small gender effect). In grade twelve, this gender effect increased nearly fivefold. Between boys and single-sex school girls, however, there is a medium gender effect showing in grade eight that did not increase at all in grade twelve.

Finally, boys as well as single-sex school girls improved their performance from grade eight to twelve significantly – with nearly identical effect sizes, indicating a uniform improvement. Girls attending a co-educative school, however, did not improve their performance at all, replicating the recent findings. Titze et al. (in press) found no performance increase between 9 years old girls and 23 years old female students whereas the respective comparison for males revealed a substantial performance increase. In the present study, in grade 12 co-educative girls performed significantly worth than both boys and girls attending the single-sex school, which might be traced back to the lack of performance improvement with age for co-educative girls obtained both here and in Titze et al. (in press).

These results are important, because the mental rotation gender effect, which is described as the largest documented cognitive sex difference, as being stable and reliable over life span (Halpern, 1989) and that is assumed to increase from childhood to adulthood (Linn & Petersen, 1986), seems to be modifiable. Although the medium effect size – comparing 12^{th} grade boys and single-sex school girls – indicated that there still is a relevant effect, the gender effect is halved compared to co-educative girls. This finding supports the assumption that girls – under certain conditions – can improve their performance so that the gender gap might be reduced substantially. Thereby it is especially noteworthy that boys and single-sex school girls improved their performance from grade eight to twelve equally.

The findings of this study are important for several reasons: The results illustrate the influence of psycho-social aspects on mental rotation performance – although biological aspects can presently not be neglected. The role of psycho-social influences, however, must not be underestimated. We provided evidence that the scholastic setting affected girls' performance to a substantial degree. In contrast to Inzlicht and Ben Zeev (2000) – who showed that the sheer temporary absence of males lead to female's improvement on cognitive tasks – we found that girls improved their performance if – and only if – they attended a single-sex school (12th grade). A gender-homogenous testing situation as realized in the present study for all groups was not sufficient since these pupils still showed the well known gender effect in favor of males. An account that might be able to explain these result is the self-to-prototype matching hypothesis (see Niedenthal, Cantor & Kihlstrom, 1985; Hannover & Kessels, 2004). If one has to opt between several alternatives, one first imagines the prototypes that would opt for each of the alternatives (i.e. the type of person who enjoys certain types of activities). The perceived similarity between one prototype and one's (favorite) self serves as an indicator for one's choice: The better the match between prototype and self, the stronger is the preferences to behave like the prototype. Thus, the self-to-prototype matching seems to be a useful cognitive strategy in many domains of social life. Regarding 8^{th} to 9^{th} graders, Hannover and Kessels (2004) found that pupils had built different prototypes according to school subjects. They pointed out: "…, the smaller the discrepancy between an individual's perception of the prototype and his or her self-image the stronger was his or her liking of the respective subject." (Hannover & Kessels, 2004, p. 63). Furthermore, Kessels (2005) showed that pupils expected girls, who excelled in physics, to be less attractive for boys. These results suggest that – especially during adolescence – development of scholastic interests is – at least partly – based upon development of identity: A girl, who wants to be perceived as (an attractive) female by their peers might have to depreciate typically male behavior/ interests. Thus, in terms of school, especially sciences will be judged negatively by girls. But Kessels and Hannover (2008) found that girls reported a better physics-related self concept (i.e. a typically "male subject") if they attended single-sex classes. The authors assumed that single-sex education leads to a less distinct accessibility of gender-related self-knowledge compared to co-educational schooling. This explanation might also be true for our results: girls attending the single-sex school were probably less prejudiced to the "male" mental rotation task. We have to realize, however, that we obtained these results only in grade twelve, but not in grade eight were puberty is most pronounced. To investigate the exclusive impact of single-sex education on mental rotation performance as a typically male task, more data are definitely needed. Our results should not only be replicated, but also expanded to other tasks but the MRT. Moreover, a validation on the basis of a longitudinal study would be quite essential.

Our research definitely should not be interpreted as saying that we should all immediately enroll our daughters in single-sex schools, although our results are in line with studies that focused on distinct academic performance tasks (see e.g., Kessels, 2007). Little is known about the impact of single-sex schools on social behavior or aggression. Additionally, the discussion, whether single-sex schooling should be maintained or whether a separated-by-sex schooling in sciences is a way to ease girls the access to sciences is widely discussed and not only an educational but also a political question. Our results, however, provide some additional food for thought, and research since little is known about the effect of pupils' age when single-sex schooling is introduced.

Acknowledgements:

This study was supported by the German Research Foundation. We thank Michael Peters, for his friendly permission to use his tests in this study. Furthermore, we thank all pupils and teachers of the Ursulinenschule Bonn-Hersel, the Theodor-Fliedner-Gymnasium Kaiserswerth, the Goerres-Gymnasium Düsseldorf and the Luisengymnasium Düsseldorf for their support and participation.

References

Bundesministerium für Bildung und Forschung (BMBF). (2008). *Berufsbildungsbericht 2008 [Education in Germany 2008]*. Bonn/Berlin: BMBF.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Cooper, L. A., & Mumaw, R. J. (1985). Spatial Aptitude. In R. F. Dillon (Ed.), *Individual Differences in Cognition* (Vol. 2, pp. 67-94). New York: Academic Press.

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

Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence, 36*, 556-563.

Halpern, D. F. (1989). The disappearance of cognitive gender differences: What you see depends on where you look. *American Psychologist, 44*, 1156-1158.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest, 8*, 1-51.

Hannover, B., & Kessels, U. (2004). Self-to-prototype matching as a strategy for making academic choices. Why high school students do not like math and science. *Learning and Instruction, 14*, 51-67.

Herwartz-Emden, L. (2007). "Schulkultur, Geschlechtersegregation und Mädchensozialisation" - eine Studie über Mädchenschulen ["Schooling culture, gender segregation and socialisation of girls" - a study about single-sex girls schools]. In L. Herwartz-Emden (Ed.), *Neues aus alten Schulen - Empirische Studien in Mädchenschule [News from old schools - empirical studies in single-sex girls schools]* Opladen/ Farmington Hills: Barbara Budrich Verlag.

Inzlicht, M., & Ben Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. *Psychological Science, 11*, 365-371.

Jansen Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition, 64*, 217-227.

Kessels, U. (2005). Fitting into the stereotype: How gender-stereotyped perceptions of prototypic peers relate to liking for school subjects. *European Journal of Psychology of Education, 20*, 309-323.

Kessels, U. (2007). Identifikation mit naturwissenschaftlichen Fächern: Ein Vergleich von Schülerinnen einer monoedukativen und einer koedukativen Schule [Identification with sciences: a comparison of girls attending a single-sex and a co-educative school]. In L. Herwartz-Emden (Ed.), *Neues aus alten Schulen - Empirische Studien in Mädchenschule [News from old schools - empirical studies in single-sex girls schools]* Oplanden/ Farmington Hills: Barbara Budrich Verlag.

Kessels, U., & Hannover, B. (2008). When being a girl matters less: Accessibility of gender-related self-knowledge in single-sex and coeducational classes and its impact on students' physics-related self-concept of ability. *British Journal of Educational Psychology, 78*, 273-289.

Linn, M. C., & Hyde, J. S. (1989). Gender, Mathematics, and Science. *Educational Researcher, 18*, 17-27.

Linn, M. C., & Petersen, A. C. (1986). A meta-analysis of gender differences in spatial ability: Implications for mathematics and science achievement. In J. S. Hyde & M. C. Linn (Eds.), *The psychology of gender: Advances through meta-analysis* (pp. 67-101). Baltimore: The Johns Hopkins University Press.

Masters, M. S., & Sanders, B. (1993). Is the gender difference in mental rotation disappearing? *Behavior Genetics, 23*, 337-341.

Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19*, 21-27.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369-377.

Nash, S. C. (1979). Sex role as mediator of intellectual functioning. In M. A. Wittig & A. C. Peterson (Eds.), *Sex-related differences in cognitive functioning* (pp. 263-302). New York: Academic Press.

Niedenthal, P. M., Cantor, N., & Kihlstrom, J. F. (1985). Prototype matching: A strategy for social decision making. *Journal of Personality and Social Psychology, 48*, 575-584.

Nuttal, R. L., Casey, M. B., & Pezaris, E. (2004). Spatial Ability as a Mediator of Gender Differences on Mathematics Tests: A Biological-Environmental Framework. In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: an integrative psychological approach* (pp. 121-142). Cambridge: Cambridge University Press.

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

Schoon, I. (2001). Teenage job aspirations and career attainment in adulthood: A 17-year follow-up study of teenagers who aspired to become scientists, health professionals, or engineers. *International Journal of Behavioral Development, 25*, 124-132.

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

Smithers, A., & Robinson, P. (2006). *The paradox of single-sex and co-educational schooling*. Buckingham: Carmichael Press.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69*, 797-811.

Titze, C., Jansen, P., & Heil, M. (in press). Mental Rotation Performance and the Effect of Gender in Fourth Graders and Adults. *European Journal of Developmental Psychology*.

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

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*, 250-270.

Footnotes

[1] All effect size and power calculations reported in this article were conducted using the G•Power 3 program (Faul, Erdfelder, Lang & Buchner, 2007).

Figure Captions

Figure 1: This figure shows the first task from version A and version B of the MRT. The participant has to mentally rotate the four sample stimuli on the right to decide whether it matches the target item or not. Two out of four alternatives match to the target item.

Figure 2: This figure shows the mean performance in the MRT as function of age and test-mode. Error bars represent standard errors respectively.

Figure 1:

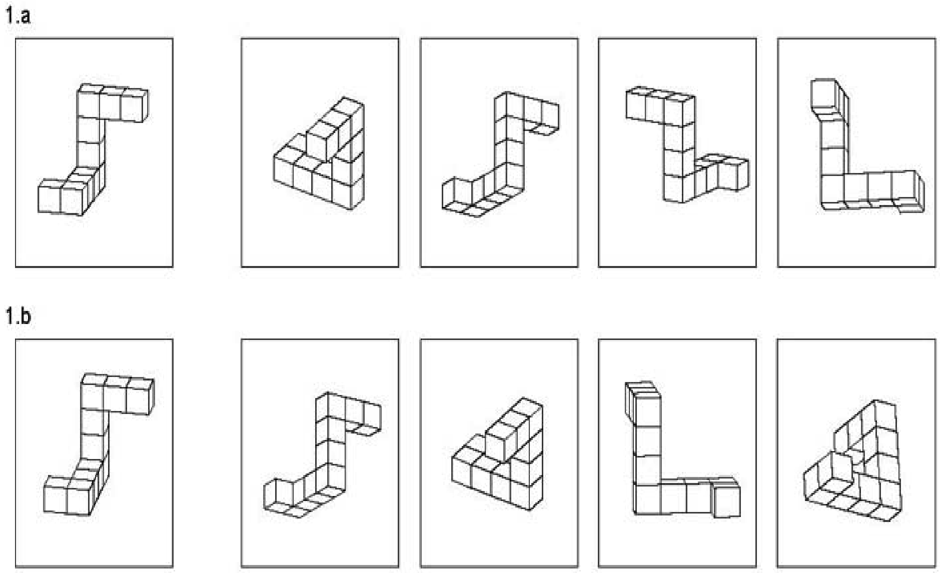
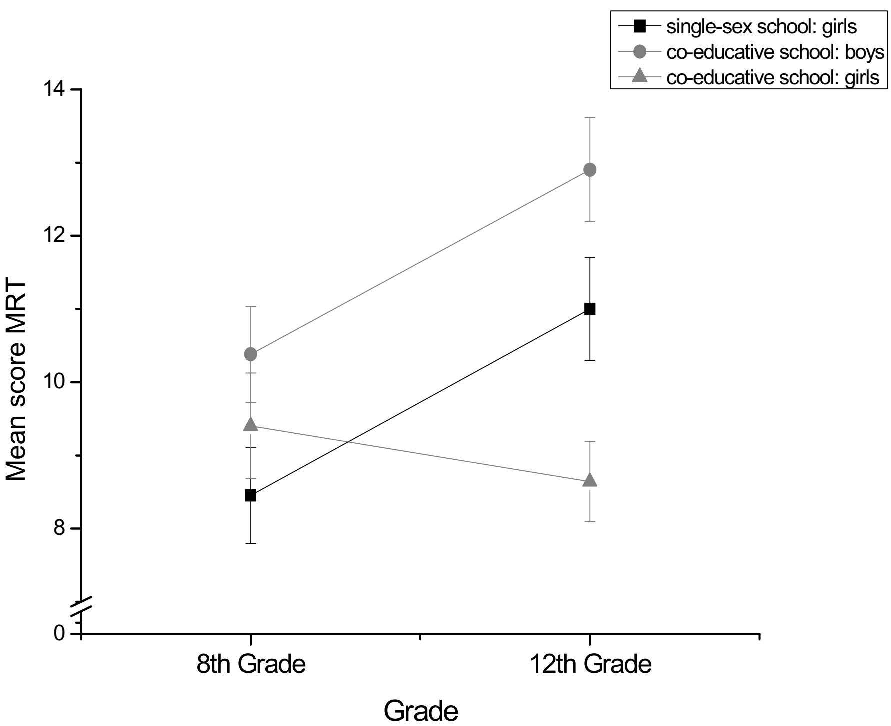


Figure 2:



Gender Differences in the Mental Rotations Test (MRT) Are Not Due to Task Complexity

Corinna Titze, Martin Heil, and Petra Jansen

Institute of Experimental Psychology, University of Düsseldorf, Germany

Abstract. Gender differences are one of the main topics in mental rotation research. This paper focuses on the influence of the performance factor task complexity by using two versions of the Mental Rotations Test (MRT). Some 300 participants completed the test without time constraints, either in the regular version or with a complexity reducing template creating successive two-alternative forced-choice tasks. Results showed that the complexity manipulation did not affect the gender differences at all. These results were supported by a sufficient power to detect medium effects. Although performance factors seem to play a role in solving mental rotation problems, we conclude that the variation of task complexity as realized in the present study did not.

Keywords: gender differences, visual-spatial cognition, mental rotation, task complexity

Whereas females outperform males on, for example, measures of verbal fluency, males outperform females on certain tests of spatial ability (e.g., Halpern, 1992). This male advantage is largest on tasks involving mental rotation (Voyer, Voyer, & Bryden, 1995), and within these, it is largest on the MRT (Mental Rotations Tests; Vandenberg & Kuse, 1978; Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson, 1995). The MRT consists of 24 items, each presenting a 3-D target block figure (taken from Shepard & Metzler, 1971) and four choice figures. Two of these are identical to the target but are rotated in depth, while two can not be matched regardless of how they are rotated. Usually (see Voyer et al., 1995), a single point is given if and only if both correct matches are identified. The test is usually presented in two sets of 12 items each, with a 3-min time constraint for each set.

Despite an enormous amount of research, the cause(s) for the gender difference, however, are still far from being understood. The explanations being offered include psychosocial ones (such as stereotype threat, Shih, Pittinsky, & Ambady, 1999; sex role identification, Signorella & Jamison, 1986; or differential experience and socialization, Baenninger & Newcombe, 1989) and biological-neuronal ones (such as sex hormone level, Imperato-McGinley, Pichardo, Gautier, Voyer, & Bryden, 1991; rate of maturation, Sanders & Soares, 1986; or cerebral lateralization, McGlone, 1980) and, for all of them, a certain amount of empirical support exists.

At the same time, however, these explanations run the risk of overgeneralization and, thus, it might help to identify in detail the empirical facts that have to be explained, especially to identify which task factors do and do not affect the size of the gender difference. For our present study, two task factors are especially relevant, i.e., time constraints and task complexity. Goldstein, Haldane, and Mitchell (1990) reported findings that the gender difference on the MRT disappears when subjects were allowed sufficient time to attempt all items or when the scoring procedure was controlled for the number of items attempted. In contrast to Goldstein et al. (1990), however, and in line with the majority of the published data (see, e.g., Delgado & Prieto, 1996; Resnick, 1993), Masters (1998) showed that the gender difference was *not* affected by time constraints. Neither the scoring method nor the time limits used modified the size of the gender difference, a result which was also obtained by Peters (2005). Peters found evidence that although females attempted fewer items than males under the standard timing condition, the magnitude of the gender difference did not change when subjects did the MRT with double the usual time allowed for the test. Thus, one can conclude that time constraints are not critical for gender differences in the MRT.

Additionally, the Peters study presented one very interesting and, hitherto, not sufficiently considered finding that might be crucial for understanding the causes for the gender difference in the MRT: In Experiment 3 of Peters (2005), a selected sample of subjects dealt with the MRT twice. These subjects were drawn from a larger sample pretested on the MRT and were selected on the basis of scoring within one standard deviation of their gender mean. As a consequence, gender differences in both the initial and the second MRT administration resulted in an effect size of Cohen's (1977) d greater than 1. In between the two MRT administrations, subjects solved a chronometric version of

DOI 10.1027/1614-0001.29.3.130
© 2008 Hogrefe & Huber Publishers

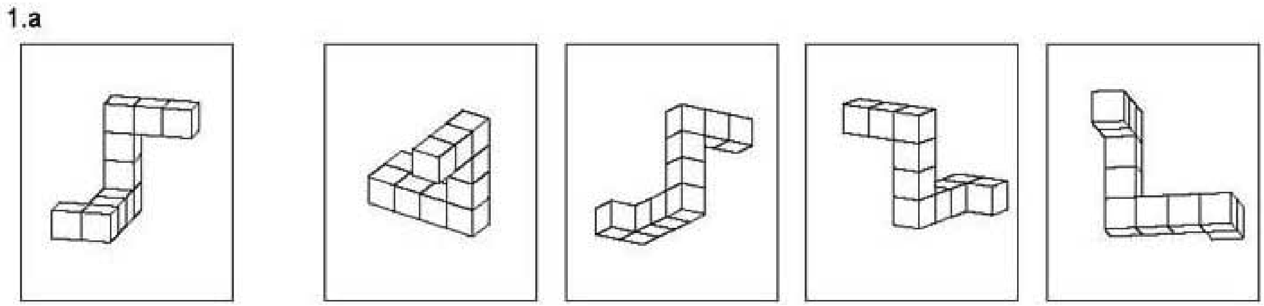


Figure 1. An example of an item used in the Mental Rotations Test. The target item is shown on the left and the four sample stimuli are presented aside. Always two of these are identical to the target item but are rotated in depth.

the mental rotation task with pairs of cube stimuli of the kind encountered in the MRT. In chronometric studies of mental rotation, two stimuli are presented with varying angular disparity, and response times and error rates are measured when participants decide whether these two match when mentally aligned. Interestingly, in this sample preselected to establish substantial and reliable gender differences in the MRT, no gender difference at all was observed for the chronometric version of mental rotation (see also Jansen-Osmann & Heil, 2007, for corroborative results). This finding raises the (important) question of the task factors that differ between the MRT and the chronometric version of mental rotation and that might be responsible for the (non)existence of gender differences in performance. Thus, identifying these task factors might be crucial to understanding the causes of the gender difference.

Obviously, the MRT and the chronometric version of mental rotation differ in a number of aspects. The MRT is a paper and pencil test that differs from a computer-administered task. On the one hand, the computer task forces participants to respond to all items whereas with the paper-pencil test, participants have greater control how they manage their time and effort. For example, they could decide to spend more time on difficult items, to skip these but return to them later, double-check the answers at the end, etc. On the other hand, only one item has to be compared with another one in the computer task, whereas in the MRT participants have to compare one item with four items. This task is much more complex than the two-alternative forced-choice task because all five items are presented at the same time, look very similar, and might give the impression of a very complex and difficult task. This impression might evoke more strongly the stereotype of being unable to solve spatial problems in women (compare Shih et al., 1999).

To investigate this phenomenon of task complexity we compared two groups of participants with one solving the MRT in the standard version but without any time limit. The second group solved the MRT with the aid of a template that, per item, breaks down the two-out-of-four alternatives choice into four consecutive two-alternative forced-choices, similar to the tasks in the computer tests.

Methods

Participants

In this study 300 subjects (150 women) participated. Their age ranged from 18 to 35 years (*M* = 24). This homogeneous group of participants was recruited on campus and needed to have the school-leaving examination to be allowed to take part in this experiment. Half of the subjects were assigned randomly to the template group; the other half to the nontemplate group.

Material and Procedure

We used the paper-pencil MRT, (Version A) redrawn by Peters et al. (1995), which was originally developed by Vandenberg and Kuse (1978) with figures created by Shepard and Metzler (1971). This test consists of two sets with 12 items that contain respectively a target item on the left side and four sample stimuli on the right. Participants had to identify those two out of four sample stimuli that show the target item in a rotated version. Figure 1 shows an example of the items used.

In the original test of Peters et al. (1995), the items were presented to the participants on four DIN A-4 sheets, with six items per sheet and a 3-min deadline to solve a set of 12 items (6 min for the entire test). Instructions were given in written form, followed by three training items so that participants became familiar with the task. The correct solutions of these training items were shown at the end of the page. We used the original test but abandoned any time limit, i.e., participants were allowed as much time as needed to solve all 24 items. Moreover, participants were explicitly instructed to attempt a solution to all items.

In the template group, participants solved the MRT with the help of a 50 × 30 cm black board template with a 5 × 20 cm horizontal whole in its middle. The template allowed participants to watch exactly one item at a time. Additionally we used two more black board templates measuring 5 × 7 cm and 5 × 4 cm, respectively. These templates were constructed to mask three out of the four sample stimuli to

© 2008 Hogrefe & Huber Publishers

ensure that subjects always compared just one sample stimulus with the target item. In this way, the two-out-of-four alternatives choice usually inherent in the MRT was broken down into four consecutive two-alternative forced-choices that subsequently were checked to fulfill the criterion of two positive and two negative choices.

The participants were tested individually. An investigator was present in each test session in both experimental groups. First, participants read the instructions and solved the three training items. Participants in the template group were not allowed to inspect all four sample stimuli before solving an item. They just saw the target item on the left side all the time and only one by one of the sample stimuli. By using the template, three of the four sample stimuli were masked by the investigator so that the participant had to compare each sample stimulus individually with the target item. When the participant had made his or her decision, this stimulus was masked and the next one was uncovered until all four sample stimuli were decided. When one item had been finished, participants checked with the help of the investigator if exactly two answers were marked as correct. If there were correct answers missing or if more than two sample stimuli were marked correct, the participant was asked to correct his or her choices but was now allowed to view all four sample stimuli at the same time, so that they had to fulfill again a two-alternative forced-choice task. To be sure that participants could not see the test as a whole the covering templates were operated by the investigator.

Statistical Analysis

The standard scoring method proposed by Peters et al. (1995) was used: One point was given if and only if both correct sample stimuli of a target item were marked correctly. Thus, participants could obtain 24 points maximum.

The design of the study involved two factors: Gender (male, female) and Experimental group (with or without template). The dependent variable was the number of correctly answered items in the MRT-A. Given a total sample size of $N = 300$ and a desired level of significance $\alpha = .05$, effects of size $d = 0.5$ (that is, medium effects as defined by Cohen, 1977) could be detected with a probability $1\text{-}\beta = .95$[1].

Results and Discussion

An ANOVA revealed a significant main effect of Gender with $F(1, 296) = 13.26, p < .001$; men gaining in mean two points more than women. The effect size of this gender difference amounted to $d = 0.42$. No main effect of factor Experimental group, $F(1, 296) = .32$, *ns*, was found. There also was no significant interaction between both factors, $F(1, 296) = .07$, *ns*. The effect sizes did not differ significantly depending upon whether participants did or did not use the complexity reducing template ($d = 0.45$ vs. $d = 0.40$). The means and their standard errors are presented in Table 1.

Table 1. Means and standard errors in the MRT as a function of subjects' gender and experimental group

Gender	Complexity of stimuli	Mean correct	Standard error
Male (*N* = 150)	with pattern (*N* = 75)	21.21	.54
	without pattern (*N* = 75)	20.76	.50
	total	20.99	.37
Female (*N* = 150)	with pattern (*N* = 75)	19.01	.56
	without pattern (*N* = 75)	18.91	.59
	total	18.99	.40
Total (*N* = 300)	with pattern (*N* = 150)	20.14	.40
	without pattern (*N* = 149)	19.83	.39
	total	19.98	.28

Discussion

The present study aimed to identify relevant task factors, such as task complexity, which might affect the size of the gender difference in mental rotation. Recently, Peters (2005) observed no gender differences at all for the chronometric version of mental rotation with pairs of cube stimuli similar to the ones used in the MRT in a sample preselected to establish substantial and reliable gender differences in exactly this MRT. The absence of gender differences in a same/different judgment with comparable stimuli in a sample that reliably produced a gender-effect size of about one standard deviation in the MRT is an intriguing finding that deserves to be explained. Moreover, the reason for this discrepancy might also turn out to be a prime candidate for explaining gender differences in spatial cognition in general.

Obviously, the MRT and the chronometric version used by Peters (2005) differ in a number of aspects, and substantial research efforts are needed to identify the relevant task factors. In this paper, we concentrated on one important aspect, i.e., the complexity of the task, especially the decision required. Whereas the MRT asks for two-out-of-four alternative choices, the chronometric versions usually use a two-alternative forced-choice task. Our aim was to realize a test situation comparable to the standard MRT as much as possible. Therefore, we used the original paper-and-pencil MRT und reduced the complexity of the experimental group from the two-out-of-four alternative choices to four consecutive two-alternative choices, which subsequently had to be checked to fulfill the criterion of two positive and two negative ones. This manipulation capitalized on the findings of, e.g., Masters (1998) and Peters (2005) that the gender differ-

[1] All power calculations reported in this article were conducted using the G·Power program (Erdfelder, Faul, & Buchner, 1996).

 © 2008 Hogrefe & Huber Publishers

ence in the MRT was not affected by time constraints. In line with these results, we obtained a reliable gender difference of medium effect size according to the definition of Cohen (1977) although (1) subjects were tested individually, (2) no time constraint was realized at all, and (3) all subjects attempted to solve all items. The complete removal of time constraints, in fact, directly asks for an individual testing situation. As a consequence, the average number of 20 problems solved correctly was substantially larger than what is usually observed with time constraints in a group-testing situation (e.g., about 12 problems in Peters, 2005). Nevertheless, although all subjects in both the control and the experimental group attempted to solve all items, men, on average correctly, solved two more problems than women did.

The manipulation of task complexity, however, did not affect the size of the gender difference at all. This conclusion is supported by the power of 96% to detect an effect of complexity. Whatever the relevant task factor is to explain the intriguing finding of Peters (2005), task complexity does not seem to be critical. More research is needed to solve this puzzle. We started with the aim of eliminating the gender difference in the MRT by introducing as few modifications to the MRT as possible, therefore, we did not split the MRT in separate two-alternative forced-choice tasks on a computer screen. However, an alternative approach might be to start from the chronometric version of the mental rotation task to identify what modifications are needed to evoke the gender difference that is absent in the chronometric version but is reliable in the MRT. With this paper we made a first contribution to the research efforts that are needed – and as a result, we have to exclude task complexity as realized here as a factor responsible for the gender difference in the MRT.

Acknowledgments

This study was supported by the German Research Foundation. We thank Michael Peters for his friendly permission to use the Mental Rotations Test (MRT) in our study. We thank our student assistants for help during data acquisition.

References

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

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Delgado, A.R., & Prieto, G. (1996). Sex differences in visuospatial ability: Do performance factors play such an important role? *Memory and Cognition, 24*, 504–510.

Erdfelder, E., Faul, F., & Buchner, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, and Computers, 28*(1), 1–11.

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition, 18*, 546–550.

Halpern, D.F. (1992). *Sex differences in cognitive abilities*. Hillsdale, NJ: Erlbaum.

Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., & Bryden, M.P. (1991). Cognitive abilities in androgen insensitive subjects – Comparison with control males and females from the same kindred. *Clinical Endocrinology, 34,* 341–347.

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition*, *64*, 217–227.

Masters, M.S. (1998). The gender difference on the Mental Rotations test is not due to performance factors. *Memory and Cognition*, *26*, 444–448.

McGlone, J. (1980). Sex differences in human brain asymmetry: A critical survey. *Behavioral and Brain Sciences, 3*(2), 215–263.

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition*, *57*, 176–184.

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

Resnick, S.M. (1993). Sex differences in mental rotations: An effect of time limits? *Brain and Cognition, 21*, 71–79.

Sanders, B., & Soares, M.P. (1986). Sexual maturation and spatial ability in college students. *Developmental Psychology*, *22*, 199–203.

Shih, M., Pittinsky, T.L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science*, *10*, 80–83.

Shepard, R.N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703.

Signorella, M.L., & Jamison, W. (1986). Masculinity, femininity, androgyny, and cognitive performance: A meta-analysis. *Psychological Bulletin*, *100*, 207–228.

Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.

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*, 250–270.

Corinna Titze

Institute of Experimental Psychology
Heinrich-Heine-University
Universitätsstr. 1
D-40225 Düsseldorf
Germany
Tel +49 211 811-2134
Fax +49 211 811-5037
E-mail corinna.titze@uni-duesseldorf.de

© 2008 Hogrefe & Huber Publishers

Original Article

Pairwise Presentation of Cube Figures Does Not Reduce Gender Differences in Mental Rotation Performance

Corinna Titze[1,2], Martin Heil[1], and Petra Jansen[2]

[1]Institute of Experimental Psychology, University of Düsseldorf, Germany
[2]Institute of Sport Science, University of Regensburg, Germany

Abstract. Gender differences still are one of the main topics in mental rotation research. Quite a number of different approaches aim to uncover the reasons for the substantial effect sizes observed. In this paper, we focus on the performance factor *task complexity*, which may contribute to gender differences. A pairwise paper-pencil presentation mode – using the original but rearranged items of the classic MRT by Peters et al. (1995) – was chosen to investigate mental rotation performance of adults. A total of 72 participants were asked to complete a complexity reduced version of the MRT: They had to complete simple "same-different" judgments without any time constraints instead of regular "two-out-of-four-alternatives" choices. Results revealed that the reduction of complexity did not affect the gender differences at all: Men outperformed women in both accuracy and speed. The reasons for these results are discussed.

Keywords: gender differences, visual-spatial cognition, mental rotation, task complexity

Introduction

In the early 1970s, Shepard and Metzler (1971) developed the first chronometric mental rotation test using two-dimensional drawings of three-dimensional cube figures rotated in depth. Participants were presented pairs of cube figures and asked to decide as quickly and accurately as possible whether these were identical or mirror-reversed versions. At that time, no results concerning gender differences were presented – although these cube figures later turned out to produce the most stable and reliable gender differences found in cognition (Voyer, Voyer, & Bryden, 1995). Vandenberg and Kuse (1978) used the Shepard and Metzler (1971) figures to develop the first paper-pencil Mental Rotations Test (MRT). Participants had to identify two correct stimuli out of four alternatives that matched the target item. Stimuli were rotated about the vertical axis.

Using this test, clear evidence for a gender difference favoring males emerged. Two meta-analyses (Linn & Petersen, 1985; Voyer et al., 1995) validated this phenomenon by reporting large effect sizes (see Cohen, 1977) for the MRT of about $d = 1.0$, which clearly exceeded effect sizes found if other mental rotation tests were used. Peters et al. (1995) redrew the original Vandenberg and Kuse's test without modifying the design and the concept of the test (see Figure 1): Participants have to identify two among the four alternatives that matched the target. In its versions A and B, stimuli were rotated about the horizontal or the vertical axis. Compared to the original Vandenberg and Kuse (1978) study, effect sizes did not change discernibly (Peters et al., 1995). In the following, we focus on Peters' version of the MRT because most actual studies applied this test rather than the Vandenberg and Kuse's (1978) original version.

Although effect sizes are impressive and various explanatory approaches for the occurrence of the gender differences were offered – starting from biological-neuronal (e.g., Hausmann, Slabbekoorn, Van Goozen, Cohen-Kettenis, & Güntürkün, 2000; Imperato-McGinley, Pichardo, Gautier, Voyer, & Bryden, 1991) to psychosocial (Moè & Pazzaglia, 2006; Shih, Pittinsky, & Ambady, 1999) – the question remains whether gender differences can be modulated by changing task factors that rely on the MRT itself. There are at least three task factors that might be relevant: The *task complexity* and the *test mode*, which are usually linked ("same-different" judgments as required in chronometric computer administered tasks vs. "two-out-of-four-alternatives" choices as demanded in the paper-pencil MRT) as well as conducting test sessions with and without *time constraints* might influence the magnitude of the gender differences found.

For the latter case, there are somewhat inconsistent findings. Goldstein, Haldane, and Mitchell (1990) found that gender differences disappeared if participants were allowed sufficient time to attempt all items. In contrast, Masters (1998) showed that gender differences were *not* affect-

© 2010 Hogrefe Publishing

DOI 10.1027/1614-0001/a000018

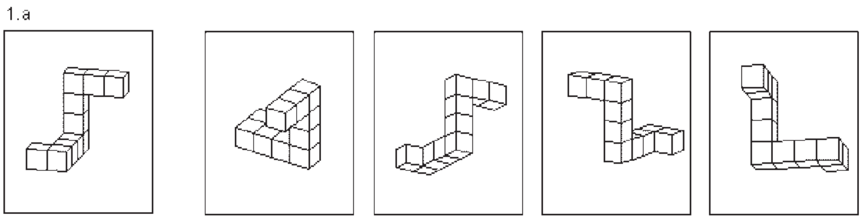


Figure 1. An example of an item used in the Mental Rotations Test (MRT-A). The target stimulus is shown on the left, and the four sample stimuli are presented aside. Two of these are identical but rotated versions of the target stimulus.

ed by time constraints (for corroborative results see, e.g., Delgado & Prieto, 1996; Resnick, 1993). Furthermore, Peters (2005) found evidence that, although females attempted fewer items than males under standard timing condition, the magnitude of the gender difference did not change when participants did the MRT with double the usual time allowed for the test.

We especially focused on the factor task complexity, i.e., the number of alternatives that have to be compared to the target item. In the paper-pencil MRT (Peters et al., 1995), participants have to compare four alternatives with the target item and to decide which two match the target item. This task is more complex then comparing, say, one alternative with one target item in a pairwise presentation task used in chronometric studies. There is evidence that gender differences decrease – or disappear altogether – if the cube figures are presented pairwise in a computer-administered chronometric task. In Peters' (2005) Study 3, participants solved a chronometric version of a mental rotation task with same-different judgments for pairs of the MRT cube stimuli, whereby no gender differences at all were observed regarding reaction times (there are no separate statistics for accuracy because it was close to 100%). Similarly, Jansen-Osmann and Heil (2007) did not find any gender differences when presenting cube figures pairwise in a chronometric study. Consequently, it might be possible to eliminate gender differences in the paper-pencil MRT by presenting the cube figures pairwise. Additionally, from another point of view it might be interesting to realize a complexity-reduced version of the MRT: Voyer and Saunders (2004) pointed out that men and women differ regarding their propensity to guess. In the standard MRT, participants have to cross out two of four answers, with men having a higher propensity to guess whereas women show a greater reluctance (i.e., resulting in more blank responses). Thus, using a complexity-reduced version where participants definitely have to decide between two alternatives might counterbalance the gender differences in guessing. Furthermore, Peters and Battista (2008) pointed out that the performance differences found in previous studies comparing a pairwise presentation to a two-out-of-four-alternatives task might be partially explained by differences in "nonuniformity of stimulus material," "required memory load," and "interindividual differences" (for details, see Peters & Battista, 2008). Therefore, it is necessary to control these factors thoroughly in future studies.

In a first attempt, we recently tried to assimilate the complexity of the paper-pencil MRT to the chronometric pairwise presentation mode (Titze, Heil, & Jansen, 2008) while introducing as few changes as possible concerning the design of the MRT. We used the MRT in its standard format and, additionally, a template that created consecutive same-different judgments. It allowed the participants to compare the target item on the left with exactly one item of the four sample stimuli. As in a chronometric task, participants had to decide whether both stimuli presented were identical or different. After participants had finished one comparison, the next sample stimulus was uncovered until all decisions were made. In this way, we aimed to break down the two-out-of-four-alternatives choice into four consecutive choices for each item. Participants were told that, although they had to do four single comparisons for each item, two out of four stimuli match the target. An investigator checked that this criterion had been fulfilled for each item. Nevertheless, this manipulation did not affect gender differences compared to a standard, nontemplate group: Men still outperformed women significantly. Unfortunately, however, it is not fully clear whether or not the reduction of complexity by the manipulation chosen was fully successful. Since the investigator was present all the time, this might have been a stressful situation that may have led to poor results – especially among the female participants.

Therefore, in this study we developed a pairwise paper-pencil-test and reduced the task complexity consistently by breaking down the two-out-of-four-alternatives choices into four consecutive "same-different" judgments – identical to the chronometric computer administered tasks. Thus, we rearranged the original stimuli of the MRT to create pairs of cube stimuli. Furthermore, no time constraints at all were applied. Finally, participants were tested individually to avoid a gender-competitive situation.

Methods

Participants

Altogether, 72 adults (36 women/36 men) participated. Their age ranged from 19 to 42 years (M = 24.25, SD = 4.8). Participants were recruited on a university campus and needed to have had a final examination to be allowed to take part in this experiment. Participants older than 45 years were not allowed to take part. Furthermore, dexterity was required. Participants were paid € 3 for their participation.

 © 2010 Hogrefe Publishing

Material and Procedure

For this study we created a modified version of the paper-pencil Mental Rotations Test (MRT, Version A) redrawn by Peters et al. (1995). The original version of the test consists of two sets of 12 items each contain a target stimulus on the left side and four sample stimuli on the right, respectively. Two of these were identical but rotated versions of the target stimulus, whereas the others did not match regardless of the rotation made. Participants had to identify the sample stimuli that matched the target one. Figure 1 shows an example.

The new paper-pencil version was constructed as follows: Instead of 24 two-out-of-four-alternatives choices (as in the original MRT), the new version consists of 24 × 4 pairwise same-different judgments, split up in two sets again. Each of the 24 original target stimuli was printed four times with one of the four original sample stimuli at a time. Afterwards, the new pairwise items were arranged in random order. Participants had to mark with a cross the *same* box beside the stimuli if both stimuli were identical and the *different* box if they were not. Figure 2 shows an example. Thus, participants had to solve 96 same-different judgments in total.

In line with the original test, instructions were given in written form. Three training items followed to familiarize the participants with the task. The correct solutions of the training items were given at the end of the page. In contrast to the original test, no time limit (i.e., usually 3 min. per set) was applied. As in chronometric studies, participants were allowed as much time as needed to solve each item. The overall time needed to finish the full test was taken by the experimenter. Participants were tested individually and were instructed to answer as quickly and accurately as possible. The completion of the test took between 6.0 and 45.8 minutes (M = 19.01, SD = 8.36).

Statistical Analysis

Two different scoring methods were used: First, one point was given for each correct **s**ame-**d**ifferent **j**udgment (SDJ) for a maximum of 96 possible points. Second, the standard scoring method proposed by Peters et al. (1995) – that is valid for the test with 24 two(**2**)-**o**ut-**o**f-four(**4**)-alternatives choices – was applied to the new version (2oo4): We rearranged the 96 pairs of items to recreate the 24 original items. One point was given if (and only if) both sample stimuli that belonged to a target stimulus were marked correctly. Thus, participants could obtain here a maximum of 24 points.

The design of the study involved Gender (male, female) as independent variable and Number of correctly answered items and Time to answer all items as dependent variables. Given a total sample size of $N = 72$ and a desired level of $\alpha = .05$, effects of size $d = 0.86$ (that is, large effects as defined by Cohen, 1977) could be detected with a probability $1 - \beta = .95$[1].

Results

The scoring methods (SDJ vs. 2oo4) showed a near-perfect correlation ($r = .96, p < .01$). Split-half reliability of both scoring methods turned out to be very satisfying ($\alpha = .91$ and $\alpha = .88$ for SDJ and 2oo4, respectively). Using the SDJ scoring method, men obtained $M = 89.03$ ($SD = 4.63$) out of 96 points, whereas women obtained $M = 80.92$ ($SD = 11.7$) points. The effect size indicated a large effect of $d = .91$. With scoring method 2oo4 men received $M = 17.89$ ($SD = 3.69$) out of 24 points, whereas women received $M = 13.56$ ($SD = 6.69$) points. The effect size indicated a large effect of $d = .80$, too.

A MANOVA revealed a significant main effect of Gender – independent of the scoring method used (SDJ: $F(1, 70) = 14.95$; $p < .001$, $\eta^2 = .176$; 2oo4: $F(1, 70) = 11.55$; $p < .001$, $\eta^2 = .142$) – with men outperforming women.

Regarding the time participants took to solve the test, we found that men performed significantly faster than women, $F(1, 70) = 1.97$; $p < .05$, $\eta^2 = .027$; men: $M = 17.63$ minutes, $SD = 6.82$ to women: $M = 20.38$ minutes, $SD = 9.55$, with a small effect size of $d = .33$. Interestingly, the time to finish the test was not correlated with either of the performance scores ($r_{SDJ} = -.18$, $p = .13$ and $r_{2oo4} = -.15$, $p = .19$).

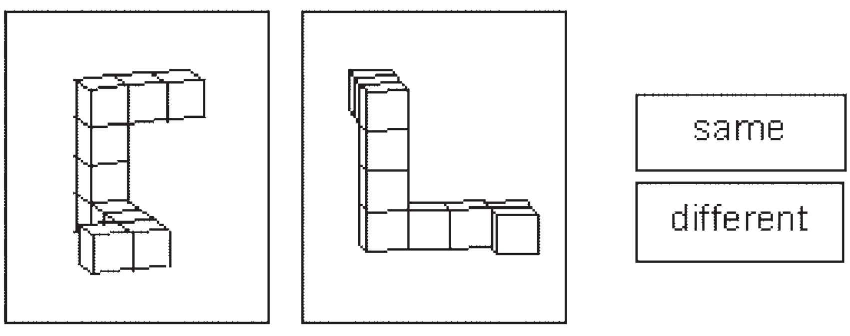


Figure 2. Item 1 from the new version of the MRT is shown. Only two stimuli have to be compared and a "same" versus "different" judgment is required.

[1] All power calculations reported in this article were conducted using the G*Power program (Faul, Erdfelder, Lang, & Buchner, 2007).

© 2010 Hogrefe Publishing

Discussion

The study shows that a reduction of task complexity – here test mode and time constraints – did not affect the gender differences in a mental rotation task. Men outperformed women regarding the number of items answered correctly independent of the scoring method used. Large effect sizes (see Cohen, 1977) of about d = .90 indicate fundamental gender differences comparable to the effects sizes reported in the meta-analysis by Voyer et al. (1995)[2]. Although both, Peters (2005) as well as Jansen-Osmann and Heil (2007), did not find gender differences using a chronometric mental rotation task with pairs of cube stimuli, in the paper-pencil version used here, gender differences did occur. Moreover, in the present study the overall time needed to solve the test differed between genders, whereas Peters (2005) as well as Jansen-Osmann and Heil (2007) did not find gender differences in the reaction times of the same-different judgments. The present finding, namely, that speed and accuracy were not correlated, however, needs further attention.

The results presented in this study are in contrast to several other findings (e.g., Peters, 2005; Jansen-Osmann & Heil, 2007). The most obvious, although trivial and therefore maybe irrelevant, difference is the *presentation mode* of the stimulus material: Pairwise comparisons of three-dimensional cube figures are used either in a computer or in a paper-pencil version. There is evidence that the presentation mode may influence test performance directly: Hänsgen (2001) pointed out that – especially regarding performance tests – the transfer from a paper-pencil test to a computer-administered test may lead to performance divergences. Consequently, the original MRT (i.e., choosing two correct alternatives among the four stimuli) should be presented as a computer-administered task. If test presentation mode does not play a role, then substantial gender differences should occur irrespective of presentation mode.

Another – more plausible – explanation is based on differences in the *stimulus material* itself: As Voyer and Hou (2006) pointed out, items used in the MRT differ in their distractor configuration: Some distractors are mirror-reversed images of the target stimulus, whereas others are totally different block configurations – something not used at all in chronometric studies. Additionally, the MRT items differ regarding their occlusion: As a result of rotation, some correct alternatives or distractors are fully or at least partly occluded, leading to a misperception of the shape. In chronometric studies (e.g., Jansen-Osmann & Heil, 2007), occlusion is usually avoided. Thus, different findings concerning gender differences in the pairwise paper-pencil versus the chronometric task may be caused by stimulus specific differences. More data are needed to investigate this hypothesis in detail.

Our aim was to develop a paper-pencil test that uses the MRT cube figures – which are well known to yield large gender differences – but is as similar as possible to the format of chronometric tests. Therefore, we used the original paper-pencil MRT and reduced the task complexity from two-out-of-four-alternatives choices to consecutive same-different judgments. With this paper we replicated our finding that a reduction of task complexity to a pairwise presentation mode does not eliminate or even affect gender differences in mental rotation performance (Titze et al., 2008). As a consequence, although gender differences in guessing behavior might affect gender differences in the original MRT (Voyer & Saunders, 2004), they do not account for a substantial part of the gender differences in mental rotation performance as such. The task factor that is accountable for different results in computer-administered vs. paper-pencil tasks still has to be identified. Finally, since researchers are often reluctant to use the MRT with children because of the complexity of the test (see, e.g., Johnson & Meade, 1987); our data suggest that the simplified version presented here might constitute a reasonable alternative to investigate the development of gender differences in mental rotation.

Acknowledgments

This study was supported by the German Research Foundation (DFG). We thank Michael Peters for his friendly permission to use the Mental Rotations Test (MRT) in our study. We especially thank Adrian Hoffmann for his help during test construction and data acquisition.

References

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Delgado, A.R., & Prieto, G. (1996). Sex differences in visuospatial ability: Do performance factors play such an important role? *Memory and Cognition, 24*, 504–510.

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

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory and Cognition, 18*, 546–550.

Hänsgen, K.-D. (2001). *Hogrefe TestSystem. Einführung in die computer-basierte Psychodiagnostik (Handbuch Band 2)* [Hogrefe TestSystem. Introduction to computer-based psychological diagnostics (Manual 2)]. Göttingen: Hogrefe.

Hausmann, M., Slabbekoorn, D., Van Goozen, S.H., Cohen-Ket-

[2] These effect sizes are found not only in the meta analysis of Voyer, Voyer, and Bryden (1995): We conducted a study with 72 participants and asked them to answer the MRT in its regular way with a regular time limit of two times 3 min. Men significantly outperformed women, $F(1, 70) = 17.09$; $p < .001$, $\eta^2 = .196$; men: $M = 12.64$; $SD = 4.91$ to women: $M = 8.28$, $SD = 3.95$) with an effect size of $d = .97$.

 © 2010 Hogrefe Publishing

tenis, P.T., & Güntürkün, O. (2000). Sex hormones affect spatial abilities during the menstrual cycle. *Behavioral Neuroscience, 114*, 1245–1250.

Imperato-McGinley, J., Pichardo, M., Gautier, T., Voyer, D., & Bryden, M.P. (1991). Cognitive abilities in androgen insensitive subjects – Comparison with control males and females from the same kindred. *Clinical Endocrinology, 34*, 341–347.

Jansen-Osmann, P., & Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain and Cognition, 64*, 217–227.

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

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

Masters, M.S. (1998). The gender difference on the Mental Rotations test is not due to performance factors. *Memory and Cognition, 26*, 444–448.

Moè, A., & Pazzaglia, F. (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16*, 369–377.

Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain and Cognition, 57*, 176–184.

Peters, M., & Battista, C. (2008). Applications of mental rotation figures of the Shepard and Metzler type and description of a mental rotation stimulus library. *Brain and Cognition, 66*, 260–264.

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

Resnick, S.M. (1993). Sex differences in mental rotations: An effect of time limits? *Brain and Cognition, 21*, 71–79.

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

Shih, M., Pittinsky, T.L., & Ambady, N. (1999). Stereotype susceptibility: Identity salience and shifts in quantitative performance. *Psychological Science, 10*, 80–83.

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*, 130–133.

Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*, 599–604.

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie experimentale, 60*(2), 91–100.

Voyer, D., & Saunders, K.A. (2004). Gender differences on the mental rotations test: A factor analysis. *Acta Psychologica, 117*, 79–94.

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*, 250–270.

Corinna Titze

Institute of Experimental Psychology
Heinrich Heine University
Universitätsstr. 1
D-40225 Düsseldorf
Germany
Tel. +49 211 8112134
Fax +49 211 8115037
E-mail corinna.titze@uni-duesseldorf.de

© 2010 Hogrefe Publishing