APPLIED COGNITIVE PSYCHOLOGY
Appl. Cognit. Psychol. **22**: 996–1013 (2008)
Published online 26 November 2007 in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/acp.1420

Durable and Generalized Effects of Spatial Experience on Mental Rotation: Gender Differences in Growth Patterns

MELISSA S. TERLECKI[1]*,
NORA S. NEWCOMBE[2] and MICHELLE LITTLE[3]

[1]*Department of Psychology, Cabrini College, USA*
[2]*Temple University, Philadelphia, USA*
[3]*Arizona State University, Tempe, USA*

SUMMARY

This study addressed questions about improvement in mental rotation skills: (1) whether growth trajectories differ for men and women with higher or lower spatial experience, (2) whether videogame training has effects on performance and leads to transfer, (3) whether effects of repeated testing or training effects are durable and (4) whether transfer is durable. Undergraduates participated in repeated testing on the MRT or played the videogame Tetris. Analyses showed large improvements in mental rotation with both repeated testing and training; these gains were maintained several months later. MRT scores of men and women did not converge, but men showed faster initial growth and women showed more improvement later. Videogame training showed greater initial growth than repeated testing alone, but final performance did not differ. Effects of videogame training transferred to other spatial tasks exceeding the effects of repeated testing, and this transfer advantage was still evident after several months. Copyright © 2007 John Wiley & Sons, Ltd.

Spatial intelligence allows us to encode and transform information about objects and their location and thus to find our way in the world and perform technical activities. That is, spatial thinking is important for functioning in both everyday activities (e.g. driving one's car, putting together a piece of furniture) and for more specialized tasks, as varied as the design of buildings, performance of surgical procedures and the solution of advanced mathematics problems. In addition, an informed citizen in an increasingly technological world must be fluent at processing spatial abstractions including graphs, diagrams and other visual representations. Research that reveals how to increase the level of spatial functioning in the population could therefore significantly improve the effectiveness of the workforce.

While spatial skill may be affected by biological factors (see Ceci & Williams, 2007 for varying opinions), it is not immutable. There is mounting evidence that both repeated testing (i.e. simple practice) and training (i.e. spatially relevant experience not assessed on the task itself) enhance spatial performance (see meta-analyses by Baenninger & Newcombe, 1989, 1995; Marulis, Liu, Warren, Uttal, & Newcombe, 2007). Practice effects on spatial tests are often substantial, and in fact, long-term practice or repeated testing can sometimes lead to massive improvement and the obliteration of pre-existing differences among groups, for

*Correspondence to: Dr. Melissa S. Terlecki, Department of Psychology, Cabrini College, 610 King of Prussia Rd, Radnor, PA 19087, USA. E-mail: mst723@cabrini.edu

example differences between children and adults (Kail, 1986). Spatial improvement can also be seen with more general spatial training rather than simply task-specific practice. For example computer and videogame usage are related to improvements in mental rotation and other related spatial tasks (e.g. De Lisi & Wolford, 2002; McClurg & Chaille, 1987; Okagaki & Frensch, 1994; Roberts & Bell, 2000; Sims & Mayer, 2002).

Despite the large number of studies of spatial practice and training, however, we still lack data about several questions that are crucial to devising effective educational interventions. One such question is whether improvement due to repeated testing and training quickly reaches some asymptotic value, or whether long-term efforts might lead to continued change. Another question is whether simple practice of a skill is the best way to create improvement, or whether more varied and less task-related interventions might work better—in the short run, in terms of maintenance of gains or in terms of transfer.

With regard to the durability of improvements, we would wish to know whether gains are maintained or fade over time. Such considerations are important to create long-term effects on learning, in comparison to temporary effects of experimental 'manipulations' (Schmidt & Bjork, 1992). The few results we have on this issue are inconclusive. On the one hand, Heil, Rosler, Link, and Bajric (1998) found continued improvement in mental rotation, compared to a control group, over a 2-week period following completion of training. On the other hand, Kass, Ahlers, and Dugger (1998), using a computerized task involving angle estimation of a ship bow (AOB) with Navy employees, found that errors increased 3 weeks after completion of training. However, because no baseline assessment was conducted, it is possible that participants were still showing some effects of their training experience.

We also know little about the generality of improvements, which is also important for real-world adaptability. Studies that have assessed transfer of spatial training have generally found only narrow transfer. For example, Stericker and LeVesconte (1982) found that group work involving physical learning aids with the Differential Aptitude Test (DAT), Group Embedded Figures Test (GEFT) and the Mental Rotations Test (MRT) increased performance on the Primary Mental Abilities (PMA) spatial test, but of course this test also involves mental rotation. As another example, experience in navigating in a computerized, virtual environment (VE) shows transfer to real-world navigation but further transfer was not assessed (Waller, Knapp, & Hunt, 2001). In a recent and well-cited study, Sims and Mayer (2002) found task-specific improvements in spatial performance following Tetris videogame playing. Performance on Card Rotations, Paper Folding and Form Board tasks did not differ from pre-test to post-test between those who played Tetris for 12 hours (across 2 weeks) and those who did not.

An additional question that has been debated in the literature on spatial skill and how to improve it is the issue of gender differences. Robust gender differences, favoring men and boys, are found on a wide variety of spatial tasks, with the most profound effects being observed on tests of mental rotation (for meta-analytic reviews, see Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). It has been suggested that experience with spatial activities may account for this difference, at least in part (e.g. Baenninger & Newcombe, 1995; Voyer, Nolan, & Voyer, 2000). For example men and boys have more computer and videogame experience than women and girls (Terlecki & Newcombe, 2005). This argument seems to suggest that female underachievement might be minimized through practice or training. However, in a meta-analysis, Baenninger and Newcombe (1989) found that men and women improve in parallel in response to practice and training, so that gender differences remain constant in size. Although some studies completed since the meta-analysis find greater gains

DOI: 10.1002/acp

for women than men (Okagaki & Frensch, 1994; Subrahmayam & Greenfield, 1996), or even the elimination of gender differences (Kass et al., 1998; Lizarraga & Ganuza, 2003; Paramswaran, 1994; Paramswaran & De Lisi, 1996; Roberts & Bell, 2000), a recent meta-analysis still does not show differential effects of practice, or repeated testing, and training on men versus women (Marulis et al., 2007).

It may be possible that men and women have different trajectories of spatial growth. Baenninger and Newcombe (1989) suggested that parallel growth might be observed early on in repeated testing or training, but that longer-term interventions might show that growth slows earlier for men (or for people of either gender with initially higher spatial ability or experience), with eventual convergence. Exploring this hypothesis requires the gathering of data over longer time periods than those commonly used in prior studies.

In the present study, we enrolled participants to continue in repeated testing or training over the course of a semester, and aimed to chart the pace of their improvements over that time through weekly test taking (which itself constituted the practice for the repeated testing groups). Such longer-term intervention is vital to ascertain growth trajectories. We sought to discover (1) whether growth trajectories differ for men and women with higher or lower levels of spatial experience, (2) whether videogame training has effects exceeding simple repeated testing improvements on the MRT, (3) whether repeated testing and training effects are durable, (4) whether repeated testing and training effects transfer and (5) whether transfer is durable.

METHOD

Materials

The Survey of Spatial Representation and Activities (SSRA) is a 17 item, open-ended and multiple choice questionnaire that was designed for this study (see Terlecki & Newcombe, 2005). The questionnaire consists of five questions related to participants' demographics, SAT scores, involvement in recreational sports and game activity and computer/videogame preferences. Following are 12 questions that assess frequency of computer, videogame and map usage as well as perceived efficacy related to their usage. For the present study, answers to the SSRA questions related to ownership of computer games and artwork software, ownership of game systems, frequency of game system usage and how skilled individuals believed they were at using/playing videogames/game systems were analysed. Previous analyses showed the SSRA to be an effective measure of computer and videogame usage (Terlecki & Newcombe, 2005).

The MRT (Shepard & Metzler, 1971; Vandenberg & Kuse, 1978) is a spatial task that involves the ability to imagine how objects will appear when they are rotated in two-dimensional (2-D) or three-dimensional (3-D) space. The MRT has 20 items (five sets of 4 items each). Each item consists of a 'criterion' figure (composed of radials of 10 connected cubes) presented to the left and four alternatives to match on the right (two of which are correct matches). Participants were timed and a maximum of 20 points was obtainable, where a point was only given if both correct options were chosen per question. This was chosen as a more stringent scoring procedure in order to reduce the potential for ceiling effects (Peters, Laeng, Latham, Jackson, Zaiyouna, & Richardson, 1995). Mental rotation problems were also randomly reordered each week to avoid order effects.

Training materials consisted of two versions of a popular spatial videogame Tetris ©, including a 2-D (semester 2) and a 3-D (semester 3) version. The training materials were

DOI: 10.1002/acp

applied (as an applet) to a website that recorded participant game playing activity, which participants could access from any location. The videogame stimuli consisted of either 2-D or 3-D configured blocks (comprised of one to three radials and three to five connected squares), which fall at varying speeds (increasing speed as game levels increase, making the game increasingly difficult). The falling blocks must be rotated in order to fit into blocks already positioned at the base of the screen. Once the surface is level, the baseline disappears and the successive falling blocks must then fit together and the process is then repeated. If the game player fails to fit a falling piece into the base at the bottom of the screen, the blocks stack up and will not be cleared. If they stack up to reach the top of the screen before the base can be cleared, the game is over.

In the 2-D version, game players watched from a horizontal screen, as if looking through a window. In the 3-D version, the viewpoint of the game player was from a sagittal plane, where it appeared as if one were standing at the top of a well and dropping objects to fall to the bottom, making the task presumably much harder.

A Tetris Game Usage Questionnaire, comprised of five-point, Likert-type questions, was also given to training group participants prior to testing to assess current and/or previous experience/usage with the game Tetris© or Tetris-like games. Questions ranged from 'never' to 'daily' in assessing usage practices (both past and present). Past and present Tetris game usage was averaged and a maximum of 5 points and a minimum of 0 points could be obtained.

The non-spatial videogame training of Solitaire (standard version) was available to all repeated testing group (control) participants in semesters 2 and 3. Solitaire comes standard on most campus computers and home PCs and involves a one-person card game aimed at ordering cards within a deck by color, suit and number.

Transfer task materials included the Guilford–Zimmerman Spatial Visualization Task (a.k.a. 'Clock Task'; Guilford & Zimmerman, 1947), which was used as a pre-test, post-test and retake measure of spatial ability transfer in semester 2. It is a 20-item test that shows a picture of a round alarm clock that is positioned at a given angle (turned about either a horizontal or vertical axis). Beside the picture of the tilted clock are instructions to then visualize a change in the clock's rotation (i.e. 'Turn right face 45°). Then, multiple choice items are presented that represent the appearance of the clock following the instruction for rotation. Participants were timed and a maximum of 20 points using a raw scoring procedure was used.

Another transfer task used was the Surface Development Test (SDT) (Ekstrom, French, Harman, & Dermen, 1976), which was used as a pre-test, post-test and retake measure of spatial ability transfer in semester 3. The SDT presents drawings of solid, 3-D forms that can be constructed by folding a given piece of paper. Within each problem, a diagram showing how that piece of paper might appear if folded (at particular edges) is presented. The object is to match letters on the 2-D surface to numbers on the constructed, solid form. Participants were timed and a maximum of 60 points using a raw scoring procedure was used.

Participants

Recruitment

Nearly 1300 undergraduates (370 men, 908 women) enrolled in an introductory psychology course (across three semesters) at an urban, northeastern university were screened for spatial experience through the spatial SSRA questionnaire (see Terlecki & Newcombe, 2005).

DOI: 10.1002/acp

Participants were of diverse ethnic backgrounds. Participants' ages ranged from 17 to 44 years, with a mean of 18.23 years ($SD = 1.83$). Screening continued in each of three semesters, in order to achieve acceptable sample sizes of students identified as *high* in spatial experience (those who scored within the upper 15% on the SSRA) and *low* in spatial experience (those who scored within the lower 15% on the SSRA), and these students were then invited to participate in a 12 week long study. The range of SSRA scores was from 0 to 18 points. High spatial experience participants (167 men, 83 women) obtained scores of 15–18, while low spatial experience participants (19 men, 218 women) obtained scores of 0–3. The distribution of gender within the high and low- experience groups was significantly different from the base rates of gender within the introductory psychology population, $\chi^2(2, N = 1765) = 208.79, p < .01$, with more men in the high spatial experience range and more women in the low spatial experience range.

Repeated testing and training sample

Table 1 shows recruitment by semester. In semesters 1 through 3, men and women were chosen to participate in a 12 week repeated testing or training study. In semester 1, 12 men and 40 women completed 12 weeks of repeated testing alone (there were no training groups in semester 1). In semester 2, 17 men and 38 women completed the 12 week study, with 31 individuals also participating in 2-D videogame training, while the remaining 24 individuals participated in a non-spatial training experience, along with repeated testing. In semester 3, 32 men and 41 women completed the 12 week study, with 41 individuals also participating in a 3-D videogame training, while the remaining 32 individuals participated in a non-spatial training experience, along with repeated testing. Participants were selected for training groups using a quasi-random assignment, which entailed randomly choosing participants within gender and high and low spatial experience groups to partake in training or repeated testing only.

Thus, the combined sample consisted of 180 participants across the three semesters: 61 men, 119 women, 86 high and 94 low spatial experience individuals, 108 repeated testing group and 72 training group individuals. All participants participated in the repeated testing of the MRT, but training group individuals also participated in training (videogame). Those who did not participate in the training served as control groups.

Within high and low spatial experience groups, gender was unevenly represented. Nearly all men were of high spatial experience (89% of men were of high spatial experience). Women were more varied in their spatial experience, but they dominated the *low* group (93% female). No differences between men ($M = 2.10$, $SD = 0.04$) and women ($M = 1.85$, $SD = 0.09$), nor between high ($M = 2.3$, $SD = 0.05$) and low spatial experience individuals ($M = 1.83$, $SD = 0.04$) were found in previous or current Tetris videogame usage (all F values < 1, p values $> .25$). No significant differences in Tetris usage were found

Table 1. Sample distribution by semester

Group	Semester 1	Semester 2	Semester 3
Repeated testing	52 (12 men, 40 women)	24 (7 men, 17 women)	32 (14 men, 18 women)
Training	—	31 (10 men, 21 women)	41 (18 men, 23 women)
Men	12 (2 low, 10 high)	17 (2 low, 15 high)	32 (3 low, 29 high)
Women	40 (27 low, 13 high)	38 (33 low, 5 high)	41 (27 low, 14 high)

Note: No training was administered during Semester 1. Low = low spatial experience, high = high spatial experience.

between men ($M = 2.10$, $SD = 0.04$) and women ($M = 1.96$, $SD = 0.06$) with high spatial experience, either.

Durability and transfer sample

A total of 79 participants (across semesters 1–3), including 28 men, 51 women, 46 repeated testing group and 33 training group individuals participated in assessment of MRT retention after 2–4 months. A total of 53 participants (across semesters 2 and 3) participated in a pre-test, post-test and retake assessment of two transfer tasks. This sample included 21 men, 32 women, 19 repeated testing group and 34 training group individuals.

Design and procedure

At the beginning of each semester, potential participants were given the SSRA as part of a packet of psychological research questionnaires administered during their first week of classes. The packet of questionnaires was completed to fulfil requirements for research credit. Those who scored *high* or *low* in spatial experience on the SSRA were contacted and invited to participate in a 12 week study, in which they would get paid ($5–$7.50 a week) for their participation. Those who agreed to participate were given the Mental Rotations Task (MRT) weekly for 12 weeks.

At each session, participants were given the MRT, face down and were reminded to choose two answers per mental rotation problem and that they had three minutes to complete the first 10 problems (first two test pages), and another 3 minutes to complete the last 10 problems (last two test pages). A 20 point scoring system was used.[1]

During week 1, participants were given the opportunity to complete several sample problems (non-identical to actual test problems), as well as witness the correct answers for those sample problems, in order to allow participants to become familiar with test stimuli. MRTs were administered in small groups and timed individually.

Participants in training groups were instructed to play Tetris while repeated testing groups (non-training, control) were instructed to play Solitaire, for 1 hour a week. Samples were across semesters 2 and 3 (see Table 1). When participants were to engage in the computer game playing was optional, as long as playtime was for an hour a week (but could be broken up into periods of at least of 15 minutes). Training group individuals played Tetris via access to our experimental website from any computer and 'logged-in' using their name and social security number as their password. The website automatically kept track of their playtime/time logged-on (regardless of pauses in game play). Also, participants were required to hand in a literal log-in sheet (corresponding to their game play times and scores), which would then be matched to their actual computerized log-in information, as a check. Access to individual participant's log-on times and scores were of restricted access. Solitaire (repeated testing, non-training control) group players were also required to play 1 hour of Solitaire, which would be available on any home PC or campus computer. Though no website log-in check was available, these participants were also required to turn in log-in sheets documenting their play. Both Tetris and Solitaire players reported and were confirmed to have played 1 hour a week for 12 weeks (12 hours total).

Durability of repeated testing and training effects was also assessed. Retake assessment of MRT performance occurred between 2 and 4 months after training ceased (depending

[1]A 20 point scoring method was used on the MRT, where one point was only given for choosing both correct answers per problem. This method was chosen to avoid ceiling effects (Voyer et al., 1995).

DOI: 10.1002/acp

when the following semester resumed).[2] Other methods were the same as previously described.

Transfer tasks (chosen to load upon similar spatial visualization and mental rotation components as the MRT; Ekstrom et al., 1976; Guilford & Zimmerman, 1947) were administered in order to discover whether or not skills learned through practice on the MRT would also transfer to improved performance on a related spatial task. It was also of interest to find out if transfer of spatial skill did occur, would increases in transfer task performance be durable as well. Transfer (and durability of transfer) was only assessed in semesters 2 and 3. Different transfer tasks were used in the two semesters to determine if transfer would differ for different spatial visualization tasks. Transfer tasks were administered at pre-test (week 1), post-test (week 12) and at retake (2–4 months following the cessation of MRT testing and videogame training). Other methods were the same as previously described.

RESULTS

Combining samples

Analyses of the combined data sets from semesters 1, 2 and 3 were undertaken in order to maximize power to detect group performance differences over time. Combining data across studies was legitimated by several considerations. First, the population and the sampling methods were equivalent across semesters 1–3. A 3×12 (semesters by weeks) repeated measures ANOVA of MRT performance showed no differences among semesters for any sex or spatial experience groups (all F values < 1, p values $> .40$). Second, a 2×12 (non-training groups by weeks) repeated measures ANOVA of MRT performance showed no differences between control group individuals who had no training and those who played Solitaire (both classified as controls) ($F < 1$). Similarly, a 2×12 (training groups by weeks) repeated measures ANOVA of MRT performance showed no differences between training group individuals who had 2-D *versus* 3-D videogame training ($F < 1$). The comparability of training groups was also supported by no significant differences in current Tetris videogame playing (within the past year) ($F < 1$) or past Tetris videogame experience (during childhood) ($F < 1$). Also, participant attrition was comparable across semesters for men (50–71%) ($\chi^2[2, N = 96] = 0.44$, $p = .48$), women (42–63%) ($\chi^2[2, N = 184 = 1.16$, $p = .26$), high spatial experience individuals (52–65%) ($\chi^2[2, N = 136] = 0.19$, $p = .59$), low spatial experience individuals (43–60%) ($\chi^2[2, N = 144] = 0.76$, $p = .30$), repeated testing group individuals (44–67%) ($\chi^2[1, N = 86 =]\ 0.86$, $p = .71$) and training group individuals (39–71%) ($\chi^2[1, N = 113] = 2.11$, $p = .10$). Note, additionally, that although dropout rates were high, they did not differ across groups of interest. Finally, no group was approaching ceiling performance on the MRT (no group reached the maximum of 20 points on the MRT and, more importantly, there were no significant differences among groups in skew that would indicate differential closeness to ceiling). An alternative 40-point scoring procedure had showed greater tendency for ceiling effects; skewness was nearly doubled across all participants (for men, especially). Also, there were significant decreases in standard deviations and some participants were reaching maximum performance on the MRT (32% of men, 9% of women, 8% of repeated testing group

[2]No differences were found between different durability samples in MRT retake performance. The variability of 2–4 months prior to retesting was due to semester breaks in academic calendar.

DOI: 10.1002/acp

(no training), and 11% of experimental group). No such results were found with the 20-point scoring method.

Comparing growth trajectories

Latent growth curve modelling, a type of structural equation modelling (SEM), was used to compare longitudinal trajectories of growth in mental rotation across all groups. SEM techniques were employed due to their enhanced ability to handle error, missing data and sensitivity to changes in growth trajectories over time, in comparison to traditional statistical techniques, such as Analyses of Variance (ANOVA) or *t*-tests. SEM techniques provide estimates of both the level and shape of growth in a construct of interest. The random intercept, part of SEM analyses, was used to estimate the average level of growth at a point of substantive interest, and random slope factors were used to measure the rate or shape of change over time. We used MPLUS software to model MRT growth (Muthén & Muthén, 2004). Preliminary model testing affirmed that a quadratic model (in comparison to a linear model), which included a slope factor that accounts for trend curvature, showed the best fit to the MRT data and was therefore retained in subsequent analyses.

Multiple group analyses were used to compare MRT growth across groups of interest (i.e., gender or training conditions). Multiple group analyses permit the simultaneous estimation of growth curves for groups of interest within a single analysis. The magnitude of parameters of interest, including estimates of mean status or mean slope change, were compared using constraint procedures. We compared four parameters of interest between groups using a series of latent growth models: pre-test status, linear slope growth, quadratic growth and post-test status (see Table 2). The relative reduction in chi-square fit between models provided a measure of the significance of the differences between groups in average performance and rate of change in performance.

Four fit indices were used to examine the fit of growth models: the chi-square (χ^2), Comparative Fit Index (CFI), Root Mean Square Error of Approximation (RMSEA) and

Table 2. Comparison of SEM model fit across groups

Measure	χ^2(df)	CFI	RMSEA	AIC	$\chi^2\Delta$ (df)
Training groups					
Unconstrained	293.36 (126)**	0.96	0.12	9545.23	
Intercept constrained	293.45 (127)**	0.96	0.12	9543.32	0.09(1)
L. slope constrained	297.46 (128)**	0.96	0.12	9545.33	4.01(1)
Q. slope constrained	297.77 (129)**	0.96	0.12	9543.64	0.31(1)
Gender groups					
Unconstrained	361.58(132)**	0.94	0.14	9520.03	
Intercept constrained	394.03(133)**	0.94	0.15	9550.47	32.45(1)**
L. slope constrained	405.46(134)**	0.93	0.15	9559.90	11.46(1)**
Q. slope constrained	391.62(133)**	0.94	0.15	9548.06	0.41(1)
3-Group					
Unconstrained	118.50(39)**	0.95	0.19	4832.64	
Intercept constrained	180.59(41)**	0.91	0.24	4950.73	62.09(2)**
L. slope constrained	199.70(43)**	0.89	0.25	4965.84	19.10(2)**
Q. slope constrained	201.68 (45)**	0.89	0.25	4963.82	1.99(2)

Note: L., linear; Q., quadratic.
$^{*}p < .05$; $^{**}p < .01$.

DOI: 10.1002/acp

Akaike's Information Criterion (AIC). The chi-square provides a measure of the degree of fit of growth factor loadings and covariances to the observed data. Chi-square values that approximate the degrees of freedom of a model and are non-significant indicate that the proposed model does not significantly differ from observed patterns in the data. However, to overcome the sample size biases of chi-square fit statistics, the CFI and RMSEA are used as 'practical fit' indices that are not as biased by sample size. CFI, whose values range from 0 to 1.00 (where CFI values of 0.95 or higher indicate that a model is an excellent fit to the data, and values between 0.90 and 0.95 are considered adequate), measures relative fit of an observed to proposed model, based on the data. The RMSEA measures a hypothesized model's estimation error given the observed data, with values of 0.05 or less indicating that a model shows a close fit to the observed data. AIC values compare difference of fit between two models. Lower AIC values indicate a better fit. Taken together, these measures give an estimate of how well a given model of observed data represents the population at large.

Effects of spatial experience and gender on MRT growth

The first set of growth models compared the rate of growth in performance among three groups of participants: high spatial experience men ($n = 54$), low spatial experience women ($n = 87$) and high spatial experience women ($n = 32$). Sample size for low spatial experience men ($n = 7$) was too small to include in analyses. Because group sample size was limited by this three-group design, MRT growth over 6 weeks was modelled (weeks 2, 4, 6, 8, 10, 12) using a quadratic slopes model.

Results of the three-group comparison showed that MRT scores differed significantly at week 2 ($\chi^2\Delta[2,N = 176] = 62.09$, $p < .001$). Fit of an unconstrained model ($\chi^2[39,N = 176] = 118.50$, $p < .001$; CFI = 0.95; RMSEA = 0.19; AIC = 4892.64) was superior to the fit of a model with intercept factors set equal across groups ($\chi^2[41,N = 176] = 180.59$, $p < .001$; CFI = 0.91; RMSEA = 0.24; AIC = 4950.73). Inexperienced women showed the lowest MRT score at week 2 ($M = 5.25$; $SE = 0.38$). Experienced men showed the highest MRT score at week 2 ($M = 11.18$; $SE = 0.66$) and experienced women showed a slightly lower mean MRT performance of 9.89 ($SE = 0.67$) at week 2. Thus, as may have been predicted, high spatial experience men were performing significantly better than high and low spatial experience women, with low spatial experience women performing the worst on the MRT.

Initial linear growth, between weeks 2 and 4, also differed significantly across the experience/gender groups ($\chi^2\Delta$ $[2,N = 176] = 19.10$, $p < .001$). The fit of a model with linear slopes constrained across groups ($\chi^2[43,N = 176] = 199.70$, $p < .001$; CFI = 0.89; RMSEA = 0.25; AIC = 4965.84) was significantly worse than a model with constrained intercepts ($\chi^2[41,N = 176] = 180.59$, $p < .001$; CFI = 0.91; RMSEA = 0.24; AIC = 4950.73). Specifically, low spatial experience women showed a slower initial MRT growth rate ($M = 0.58$; $SE = 0.24$) compared to either high spatial experience men ($M = 1.92$; $SE = 0.27$), or high spatial experience women ($M = 1.11$; $SE = 0.52$). Follow-up multiple group contrasts affirmed that the initial growth rate of low spatial experience women differed significantly from the initial growth rates of high spatial experience men and women ($\chi^2\Delta$ $[1,N = 176] = 18.46$, $p < .001$), but the initial MRT growth rate did not significantly differ between experienced men and women ($\chi^2\Delta$ $[1,N = 176] = 0.64$, $p > .05$). Thus, initial MRT growth was similar between high spatial experience men and women, and similar to repeated testing group individuals (without added videogame training), women with low levels of spatial experienced showed the slowest initial growth in MRT ability.

DOI: 10.1002/acp

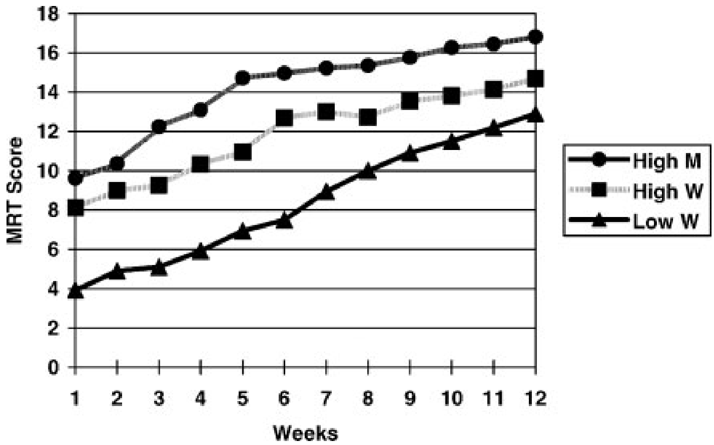

Figure 1. MRT growth in high and low spatial experience men and women. *High M* refers to high spatial experience group men, *High W* refers to high spatial experience group women and *Low W* refers to low spatial experience group women

The multiple group comparison suggested that the three groups did not differ in quadratic growth ($\chi^2\Delta[2,N = 176] = 1.99$, $p > .05$). However, comparison of quadratic slopes showed that while high spatial experience men's MRT growth was slowing by the end of the study ($M = -0.16$; $SE = 0.04$; $t = -3.62$, $p < .001$), low spatial experience women's MRT growth rate increased slightly ($M = 0.09$; $SE = 0.04$; $t = 2.65$, $p < .01$) and high spatial experience women's growth rate did not change by the end of the study period ($M = -0.03$; $SE = 0.08$; $t = -0.37$, $p > .05$). In comparison to initial levels of MRT growth, growth in experienced men and women had begun to level off by the 12th week, while inexperienced women continued to grow (see Figure 1).

Effects of training

Small effects were found between repeated testing and training groups in MRT performance, but were not reliable longitudinally. As expected, results showed that latent growth in mental rotation did not differ at pre-test between repeated testing and training groups, as indicated by a non-significant chi-square difference when pre-test intercepts were fixed to be equal ($\chi^2\Delta[1,N = 180] = 0.10, p > .05$), as compared to a full model. Further, the fit of a full model ($\chi^2[126,N = 180] = 293.36$, $p < .0001$; CFI = 0.96; RMSEA = 0.12; AIC = 9545.23) was not superior to a model with constrained pre-test status ($\chi^2[127,N = 180] = 293.45, p < .001$; CFI = 0.96; RMSEA = 0.12; AIC = 9545.32).[3] This non-significant chi-square difference was reflected in nearly equal pre-test factor means for the repeated testing group ($M = 6.97$; $SE = 0.61$) and the training group ($M = 6.74$; $SE = 0.41$). Thus, as expected, individuals who would have videogame training did not differ from those exposed only to repeated testing on the MRT, at the beginning of the study.

Initial growth in MRT performance, however, was slightly different between repeated testing and control groups. Constraining linear slopes across groups significantly changed

[3]The RMSEA values for all growth models were above 0.10, reflecting unexamined heterogeneity in the models, although other practical fit indices were acceptable (i.e., the CFI). This discrepancy in the fit indices was likely related to the fact that a large number of observation periods were modelled and sample size was limited for some group comparisons (MacCallum, Kim, Malarkey, & Kiecolt-Glaser, 1997).

DOI: 10.1002/acp

the model fit of a constrained intercepts model ($\chi^2\Delta[1, N = 180] = 4.01$, $p > .05$), thus indicating that the initial rate of change in mental rotation was significantly different between repeated testing and training groups. However, the fit of a constrained pre-test status model was only slightly superior (AIC = 9543.32) to a model with both intercepts and linear slopes fixed to be equal across groups (AIC = 9545.33) owing to the small difference in this parameter between groups. The repeated testing group showed slower initial growth (*M* linear slope = 0.63; *SE* = 0.11) than the training group (*M* linear slope = 0.94; *SE* = 0.10). Thus, although the effect was small in magnitude, training group individuals showed greater initial rates of growth in mental rotation performance.

Repeated testing and training groups, however, did not differ in rate of growth on the MRT, long term. To assess differences in quadratic decline across groups, we compared the chi-square fit of a reduced model with equal intercepts and linear slopes to a fully constrained model which showed no difference in chi-square fit ($\chi^2\Delta[1, N = 180] = 0.31$, $p > .05$). Thus, training and repeated testing groups did not differ in their rate of learning acceleration. This was indicated by only very small differences in quadratic growth between the repeated testing group (*M* quadratic slope = −0.01; *SE* = 0.11) and the training group (*M* quadratic slope = −0.03; *SE* = 0.01).

Follow-up SEM analyses examined whether final status on mental rotation (i.e. the estimate of week 12 mental rotation) differed significantly across repeated testing/training groups. A full model was compared to a model with mean final status constrained to be equal across repeated testing and training groups. Results showed that final status (post-test) on mental rotation did not reliably differ across repeated testing and training groups, as indicated by a null chi-square difference test ($\chi^2\Delta[1, N = 180] = 1.43$, $p > .05$). Thus, by the 12th week of study, those with videogame training did not significantly differ from repeated testing group individuals in mental rotation ability. However, what is more promising is training effects on transfer.

Effects on transfer

Because transfer task performance did not significantly differ between the Clock Task and the SDT (F [1,52] = 4.98, $p = .46$), transfer assessments at pre-test, post-test and retake were pooled and analysed. Transfer task scores were computed as percentages correct in order to compare across tasks. A $2 \times 2 \times 3$ (training groups by gender by testing time; pre-test, post-test and retake) ANOVA showed that the main effects for testing time (F [2,50] = 89.97, $p < .01$, $\eta^2 = 0.14$) and gender (F [2,50] = 92.65, $p < .01$, $\eta^2 = 0.18$) were significant, as was the main effect for training (F [2,50] = 12.56, $p < .05$, $\eta^2 = 0.09$). Paired comparisons across testing times showed that transfer effects were both evident and durable.

An interaction between testing time and training group was also found (F [2,50] = 5.56, $p < .05$, $\eta^2 = 0.08$). Post-hoc contrasts showed that training and repeated testing group individuals did not differ at pre-test, but in agreement with separate analyses, the training group significantly outperformed the repeated testing group on transfer tests at post-test (F [2,50] = 15.63, $p < .01$, $\eta^2 = 0.20$) and again 2–4 months later at retake (F [2,50] = 11.99, $p < .05$, $\eta^2 = 0.18$) (see Figure 2). These effects were small to moderate in size.[4]

[4]Partial eta squared (η^2) is an effect size measure (used in compatibility with the SPSS statistical program) that is used with analysis of variance techniques to estimate the total variability attributable to a factor, while taking error variance into account. Effect sizes for eta squared range from 0.10 (small), 0.30 (moderate) to 0.50 (large), when df = 1 (Cohen, 1973). If more than two groups are being compared in an ANOVA, eta squared as an estimate of effect size becomes overestimated (Rosenthal & Rosnow, 1991).

DOI: 10.1002/acp

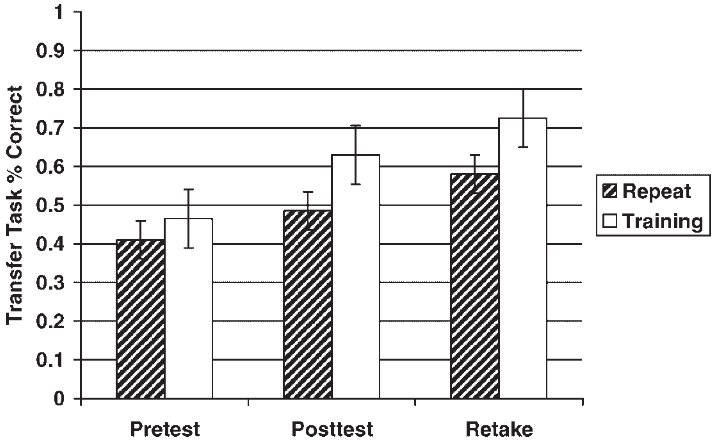

Figure 2. Durability of transfer in repeated testing and training groups. Pre-test occurred at week 1, post-test occurred at week 12 and retake occurred 2–4 months following the cessation of repeated testing or training. *Repeat* refers to repeated testing group

Transfer performance, across all participants, showed improvements that were significant and durable. Across both transfer tasks (Clock Task and SDT), post-test performance (week 12) was found to be significantly better than pre-test performance (week 1) ($d = 0.11$, $p < .01$). In addition, retake performance (2–4 months later) was actually significantly better than post-test performance (week 12) ($d = 0.11$, $p < .05$). Hence, retake performance on transfer tasks was also significantly better than pre-test performance ($d = 0.22$, $p < .01$).

Durability of MRT improvements

In order to evaluate durability of MRT performance, a $2 \times 2 \times 3$ (training groups by gender by testing time; pre-test, post-test and retake) ANOVA was conducted. The main effects for

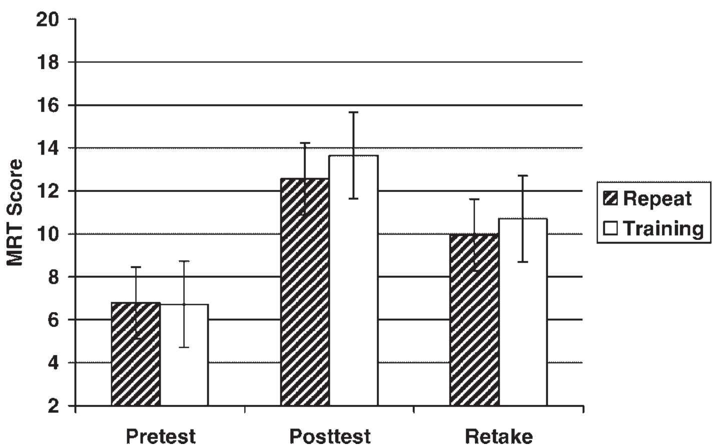

Figure 3. Durability of repeated testing and training effects. Pre-test occurred at week 1, post-test occurred at week 12 and retake occurred 2–4 months following the cessation of repeated testing or training. *Repeat* refers to repeated testing group

DOI: 10.1002/acp

testing time (F [2,177] = 61.25, $p < .01$, $\eta^2 = 0.19$) and gender (F [2,177] = 109.323, $p < .01$, $\eta^2 = 0.18$) were significant and moderate in size. Consistent with the SEM analysis, although Figure 3 shows a small advantage for the training group at post-test and retake, the main effect of training versus repeated testing groups was non-significant (F [2,177] = 78.83, $p = .09$). No interactions were found, however, MRT performance was durable.

Across all participants, post-test performance on the MRT (week 12) was significantly better than pre-test performance (week 1) ($d = 6.53$, $p < .01$) as one would expect given the SEM analysis. Crucially, although retake scores were significantly lower than post-test scores ($d = 2.81$, $p < .01$), retake performance (2–4 mos. later) was still significantly better than pre-test performance ($d = 3.72$, $p < .01$), showing that the improvements in MRT performance lasted over several months. Taken together with transfer effects, repeated testing and training effects proved to be durable and generalizable across both males and females of varying spatial experience.

GENERAL DISCUSSION

The present study has helped to answer important questions regarding the nature of spatial experience on mental rotation performance, and how practice or repeated testing and training effects vary across genders and other spatial tasks. We found that the growth trajectory for mental rotation skills during long-term repeated testing and training continued to show improvement over the course of a whole semester; that these improvements remained evident after several months; that videogame training, while not elevating MRT performance above the effects of repeated testing, led to significantly enhanced transfer compared to repeated testing alone; and, that men's and women's growth trajectories differed, especially for women with initially low experience, so that rapid gains took longer to be evident for the group most in need of them.

Growth trajectories as a function of gender and prior spatial experience

The most important goal of this study was to address gender differences in mental rotation performance and growth, and to determine whether repeated testing or training would have an effect on gender differences in mental rotation ability, as a function of pre-existing spatial experience. As evident in the literature, large gender differences in spatial ability, favoring men, still exist; most markedly in mental rotation ability (for meta-analytic reviews, see Linn & Petersen, 1985; Voyer et al., 1995), which many consider biologically determined and immutable. We found that, comparable to the results of prior work as reflected in meta-analyses, differences in mental rotation performance between men and women continued regardless of intervention, although differences in growth curves differed between men and women with varying levels of pre-existing spatial experience. Initial SEM analyses showed that men exhibited a faster initial burst of improvement but then began to level off over time, while women showed slower initial increases in MRT performance, but then continued to grow steadily and exhibited greater MRT growth by the end of the 12th week of study. The three-group SEM analysis, which compared MRT growth of high spatial experience men and high and low spatial experience women, elucidated this pattern by showing that the later increase in growth characterized low spatial experience women in particular (see Figure 1). We explored the possibility of asymptotic performance being exhibited by high spatial experience men and found this not

DOI: 10.1002/acp

to have affected our results. Further examination of this group showed their average performance to be more than one standard deviation away from the maximum score on the MRT. Although skew analyses showed all groups to be performing towards the high end of performance on the MRT by post-test at week 12, none were performing at their maximum potential, which leaves room for improvement for both men and women with varying levels of spatial experience.

Also promising is that when MRT performance of women who had videogame training were compared to men who had no training and only repeated testing on the MRT, regardless of pre-existing spatial experience, we found the difference at post-test (week 12) to be only marginally significant, with a small effect size (F [1,41] = 9.86, $p = .053$, $\eta^2 = 0.09$). Perhaps with an even longer intervention, convergence may be obtained between men and women, given that women are receiving extra spatial training.

These findings are of practical importance. They suggest that women with low spatial experience may be a difficult group to engage in educational interventions, because it may be discouraging both to them and to teachers to see slow initial improvements. Such disappointment may be especially pronounced if it is seen that initially able individuals are showing rapid gains. However, it may also be that high-experience men and women experienced the same slow growth when they were of low experience, and motivation perhaps was the key to greater growth and cumulative experience. It may be that men and boys are more motivated to gain experience in these tasks, so that the issue becomes how to get women and girls more interested in spatial activities, such as videogame playing.

Our data suggest that disappointment due to slow initial gains may be premature. Persistence will eventually lead to substantial improvement. As suggested, it is possible that interventions more intense and/or longer than that implemented in the present study would lead to eventual convergence of ability levels. Future studies will have to invest in more intense longitudinal designs, as well as recruit a sample of low spatial experience males (if possible), in order to discover if similar growth patterns apply. Also, study of motivational factors in both men and women, boys and girls would be helpful in further understanding why differences exist between men and women in spatial experiences, such as videogame playing, and engagement in spatial activities more generally.

Direct and transfer effects of videogame training

The effects of videogame training were mixed, although some lasting training effects were found. SEM results showed that individuals with training exhibited an initial burst in growth (in comparison to repeated testing group individuals), but this effect disappeared by the 12th week, with no final differences between repeated testing and training group individuals. Limitations due to lack of power and unexplained heterogeneity in individual trajectories may have prevented this effect from reaching significance, as well as the great overshadowing effect of the substantial improvements due to repeated testing.

Effects of videogame training on transfer were more impressive. The greater spatial experience experienced by the videogame training group led to transfer effects that were maintained over several months (see Figure 2), even after exposure to training materials had ceased. Such transfer, especially such wide transfer, has been elusive in prior studies (Green & Bavelier, 2003, 2006). For instance, Green and Bavelier (2003) have found experience with action videogames, specifically, to transfer to a number of visuospatial tasks, while experience with Tetris did not transfer. Our results suggest broad transfer from Tetris videogame playing in that improvements in the Clock Task and SDT bear little resemblance to Tetris stimuli. However, one must question whether the underlying mechanisms are the same. We might consider that participants who did not experience videogame training had mental rotation practice, or repeated testing, in one context (the MRT), while training group individuals had mental rotation practice in two contexts (the MRT and Tetris videogame playing). It may be that training group individuals performed better than repeated testing group individuals because they had more experience using the same or similar skills (e.g. mental rotation) and/or were using them in different contexts, and thus would be more likely to transfer that ability to other novel tasks. It has been suggested that encouraging different types of practice, even if the underlying mechanisms are the same, may produce greater and wider transfer and thus, more efficacious learning (Schmidt & Bjork, 1992).

Note that our lack of retest controls (i.e. a sample to compare pre- and post-test transfer task performance for people not exposed to repeated testing or training) compromises any interpretation of transfer of practice, but the difference between the repeated testing and training groups is unaffected, because retesting effects would be equal across those groups. Thus, our transfer task results suggest that videogame training did have an impact above and beyond simple practice, or repeated testing on the MRT.

The durability of long-term growth

SEM revealed that growth in MRT performance continued over the course of a semester, although a quadratic trend was found that suggested that growth in MRT ability flattened somewhat across the 12 weeks of study, with the magnitude of improvements in MRT performance decreasing by the end of the study. However, mental rotation ability continued to improve (at some rate) for all participants and overall improvements were large. Further, although MRT performance decreased at retake (2–4 months after repeated testing and training), it still remained at a level that represented a 20% improvement over initial MRT performance.

Transfer effects were also still evident 2–4 months following the cessation of repeated testing and training. In fact, all groups continued to improve on transfer tasks at retake (see Figure 2). The improvements due to training continued to be evident well after exposure to task-specific materials had ended.

These results show that spatial ability is malleable regardless of gender or previous spatial experience, which is especially important for low spatial experience women, and that the effects of training with such materials can be long lasting. This demonstration is vital to the idea that all individuals can potentially improve their spatial skills given appropriate practice or training, and that superior ability is not a prerequisite for success (Newcombe, Mathason, & Terlecki, 2002). The data also suggest the importance of sustained and distributed training and education for spatial skill.

Future directions

Although SEM has several benefits over traditional statistical techniques, such as more power to detect group differences in longitudinal patterns of growth, and the ability to handle missing data, the SEM model fits were not optimal, which suggests large variation in performance that our models could not capture. This heterogeneity in growth may have been a result of sample size limitations, lack of measurement reliability over time or inherently different growth curves within unexamined subgroups, along with other potential omitted variable biases (i.e. 'control' variables that were not included in the model). Although a quadratic model fits our data best, some individual growth patterns showed erratic trends across the 12 weeks of study. Different, more consistent and stable results may have been obtained had a larger sample size been used. A larger sample may have also improved some of our parameter estimates, such as the RMSEA (MacCallum et al., 1997). Future research in longitudinal growth trajectories should consider some of the issues aforementioned in designing such complex statistical analyses.

Second, a wider range of measurement tools should be employed to add to the reliability, validity and generalizability to this study. We used only one measure, the SSRA, to classify our groups of high and low spatial experience and also used only one measure, mental rotation through the MRT, to assess spatial ability. Although previous research has found the SSRA to be a valid tool for spatial experience assessment (Terlecki & Newcombe, 2005), adding other standardized measures with surveys of a wider range of spatial experiences would be advantageous. Also, assessing performance on other tasks that involve spatial ability, such as working memory, spatial reasoning and speeded perceptual tasks, could enhance our understanding of the breadth of group differences and similarities in performance. It is difficult to conclude whether increases in performance may reflect specific learning related to our tools, rather than general improvements in spatial skills.

Longitudinal studies are often hampered by attrition. Generalizability claims lay subject to high dropout rates and unequal sample sizes. In our study, participant attrition was difficult to control given the long-term commitment participants had to have had to complete the 12-week study. We did not find differences in attrition between training, spatial experience or gender groups and we did attempt to keep participation high by providing compensation throughout the 12 weeks of each semester.

We acknowledge that the small number of low spatial experience men also creates a problem for interpretation. However, this sample is rare in the population and thus is not artifactual to this study. Future studies on such topics should specifically seek out such a sample, though we guess low spatial experience males may be a difficult population to find.

Finally, a more aggressive training regimen may enhance the potential for training effects between groups. Videogame training lasting for more than 1 hour a week may have larger effects. Given the longitudinal nature of our study, we found it difficult to ask participants to do any more than an hour, but future studies should consider longer, more intense periods of distributed training. Also, research suggests variable training to support retention and transfer better than constant or regular training (Schmidt & Bjork, 1992), as in our study, participants could play Tetris anytime they wanted, as long as it was in (at least) 15 minute blocks. Randomized training may have a greater long-term effect on learning as well.

CONCLUSIONS

This study confirms the significant impact of long-term practice or repeated testing, and the potential of training in improving mental rotation performance. Although neither mental rotation practice nor videogame training reduced gender differences in spatial skill, the shape of the growth curves suggested that longer interventions may be needed to achieve this goal. In any case, it is socially and practically important to know that practice and training can dramatically improve mental rotation ability, regardless of previous spatial experience or gender. Even those who already had reported high levels of spatial

DOI: 10.1002/acp

experience did not reach ceiling on the MRT. Thus, we are optimistic that there exists great room for improvement in spatial ability, in both men and women.

Perhaps more importantly, these effects can last over several months and the effects of videogame experience show results that are not task-specific. Training effects transferred to other spatial tasks which were not subject to practice effects (or minimally so). This finding is crucial to our goals in creating long-term improvements in real-world spatial skills.

In summary, differences in levels of initial spatial experience do not constrain growth, although low-experience individuals may take some time to begin to show spatial gains. Thus, success in spatial performance is within reach for both men and women with appropriate experience, education and training. Videogame training may be an effective method to achieve this objective; more training studies are needed to confirm its utility.

ACKNOWLEDGEMENTS

This research was supported by grants 0087516 and 0337360 from the ROLE program of the National Science Foundation.

REFERENCES

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

Baenninger, M., & Newcombe, N. (1995). Environmental input to the development of sex-related differences in spatial and mathematical ability. *Learning and Individual Differences*, *7*, 363–379.

Ceci, S., & Williams, W. (2007). *Why aren't more women in science? Top researchers debate the evidence*. Washington, DC: American Psychological Association.

Cohen, J. (1973). ETA-squared and partial ETA-squared in fixed factor ANOVA designs. *Educational and Psychological Measurement*, *33*, 107–112.

De Lisi, R., & Wolford, J. (2002). Improving children's mental rotation accuracy with computer game playing. *Journal of Genetic Psychology*, *163*, 272–283.

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

Green, C., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.

Green, C., & Bavelier, D. (2006). Enumeration versus object tracking: Insights from video game players. *Cognition*, *101*, 217–245.

Guilford, J., & Zimmerman, W. (1947). *Guilford–Zimmerman Aptitude Survey*. Orange, CA: Sheridan Psychological Services.

Heil, M., Rosler, F., Link, M., & Bajric, J. (1998). What is improved if mental rotation task is repeated- the efficiency of memory access, or the speed of a transformation routine? *Psychological Research*, *61*, 99–106.

Kail, R. (1986). The impact of extended practice on rate of mental rotation. *Journal of Experimental Child Psychology*, *42*, 378–391.

Kass, S., Ahlers, R., & Dugger, M. (1998). Eliminating gender differences through practice in an applied visual spatial task. *Human Performance*, *11*, 337–349.

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

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

MacCallum, R., Kim, C., Malarkey, W., & Kiecolt-Glaser, J. (1997). Studying multivariate change using multilevel models and latent curve models. *Multivariate Behavioral Research*, *32*, 215–253.

DOI: 10.1002/acp

Marulis, L., Liu, L., Warren, C., Uttal, D., & Newcombe, N. (2007). *Effects of training or experience on spatial cognition in children and adults: A metaanalysis.* Poster presented at the biennial meeting of the Society for Research in Child Development, Boston, MA.

McClurg, P., & Chaille, C. (1987). Computer games: Environments for developing spatial cognition? *Journal of Educational Computing Research*, *3*, 95–111.

Muthén, L., & Muthén, B. (2004). *Mplus: The comprehensive modeling program for applied researchers. User's guide* (3rd ed.). Los Angeles, CA: Muthén & Muthén.

Newcombe, N., Mathason, L., & Terlecki, M. (2002). Maximization of spatial competence: More important than finding the cause of sex differences. In A. McGillicuddy-De Lisi, & R. De Lisi (Eds.), *Biology, society and behavior: The development of sex differences in cognition* (pp. 183–206). Westport, CT: Ablex.

Okagaki, L., & Frensch, P. (1994). Effects of videogame playing on measures of spatial performance: Gender effects in late adolescence. *Journal of Applied Developmental Psychology*, *15*, 33–58.

Paramswaran, G. (1994). Gender differences in horizontality performance before and after training. *The Journal of Genetic Psychology*, *156*, 105–113.

Paramswaran, G., & De Lisi, R. (1996). Improvements in horizontality performance as a function of type of training. *Perceptual and Motor Skills*, *82*, 595–603.

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.

Roberts, J., & Bell, M. (2000). Sex differences on a computerized mental rotation task disappear with computer familiarization. *Perceptual Motor Skills*, *91*, 1027–1034.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research* (2nd ed.). Boston, MA: McGraw-Hill.

Schmidt, R., & Bjork, R. (1992). New conceptualizations of practice: Common principles in three paradigms suggest new concepts for training. *Psychological Science*, *4*, 207–217.

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

Sims, V., & Mayer, E. (2002). Domain specificity of spatial expertise: The case of the video game players. *Applied Cognitive Psychology*, *16*, 97–115.

Stericker, A., & LeVesconte, S. (1982). Effect of brief training on sex-related differences in visual-spatial skill. *Journal of Personality and Social Psychology*, *43*, 1018–1029.

Subrahmayam, K., & Greenfield, P. (1996). Effect of videogame practice on spatial skills in girls and boys. In P. Greenfield, & R. Cocking (Eds.), *Interacting with video* (pp. 95–114). Norwood, NJ: Ablex.

Terlecki, M., & Newcombe, N. (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., & Kuse, A. (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. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*, 250–270.

Waller, D., Knapp, D., & Hunt, E. (2001). Spatial representations of virtual mazes: The role of visual fidelity and individual differences. *Human Factors*, *43*, 147–158.

DOI: 10.1002/acp