# The computer game training effect for women may depend on initial spatial ability scores

Robert Iversen

Master assignment in Psychology

Tromsø, 2010

Supervisor: Susanne Wiking

Department of Psychology
Faculty of Health Sciences
University of Tromsø

# Acknowledgements

First I would like to thank my dear supervisor Susanne Wiking. This I would like to do for several reasons, first of all for meeting with me and helping me find the topic I could work on in my master thesis. Later you helped me find new relevant articles, let me borrow your lab, program the computerized tests we used, and fund the whole project. Thank you for all of your advice and commentary on the previous drafts of the assignment. During this whole period you supported and encouraged me, many meetings and e-mails left me smiling and in good spirit. Thank you Susanne, I would not have made this project without you and your help.

Secondly I would like to thank all the participants that did give their precious time to participate in this project, especially those that did get "the computer training" whom actually ended up with a total of 10 hours each spent on the lab. To all of you; Thank you, we would not have any valid data, or for that matter a complete thesis without you.

My third thank goes to Morten Øvervoll for giving of his lecture time to recruitment for the project. As I was busy with the testing at the time of the particular lecture my wife and at the time six weeks old daughter went there to ask participants to sign up.

This brings us to the fourth and final thanks that goes to my wife and our little daughter. In the beginning of this period, of one and a half years, my wife was pregnant and with about one year left we became happy parents of a baby girl. Thank you to my wife, Birgitte. First of all for always being there for both me and our daughter, secondly for going to that psychology class to recruit and last but not least for giving me the time to work on this project. This project would not have come far without you. I would thank our daughter, Elizabeth, for being nice with your parents and cheering us up with your beautiful smile.

Kvaløysletta, 21.04.2010

Robert

Sammendrag

I dette forsøket ønsket vi å utforske hva det er i spill som kan forbedre spatiale evner. Tidligere forskning har vist at action spill kan forbedre spillernes score på Mental Rotation test (MRT), mens det er funnet bevis både for og i mot at puzzle spill kan gjøre det samme. Vi brukte tre forskjellige spill, og en kontroll gruppe, med totalt 32 deltakere matchet utover disse fire gruppene. Spillene var Medal of Honor: Pacific Assault, som har blitt brukt som action spill i tidligere forsøk; Portal, som er merket som et action/puzzle spill; og Supreme Commander, merket som et sanntids strategi spill. I tillegg til MRT brukte vi også Visual Patterns test (VPT) og Corsi Block test (CBT) for å utforske treningseffekten på spatiale ferdigheter. Ingen treningseffekt av spill ble funnet på de tre testene. Alle gruppene, inkludert kontroll gruppen, forbedret scorene på MRT og VPT, men det var ingen forskjell mellom spillgruppene. Resultatene indikerer videre at bare de matchede kvartettene med lav pre-test score på MRT forbedret scorene sine, mens de matchede kvartettene med middels og høy score presterte på samme nivå på de etterfølgende testene.

Nøkkelord: dataspill, spatiale evner, Mental Rotasjons Test, pretest scorer, trenings effekt.

Robert Iversen
(Student)

Susanne Wiking
(Veileder)

## Abstract

In this project we tried to explore what it is in games that may enhance spatial abilities. Previous research has shown that action games may enhance gamers' scores on the Mental Rotation test (MRT), while evidence is found both for and against that puzzle games could do the same. We used three different games, and one control group, with a total of 32 participants matched over these four groups. The games were Medal of Honor: Pacific Assault, which has been used as an action game in previous studies; Portal, which is labeled as an action/puzzle game; and Supreme Commander, labeled as a real-time strategy game. In addition to the MRT, we also explored the training effect on spatial abilities with the Visual Patterns test (VPT) and the Corsi Block test (CBT). No training effect was found for any of the games on any of the tests. All game groups, including the control group improved their scores on the MRT and the VPT, but there was no difference between the game groups. The results further indicate that only the matched quartets with a low pre-test score on the MRT improve their scores, while those with higher pre-test score perform at the same level on subsequent testings.


Keywords: computer games, spatial abilities, Mental Rotation Test, pretest scores, training effect.


Robert Iversen

(Student)

Susanne Wiking

(Supervisor)

# Preface

The idea for this thesis was developed in cooperation with the supervisor, as the student had a keen interest for both cognitive psychology and computer games. There have been several articles about improvement in spatial abilities due to computer game training, but none of them gives an explanation to why the spatial abilities improve. With that question in mind we set out to test our thesis.

The first semester was used to read up on the topic and plan the experiment, while the recruitment started early the second semester. The recruitment survey was made by the student, while the supervisor helped arrange a sign up in class of the first year psychology students. The experimentation phase was planned in cooperation with the supervisor and administered by the student. For the experiment we used two computerized tests, both programmed by the supervisor. The experiment took place in the supervisor's lab at the University of Tromsø during the second and third semester. Analysis of the data was done in cooperation with the supervisor. The student wrote the whole assignment and got guidance from the supervisor during the writing process.

Computer games have continually grown in popularity since their early release in the 1970s, in fact the game industry's revenue surpassed that of the movie industry some years back. According to eMarketer (2008) the revenue of the US video game industry in 2007 was 35,6 billion dollars, and it was estimated to have a stunning revenue of 63,2 billion dollars by 2013 (eMarketer, 2008). Recent research has suggested that spatial abilities might improve from training with computer games (Green & Bavelier, 2003; Feng, Spence, & Pratt, 2007; Cherney, 2008; Li, Polat, Makous, & Bavelier, 2009).

In an attempt to replicate the findings of training effects using computer games and elucidate what aspects of computer games that gives the effects we tested spatial abilities both before and after training. To measure visuospatial abilities the Mental Rotation test (MRT), the Visual Patterns test (VPT) and the Corsi Block test (CBT) were used. MRT has been used to measure the ability to mentally rotate a three dimensional figure without reference to one self (Hegarty and Waller, 2005). Both the VPT and the CBT measure short term visual memory (Della Sala, Grey, Baddeley, Allamano & Wilson, 1999; Milner, 1971). As for the training we used the three different computer games: Portal, Supreme Commander and Medal of Honor. In addition we had a control group that was tested at the same times as the game groups, but received no training. We included the control group to ensure that a possible training effect was not due to other factors, such as re-testing.

Spatial abilities relate to the way people process and present spatial information (Shah & Miyake, 2005). Miyake, Friedman, Rettinger, Shah and Hegarthy (2001) found that spatial abilities are related to working memory. According to the model of working memory that Baddeley and Hitch released in 1974, there are three components; the phonological loop responsible for storing and processing verbal and phonological information; the visuospatial sketchpad responsible for encoding visuospatial information; and the central executive that acts as a control center required for more complex processing (Repovs & Baddeley, 2006). Later research led to the addition of the episodic buffer component to the model, a component serving as temporary storage capacity for the phonological loop, the visuospatial sketchpad and long term memory. The central executive is in control of all the other components and uses the episodic buffer to integrate data from the different components in order to make the representation complete (Baddeley, 2000). Miyake et al. (2001) tested the relation between working memory and spatial ability factors (spatial visualization, spatial relation and perceptual speed). They found that all three of the spatial ability factors placed demands on the visuospatial sketchpad as temporal storage, but the factors differed in terms of demands placed on the central executive. Spatial visualization demanded the most of the central executive, while perceptual speed demanded the least (Miyake et al., 2001).

Previous researchers have used a range of different tests when testing for training effects from computer games, these include the Useful-Field-of-View (UFOV) (Green & Bavelier, 2003; Feng et al., 2007), the Mental Rotation test (MRT) (Feng et al., 2007; Cherney, 2008), the Card Rotation task (Cherney, 2008) and the Contrast Sensitivity test (Li et al., 2009). The question is, however, whether all these tests really measure the constructs they are assumed to, and if they are “precise” enough to be used at all.

The UFOV was one of the tests used by Green and Bavelier (2003) when they showed that playing an action video-game improves spatial attention. They did five different experiments, the first four using people who reported being either video-game players (VGPs) or non-video-game players (NVGPs), and the last experiment using only NVGPs. For the first four experiments they tested whether there were differences in performance between VGPs and NVGPs on the Flanker Compatibility task, Enumeration task, UFOV and Attentional Blink task. In the fifth experiment they trained their NVGPs using the action game Medal of Honor: Pacific Assault; this was chosen because they found it similar to the games the VGPs reported playing. The control group was trained with the game Tetris, which they assumed would not affect the spatial abilities because it only needs focusing on one object at a time. Both before and after training they tested their subjects using the UFOV, Attentional Blink task and the Enumeration task. Green and Bavelier (2003) found that the players outperformed the non-players on all of the tests, and when receiving training it were those training on the action video-game that improved the most. They also improved on the UFOV outside of the angle they trained.

The UFOV is assumed to measure attentional resources and their spatial distribution (Green and Bavelier, 2003). The task is to detect a target stimuli among distractors and remember in which of eight directions the stimuli appeared. We believe the UFOV might be a measure of visuospatial working memory rather than general attention. This is the main reason we chose the VPT and the CBT, which measures the simultaneous and the sequential visuospatial working memory respectively.

Furthermore, one can argue that the UFOV might not be a good test because of its timeframes. In the UFOV test the stimuli is shown for 10-30 milliseconds (Feng, et al., 2007). Is it even possible for the regular computers utilized in psychology labs today to measure and maintain these time requirements? If they are running a screen in 50Hz then the refresh rate itself is 20ms, that is the double of the minimum time requirement in the test. Since it is such a small amount of time it might be unlikely that this test is consistent, or rather that the computers are able to keep up with the short timeframes. That means that the subjects will have different amounts of time to look at the stimuli and the other items of the test, and hence have a different basis for their performance and get different results that cannot be predicted.

Feng, Spence and Pratt (2007) based much of their research on the experiments done by Green and Bavelier (2003). They chose the same action video game to train their participants, and they used the UFOV test, in addition to the MRT. Feng et al. (2007) showed that gender differences in mental rotation could be reduced by video-game playing. Males have been reported to outperform females on the MRT, this were also the case on the pre-tests in Feng et al.'s (2007) study. They were able to show that women might have more to gain from playing than men do. An interesting question here would be what happens if someone tests men with the same pre-score as the females, to see if they then gain as much as the females or if their level of gain is truly gender specific.

In the first experiment Feng et al. (2007) tested for group differences in spatial attention, the groups being based on the subjects' gender, if they played video-games and what field of study they were attending. In this study they used only the UFOV, and found that the VGPs outperformed the NVGPs, the science students outperformed the arts students and males did better than females. In the second experiment Feng et al. (2007) added the MRT, and chose only non-players as subjects. They trained their subjects using the same action game as Green and Bavelier (2003) for the experimental group, while the control group was trained using a 3D puzzle game (Feng et al., 2007). By doing so they found that after only ten hours of training, spread over a period of five weeks, the ones playing the action game had improved their performance on both the UFOV and the MRT. The improvements were greater for women than men in both tests and lead to diminished gender differences. The control group playing a 3D puzzle game showed no improvement (Feng et al., 2007).

One can argue that Feng et al.'s (2007) recruitment procedure may be a source of weakness. They posted an advertisement that described the preferred characteristics, and honored their subjects with $10 per hour. This might seem easy bucks for anyone, and students might be tempted to lie about their gaming habits to earn some money. For instance, you could have a player who lied about not playing, then received money for doing his or her hobby. This possible recruiting source of error might not have been evident in their first experiment as it called for both players and non-players (Feng et al., 2007). To avoid this possible pitfall ourselves, we decided to recruit through an internet questionnaire about gaming habits, not revealing who we were seeking. We were however forced to recruit some participants through other means as we got too few who fit our requirements in terms of no computer gaming within the last year.

The findings of Cherney (2008) contrasts with those of Feng and colleges (2007). She found that practicing both the 2D game Tetrus and the 3D game Antz significantly increased performance both on MRT and Card Rotation test. Cherney (2008) also tested whether the administration of the training influenced the results. By letting some of the participants train intensively one hour per day for three consecutive days, and the rest in three one-hour sessions over more than two weeks, she found that intensive (massed) practice was the most effective (Cherney, 2008). One can argue that this difference might be due to the fact that the time span between the two tests was shorter when doing the massed practice, meaning the participants might remember test items better and hence show more improvement. To ensure that the difference in massed versus distributed practice is true, the time span between pre- and post-tests must be the same for both groups. In total her participants trained four hours, and even though this constitutes less training than what was used in the previously mentioned studies, she was able to conclude that both massed and distributed training had positive effects on both MRT and Card Rotation tests (Cherney, 2008).

The previous research did not focus on finding out why playing computer games should improve spatial abilities, or what aspects of the games that made the diffference. They focused on whether there was a difference between action games or non- action games (Feng et.al, 2007; Green & Bavelier, 2003). Both Green and Bavelier (2003) and Feng et.al (2007) found that only action games gave training effects. We suggest that there is a need for a different classification of games for research purposes than the traditional division into genres. The genre labeling is done by the developer of a game and it is supposed to help the consumer to decide if the game is their cup of tea or not. The labeling is unlikely to help a researcher understand why spatial abilities improve when playing. The traditional genres consists of different elements; for example puzzle games are games that presents puzzles that you need to solve finding smart solutions, like Tetris; action games provides action with shooting, explosions and mosstly a high tempo, like Medal of Honor. You also have strategy games where you are to use tactics in order to successfully lead your army, your company or your nation to victory, like Supreme Commander.There  are other genres as well, such as car games, flight simulator games, roleplaying games (RPGs) and massive multiplayer online games (MMOs). Even though there are very spesific expectations to what each of the genres should contain, the expectations might be different from developer to developer, publisher to publisher and person to person. The lack of confirmity and consistency makes the classification unsuitable to use for research purposes. Another problem for research purposes is that a game might be labelled with more than one genre, and also might be labelled differently depending on where you look.

We hypothesized that it might be the first person perspective of the action games, rather than the explosions and killings (the elements that makes it an action game), that contributes to the training effect. By training one group with the action game Medal of Honor, one group with the action puzzle game Portal and a third group with the strategy game Supreme Commander we hoped to redefine training effects on spatial abilities. Since Portal is labelled as an action puzzle game, but still is very similar to an action game in terms of elements such as the perspective where you see the environment through the eyes of your figure, in first person, and the manuvering in a three dimentional environment, it was ideal to use. If the participants training with Portal showed the same improvement as those training with Medal of Honor we would have revealed that it is not the genre of the game that determine if spatial abilities would improve or not. Supreme Commander, the strategy game, has action game elements such as big explosions, but a very different perspective where you see everything from above in bird's eye view. If the participants training with this game showed the same improvement as the ones training with Medal of Honor it could be the action elements rather than the perspective that ensured improvement.

The important thing to remember with that hypothesis is that there are more to games than action or perspective that might influence spatial abilities. Other aspects that might make a difference is the graphics in terms of realism, how the colors are used, whether there is a lot of contrasts and what resolution the game is run in. Cherney (2008) found training effects using both a 2D game and a 3D game, showing that the training effect is independent of this aspect of a game. Another factor might be the gameplay, meaning how fun the game is for the player. This is a very individual factor, as what is fun for one person might not be for another. It might also depend on the individual skills of the player, for example for a person with poor coordination shooting games that require presision might be boring, while puzzle games might be fun. Every game tells a story, where you as a player might decide more or less what happens. The story and gameplay is mostly relevant because they might determine how much the player is captured by the game, and one might assume that a player that is very captivated puts more effort in the game and hence get better training effects. We will not be able to address all these other aspects of games as it would require a bigger and more extensive study. But an interesting alternative explanation to the training effect is that it is not something in the game itself that ensures the improvement but rather something about the individual playing the game that does.

Who are the gamers, the ones assumed to be gaining spatial benefits from games? A report from BBC back in 2005 shows that from age 6 to 65 an astonishing 59% of the British population have played videogames at least once the last 6 months (Pratchett, 2005). It also showed, maybe a bit surprisingly, that the percentage of women and men that played was about the same with 45% and 55% respectively (Pratchett, 2005). One might assume though that women and men play different kinds of games. Youth was shown to have the highest percentage of gamers and it seemed to decline over the years. The average gamer was found to be about 30 years of age and this age seems to be rising (Pratchett, 2005). These two things, that the percentage of gamers decline at higher ages and that the average gamer is getting older, can give us the assumption that more and more older people are trying games and that the generations that are used to games carry the tradition with them to later stages of life.

## Method

### Survey and selection

An Internet survey was used to recruit participants to our study (see Appendix 1). The survey was distributed to all student e-mails at the University of Tromsø, and about 900 answered the survey. Of the 900, only 94 did not play videogames. We also did recruitment by going to lectures on the first year of psychology, and 14 students were recruited this way. In the pre-tests we ended up testing a total of 58 participants, with an age distribution from 18-55 years.

The sex distribution of the internet survey was balanced. However of those we could use from the internet survey the distribution was heavily skewed toward females (80 of 94). We decided to recruit females for the actual study. The participants' fields of study varied since all students at the University of Tromsø got the mail for the internet survey.

### Participants

The study included 32 participants, of which 8 were recruited through sign-up in class and the rest were recruited through the internet survey. The participants were all female, and their age ranged from 18 to 38 (mean 26,4 years). We matched the participants into quartets based on their scores on the pre-tests. This matching involved finding 4 participants that had as similar as possible scores over the three tests. We did this to be able to assign three of them to each of the games and the last one into the control group. Thus, we had eight participants on each of the games and eight participants in the control group.

All participants gave their informed consent. As a reward participants received one scratching ticket for taking the pre-tests and one for each completion of the post-tests. Those that went through the videogame training also received a gift certificate for NOK 500,-.

## Materials

For spatial ability testing we used three different tests. These were the Mental Rotation Test (MRT), the Visual Patterns Test (VPT) and the Corsi Block Test (CBT). To test computer anxiety we used the Computer Anxiety Rating Scale (CARS).

In the original procedure of Shepard and Metzler (1971), participants were given the task to look at two figures and decide if the figures were the same or not. In this task one of the figures could be rotated. We constructed a computerized version of the MRT test based upon Shepard and Metzler's images (see Figure 1). The participants had to evaluate a total of 100 image pairs. We also needed a second version of the test with the same difficulty for the second time the participants were tested. This version was constructed by mirroring one of the figures in each of the pairs, making the ones that were similar not similar and vice versa. The MRT was programmed using E-prime software version 1.1 SP3 (Schneider, Eschman, & Zuccolotto, 2002).

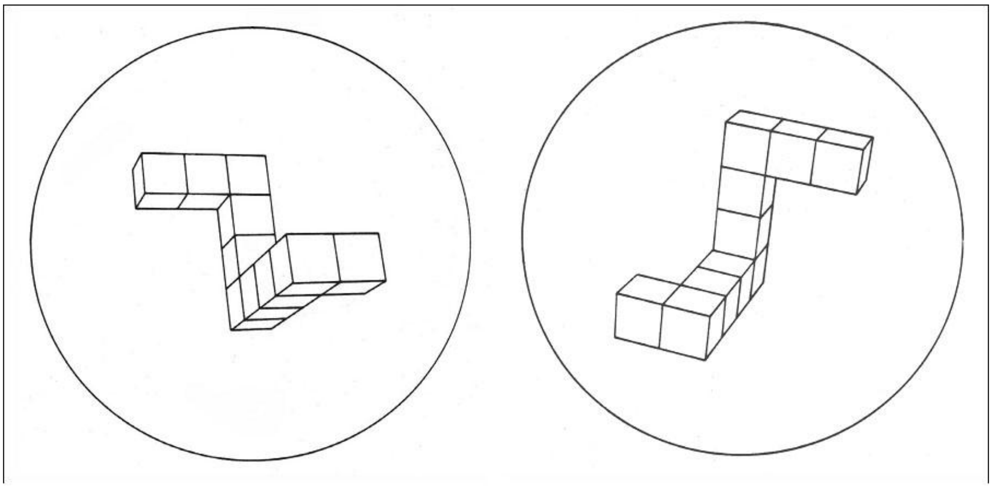

**Figure 1 Example of a MRT item**

The next test we used was the Visual Patterns Test (VPT). The original VPT (Della Sala, Grey, Baddeley, & Wilson, 1997) is a test where the participant gets to see a matrix filled to half with black squares and to the other half with white squares. The dark squares are spread randomly within the matrix. The participants' task is to look at the matrix for 3 seconds, wait for 10 seconds and then mark where the black squares were on an empty matrix. The test starts with two black and two white squares working itself up to 15 of each. The version of the VPT we constructed differed from the original in three main ways.

First, our VPT was semi-computerized in order to make it more relevant for videogame playing. The matrix display was programmed using E-prime software version 2.0 (Schneider, Eschman, & Zuccolotto, 2002). Secondly, the participants were not allowed to look at the empty matrix, while they had to wait the 10 seconds. The third and final difference was that we constructed the VPT so that the matrices were always symmetrical, 2x3, 3x4, 3x6, 4x6 and 5x6. When the number of black squares had reached half of the total squares in a matrix and the participant went on to the next level, the whole matrix was switched to the bigger one. Figure 2 shows examples of the smallest and largest VPT items.

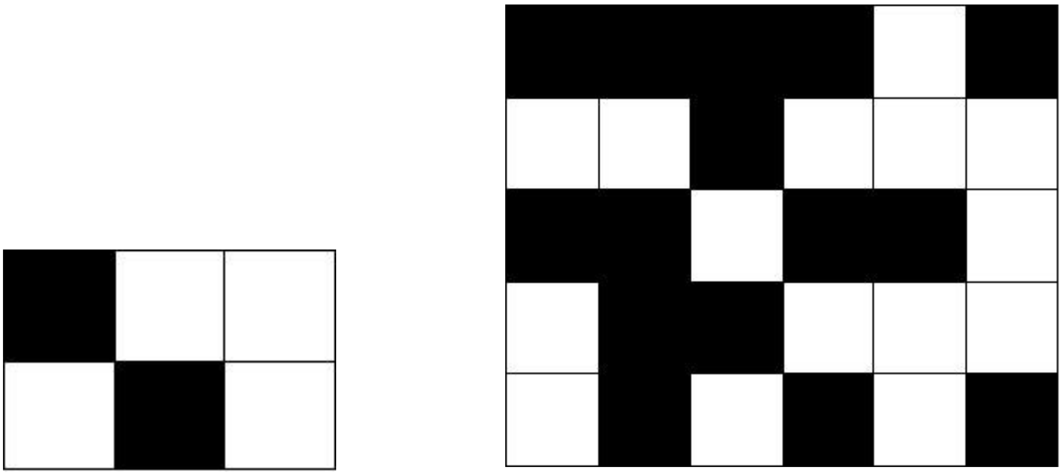

Figure 2 Examples of the smallest (2x3) and largest (5x6) VPT items.

Finally, the standardized version and procedure of the Corsi Block Test (CBT) was used (Milner, 1971). Both the VPT and the CBT ends when the participant fails more than one sequence on the same level, and the last successful level was recorded as the individual score. The simultaneous sequences in the VPT range up to 15, while the sequential sequences in the CBT range up to 9. All the sequences used in this study were constructed by means of the “randperm” function in MATLAB.

In this project three computers were used for gaming, and one of them was also used for the tests. The computer used for the MRT and the VPT in addition to Medal of Honor was a single core computer at 3.00 GHz, 1GB ram and ATI Radeon x600 graphics card. Medal of Honor was run with 800x600 resolution at 75Hz. The computer used for Portal was a dual core computer at 3.00 GHz, 3GB ram and ATI radeon 4870x2 graphics card. Portal was run with 1600x1050 resolution at 75 Hz. The computer used for Supreme Commander was a single core computer at 3.00 GHz, 1GB ram and Radeon x300 graphics card. Supreme

Commander was run with 1024x768 at 60Hz.

The original CARS test (Heinssen, Glass & Knight, 1987) consisted of 19 items, but the version we used was the shortened version by Miller & Rainer (1995). The seven items of the test were translated to Norwegian (See Appendix 2).

Procedure

When participants came in for the tests they did the MRT first, and then they did the VPT while they were still seated by the computer. The CBT was completed last.

In the MRT the two figures were shown simultaneously on the screen. For identical the participant pushed the “v”-key (valid) and for mirrored the “n”-key (non-valid) was pushed. If the participant did not solve the task within 10 seconds the test moved on to the next set of figures. There were given five practice trials with feedback before the 100 real trials without feedback.

In the VPT the participant saw a matrix with some black and white squares for 3 seconds on the screen, and then they had to wait for 10 seconds before a beep came from the computer. When the beep came they had to remove a colored sheet of paper and write down their answer in an empty matrix on paper.

In the CBT the black board with nine cubes was placed between the participant and the experimenter. Participants had to repeat the sequences given on the boxes by the experimenter, starting at a sequence of 2 with 3 trials where at least 2 have to be correct to reach the next level of difficulty.

The CARS test was administered as a paper-pencil test given after completing the CBT in the testing five months later. The participants were to answer seven items regarding computers and their feelings towards them. They answered on a Likert type scale from one, “strongly disagree”, to five, “strongly agree”.

The participants in the game groups had to play 2 hours each week over a total of four weeks. Most of the participants were tested with the pre-tests the week before they started playing, and again both the week after playing and about 5 months after this training. For a few participants we had to make exceptions, and skip training for one week, transferring the training session to the next week. This happened if the participant got sick and could not come for training, or when the participant had to travel for one week. All participants completed the tests and training parts within schedule give or take one week. The control group came back for post-tests only, but they did so after the same time (five weeks and five months) as the game groups. Two of the participants were unable to come in for the testing after 5 months.

After the tests were finished, the participants received their rewards and were debriefed and thanked for their participation.

Data analysis

When analyzing the MRT results, the accuracy and mean response times were calculated. Accuracy was given by the proportion of correct items on the test. Response time means were calculated for each individual, using only the response times for correctly answered items. Response times that were more than three standard deviations from the individual mean were deleted from the dataset, in three iterations.

The two participants missing in the testing after five months happened to be in the same quartet. Therefore we excluded their whole quartet when analyzing the difference between the first post-test and the one after five months, and when analyzing all three testings together.

## Results

When the participants came in for the post-test all those who had played were asked a few questions, where they were to answer using the Likert scale one to five, one being very little and five being very good. They were asked how much they liked the game and how well they felt they understood their task. The overall average for how well they liked the game was M=2.50; SD=1.063; n=24. Was there any game that was better liked than the others? Participants liked Portal the most (M=3.12; SD=.835; n=8) and Supreme Commander the least (M=2.0; SD=.926, n=8), but this difference failed to reach significance (F=2.657; p=.094). The overall average understanding of their task was M=2.88; SD=1.116; n=24, and there was no significant difference between the games (F=2.358; p>.10).

One of the players managed to complete the whole game of Portal and was provided with a new challenge in form of some of the earlier completed stages but with a more diffucult soultion provided by the developers of the game. As for the rest, their stage when completing the training phase varied from being half way through to being at the last stage. Among those playing Supreme Commander and Medal of Honor none were able to complete the whole game, but all reached sufficient levels and improved their ingame skills.

*Mental Rotation Test.* Participants generally improved their score on the second MRT test compared with the first. The total average of spatial accuracy on the pre-test MRT (M=

.7415; SD=.10421; n=32) differed from the total average of spatial accuracy on the post-test MRT (M=.7992; SD= .08591; n=32 ), showing a significant [F(1,28)=14.376; p=.001; $\eta_p^2$=.34] improvement in spatial accuracy from the first to the second test. The total average in spatial accuracy after five months was M=.7920; SD=.07966, n=28, and did not differ from the post-test (F<1; p>.10) which had a total average of M=.7842; SD=.08098, n=28 when excluding the quartet with missing participants on the five month later test. Looking at all three testings, the total average significantly improved [F(2,48)=14.144; p<.001; $\eta_p^2$=.371]. The average for each of the game groups is presented in Figure 3. There were no significant differences between the game groups (F<1; p>.10).

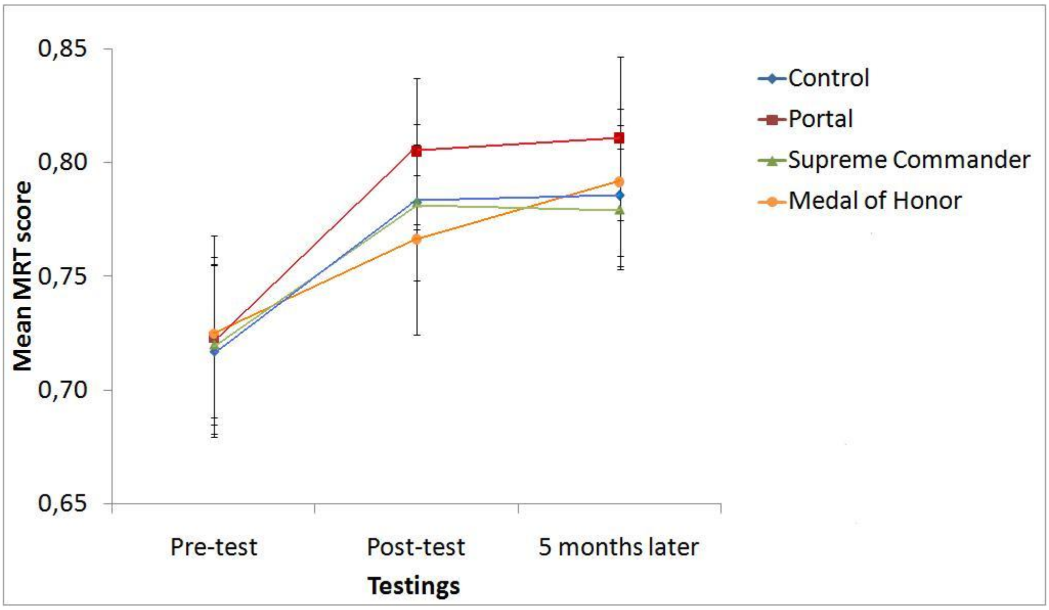


Figure 3 Mean MRT score over the game groups, n=28

The one thing that affected improvement in spatial accuracy in this study was not computer games, but the score on the pre-test. The matched quartets, where all the participants were included and matched based on pre-test scores, were divided into three different spatial ability groups based on the mean pre-test scores of the quartets. The three matched quartets with a mean MRT score lower than 0.7 were in the low score group. The two quartets with mean scores between 0.7 and 0.8 were in the middle score group and the three quartets with mean scores above 0.8 were in the high score group. The low score group was the group that improved, while the middle and high score groups did not improve, as seen in Figure 4. This interaction was significant with [F(2,29)=6.299; p=.005; $\eta_p^2$=.303; n=32] from the pre-test to the post-test. When including the testing after five months and excluding the quartet with missing participants the interaction was still significant [F(4,50)=3.598; p=.012; $\eta_p^2$=.224; n=28].

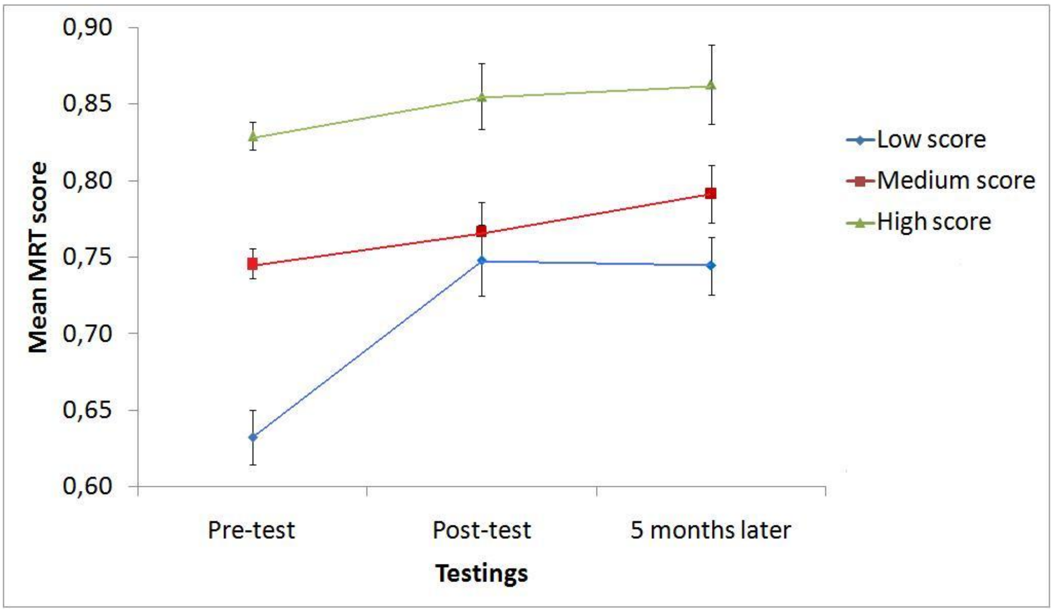


**Figure 4 Mean MRT over the spatial ability groups, n=28.**

The spatial response time total average on the pre-test (M=4390.8; SD=878.4; n=32) was significantly different [$F(1,28)=16.836$; $p\approx.000$; $\eta_p^2= .38$] from the total average of spatial response time on the post-test (M=3797.3; SD=867.6; n=32). When excluding the quartet with missing participants the response time total average on the first post-test was M=3843.0; SD=847.0; n=28, while the test after five months had M=3342.3; SD=937.6; n=28 in total average. The post-test and the test after five months differed significantly in response time [$F(2,24)=30.007$; $p<.001$; $\eta_p^2=.556$]. When including all three testings the difference remained significant [$F(2,48)=31.194$; $p<.001$; $\eta_p^2=.565$]. The average response times of each game group are similar, and multivariate methods showed no significant difference between them ($F<1$; $p>.10$).

*Visual Patterns Test.* On the pre-test VPT the total average was M=6.75; SD=2.11; n=32, while the total average on the post-test VPT was M=7.94; SD=1.966; n=32. The improvement from the pre-test was significant [$F(1,28)= 10.845$; $p=.003$; $\eta_p^2= .28$], but there was no difference between the game groups ($F<1$; $p>.10$). When testing after five months the total average VPT score was M=7.64 ; SD=1.254 ; n=28, a result that is not significantly different ($F<1$; $p>.10$) from the post-test result when excluding the quartet with missing participants M=7.61; SD=1.833 ; n=28. When including all three testings, however, there is an overall significant difference [$F(2,48)=12.674$; $p<.001$; $\eta_p^2=.346$], but still no difference between the game groups ($F<1$; $p>.10$) as seen in Figure 5.

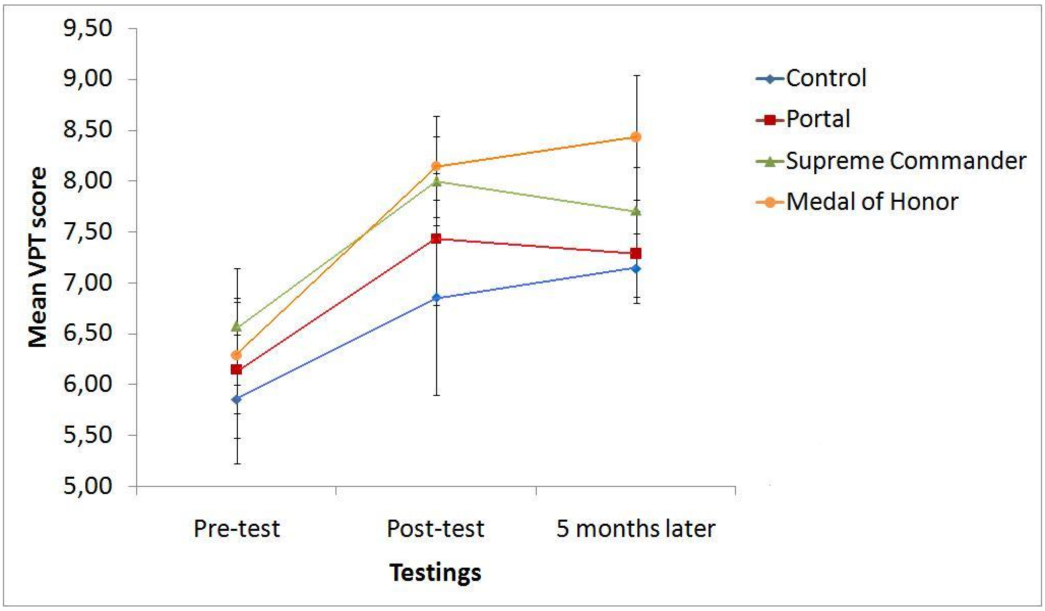


Figure 5 Mean VPT score over the game groups, n= 28.

*Corsi Block Test.* The CBT was the only test where there was no significant difference ($F<1$; $p>.10$) between the total average scores on the pre-test (M=6.03; SD=.897, n=32) and on the post-test (M=6.19; SD=1.176, n=32), nor any difference between the game groups ($F<1$; $p>.10$). Five months later the result remains the same, with no difference ($F<1$; $p>.10$) between the post-test M=6.14; SD=.9; n=28 and the test after five months M=5.93; SD=.813; n=28. When including all three testings there is still no difference between the testings, nor between the game groups ($F<1$; $p>.10$). The average scores of the game groups are presented in Figure 6.

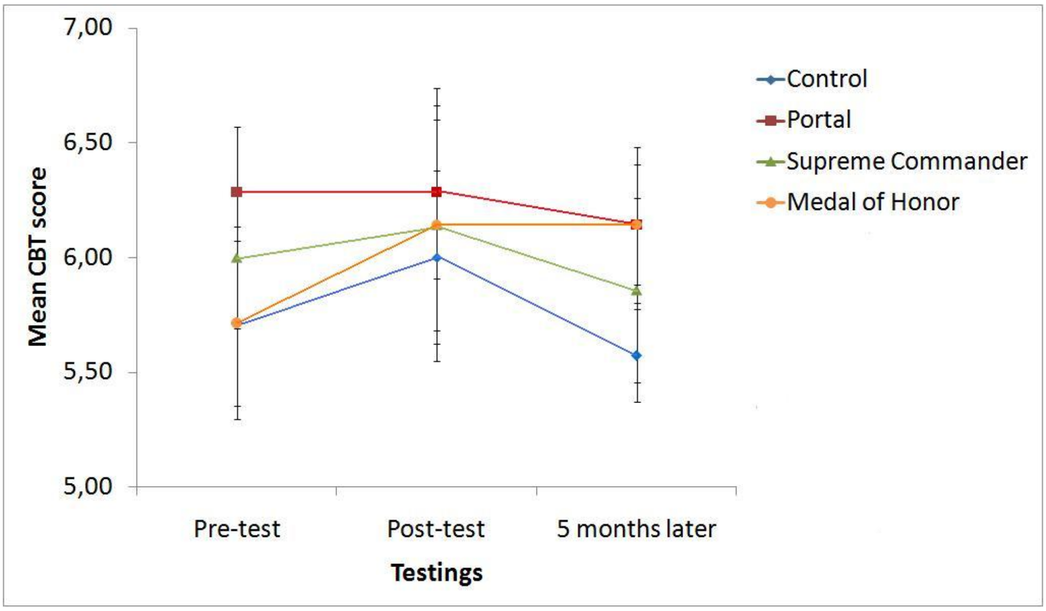


Figure 6 Mean CBT score over the game groups, n=28.

The CARS test revealed that the control group had tendency toward a higher score on the “high anxiety” items of the test than did the other game groups [$F(3,24)=2.516$; $p=.082$], but could not be separated from the other game groups on post-hoc testing ($p>.05$, Tukey B). The mean score of the game groups are presented in Table 1. On the “low anxiety” items of the CARS test there were no such difference [$F(3,23)<1$; $p>.10$].

We also tested the difference between the three spatial ability groups (low, medium and high) on CARS, and there was no difference on either high anxiety items ($F<1$; $p>.10$) or low anxiety items ($F<1$; $p>.10$).

Table 1. Mean score on “High anxiety” items of CARS

| Game group | Mean | SD |
|---|---|---|
| Control | 8.86 | 3.132 |
| Portal | 6.57 | 2.070 |
| Supreme Commander | 6.29 | 1.890 |
| Medal of Honor: Pacific Assault | 6.00 | 1.155 |

NOTE: *n=28*

## Discussion

The purpose of this study was to elucidate what it is in an action game that ensures a training effect in spatial abilities. We hypothesized that it might be due the first person perspective of an action game. This was tested using three different tests of spatial abilities, three different computer games and a control group. All participants were tested and matched into quartets and then assigned to either a game group or the control group. Then the game groups received training on their respective computer games before all groups, including the control group, were tested again, both one week after the game training and about five months later. The three different computer games were chosen so that if one or two, but not the other gave a training effect we might be able to decide whether it could be related to the perspective or not.

We had predicted an improvement in spatial ability tests for those receiving training on either of the games, but not in the control group. The results did not support this main hypothesis. Even though the overall average improved on both the MRT spatial accuracy and the VPT, there was no difference between the game groups and the control group. Instead, there was a difference between the spatial ability groups when it came to spatial accuracy on the MRT, providing evidence that only the ones with poor MRT scores might improve in subsequent testing.

There are at least three possible explanations for the fact that those with low pre-test MRT scores improved the most. First, it might be due to the fact that they have a larger potential for improvement than those with a higher pre-test MRT scores. This does not seem likely, as both the participants with medium and those with high scores failed to improve. Another possibility is that these participants experienced anxiety during the pre-test. Experiencing anxiety might limit their test-performance and hence give a low pre-test score. It is likely that such anxiety would be reduced during the post-test, as the testing is a more familiar situation by then. Females are more likely to experience this form of computer anxiety (Broos, 2005), and the present study had only female participants. We tested the participants' computer anxiety at the last session, and there was no difference between the spatial ability groups at this point. The third potential explanation is that the low score group might be less familiar with the computer than the other groups, and hence struggle to complete the computerized MRT, while at the post-test they know how to manage the computer and can fully concentrate on the mental rotation. The latter explanation is supported by the findings of Roberts and Bell (2000) who showed that letting participants familiarize with the computer before taking a two dimensional Mental Rotation task eliminated the previously reported gender differences.

The results of Feng et al. (2007) might be a consequence of them using participants with low pre-test scores. Their reported MRT means are somewhat different than the ones reported in this study due to the fact that they used a computerized version of the Vandenberg & Kuse (1978) MRT, while we used a computerized version with the Shepard and Metzler (1971) stimuli. Both tests present a three dimensional figure, the main difference between the two tests is how many alternatives are presented. The Shepard and Metzler (1971) procedure presents only one alternative where the participant is to decide whether the figures are the same, while the Vandenberg & Kuse (1978) presents four alternatives, where the task is to decide which two of the alternatives that are the same as the originally presented figure. In the Feng et al. (2007) study they presented 24 sets of figures, that gives a maximum score of 24. The MRT used in this study presents 100 figures, giving a maximum score of 1, or 100 percent correct. If a participant answered all items by guessing in the Feng et al. (2007) study the score would be four, as there is one sixth chance of guessing the correct answer on each set. In this study the result by guessing is 0.5, as there is a 50 percentage chance of guessing the correct alternative on each set.

Feng et al. (2007) reported the mean square root transformation of the number of correct items. For females it was 2 on the pre-test, giving a real mean of 4, while the mean square root transformation of the number of correct items for males was 2.9, giving a real mean of 8.41. For females the mean score is the exact score one would expect from guessing, while the score was better for the males. In accordance with the findings of the present study the low pre-test scores of Feng et al. (2007) makes it more likely that their participants would improve their score. The low pre-test scores, and the fact that the females started out with a lower score than did the men, might explain some of the sex difference in improvement that they found.

The fact that the Shepard and Metzler (1971) based MRT and the Vandenberg and Kuse (1978) based MRT are two completely different tests, makes it difficult to compare the scores. Since the figures they use are of the same format however, it would be reasonable to compare scores if you take the differences between the tests into consideration. The differences are the possible max score and the chance of guessing correctly, as there are different amounts of alternatives. We hereby propose a formula for transformation from the Vandenberg and Kuse based MRT score to the Shepard and Metzler based MRT score.

$$SM\ score = \frac{VK\ real\ score - VK\ mean\ guessing\ score}{\left(\frac{Max\ VK\ score - VK\ mean\ guessing\ score}{Max\ SM\ score - SM\ mean\ guessing\ score}\right)} + SM\ mean\ guessing\ score$$

The “VK real score” is the score that a participant had on a Vandenberg and Kuse based MRT, while the “VK mean guessing score” is the score one can get by guessing on all the items of the test. The “VK mean guessing score” is found by multiplying the chance of guessing both alternatives correctly with the max score of the Vandenberg and Kuse MRT. “SM mean guessing score” is the score one can get from guessing on the Shepard and Metzler based MRT, and is found by multiplying chance of guessing correctly with the max score. The first thing this formula does is to subtract what you can get from guessing from the “VK real score”, so that you only have what skill contributed with to the test result. The next thing the formula does is to divide the real skill contribution with the relation between the max possible skill contribution for both tests. The max possible skill contribution for each test is found by subtracting the mean guessing score from the max score. By doing this we have transformed the real VK skill contribution into the skill contribution score for the SM test. The only thing left to do is to add the “SM mean guessing score” and you have transformed the “VK real score” into its equivalent “SM score”.

Let's take a closer look at the MRT results from Cherney (2008), who used the Vandenberg and Kuse procedure with a max possible score of 20. Cherney (2008) reported three different female means, one for the group playing a three dimensional game (M=7.4), one for the group playing a two dimensional game (M=7.5) and one for the control group (M=10.5). This is interesting because if you use the transformation formula, the two game groups have a SM mean score of 0.62 and 0.63 which means they belong in the low score group in the present study, while the control group has a SM mean score of 0.72 which means it belong in the middle score group. As the results of the present study indicate, only the low score groups will improve on the MRT, and it might be the pre-test score rather than the computer game playing that ensured the difference between the game groups and the control group in the Cherney (2008) study. The fact that the male MRT pre-test scores reported by Cherney (2008) all were above 11, that equals 0.73 and the middle score group might explain why the males showed no improvement on the MRT.

An important question is whether the improvement on the MRT is due to the game training or only the effect of re-testing. Cherney and Neff (2004) showed that participants that had previously completed a Mental Rotation test scored significantly better than those who had not. The overall improvement in the MRT scores in the present study might be explained by the effect of having taken the test once before, supported by the fact that there were no differences between the game groups and the control group. However, the control group was higher than the game groups in computer anxiety, and may therefore have benefited more from the lower anxiety in additional testings.

The result of the CARS test suggests that those in the control group avoided being in one of the playing game groups because of computer anxiety. Due to a lack of participants rotations were made within the matched quartets. If one of the participants was offered participation in a game group but declined the offer, they were asked if they could be in the control group instead, and the participant originally in the control group was offered a place in a game group. In retrospect, this recruitment procedure was unfortunate, as we do not know how much the individual computer anxiety influenced performance on subsequent testings. The ideal situation would have been to have a bigger pool of participants to match from, so that if one said no to participate in a game group we could recruit someone new instead of rotating within the quartet.

A low number of participants is a common difficulty and shortcoming of these kinds of studies. In the Feng et al. (2007) study they had 48 participants in the first experiment, but only 20 in the second where they trained the participants using an action game. The present study had a total of 32 participants, but divided into a larger number of groups. A lack of participants makes it difficult to generalize, yet there will be more and more difficult to find participants for such studies as the report from BBC pointed out that more and more people are trying computer games (Pratchett, 2005). In addition, the upcoming generations will have had both computers and computer games as a part of their childhood environment.

The difference in recruiting between the present study and the Feng et al. (2007) study may have resulted in two different samples. Theirs being one of non-video game players that knew they would be playing a computer game, and hence wanted to play the required time for the study. The present study on the other hand consisted mainly of participants that did not know what to expect when they answered an Internet survey about videogame playing. The difference in sample may have contributed to the difference in findings as it might have been different motivations between the participants of the two studies.

Feng et al. (2007) argued that their research has practical implications for attracting people to mathematical and engineering sciences. One can argue that the present study showed that only those with low MRT scores would improve their scores and hence lessen the practical implications of their findings. The next question then would be whether video game training could improve the spatial ability enough to make the mathematical and engineering sciences attractive and manageable to people.

Whether such an improvement is possible is left for future research to answer. But first of all one should put more effort in finding out whether the game training effects on spatial abilities is real. This study was unable to replicate the findings of Feng et al. (2007) and Cherney (2008), and found no differences between the control group and the game groups. In case the effect is real, research should focus on why computer games may enhance spatial abilities. What is it in the games or about the gamer that ensures the improvement? The third suggestion we would like to make is to explore why only the low score spatial ability group improved their MRT scores.

The present study brings attention to a possible limitation for improvement in spatial abilities, as it may be only those with a poor basis that can improve. How this affects us in the daily life might depend on the unique situation each of us is in. If you are one of those always doing good in rotation tasks when they show up as “brainteasers” in the newspaper it might mean you are as good as you’ll ever be. But if you are struggling solving such a puzzle, don’t despair - you might be able to improve if you keep trying. If you are a fellow researcher you should be sure to take the pre-test scores into account when analyzing your results.

# References


Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423.

Broos, A. (2005). Gender and information and communication technologies (ICT) anxiety: Male self-assurance and female hesitation. *Cyberpsychology & Behavior*, *8*, 21-31.

Cherney, I. D. (2008). Mom, let me play more computer games: They improve my mental rotation skills. *Sex Roles*, *59*, 776-786 .

Cherney, I. D., & Neff, N. L. (2004). Role of strategies and prior exposure in mental rotation. *Perceptual and Motor Skills*, *98*, 1269-1282.

Della Sala, S., Grey, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo-spatial memory. *Neuropsychologia*, *37*, 1189-1199.

Della Sala, S., Grey, C., Baddeley, A., & Wilson, L. (1997). *The Visual Patterns Test: A new test of short-term visual recall*. Feltham, Suffolk: Thames Valley Test Company.

eMarketer. (2008, August 25). *Video Game Revenues Keep Rising*. Retrieved May 14, 2009, from website for eMarketer.com: http://www.emarketer.com/Article.aspx?R=1006495

Feng, J., Spence, I., & Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. *Psychological Science*, *18*, 850-855.

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

Green, C. S., & Bavelier, D. (2007). Action-video-game experience alters the spatial resolution of vision. *Psychological Science*, *18*, 88-94 .

Heinssen, R.K., Glass, C.R., & Knight, L.A. (1987). Assessing computer anxiety: Development and validation of the computer anxiety rating scale. *Computers in Human Behavior, 3*, 49-59.

Li, R., Polat, U., Makous, W., & Bavelier, D. (2009). Enhancing the contrast sensitivity function through action video game training. *Nature Neuroscience*, *12*, 549-551.

Miller, M. D., & Rainer, R. K. (1995). Assessing and improving the dimensionality of the computer anxiety rating scale. *Educational and Psychological Measurement, 55*, 652-657.

Milner, B. (1971). Interhemispheric differences in the localization of psychological processes in man. *British Medical Bulletin, 27,* 272-277.

Miyake, A., Friedman, N. P., Shah, P., Rettinger, D. A., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General*, *130,* 621-640.

Pratchett, R. (2005, December). *Gamers in the UK: Digital play, digital lifestyles.* Retrieved May 14, 2009, from website for BBC: http://open.bbc.co.uk/newmediaresearch/files/BBC_UK_Games_Research_2005.pdf

Repovs, G., & Baddeley, A. (2006). The multi-component model of working memory: Explorations in experimental cognitive psychology. *Neuroscience, 139*, 5-21.

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

Shah, P., & Miyake, A. (2005). *The Cambridge Handbook of Visuospatial Thinking.* New York: Cambridge University Press.

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

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburg: Psychology Software Tools Inc.

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

# Appendix 1 – Recruitment survey

## Computer gaming habits

### 1. Intro

Kjære deltaker

Vi er her interessert i å se på vaner knyttet til spilling på pc eller konsoller.
Formålet videre med denne undersøkelsen er å bruke den som en forundersøkelse for å se om du er en egnet deltaker videre i vår studie.

Robert Iversen,
Masterstudent ved Universitetet i Tromsø
----

Dear participant

We are interested in habits people have regarding computer- and console gaming.
The further purpose of this study is to use it as a pre-survey to see if you may be a fit participant in the next step of the study.

Robert Iversen,
Masterstudent ved Universitetet i Tromsø

## Computer gaming habits

### 2. Vaner/Habits

*** 1. Kjønn/Sex**

- Kvinne/Woman
- Mann/Male

**2. Hvilket fagfelt er du tilknyttet?**
**What kind of education are you doing?**

- Arkitektur/Architecture
- Ergoterapi eller Fysioterapi/Ergonomics or Physiatrics
- Estetikk eller Dans eller Drama eller Kunst/Aesthetics or Dance or Drama or Art
- Filosofi/Philosophy
- Fiskeri og sjøfart/Fishing and shipping
- Geografi/Geography
- Geologi/Geology
- Helse og sosialfag eller samfunnsfag/Health and social studies
- Historie eller Arkeologi/History or Archaeology
- Ingeniør/Engineering
- IT og Kommunikasjon/IT and Communication
- Jus/Law
- Media/Media
- Medisin eller Odontologi eller Farmasi/Medicine or Odontology or Pharmaceutics
- Pedagogikk eller Lærer/Pedagogics or Teacher
- Psykologi/Psychology
- Realfag/Subjects in the sciences
- Språk eller Tolk eller Litteratur/Language or Interpreter or Literature
- Sykepleie eller Radiograf eller Jordmor eller Vernepleie/Nursing or Radiologist or Midwife or Social educator
- Teologi eller Religion/Theology or Religion
- Økonomi og ledelse/Economics and Leadership

Annet?/Other?

*** 3. Har du noen gang regelmessig spilt dataspill eller konsoll spill?**
**Have you ever played computergames or consolegames on a regular basis?**

- Ja/Yes
- Nei/No

*** 4. Hvor lenge siden er det du sist spilte?**
**How long is it since you last played?**

- 0-7 dager/days
- 1-4 uker/weeks
- 1-6 månder/months
- 7-12 månder/months
- Over 12 månder/Over 12 months
- Aldri spilt/Never played

## Computer gaming habits

*** 5. Hvor mye spiller du? (i timer pr uke)**
**How much do you play? (In hours pr week)**

- ◯ 0
- ◯ 1-3
- ◯ 4-7
- ◯ 8- eller mer/or more

## Computer gaming habits

### 3. Final

Hvis du er en av deltakerne vi kan bruke videre i studien trenger vi kontakt info.

---

If you are one of the participants we can use later in the study, we need contactinfo.

**1. Navn/Name**

**2. Telefonnummer/Telephonenumber**

*** 3. E-post/E-mail**

Appendix 2 – Computer Anxiety Rating Scale (CARS)

## CARS-7

**Sett en ring rundt ett av tallene på skalaen for å vise hvor godt hvert utsagn passer for deg.**

| | | **Svært uenig** | | | | **Svært enig** |
|---|---|---|---|---|---|---|
| 1. | Jeg nøler med å bruke en datamaskin fordi jeg er redd for å gjøre feil som ikke kan rettes opp. | 1 | 2 | 3 | 4 | 5 |
| 2. | Jeg føler meg usikker på min evne til å tolke en datautskrift. | 1 | 2 | 3 | 4 | 5 |
| 3. | Jeg har prøvd å unngå datamaskiner fordi de er ukjente og litt skumle for meg. | 1 | 2 | 3 | 4 | 5 |
| 4. | Jeg har vanskelig for å forstå de tekniske aspektene med datamaskiner. | 1 | 2 | 3 | 4 | 5 |
| 5. | Utfordringen med å lære om datamaskiner er spennende. | 1 | 2 | 3 | 4 | 5 |
| 6. | Jeg ser frem til å bruke en datamaskin på jobben min. | 1 | 2 | 3 | 4 | 5 |
| 7. | Hvem som helst kan lære seg å bruke en datamaskin om de er tålmodig og motivert. | 1 | 2 | 3 | 4 | 5 |