THE EMERGENCE OF COGNITIVE SEX DIFFERENCES DURING ADOLESCENCE: A LONGITUDINAL STUDY

Heather H. Hill

A Thesis Submitted to the
University of North Carolina at Wilmington in Partial Fulfillment
Of the Requirements for the Degree of
Master of Arts

Department of Psychology

University of North Carolina at Wilmington

2004

Approved by

Advisory Committee

Dr. Julian Keith

Dr. Dale Cohen

Dr. William Overman

Chair

Accepted by

Dean, Graduate School

TABLE OF CONTENTS

ABSTRACT

Cognitive sex differences have consistently been found in adulthood, with males excelling on visual-spatial tasks and females excelling on manual dexterity tasks. Although these differences are found in adulthood, they rarely exist before adolescence. The present study aims to document the emergence of cognitive sex differences in an adolescent sample. The first year of the study began when the adolescent sample was in the 7th grade. Cognitive measurements continued in the same sample once a year for five consecutive years. Participants were administered a battery of sex-sensitive tasks known to show sex differences in adulthood. In order to compare adult performance with adolescent performance, college-aged males and females were also tested once on the same battery of cognitive tasks.

ACKNOWLEDGEMENTS

I would like to begin by giving a very special thanks to my mentor, Dr. William Overman, for his support and guidance throughout this project. His encouragement and recommendations were invaluable to me.

I would like to thank my committee members, Dr. Dale Cohen and Dr. Julian Keith, for all of their assistance and suggestions.

My thanks also go to my father for giving me the strength and courage to live my dreams. His continual support has greatly contributed to the person I am today. I would also like to thank my mother for always believing in me.

Many thanks go to Mary Engelstatter and Adam Bennett. I honestly do not know what I would have done without the both of you.

Finally, a very special thanks goes to Michael Mivshek for his continual love and support. Words cannot express the thanks I owe to you for standing behind me throughout my graduate career.

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

A considerable amount of evidence suggests that cognitive sex differences that exist during adulthood do not exist before adolescence. Since adolescence is the time period during which cognitive sex differences emerge, events during puberty must play a vital role in the development of these sex differences. In adulthood, males excel on visual-spatial tasks measuring spatial perception and mental rotation (Linn and Peterson, 1985), while females excel on certain verbal tasks, including verbal fluency tasks, synonym generation tasks, and perceptual speed and accuracy tests (Halpern & Tan, 2001). Females also consistently outperform males on precision manual tasks (Kimura, 1992).

Many theories have been postulated about the etiology of cognitive sex differences, including biological, social learning, and evolutionary theories. Consequently, many studies have focused on documenting cognitive sex differences; however, few attempts have been made to longitudinally study the development of these differences. The proposed research longitudinally tracked the emergence of cognitive sex differences during adolescence.

Implications for Studying Cognitive Sex Differences

Females show superior performance on mathematical items during the early elementary and middle school years when math tasks involve learning mathematical facts and arithmetic calculations (Engelhard, 1990; Hyde, Fennema, & Lamon, 1990). However, sometime before adolescence, when the content of the mathematical items becomes more visual-spatial, such as contained in the topics of geometry, trigonometry, and calculus, the advantage shifts to males, who maintain their superior performance into adulthood and old age (Engelhard, 1990; Halpern & Tan, 2001; Hyde, Fennema, & Lamon, 1990). The male advantage is found on college entrance examinations, such as the Scholastic Aptitude Test-Mathematics (SAT-M). College-bound females show a disadvantage on the SAT-M by an average of 34 points when compared with college-bound males (The College Board, 2002). This has important social implications for college-bound females because these tests affect college acceptance (Halpern & Tan, 2001), and they are predictive of later performance in college science and mathematical courses (Benbow and Lubinski, 1993).

Benbow and Stanley (1980) and Benbow and Lubinski (1993) stated that sex differences in mathematical ability are coupled with sex differences in spatial ability. Both of these reports conclude that sex differences in mathematical performance result from superior male mathematical ability, which in turn results from greater male ability on spatial tasks. Additionally, Casey, Nuttall, Pezaris, and Benbow (1995) found a significant relationship between the spatial skill of mental rotation and mathematical ability in females. They also found that when the spatial skill of mental rotation was adjusted for, significant sex differences on the Scholastic Aptitude Test-Mathematics (SAT-M) were eliminated. Casey et al. stated that it is not clear why sex differences in mental rotation serve as a mediator for sex differences in math ability. However, one suggestion is that these differences are caused by a difference in processing styles. For example, Pezaris and Casey (1991) found that boys had a greater dependence on spatial strategies when solving mental rotation problems than the majority of the girls. Additionally, Pezaris and Casey found a subset of females who, like the male students, depended on spatial strategies to a greater extent; therefore, investigating individual differences within females may be the best way to improve female math skills. It is possible that female students who apply spatial strategies may have an advantage when solving both mental rotation and math problems (Casey et al.). These findings have important implications for the career fields of mathematics, science, and engineering because these fields require spatial and mathematical skills and are all underrepresented by females (Halpern, 2000).

The male disadvantage in certain language abilities also has important implications for studying cognitive sex differences. For the past 25 years, males have scored significantly lower than females on the reading portion of The National Assessment of Educational Progress (NAEP), which is the only nationally representative and continuing assessment of American students' knowledge and performance in various subject areas. Males also scored significantly lower on the writing portion of the NAEP, and the percentage of females at or above the *basic*, *proficient*, and *advanced* level of the writing portion exceeded that of males in the 1998 assessment (U.S. Department of Education, Office of Educational Research and Improvement, & National Center for Educational Statistics, 1999). Additionally, stuttering, a disability of fluent speech is predominately a male disorder (Halpern, 2000), with males being three to four times more likely to have this problem (Skinner & Shelton, 1985). Males are also four to five times more likely to have developmental dyslexia than females (Stein, 1994). Sex differences also appear following brain damage. For example, with lesions to the left hemisphere and strokes, males suffer from more speech disorders and more impairment on verbal tasks than females (Witelson, 1976; Kimura, 1987).

It is essential that we truly identify and understand these cognitive sex differences so that we can provide additional or different types of education that benefit both males and females.

Evidence of Sex Differences in Spatial Abilities

Linn and Peterson (1985) conducted a meta-analysis on the nature, magnitude, and first occurrence of sex differences in spatial ability. They divided spatial ability into three categories based on the similarities in the processes that participants use for individual items. *Spatial perception* was defined as the spatial relationships that participants use with respect to the orientation of their bodies, while ignoring distracting information. Examples of spatial perception tasks are the Rod and Frame Test (RFT) and the Water Level task. *Mental rotation* was defined as the ability to quickly and accurately rotate two and three-dimensional figures. Examples include the Shepard-Metzler Mental Rotation Test, the Flags and Cards from the French kit, and Primary Mental Abilities (PMA) space. The final category that Linn and Peterson defined was *spatial visualization*. Tasks that measure spatial visualization involve complex multistep manipulations of spatially presented information. These tasks may involve measures of mental rotation and spatial perception; however, they are different in the sense that there are many possible solution strategies. Examples of spatial visualization tasks are the Embedded Figures Test, Hidden Figures, Paper Form Board, Surface Development, Block Design from the Wechsler Adult Intelligence Scale, Differential Aptitude Test, and Guilford-Zimmerman spatial visualization (Linn & Peterson).

Mental Rotation

Linn and Peterson (1985) found the largest sex differences for mental rotation (MR) using the Shepard-Metzler Mental Rotation Test, the Flags and Cards from the French kit, and Primary Mental Abilities (PMA) space. They found that sex differences for MR emerged around age 10 or 11 and that sex differences remain consistent across ages. They also noted that the magnitude of the differences can range from about one-quarter of a standard deviation unit to almost an entire standard deviation unit, depending on the test being used.

Voyer, Voyer, and Bryden (1995) conducted a more recent meta-analysis of sex differences in spatial ability. They found that measures of MR show the most reliable sex differences, the largest effect size, and on average, males outperform females by 0.6 standard deviation units. In addition, they found that the differences in MR are found consistently between the ages of 13 and 18 and that the magnitude of sex differences increases with age.

Spatial Perception

Linn and Peterson (1985) found medium sex differences for tasks measuring spatial perception. The tasks used to measure spatial perception were the Rod and Frame Test and the Water Level task. Linn and Peterson noted that spatial perception can be measured reliably in 4-year-olds. However, it has been suggested that these tasks measure a different dimension when they are tested in older children, for example, Piaget and Inhelder's (1956) study of water level suggested that younger children view the task as a memory task, whereas older children can reason about the variables. Maccoby and Jacklin (1974) and Linn and Peterson both found that males outperform females on spatial perception tasks beginning around age 7 or 8 and that these differences persist across the life span. Linn and Peterson also noted that effect sizes were much larger for those over 18 years of age, suggesting that sex differences in spatial perception may increase beyond adolescence. Linn and Peterson concluded that males outperform females by one-third of a standard deviation when under the age of 18 and those above 18 outperform by two-thirds of a standard deviation.

Voyer, Voyer, and Bryden's (1995) meta-analysis findings on spatial perception tasks were very similar to Linn and Peterson's (1985) findings. They found that males outperform females by 0.4 standard deviation units, the magnitude of sex differences increases with age, and that differences in spatial perception are found consistently between the ages of 13 and 18.

Spatial Visualization

Linn and Peterson's (1985) and Voyer, Voyer, and Bryden's (1995) meta-analyses findings were highly variable and sex differences were not consistently found. The effect sizes were small and were homogenous across the life span. Maccoby and Jacklin (1974) also found no consistent sex differences in spatial visualization.

Evidence of Sex Differences in Verbal Abilities

There are a number of reports of female advantages in certain verbal abilities (Hampson & Kimura, 1992; Halpern & Tan, 2001; Maccoby & Jacklin, 1974). Female superiority has been found on verbal fluency tasks and synonym generation tasks (Halpern & Tan, 2001). Hampson and Kimura (1992) reported that the female advantage also exists in perceptual speed, which consists of rapidly identifying matching items, and in precision manual tasks, such as placing pegs in designated holes on a board.

Maccoby and Jacklin (1974) suggest that sex differences in verbal ability are not consistently found before adolescence. They suggest that around age 10 or 11 girls first start outperforming boys on certain verbal tasks and concluded that this superiority continues into adulthood. They reported that the female advantage ranges from about 0.1 to 0.5 standard deviation units, with the average being 0.25 standard deviation units. Similarly, Hyde's (1981) meta-analysis on cognitive sex differences reported an effect size of d = .24 for verbal abilities. Hyde and Linn (1988) also conducted a meta-analysis on gender differences in verbal ability and reported a weighted mean effect size of d = .11. They stated that this small effect size only indicated a slight female advantage in performance. However, it is important to remember that verbal performance deficits, such as, dyslexia, mental retardation, and stuttering are more prevalent among males (Collaer & Hines, 1995; Halpern, 2000; Hyde and Linn, 1988; Maccoby & Jacklin, 1974; Stein, 1994). Consequently, many lower-scoring boys might be removed from the regular classroom, therefore, omitting their verbal performance test results. The result of this would be a reduction in the effect size for sex differences in verbal performance (Hyde & Linn, 1988) and a possible explanation for Hyde and Lynn's (1988) small effect size.

Theories Explaining Cognitive Sex Differences

Several theories have been postulated about the etiology of cognitive sex differences, including biological, social learning, and evolutionary theories.

Biological Theories

A considerable amount of research exists on the biological determinants of cognitive sex differences, with the majority of the theories and research focusing on the effect of hormones on cognitive abilities.

Prenatal exposure to androgens has been shown to have an effect on cognitive abilities. Congenital adrenal hyperplasia (CAH) is an autosomal recessive disorder associated with exposure to abnormally high levels of androgens during fetal development. Resnick, Berenbaum, Gottesman, and Bouchard (1986) administered a battery of cognitive tasks to both males and females with CAH. They used siblings and cousins to serve as controls because they provide the best control for common genes and environmental background. The Card Rotations and the Mental Rotations tasks were used to measure spatial orientation, while Paper Form Board, Hidden Patterns, and Paper Folding tasks were used to measure spatial visualization. The CAH females scored significantly better on three of the five spatial tasks when compared to the unaffected relatives, thus providing evidence that prenatal androgen exposure is associated with enhanced spatial abilities for females. Resnick et al. noted that CAH females showed a trend towards greater participation in spatial manipulation activities when compared to female controls. However, the differences in activities between CAH girls and control girls did not account for observed differences in spatial ability because no significant correlation between spatial manipulation activity and spatial ability was found.

Berenbaum and Hines (1992) also found that girls with CAH from ages 3 to 8 showed increased play with boys' toys and decreased play with girls' toys when compared with their unaffected female relatives, therefore, suggesting that these girls are both masculinized and defeminized. Berenbaum (1999) extended these findings when she found that CAH adolescent girls from ages 9 to 19 years old differed significantly from their unaffected sisters on sex-typed activities and interests. Girls with CAH continued to show increased interest in male-typical activities and careers and reduced interest in female-typical activities and careers in adolescence, therefore strengthening the argument that prenatal exposure to androgens had a direct effect on their behavior rather than social experiences.

Another hormonal theory suggests that an optimal amount of androgens are needed in puberty for spatial skills to develop. Hier and Crowley (1982) studied 19 men with idiopathic hypogonadotropic hypogonadism, a disorder characterized by a deficiency of androgens during puberty. These men provide useful information to hormonal theories because they were reared as normal males until puberty; therefore, eliminating any social or educational differences. Hier and Crowley used the Block Design subtest from the Wechsler Adult Intelligence test, the Embedded Figures Test, and the Space Relations subtest from the Differential Aptitude Tests to measure spatial ability. They found that the men with who acquired the disorder after puberty and control men did not differ significantly on all three spatial tasks; however, the men who had the disorder during puberty scored significantly lower than both groups. Additionally, they found that the more severe the androgen deficiency, the more severe the spatial impairment on the Block Design and the Embedded Figures test. They also attempted to determine if hormone-replacement therapy would improve performance on the spatial tasks; however, they failed to find an effect. The findings from this study indicate that there is an optimal amount of androgens needed in puberty for spatial skills to develop. However, if androgen levels are not normal during puberty then impaired spatial skills will persist and are not correctable. Females with Turner's Syndrome and males with androgen insensitivity provide additional support for this theory, given that both of these groups have abnormally low levels of androgens during puberty and impaired spatial skills (Nyborg, 1983; Halpern, 2000).

Peterson (1976) conducted a longitudinal study on normal adolescent males and females and also found a relationship between the amount of androgens available during puberty and the development of cognitive sex differences. She inferred the quantity of sex hormones by rating photographs of the adolescents' development of secondary sex characteristics. Hormonal influence measures were based on the Bayer and Bayley (1959) ratings of physical development, and the Tanner (1962) scale. The Block Design from the Wechsler-Bellvue Test and Space from the Primary Mental Abilities Test measured spatial ability, while fluent production was measured by Digit Symbol from the Wechsler-Bellvue Test and Word Fluency from the Primary Mental Abilities Test. She found that physical variables had no relationship to females' fluent production. However, spatial ability was significantly related to females' masculinity, with the more physically masculine females exhibiting the best spatial skills. She also found that less physically masculine males had better spatial abilities than fluent production, while the more physically masculine males had better fluent production than spatial abilities. Therefore, she hypothesized that higher spatial ability was associated with more androgens for females and less androgens for males. Support for this position was also provided by Maccoby (1966) when she reported that boys with high spatial ability were rated as less masculine by their peers than boys with lower spatial ability.

Nyborg (1983) proposed a hormonal theory known as the *optimal estrogen range* (OER) theory. Nyborg hypothesized that estradiol (E_2), the biologically most active estrogen, was the main biological agent mediating spatial ability. He suggested that there is an optimal range of cerebral E_2 needed for individuals to exhibit maximum spatial skills and postulated that too high or too low levels of E_2 impair spatial skills. The OER theory states that the relatively low spatial ability of women compared to men is caused by a surge in plasma E_2 around puberty, which brings their E_2 levels above the optimal range. Nyborg also proposes an explanation for the more feminized male and the more masculinized female possessing the best spatial skills. He states that the more feminized male has slightly higher levels of plasma E_2 then most men. According to the OER theory, most men are situated just below the optimal range of central E_2 values; therefore, the slightly increased E_2 levels in the feminized men brings their E_2 levels within the optimal central E_2 range for the enhancement of spatial ability. Similarly, the more masculinized females have slightly lower levels of plasma E_2 then most women, therefore, their lower E_2 levels brings them within the optimal central E_2 range for the enhancement of spatial ability compared with most other women. Nyborg provided support for his theory by citing research that showed a dramatic increase in the usual impaired spatial skills of women with Turner's Syndrome when they were treated with E_2.

Experimental manipulation of hormones in humans offers strong support for the effect of hormones on cognitive abilities. Van Goozen, Cohen-Kettnis, Gorren, Frijda, and Van de Poll (1995) were able to do this by testing the cognitive abilities of male-to-female (M>F) and female-to-male transsexuals (F>M), both before and after cross-sex hormone treatments. Anti-androgen and estrogen treatment was administered to the M>F transsexuals, while testosterone was administered to the F>M transsexuals. Spatial ability was measured by the Cards Rotations test, while verbal fluency was measured by the Word Production, Categories, and Sentence Production tests. Before the cross-sex hormone treatment, both groups performed consistently with their biological sex on the sex-specific tasks. However, after treatment, the F>M group improved their performance on the spatial task and decreased their performance on the verbal fluency tasks. This suggested that the testosterone influenced their enhanced spatial performance. In the M>F group, the estrogen treatment and the androgen deprivation resulted in poorer performance on the spatial task and enhanced performance on the verbal fluency tasks. It should be noted that Van Goozen, Gorren, Slabbekoorn, Sanders, and Cohen-Kettenis (2002) failed to replicate the findings of the activating effect of hormones. Van Goozen et al. studied homosexual transsexuals and failed to find an effect of sex hormones on the visual-spatial abilities of homosexual transsexuals. These transsexuals scored in the direction of the opposite sex before cross-sex hormone treatment. Therefore, the authors suggest that failure to find an effect could be due to these transsexuals reaching a ceiling effect in performance. They also suggest that the failure to find an effect could be due to the inclusion of only right-handed transsexuals and controls, or due to the exclusion of non-homosexual transsexuals.

Waber (1976, 1977) suggested that sex differences in verbal and spatial abilities are due to differences in maturation rate. She found that regardless of sex, late maturing adolescents performed superior when compared to early maturing adolescents on spatial tasks, and late maturing adolescents performed superior on spatial tasks than on verbal tasks. She also found that early maturing adolescents performed better on tests of verbal abilities than on spatial abilities. In addition, the later maturing youths of both sexes showed a stronger ear advantage on the dichotic listening task, which often indicates greater hemispheric lateralization. These findings suggest that late maturation is associated with enhanced spatial skills and strong hemispheric lateralization, while early maturation is associated with enhanced verbal abilities and weaker hemispheric lateralization. Waber (1976, 1977) measured maturation rate according to the Tanner (1962) criteria for staging secondary sexual characteristics, which includes ratings of breasts and pubic hair development for girls and genital and pubic hair development for boys. Height, weight, and the girls' approximate menarcheal age were also recorded. Participants were classified as early maturers if their chronological age was at least 1 standard deviation below the mean age for their stage of sexual development and were classified as late maturers if their chronological age was at least 1 standard deviation above the mean. Verbal abilities were measured by the Digit Symbol subtest of the Wechsler Intelligence Scale for Children (WISC), the Color-or-Naming subtest of the Stroop Color Word test, and the Word Fluency subtest of the Primary Mental Abilities (PMA) test, while she used the Block Design subtest of the WISC, the Embedded Figures test, and the Spatial Abilities subtest of the PMA to measure spatial ability. Waber's explanation may account for the fact that girls, on average, mature earlier than boys, thus explaining the differences in cognition. Sanders and Soares (1986) provided supporting evidence for Waber's (1976, 1977) theory when they found that that timing of pubertal events and spatial ability were significantly related in college students. They found that both male and female participants that classified themselves as later maturers scored significantly higher on the Mental Rotation test than those that classified themselves as early maturers.

Nyborg (1983) conducted a literature review and found general trends in the data that suggest that early maturing adolescents score slightly higher on some tests of spatial ability; however, these trends tend to decrease in early post-adolescence. Therefore, the data suggests that late maturing adolescents' score lower than early maturing adolescents at first, but then after puberty the late maturing adolescents take the lead and score higher than early maturers. The latter portion confirmed Waber's (1976, 1977) results. This advantage for the later maturing adolescents also continues through adulthood. Nyborg notes that some studies could not replicate the relationship between spatial ability and physical maturation. However, he suggests that interpretation of these results are complicated due to the fact that different studies used different criteria for late and early maturing adolescents, extreme groups were used in some studies, and not all studies used the same age of participants defined as late and early maturing adolescents.

An additional theory proposed by Nyborg (1983) suggests that low spatial ability found in adult women may be caused by these skills leveling off in early adolescence, while on the other hand; boys continue to increase in their spatial performance after adolescence. He further suggests that girls may actually decline in their spatial performance after puberty. Witkin, Goodenough, and Karp (1967) provided evidence for the latter theory when they found a postpubertal decline in spatial ability in girls, using the Rod-and-Frame Test in a cross-sectional study. Witken et al. concluded that this observation might have been due to artifacts in the selection of their participants, however, Nyborg observed a similar postpubertal decline on Embedded-Figures Test scores and on Money's Road-Map Test scores in girls, which could not be ascribed to selection artifacts. Additionally, Noble, Baker, and Jones (1964) found that males and females perform similarly on the United States Air Force (USAF) Discrimination Reaction Time apparatus until the age of 16, after which females began a linear decline in their performance.

Social Learning Theories

Social learning theories have also attempted to explain sex differences in cognition. Most social learning theories explain that sex differences are due to differential rearing of girls and boys. These theories also suggest that socialization affects cognitive sex differences. Thus, children are said to exhibit behavior that is appropriate for their gender and learn to do so through reinforcements from societal norms (Nyborg, 1983). Evidence for these theories is displayed in the fact that boys are more likely than girls to receive reinforcement for completing difficult math problems, and girls receive less praise than boys for correct answers in mathematical classes (Stage & Karplus, 1981). In addition, girls receive more encouragement to participate in verbal activities, such as reading, while boys are encouraged to participate in spatial activities, such as playing with building blocks (Halpern, 2000).

Social learning theories also contend that sex differences in cognitive abilities are due to differential life experiences. For example, boys generally move more freely around their surroundings than girls (Nyborg, 1983), and boys touch unfamiliar objects more often than girls (Adams & Bradbard, 1985). According to these theories, locomotion and early life exploration of unfamiliar objects furthers spatial abilities; therefore, since girls are subjected to more restrictions in moving around, it is reasoned that sex-specific rearing hampers the development of spatial ability in females (Nyborg, 1983). Social learning theories also state that more males than females participate in spatial activities (Halpern, 2000; Newcombe, Bandura, & Taylor, 1983). However, these findings do not prove that boys have better spatial abilities because they participate in more spatial activities. There are numerous other explanations; for example, it is also possible that boys engage in more spatial activities because they have better spatial skills. Baenninger and Newcombe (1989) conducted a meta-analysis on spatial abilities and spatial activities and concluded that there is a weak relationship between the two, and they also noted that the magnitude of the effect is the same for females and males. Therefore, both sexes benefit equally from participating in spatial activities.

Although social learning theories have attempted to explain sex differences, they have also been criticized. One of the criticisms comes from the view that girls and boys are reared differently. Maccoby and Jacklin (1974) found very little supporting evidence that girls and boys were reared differently and that one sex received more reinforcements than the other. Nyborg (1983) also noted that they fail to account for the absence of sex differences found before puberty and the emergence of the sex differences later. The impaired spatial skills of males with idiopathic hypogonadotropic hypogonadism also weaken the arguments made by social theories because these males are reared as normal males, which eliminates any social or educational differences.

Evolutionary Theories

Evolutionary theories have also attempted to explain sex differences in cognitive abilities. Evolutionary theories base their hypotheses on the division of labor in earlier human societies. These theories state that men engaged in hunting and foraging at a distance; therefore, men needed better spatial skills than women, who primarily performed the gathering tasks. According to this theory, men evolved to be genetically superior in spatial ability because they had to be able to maintain their orientation while hunting in unfamiliar territories (Halpern, 2000). Other typical "male" activities, such as weapon manufacturing, fishing, and boat-building also required superior spatial ability (Kimura, 1987). Women, on the other hand, needed to have an advantage in activities near the home base. Typical "female" activities included, gathering plants and seeds for food, making clothing, making pots, and tending to the children. Gathering food included remembering the location of edible plants from season to season; therefore, evolutionary theories argue that women needed superior location memory (Halpern, 2000; Kimura, 1987). Eals and Silverman (1994) provided support for this theory when they found that women were better at tasks that require memory for locations, while men are better at tasks that require mental transformations of spatial displays. The other "female" activities require a female advantage in fine manual skills and perceptual speed, and many reports have confirmed that females are superior to males in these abilities (Halpern & Tan, 2001; Hampson & Kimura, 1992; Kimura, 1987). These theories have also been criticized; for example, women also needed spatial skills in order to gather and travel long distances to find food. Furthermore, there were many hunter-gatherer societies in which women hunted (Halpern, 2000).

Although all of these have been criticized in one way or another, each contributes to our understanding and knowledge of the causes of sex differences in cognitive abilities. Therefore, each theory puts us one step closer to developing additional or different types of education that benefit both males and females.

Hypotheses

Research on cognitive sex differences clearly indicates that there are differences in performance on certain cognitive tasks in adults. The evidence also suggests the cognitive sex differences that exist during adulthood do not exist before adolescence.

The proposed study will attempt to longitudinally track the emergence of cognitive sex differences during adolescence. Consistent with the current scientific literature on cognitive sex differences the following predictions are made:

Hypothesis I: Sex differences will emerge and increase in each of the sex-sensitive cognitive tasks during the five years of this study. During the fifth year of this study, adolescent males and adolescent females will differ on each of the sex specific cognitive tasks. Specifically, males will outperform females on the spatial perception and mental rotation tasks, and females will outperform males on the precision manual task.

Hypothesis II: There will be sex differences in each of the cognitive tasks in the adult sample. Performance on each of the cognitive tasks will be comparable for the year 5 males and the adult sample males. Likewise, performance on each of the cognitive tasks will be comparable for the year 5 females and the adult sample females.

METHOD

Participants

Adult Participants

Participants were undergraduate students at the University of North Carolina at Wilmington (UNCW) in the 1999 school year. A total of 69 participants (36 females and 33 males) completed the experiment. The ages of the participants ranged from 18-45 (M = 20.23). The mean age of the female participants was 19.89 and the mean age of the male participants was 20.61. The racial composition was 87% Caucasian, 10.1% African American, and 2.9% Hispanic. While this sample is not representative of the general population we did not pre-select or discriminate due to ethnicity.

Adolescent Participants

Participants were students attending high school in the New Hanover County school district that had participated in the study for the past four years. The first year of the study began with a total of 149 adolescents. In the first year of the study the adolescents were in the seventh grade and ranged from 12-13 years of age. In the fifth and final year of the study, 68 of the original 149 participants remained (32 males and 36 females). Only individuals who participated in all test years will be included the following description and analyses. In the fifth year of the study the ages of the participants ranged from 16-17 (M = 16.26). The female participants had a mean age of 16.25 and the male participants had a mean age of 16.28. The racial composition was 92.6% Caucasian, 5.9% African American, and 1.5% Hispanic. While this sample is not representative of the general population we did not pre-select or discriminate due to ethnicity.

Procedure

Adult Procedure

All adult participants were solicited via bulletin board posting in the UNCW psychology department and all received credit in their Introductory Psychology class. All testing appointments were scheduled at either 5:00 pm or 6:00 pm. Unless otherwise stated, all remaining experimental procedures were the same for the adult and the adolescent participants.

Adolescent Procedure

The University of North Carolina at Wilmington institutional review board approved all procedures. Letters requesting the parents' permission to participate in the study were mailed to the participants along with a brief description of the study. After receiving the parents' permission, testing appointments were scheduled so that the participants returned the same time of day and approximately the same week as former test years.

Prior to testing, both parents and children completed an informed consent (see Appendix A). Parents also completed a brief questionnaire regarding their socio-economic status and profession (see Appendix B). Participants completed an information sheet regarding their sex, ethnicity, height, weight, and current medications (see Appendix C). Females also completed an information sheet concerning their menstrual cycle and current birth control being used (see Appendix C). Each participant was then administered a battery of four cognitive tasks in the same order. The selected tasks have shown sex differences in adulthood, with three of the tasks demonstrating a male advantage and the remaining task demonstrating a female advantage.

Upon completion of the testing session participants received a $5.00 gift certificate for Millennium Music, a t-shirt from a local surf shop, and various coupons from local businesses. Parents that accompanied their children to the testing session were also compensated with a $5.00 gift certificate from Wal-Mart/Sam's Club. The testing session lasted approximately one hour.

Materials

Each participant received the cognitive tasks in the following order:

Task 1: Piagetian Water Level Task. (Piaget & Inhelder, 1956). This task measures spatial perception. Participants were given a pencil and a ruler. They were then shown a picture of a glass jar filled halfway with water (see Figure 1). They were told that the rubber stopper on the glass jar does not let any water spill out of the jar. They were also told that the box underneath the glass jar represented the table which the glass jar rests on. Participants were then instructed to draw a line across the glass jar to represent where the water level would be if the jar were half full with water. The participants were given seven different views of the glass jar each tilted at an angle of 45º, 90º, 135° or 180º. Four of the views were turned to the left and the other four views were turned to the right. The correct response for this task is a horizontal line regardless of the tilt of the jar. The dependent measure for this task was the mean number of degrees off from a horizontal line for all seven views. This task has been shown to have a male advantage in several studies (Piaget & Inhelder, 1956; Linn & Peterson, 1985).

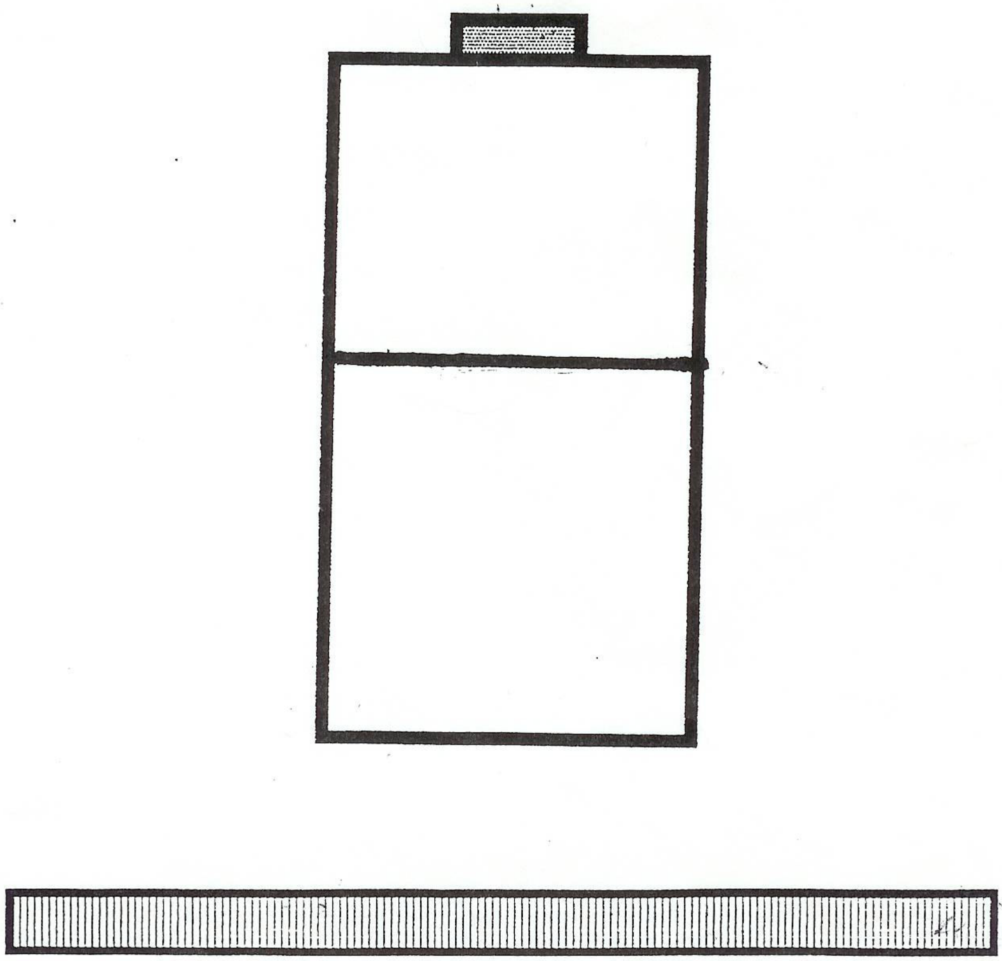


Figure 1. Glass Jar Participants are shown in the Water Level Task.

Task 2: The Flag Task. (Thurstone, 1938). In this two-dimensional mental rotation task, participants were shown two pairs of paper cutout American flags side by side. One of the pairs of flags showed the same image on both cutouts, while the other pair showed two American flags that were mirror images of one another (see Figure 2). Participants were able to rotate these paper cutout flags to determine if the pair of flags were the same images of each other or if the pair of flags were the mirror images of one another. Before the testing trials began, participants were given four practice trials. In the practice trials, participants were asked to verbally state whether the pair of flags represented the same image or if the flags were mirror images of each other. If participants responded incorrectly on the practice trials, the experimenter verbally corrected the participant and rotated the paper cutouts to visually show the participant the correct answer. It should be noted that the adult sample did not view these paper cutouts. They were only shown examples on a sheet of paper; therefore, they were not able to physically rotate the paper cutouts.

Participants were then given 48 pairs of flags rotated at various angles, 24 were mirror images of each other and the other 24 were the same sides of each other. Participants were instructed to put a plus sign (+) next to the pairs of flags that were the same sides of each other or a minus sign (-) next to the pairs of flags that were mirror images of each other (see Figure 3). Each participant had three minutes to complete this task. The dependent measure was the number of correct responses. Males have been shown to excel on this task (Voyer, Voyer, & Bryden, 1995).

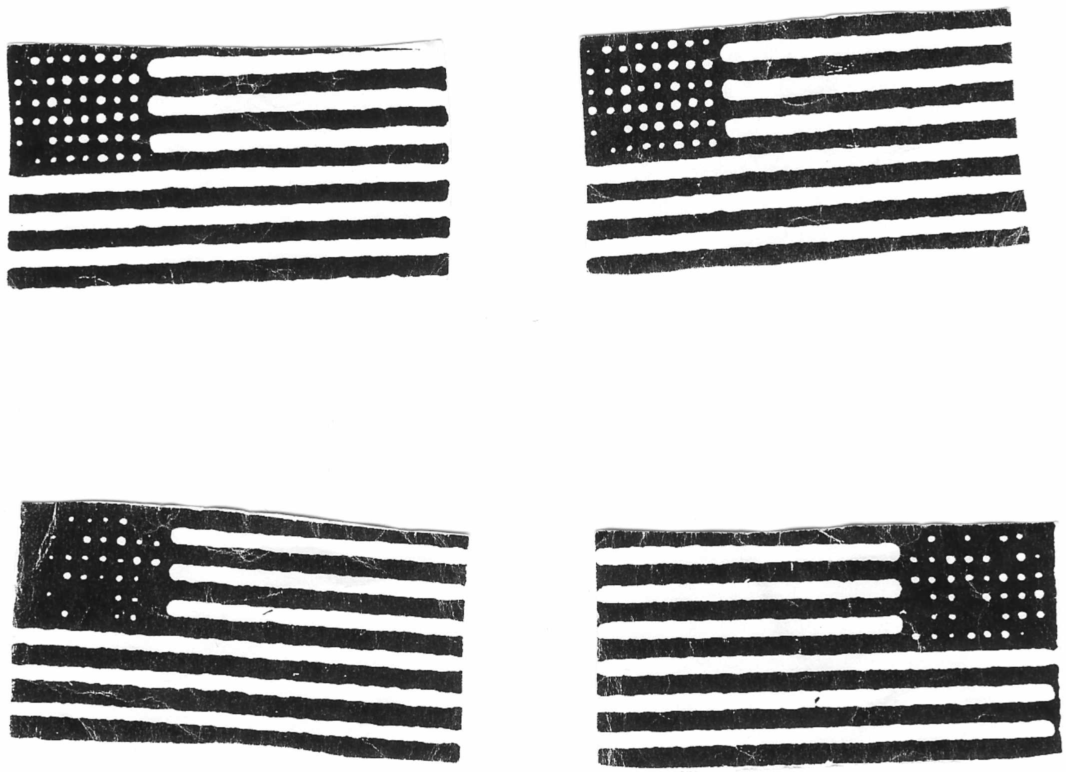


Figure 2. Paper Cutout Flags that Participants Physically Rotate before Beginning the Test Trials in the Flags Task.

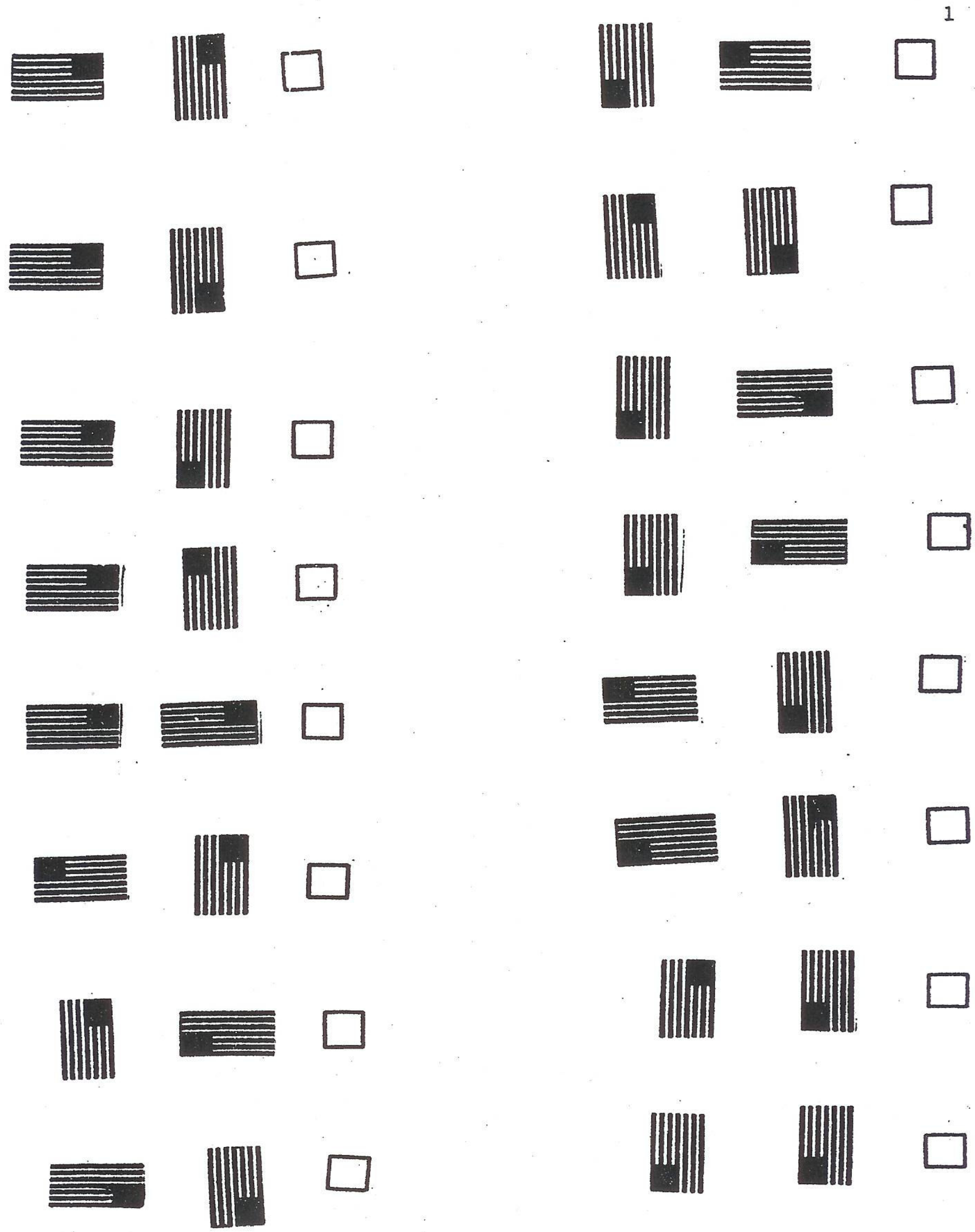


Figure 3. Response Sheet Given to Participants in the Flags Task.

Task 3: Purdue Pegboard. (Tiffin, 1968). This task measures manual dexterity skills. Participants were instructed to insert metal pegs in holes on a pegboard as quickly and as accurately as possible. Participants were told to perform this task first with their dominant hand, then with their non-dominant hand, next with both hands, and finally, they were instructed to construct an assembly of pegs, washers, and collars. Participants had 30 seconds to complete the first three conditions and one minute to complete the assembly. The dependent measure was the number correctly inserted pegs in the allotted time. Females have been shown to have an advantage on this timed finger dexterity task (Hampson & Kimura, 1992).

Task 4: Ratcliff Mannequin (Basketball Man Task) (Ratcliff, 1979). This task measures two and three dimensional mental rotation abilities. Participants were shown a picture of a man on a computer screen. The man was holding a basketball in either his left or right hand. The participants had ten seconds to mentally rotate the man and determine which hand the man was holding the ball in. They made their responses by pressing one of two keys on a keyboard. They were instructed to press the left key if the man was holding the ball in his left hand or to press the right key if the man was holding the ball in his right hand. Figures were presented either from a front view to assess three-dimensional rotation or from a back view to assess two-dimensional rotation (see Figure 4). There were 24 trials in which the figures were rotated at 0º, 45º, 135º, or 180º. The dependent measures were the latency of responses, the error rates, and processing as a function of rotation angle (slope). These types of mental rotation tasks have consistently shown a male advantage (Voyer, Voyer & Bryden, 1995).

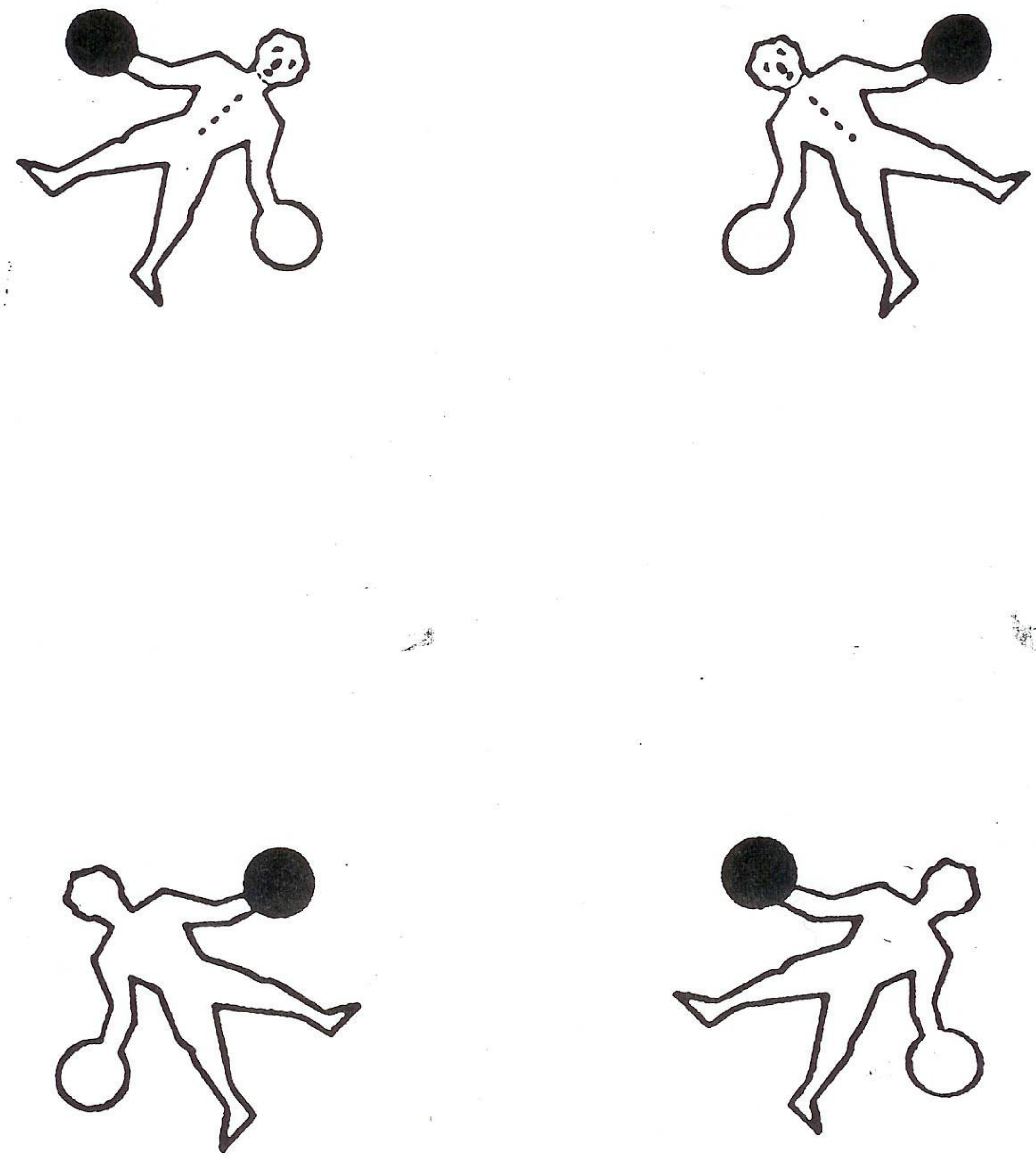


Figure 4. Basketball Man Task from Both Front and Back Views.

RESULTS

Hypothesis I

Hypothesis I stated that sex differences would emerge and increase in each of the sex-sensitive cognitive tasks during the five years of this study. Specifically, males were predicted to outperform females on the Water Level Task, the Flags Task, and the Ratcliff Mannequin (Basketball Man Task), while females were predicted to outperform males on the Purdue Pegboard Task. In order to evaluate Hypothesis I the following analyses were conducted. Only those individuals who participated in all five years of the study will be included in the following descriptions and analyses.

Male-Sensitive Tasks

Water Level Task. Exploratory analysis revealed that the frequency distributions for the 45° angle condition and the 135° angle condition were almost identical; therefore, these conditions were combined for all further analyses. For each participant, the scores from each angle of rotation were averaged to form a single score for the combined 45°/135° angle condition, the 90° angle condition, and the 180° angle condition. A 2 (sex) x 5 (test time) mixed factorial ANOVA was then conducted on the mean degrees off from horizontal for each of the three angle conditions.

Means and standard deviations for the combined 45°/135° angle condition of the Water Level Task are presented in Table 1 by sex and test time. As predicted, the combined 45°/135° angle condition yielded a main effect of sex such that males (M = 12.20°, SD = 15.00) significantly outperformed females (M = 21.48°, SD = 15.00), $F(1, 60) = 5.88$, $p = .018$, $MSE = 1125.51$. A main effect of test time was also found to be statistically significant, $F(4, 240) = 26.49$, $p = .000$, $MSE = 98.29$. Pairwise comparisons

Table 1

Mean Degrees from Horizontal for the Combined 45°/135° Condition of the Water Level Task by Sex and Test Time.

Test Time	Males	Females	p-value
Year 1	22.64(20.09)	31.29(19.03)	0.087
Year 2	13.48(17.23)	22.94(19.11)	0.047
Year 3	10.76(13.90)	22.63(20.84)	0.012
Year 4	7.63(12.28)	15.80(18.00)	0.046
Year 5	6.47(10.08)	14.74(18.39)	0.038

Values presented are means with standard deviations in parentheses.

revealed that year 1 (M = 27.39°, SD = 19.83) performed significantly worse than all other test years (all p's = .000). Year 2 (M = 18.67°, SD = 18.75) and year 3(M = 17.27°, SD = 18.86) did not significantly differ from one another ($p > .05$). However, year 2 and year 3 performed significantly worse from year 4 (M = 12.11°, SD = 16.09) and year 5 (M = 11.00°, SD = 15.65) (all p's $\leq .001$). Year 4 and year 5 did not significantly differ from one another ($p > .05$) and both of these test years significantly outperformed all other test years (all p's $\leq .001$) (see Figure 5). The interaction between sex and test time failed to reach statistical significance, $F(4, 240) = .37$, $p = .831$, $MSE = 98.29$.

When the data were sorted by test year, there was a marginally significant effect of sex, such that the year 1 males (M = 22.64°, SD = 20.09) significantly outperformed the year 1 females (M = 31.29°, SD = 19.03), $F(1, 60) = 3.02$, $p = .087$, $MSE = 380.70$. The year 2 males (M = 13.48°, SD = 17.23) significantly outperformed the year 2 females (M = 22.94°, SD = 19.11), $F(1, 60) = 4.11$, $p = .047$, $MSE = 334.52$. The year 3 males (M = 10.76°, SD = 13.90) also significantly outperformed the year 3 females (M = 22.63°, SD = 20.84), $F(1, 60) = 6.65$, $p = .012$, $MSE = 325.65$. Additionally, the year 4 males (M = 7.63°, SD = 12.28) significantly outperformed the year 4 females (M = 15.80°, SD = 18.00), $F(1, 60) = 4.16$, $p = .046$, $MSE = 246.13$ and the year 5 males (M = 6.47°, SD = 10.08) significantly outperformed the year 5 females (M = 14.74°, SD = 18.39), $F(1, 60) = 4.52$, $p = .038$, $MSE = 231.68$ (see Figure 6).

As previously stated, the frequency distributions for the 45° angle condition and the 135° angle condition were almost identical; therefore, these conditions were combined for all further analyses. Examination of the mean degrees from horizontal in these combined angle conditions revealed a bimodal distribution. This bimodal

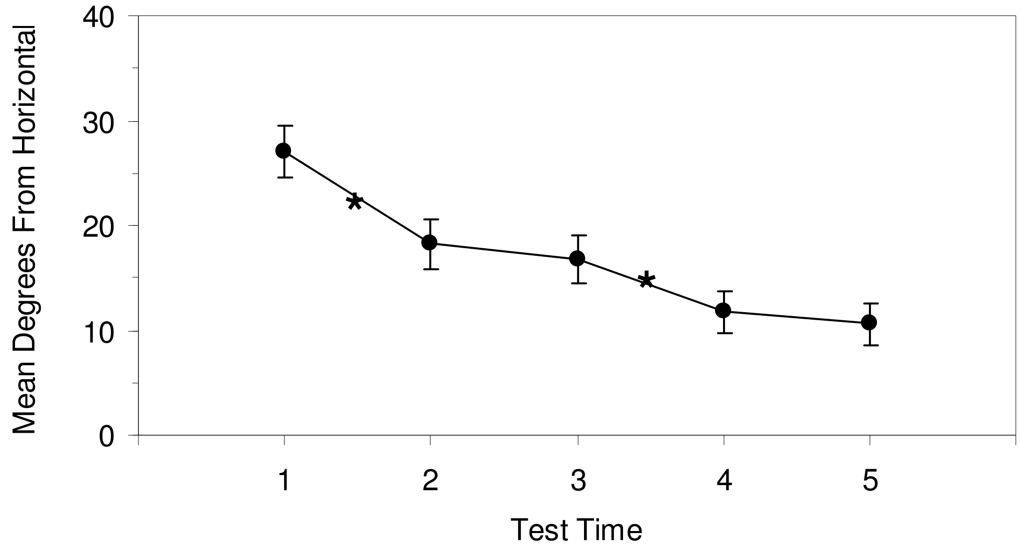


Figure 5. Mean Degrees from Horizontal in the Combined 45°/135° Angle Condition of the Water Level Task by Test Time. Asterisks indicate significant differences between test years.

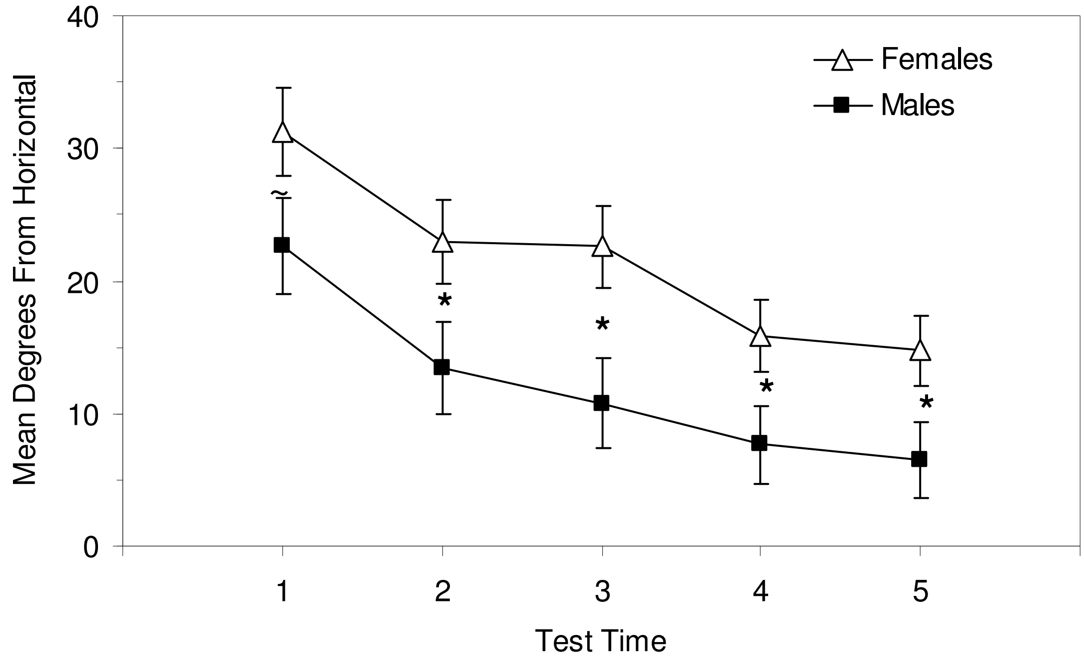
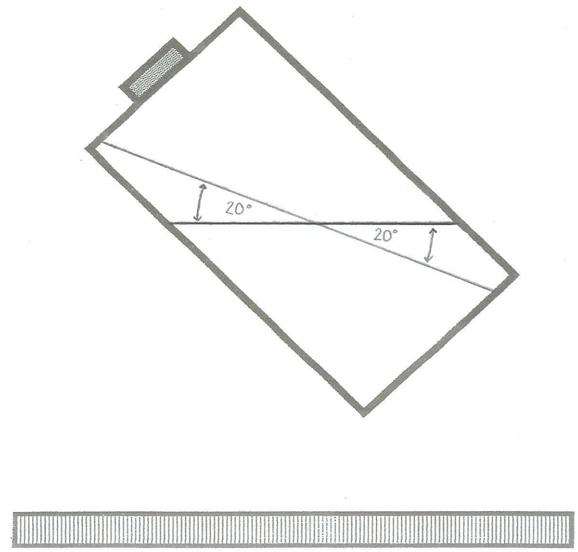


Figure 6. Mean Degrees from Horizontal for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time. Asterisks indicate significant differences and the ~ symbol indicates marginally significant differences.

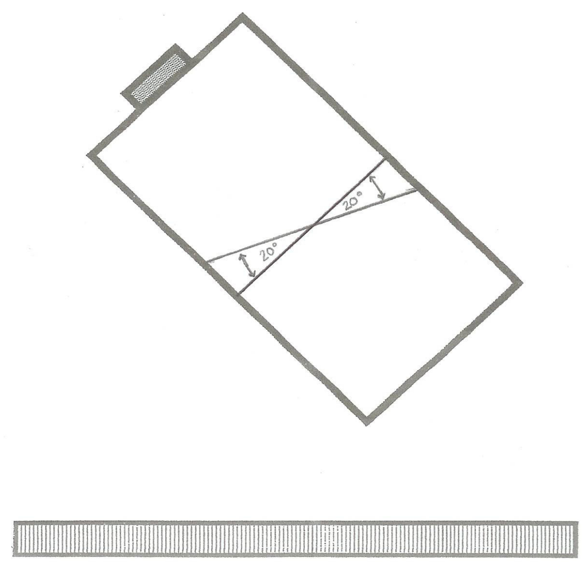
distribution revealed one cluster of scores consisting of almost no errors (0° from horizontal) and another cluster of scores consisting of many errors, with an average of 45° from horizontal. This bimodal distribution led to the following conclusions.

Traditionally, the only correct answer for the Water Level Task was a horizontal line. Participants who drew a horizontal line were using the external context to make their response on where the water line should be drawn. The external context refers to the table that the glass jar sits on. The bimodal distribution in the present analyses revealed that one group of participants did in fact rely on the external context (the horizontal table) to make their responses. However, another group of participants relied on the internal context of the glass jar to make their responses. This internal context refers to relying on the 45° angle top and bottom of the glass jar to make a response, rather than relying on the external context of the table. When the participants relied on the internal contexts, they drew a 45° line instead of the correct horizontal line.

It is important to remember that there were seven total views of the glass jar. Four of these views fall into the combined 45°/135° angle conditions, with two of the views being tilted at the 45° angle and the remaining two views being tilted at the 135° angle. The dependent variable for each of these four views was the mean degrees off from a horizontal line. When the data for each of these views fell in the 0° ± 20° range from horizontal, the response was labeled an "external response." The decision was labeled an "internal response" when the data fell into the 25°- 65° range from horizontal (± 20° from 45° from horizontal) (see Figure 7). The following results discuss the findings when the data was coded according to the type of responses the participants made (either external or internal).



a.



b.

Figure 7. a.) The 0 ± 20° Angles Ranges (from horizontal) that Constituted an "External" Response
b.) The 25° - 65° (± 20° from 45°) Angle Ranges (from horizontal) that Constituted an "Internal" Response

The following analyses coded the responses for males and females according to the total percentage of external and internal responses for all four views. The total percentage of external responses by sex and test time can be found in Table 2. The total percentage of internal responses by sex and test time can be found in Table 3. Log transformations were made to test time (1-5) and to the total percentage of external responses made by males and females and the total percentage of internal responses made by males and females for all analyses.

Regressions were performed separately for males and females to determine if the percentage of external/internal responses were a function of test time for either sex. Slopes obtained from these regression analyses were used to obtain estimates of the increase or decrease in percentage of external and internal responses as a function of test time. The intercepts indicate the percentage of external and internal responses made by males and females in the first test year.

For females, a regression analysis revealed that test time was a highly significant predictor of the total percentage of external responses made, $\underline{F}(1, 3) = 60.35$, $\underline{p} = .004$, $\underline{R}^2 = .95$, $\underline{MSE} = .001$, with test time accounting for 95% of the variance in the total percentage of external responses made by females. The slope for the females ($\underline{b} = .354$) was also statistically significant, $\underline{t}(4) = 7.67$, $\underline{p} = .004$, indicating that females significantly increased their total percentage of external responses made across time.

For males, a regression analysis revealed that test time was a highly significant predictor of the total percentage of external responses made, $\underline{F}(1, 3) = 70.56$, $\underline{p} = .004$, $\underline{R}^2 = .96$, $\underline{MSE} = .000$, with test time accounting for 96% of the variance in the total percentage of external responses made by males. The slope for the males ($\underline{b} = .273$) was

Table 2

Total Percentage of External Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time.

Test Time	Males	Females
Year 1	61.0%	42.0%
Year 2	79.0%	59.5%
Year 3	88.0%	61.0%
Year 4	92.0%	73.0%
Year 5	94.0%	74.5%

Table 3

Total Percentage of Internal Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time.

Test Time	Males	Females
Year 1	29.0%	51.0%
Year 2	18.0%	31.0%
Year 3	8.0%	27.0%
Year 4	7.0%	23.0%
Year 5	6.0%	24.0%

also statistically significant, $t(4) = 8.40$, $p = .004$, indicating that males significantly increased their total percentage of external responses made across time.

Figure 8 displays the total percentage of external responses made for both males and females by test time. The trendlines shown in the figure demonstrate that both sexes made more external responses as they aged. Analyses comparing the slopes of the regression lines revealed that males ($b = .273$) and females ($b = .354$) did not significantly differ from one another, $t(8) = 1.44$, $p = .189$. This means that both sexes increased at the same rate regarding the total percentage of external responses made across time. However, analyses comparing the intercepts of the regression lines revealed that males ($a = .630$) made significantly more external responses than females ($a = .433$) in the first test year, $t(8) = -6.01$, $p = .000$.

For females, a regression analysis revealed that test time was a highly significant predictor of the total percentage of internal responses made, $F(1, 3) = 43.65$, $p = .007$, $R^2 = .94$, $MSE = .002$, with test time accounting for 94% of the variance in the total percentage of internal responses made by females. The slope for the females ($b = -.490$) was also statistically significant, $t(4) = -6.61$, $p = .007$, indicating that females significantly decreased their total percentage of internal responses made across time.

For males, a regression analysis revealed that test time was a highly significant predictor of the total percentage of internal responses made, $F(1, 3) = 65.53$, $p = .004$, $R^2 = .96$, $MSE = .005$, with test time accounting for 96% of the variance in the total percentage of internal responses made by males. The slope for the males ($b = -1.048$) was also statistically significant, $t(4) = -8.09$, $p = .004$, indicating that males significantly decreased their total percentage of internal responses made across time.

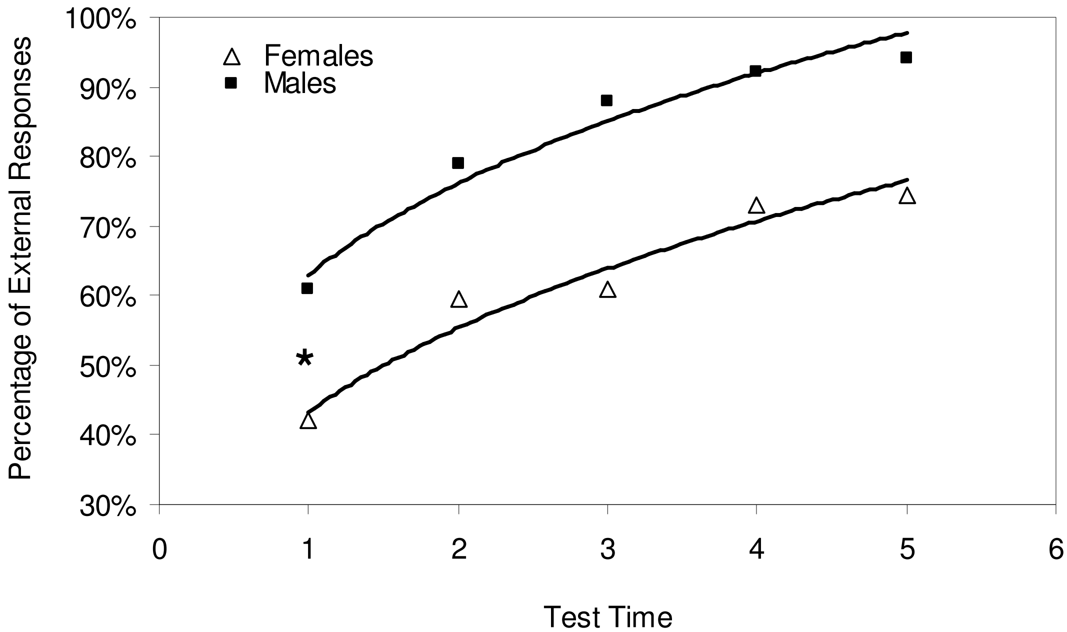


Figure 8. Total Percentage of External Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time. The power trendlines demonstrate the accelerating nature of the total percentage of external responses made. The asterisk indicates a significant difference between males and females in the percentage of external responses made in the first test year (intercepts). The slopes of the lines do not significantly differ from one another.

Figure 9 displays the total percentage of internal responses for both males and females by test time. The trendlines shown in the figure demonstrate that both sexes made fewer internal responses as they aged. Analyses comparing the slopes of the regression lines revealed that males ($\underline{b}$ = -1.048) decreased in their total percentage of internal responses at a faster rate than the females ($\underline{b}$ = -.490), $\underline{t}(8) = 3.74$, $\underline{p} = .006$. Analyses comparing the intercepts of the regression lines revealed that females ($\underline{a}$ = .477) made significantly more total internal responses than males ($\underline{a}$ = .305) in the first test year, $\underline{t}(8) = 2.69$, $\underline{p} = .028$.

The following analyses coded the responses according to the percentage of male participants and the percentage of female participants that made all (4/4) external and all (4/4) internal responses for all four 45°/135° angle views in each year. The percentage of participants that made all external responses can be found in Table 4 by sex and test time. The percentage of participants that made all internal responses can be found in Table 5 by sex and test time. Log transformations were made to test time (1-5) and to the percentages of male and female participants who made all external and all internal responses for all analyses.

Regressions were performed separately for males and females to determine if the percentage of participants making all external or all internal responses were a function of test time for either sex. Slopes obtained from these regression analyses were used to obtain estimates of the increase in the percentage of male and female participants making all external and all internal responses as a function of test time. The intercepts indicate the percentage of male and female participants making all external and all internal responses in the first test year.

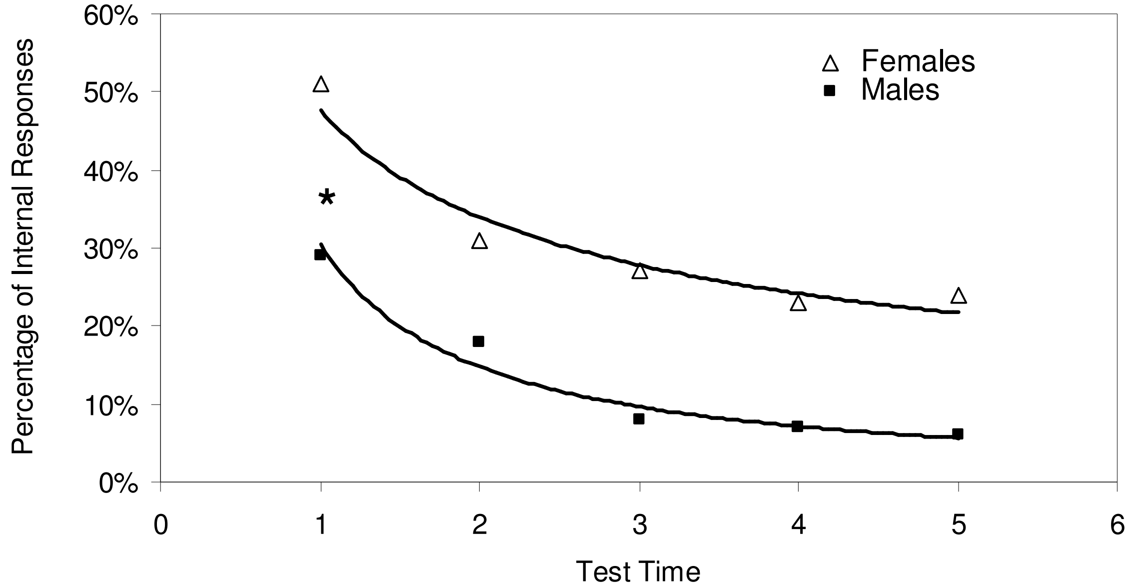


Figure 9. Total Percentage of Internal Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time. The power trendlines demonstrate the decelerating nature of the total percentage of internal responses made. The asterisk indicates a significant difference between males and females in the percentage of external responses made in the first test year (intercepts). The slopes of the lines also significantly differ from one another, with males decelerating in the total percentage of internal responses made at a faster rate than the females.

Table 4

Percentage of Participants in the Combined 45°/135° Angle Condition of the Water Level Task Making All External Responses by Sex and Test Time.

Test Time	**Males**	**Females**
Year 1	57.0%	32.0%
Year 2	75.0%	47.0%
Year 3	82.0%	44.0%
Year 4	89.0%	65.0%
Year 5	89.0%	71.0%

Table 5

Percentage of Participants in the Combined 45°/135° Angle Condition of the Water Level Task Making All Internal Responses by Sex and Test Time.

Test Time	**Males**	**Females**
Year 1	29.0%	41.0%
Year 2	18.0%	29.0%
Year 3	7.0%	24.0%
Year 4	7.0%	23.0%
Year 5	4.0%	23.0%

For females, a regression analysis revealed that test time was a significant predictor of the percentage of female participants making all external responses, $F(1, 3) = 23.17$, $p = .017$, $R^2 = .89$, $MSE = .003$, with test time accounting for 89% of the variance in the percentage of female participants making all external responses. The slope for the female participants ($b = .475$) was also statistically significant, $t(4) = 4.81$, $p = .017$, indicating that the percentage of female participants making all external responses significantly increased across time.

For males, a regression analysis revealed that test time was a highly significant predictor of the percentage of male participants making all external responses, $F(1, 3) = 68.37$, $p = .004$, $R^2 = .96$, $MSE = .000$, with test time accounting for 96% of the variance in the percentage of male participants making all external responses. The slope for the male participants ($b = .285$) was also statistically significant, $t(4) = 8.27$, $p = .004$, indicating that the percentage of male participants making all external responses significantly increased across time.

Figure 10 displays the percentage of participants that made all external responses by sex and by test time. The trendlines shown in the figure demonstrate that the percentage of participants making all external responses increased for both sexes as the participants aged. Analyses comparing the slopes of the regression lines revealed that males ($b = .285$) and females ($b = .475$) did not significantly differ from one another, $t(8) = 1.82$, $p = .107$. This means that participants of both sexes increased at the same rate regarding the percentage of participants that made all external responses. However, analyses comparing the intercepts of the regression lines revealed that the percentage of male participants ($a = .589$) making all external responses was significantly greater than

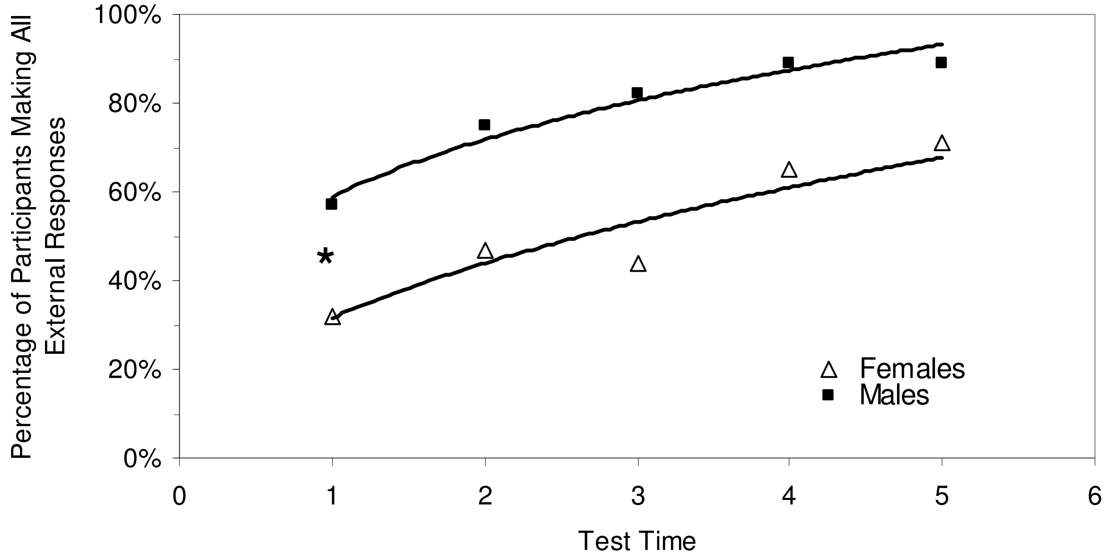


Figure 10. Percentage of Participants Making All External Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time. The power trendlines demonstrate the accelerating nature of the percentage of participants making all external responses. The asterisk indicates a significant difference between the percentage of male and female participants making all external responses made in the first test year (intercepts). The slopes of the lines did not significantly differ from one another.

the percentage of female participants (a = .316) making all external responses, in the first test year, $t(8) = -5.36$, $p = .001$.

For females, a regression analysis revealed that test time was a highly significant predictor of the percentage of female participants making all internal responses, $F(1, 3) = 45.96$, $p = .007$, $R^2 = .94$, MSE = .001, with test time accounting for 94% of the variance in the percentage of female participants making all internal responses. The slope for the female participants ($b = -.377$) was also statistically significant, $t(4) = -6.78$, $p = .007$, indicating that the percentage of female participants making all internal responses significantly decreased across time.

For males, a regression analysis revealed that test time was a highly significant predictor of the percentage of male participants making all internal responses, $F(1, 3) = 46.06$, $p = .007$, $R^2 = .94$, MSE = .010 , with test time accounting for 94% of the variance in the percentage of male participants making all internal responses. The slope for the male participants ($b = -1.223$) was also statistically significant, $t(4) = -6.79$, $p = .007$, indicating that the percentage of male participants making all internal responses significantly decreased across time.

Figure 11 displays the percentage participants that made all internal responses by sex and by test time. The trendlines shown in the figure demonstrate that the percentage of participants making all internal responses decreased for both sexes as the participants aged. Analyses comparing the slopes of the regression lines revealed that the percentage of male participants making all internal responses ($b = -1.223$) decreased at a faster rate than the percentage of female participants making all internal responses ($b = -.377$), $t(8) = 4.49$, $p = .002$. Analyses comparing the intercepts of the regression lines revealed that

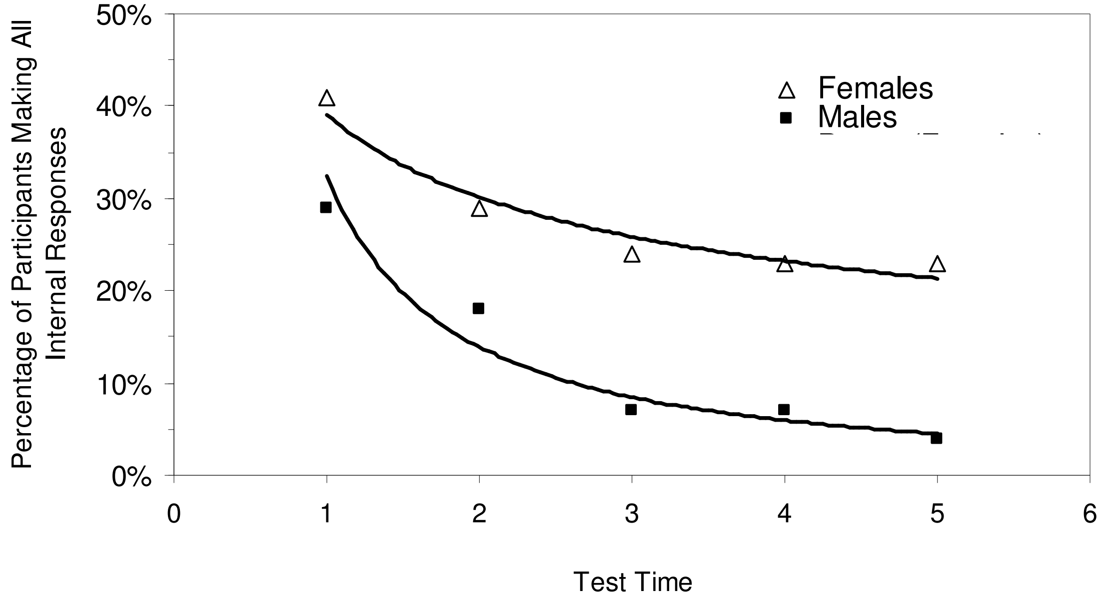


Figure 11. Percentage of Participants Making All Internal Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex and Test Time. The power trendlines demonstrate the decelerating nature of the percentage of participants making all internal responses. The intercepts did not significantly differ from one another, meaning that the percentage of male and female participants making all internal responses did not significantly differ from one another in the first test year (intercepts). The slopes of the lines did significantly differ from one another, with the percentage of male participants making all internal responses decelerating at faster rate than the females.

the percentage of male participants (a = .324) making all internal responses did not significantly differ from the percentage of female participants (a = .391) in the first test year, t(8) = .899, p = .395.

Means and standard deviations for the 90° angle condition of the Water Level Task are presented in Table 6 by sex and test time. In the 90° angle condition, a main effect of sex, such that males (M = 12.61°, SD = 30.75) outperformed females (M = 27.39°, SD = 30.75), was found to be marginally significant, F(1, 60) = 3.55, p = .065, MSE = 4728.49. A main effect of test time reached statistical significance, F(4, 240) = 7.35, p = .000, MSE = 476.15. Pairwise comparisons revealed that year 1 (M = 33.35°, SD = 42.26) performed significantly worse than all other test years (all p's < .05). Year 2 (M = 21.22°, SD = 37.88) performed significantly worse from year 4 (M = 14.52°, SD = 33.37) (p < .05). Year 2, year 3(M = 18.34°, SD = 35.85), and year 5 (M = 16.12°, SD = 34.55) did not significantly differ from one another (p > .05). Year 3, year 4, and year 5 did not significantly differ from one another (all p's > .05) (see Figure 12). The interaction between sex and test time failed to reach statistical significance, F(4, 240) = .44, p = .782, MSE = 476.15.

When the data were sorted by test year, results revealed that year 1 females (M = 39.82°, SD = 43.35) and year 1 males (M = 25.50°, SD = 40.25) did not significantly differ from one another, F(1, 60) = 1.79, p = .186, MSE = 1762.74. Year 2 females (M = 25.28°, SD = 40.15) and year 2 males (M = 16.29°, SD = 35.00) also did not significantly differ from one another, F(1, 60) = .864, p = .356, MSE = 1438.12. However, year 3 males (M = 8.14°, SD = 24.63) did significantly outperform the year 3 females

Table 6

Mean Degrees from Horizontal for the 90° Condition of the Water Level Task by Sex and Test Time.

Test Time	Males	Females	p-value
Year 1	25.50(40.25)	39.82(43.35)	0.186
Year 2	16.29(35.01)	25.28(40.15)	0.356
Year 3	8.14(24.63)	26.74(41.46)	0.041
Year 4	6.43(23.60)	21.19(38.74)	0.083
Year 5	6.68(23.54)	23.90(40.19)	0.050

Values presented are means with standard deviations in parentheses.

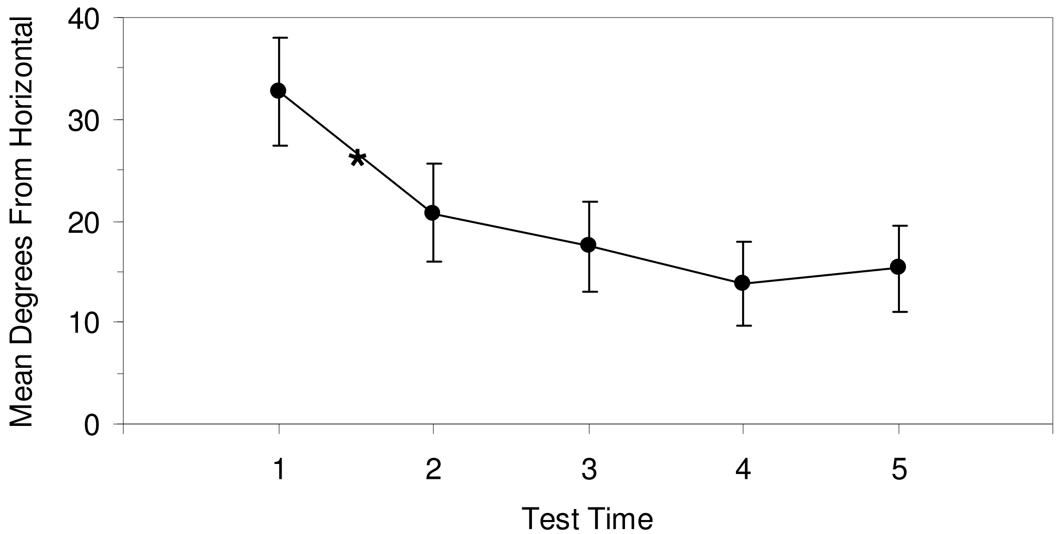


Figure 12. Mean Degrees from Horizontal in the 90° Angle Condition of the Water Level Task by Test Time. Asterisks indicate significant differences between test years.

(M = 26.74°, SD = 41.46), F(1, 60) = 4.36, p = .041, MSE = 1218.18. A marginally significant effect of sex was found in year 4, such that year 4 males (M = 6.43°, SD = 23.60) outperformed the year 4 females (M = 21.19°, SD = 38.74), F(1, 60) = 3.11, p = .083, MSE = 1076.25. Year 5 males (M = 6.68°, SD = 23.54) significantly outperformed year 5 females (M = 23.90°, SD = 40.19), F(1, 60) = 4.00, p = .050, MSE = 1137.78 (see Figure 13).

The main effect of sex in the 180° angle condition failed to reach statistical significance, F(1, 60) = .09, p = .765, MSE = 1.42 (see Figure 14 for results by sex and test time). However, a main effect of test time was found to be statistically significant, F(4, 240) = 14.75, p = .000, MSE = 1.39. Pairwise comparisons revealed that year 1 (M = 1.47°, SD = 2.30) performed significantly worse than all other test years (all p's ≤ .001). Year 2 (M = .11°, SD = .58), year 3(M = .19°, SD = .62), and year 4 (M = .08°, SD = .42) did not significantly differ from one another (all p's > .05). Finally, year 4 significantly outperformed year 5 (M = .35°, SD = .85) (p = .024), but year 5 did not significantly differ from year 2 and year 3 (all p's > .05) (see Figure 15). The interaction between sex and test time did not yield statistically significant results, F(2, 240) = .32, p = .870, MSE = 1.39.

Flags Task. A 2 (sex) x 5 (test time) mixed factorial ANOVA was conducted on the dependent variable of the number of correct responses out of 48 trials. Means and standard deviations for the Flags Task are presented in Table 7 by sex and test time.

The main effect of sex failed to reach statistical significance; however, there was a trend in the expected direction with males averaging higher scores across all five test years, F(1, 64) = 2.09, p = .153, MSE = 180.03 (see Figure 16 for results by sex and test

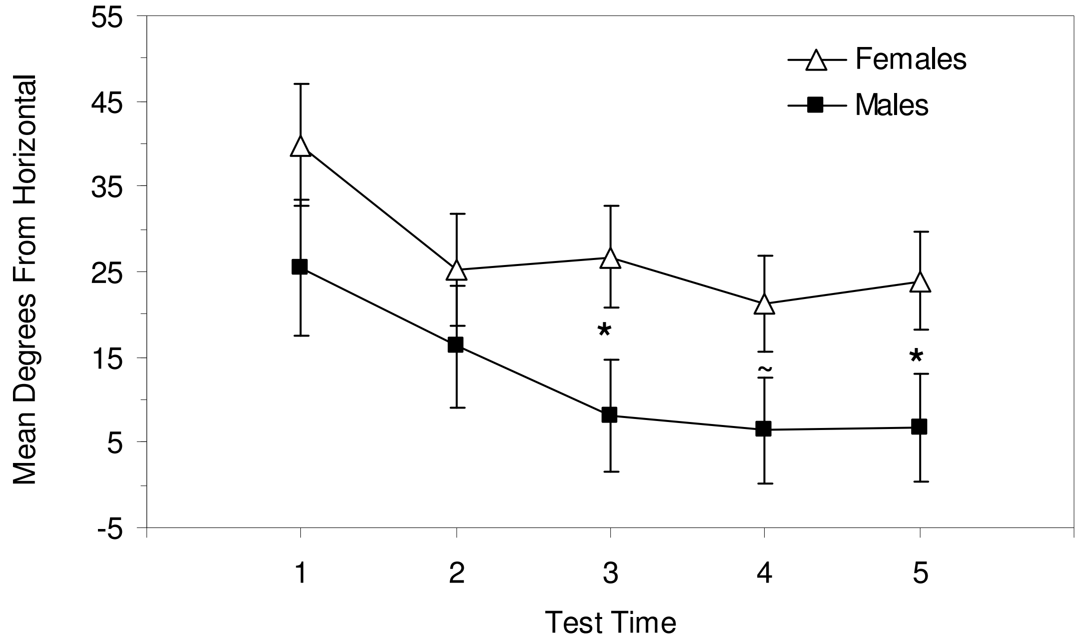


Figure 13. Mean Degrees from Horizontal in the 90° Angle Condition of the Water Level Task by Sex and Test Time. Asterisks indicate significant differences and the ~ symbol indicates marginally significant differences.

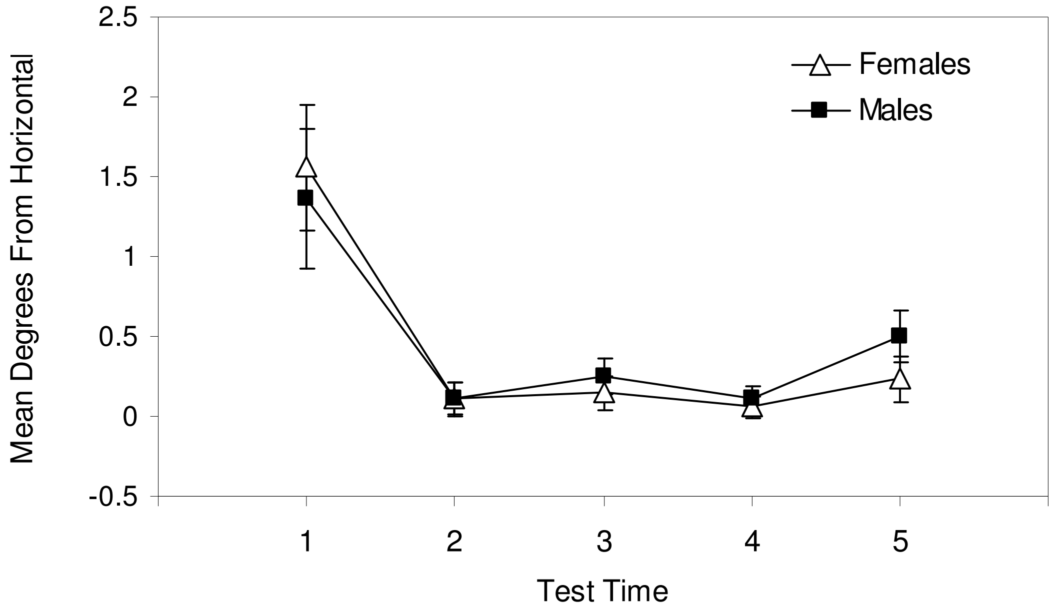


Figure 14. Mean Degrees from Horizontal in the 180° Angle Condition of the Water Level Task by Sex and Test Time.

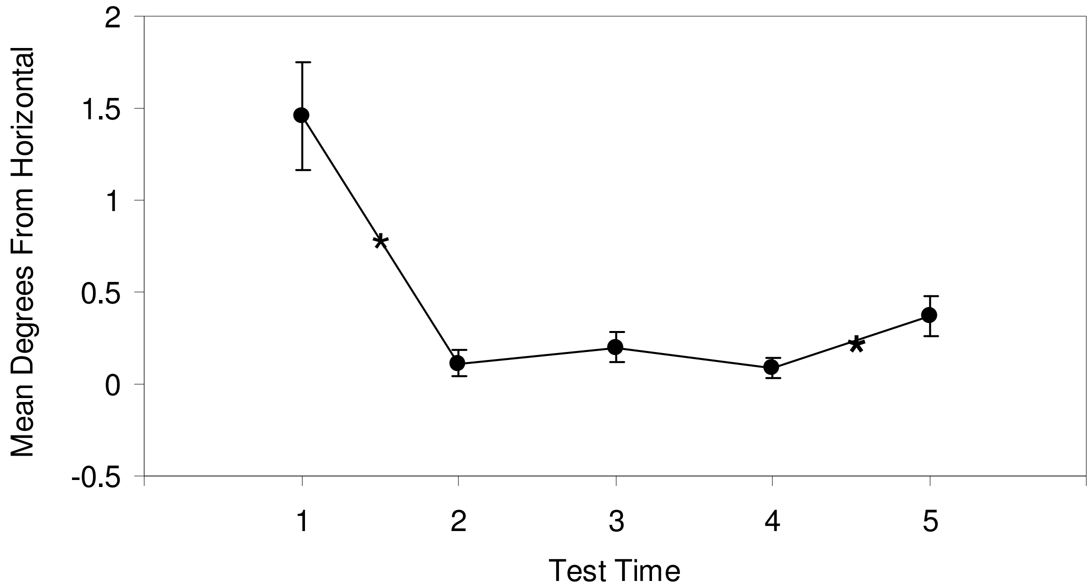


Figure 15. Mean Degrees from Horizontal in the 180° Angle Condition of the Water Level Task by Test Time. Asterisks indicate significant differences between test years.

Table 7

Mean Number of Correct Responses on the Flags Task by Sex and Test Time.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	38.73	9.87	30	35.06	10.16	36
Year 2	42.33	6.70	30	40.06	7.34	36
Year 3	43.20	6.35	30	40.94	7.44	36
Year 4	44.50	4.32	30	43.11	5.92	36
Year 5	44.93	4.03	30	43.81	5.38	36

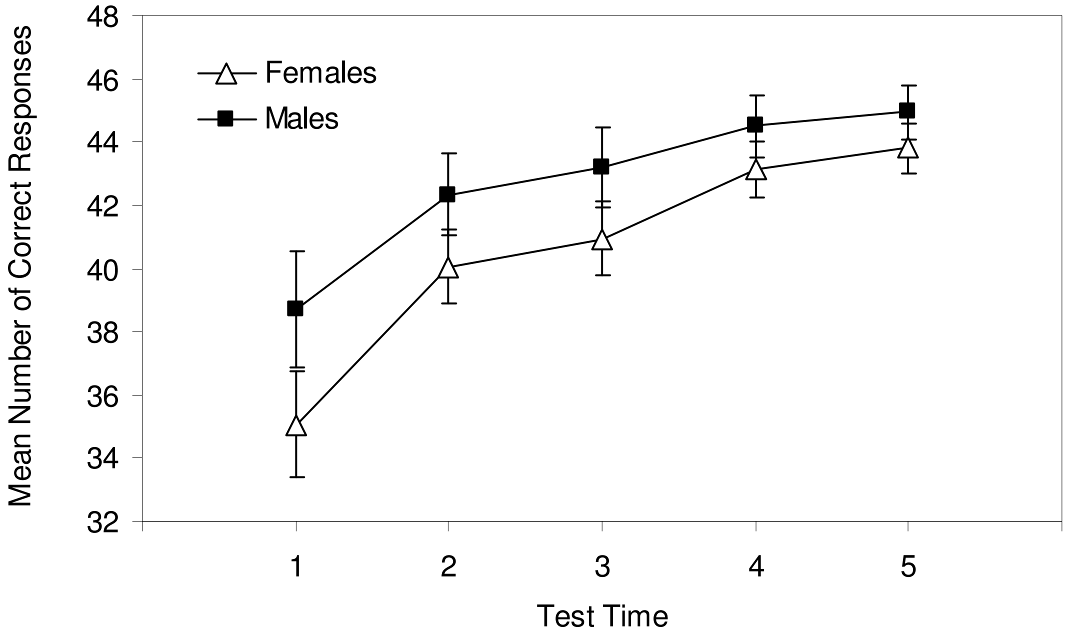


Figure 16. Mean Number of Correct Responses on the Flags Task by Sex and Test Time.

time). A significant main effect of test time was found, $\underline{F}(4, 256) = 32.87$, $\underline{p} = .000$, $\underline{MSE} = 17.44$. Pairwise comparisons revealed that year 1 ($\underline{M} = 36.72$, $\underline{SD} = 10.12$) performed significantly worse than all other test years (all $\underline{p}$'s = .000). Year 2 ($\underline{M} = 41.09$, $\underline{SD} = 7.09$) and year 3 ($\underline{M} = 41.97$, $\underline{SD} = 7.01$) did not significantly differ from one another ($\underline{p} > .05$) and performed significantly worse from year 4 and year 5 (all $\underline{p}$'s < .01). Year 4 ($\underline{M} = 43.74$, $\underline{SD} = 5.26$) and year 5 ($\underline{M} = 44.32$, $\underline{SD} = 4.81$) did not significantly differ from one another ($\underline{p} > .05$) and significantly outperformed all other test years (all $\underline{p}$'s < .01) (see Figure 17). The interaction between sex and test time failed to reach statistical significance, $\underline{F}(4, 240) = .94$, $\underline{p} = .444$, $\underline{MSE} = 17.44$.

Basketball Man Task. Separate regressions were performed for each participant in each year on the front (three-dimensional) and the back (two-dimensional) rotation conditions. The slopes, intercepts, R^2 values, and error rates from these regressions were also averaged for each participant in each year to form a combined front and back score. Slopes obtained from these regression analyses were used to obtain estimates of the increase in reaction time as a function of rotation angle (0°, 45°, 135°, & 180°). The slopes, intercepts, R^2 values, and error rates for the front, the back, and the combined conditions were then entered into a 2 (sex) x 5 (test time) mixed factorial ANOVA.

Participants with error rates greater than .2 were omitted from the slope, intercept, and R^2 value analyses. Mean scores for the participants that were included in the present analyses can be found in Tables 8-12 by sex and test year. Only significant findings are reported in the following analyses.

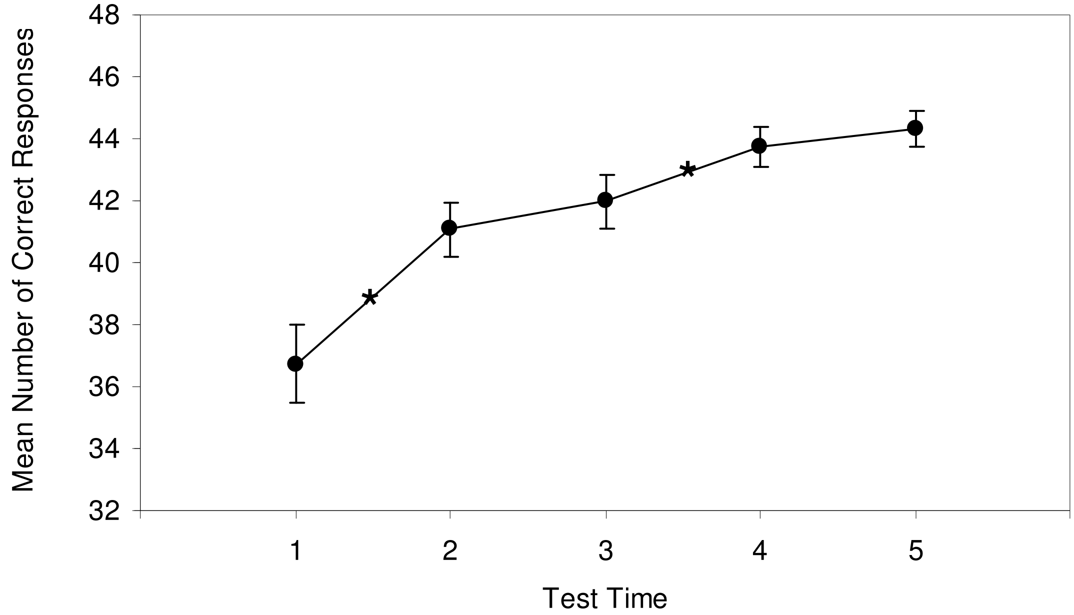


Figure 17. Mean Number of Correct Responses on the Flags Task by Test Time.

Asterisks indicate significant differences between test times.

Table 8

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 1

Year 1	Females				Males			

Front View (3-D)

	Slope	Intercept	R^2	Errors	Slope	Intercept	R^2	Errors
Mean	0.009	1.626	0.331	0.101	0.005	1.586	0.234	0.069
SD	0.005	0.810	0.227	0.166	0.005	0.503	0.207	0.084
N	19	19	19	28	18	18	18	24

Back View (2-D)

	Slope	Intercept	R^2	Errors	Slope	Intercept	R^2	Errors
Mean	0.004	1.474	0.313	0.092	0.006	1.116	0.411	0.056
SD	0.007	0.844	0.229	0.156	0.003	0.280	0.190	0.097
N	21	21	21	28	18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

Table 9

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 2

Year 2	Females	Males

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.004	1.531	0.253	0.098		0.004	1.509	0.225	0.108
SD		0.006	0.574	0.239	0.126		0.004	0.901	0.188	0.160
N		19	19	19	28		18	18	18	24

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.004	1.205	0.286	0.062		0.004	1.020	0.342	0.042
SD		0.002	0.537	0.171	0.105		0.004	0.260	0.242	0.095
N		21	21	21	28		18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

Table 10

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 3

Year 3	Females	Males

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.006	1.368	0.340	0.074		0.004	1.420	0.265	0.038
SD		0.006	0.448	0.233	0.139		0.004	0.586	0.204	0.055
N		19	19	19	28		18	18	18	24

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.004	1.263	0.352	0.068		0.003	0.996	0.354	0.014
SD		0.004	0.829	0.243	0.165		0.003	0.256	0.227	0.040
N		21	21	21	28		18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

Table 11

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 4

Year 4	Females					Males			

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.004	1.288	0.328	0.080		0.002	1.207	0.192	0.080
SD		0.004	0.381	0.211	0.142		0.006	0.404	0.207	0.131
N		19	19	19	28		18	18	18	24

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.003	0.958	0.280	0.092		0.003	0.912	0.320	0.045
SD		0.003	0.325	0.195	0.149		0.003	0.390	0.241	0.123
N		21	21	21	28		18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

Table 12

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 5

Year 5	Females	Males

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.003	1.147	0.280	0.101		0.003	1.272	0.243	0.056
SD		0.003	0.600	0.204	0.175		0.003	0.577	0.257	0.068
N		19	19	19	28		18	18	18	24

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.003	0.811	0.296	0.110		0.003	0.877	0.305	0.038
SD		0.002	0.222	0.185	0.190		0.001	0.213	0.185	0.070
N		21	21	21	28		18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

No sex effects were found to be statistically significant (all p's > .05).

A significant main effect of test time was found in the slopes of the front condition, F(1, 140) = 5.79, p = .000, MSE = .000. Pairwise comparisons revealed that year 1 and year 3 (M = .0050, SD = .0056) did not significantly differ from one another (p > .05). Year 1 (M = .0071, SD = .0053) had significantly steeper slopes than year 2 (M = .0044, SD = .0052), year 4 (M = .0033, SD = .0050), and year 5 (M = .0026, SD = .0032) (all p's ≤ .009). Year 2, year 3, and year 4 did not significantly differ from one another (all p's > .05). Year 2 and year 3 had significantly steeper slopes than year 5 (all p's ≤ .048). Year 4 and year 5 did not significantly differ from one another, (p > .05) (see Figure 18).

A significant main effect of test time was found in the intercepts of the front condition, F(4, 140) = 5.90, p = .000, MSE = .182. Pairwise comparisons revealed that year 1 (M = 1.61, SD = .67) and year 2 (M = 1.52, SD = .74) did not significantly differ from one another (p > .05); however both of these test years had significantly longer reaction times than year 4 (M = 1.25, SD = .39) and year 5 (M = 1.21, SD = .58) (all p's ≤ .006) Year 3 (M = 1.39, SD = .51) did not significantly differ from any other test year (all p's > .05). Year 4 and year 5 did not significantly from each other (p > .05) (see Figure 19).

A significant main effect of test time was found in the slopes of the back condition, F(4, 148) = 3.84, p = .005, MSE = .000. Pairwise comparisons revealed that year 1 (M = .0049, SD = .0052), year 2 (M = .0041, SD = .0031), and year 3 (M = .0036, SD = .0034) did not significantly differ from one another (all p's > .05). Year 1 had significantly steeper slopes than year 4 (M = .0032, SD = .0025) and year 5 (M = .0026,

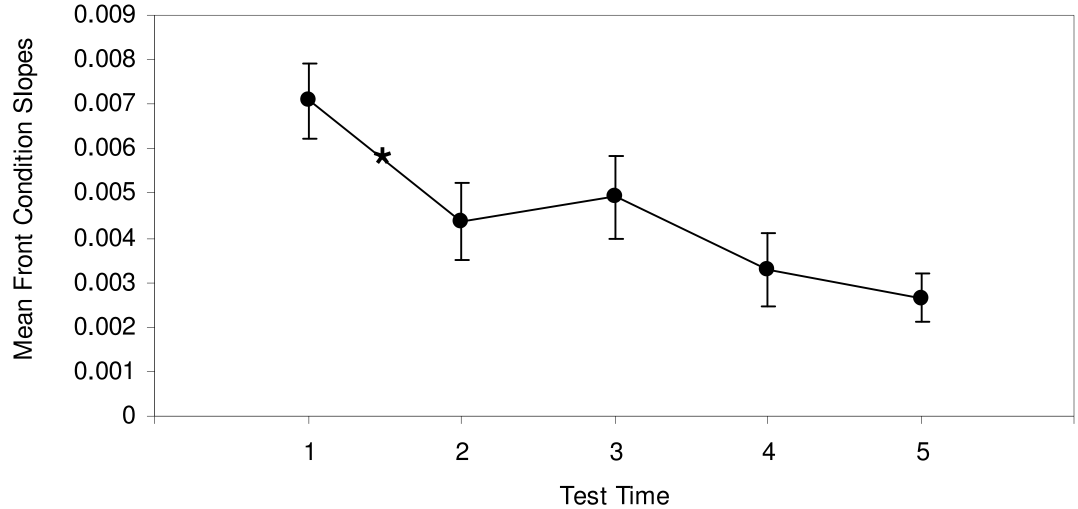


Figure 18. Mean Front (3-dimensional) Slopes on the Basketball Man Task by Test Time. Asterisks indicate significant differences between test times.

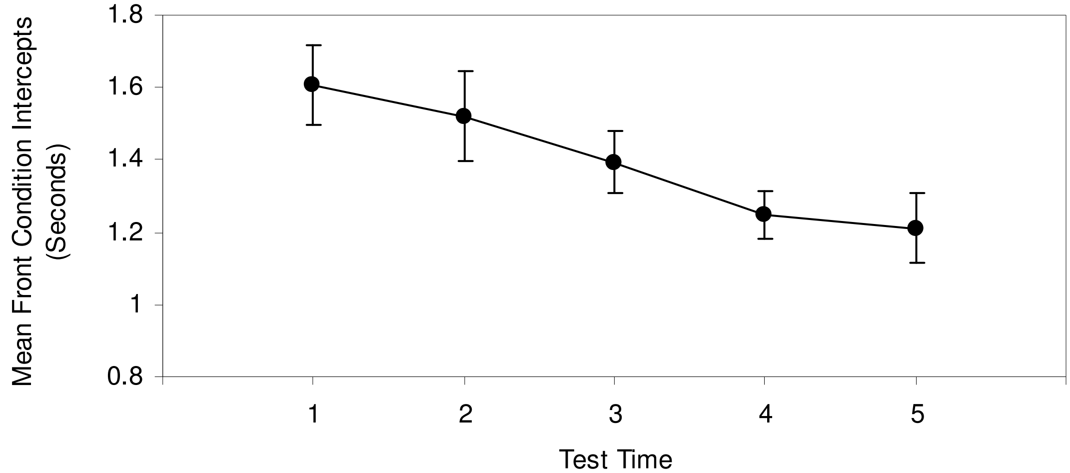


Figure 19. Mean Front (3-dimensional) Intercepts on the Basketball Man Task by Test Time.

SD = .0015) (all p's ≤ .039). Year 2 had significantly steeper slopes than year 5 (p = .003). Year 3 did not significantly differ from any other test year (all p's > .05). Year 4 and year 5 did not significantly from each other (p > .05) (see Figure 20).

A significant main effect of test time was also found in the intercepts of the back condition, F(4, 148) = 8.82, p = .000, MSE = .138. Pairwise comparisons revealed that year 1 (M = 1.31, SD = .67), year 2 (M = 1.12, SD = .44), and year 3 (M = 1.14, SD = .64) did not significantly differ from each other (all p's > .05). Year 1 and year 2 had significantly longer reaction times than year 4 (M = .94, SD = .35) and year 5 (M = .84, SD = .22) (all p's ≤ .017). Year 3 and year 4 did not significantly differ from one another (p > .05) and year 4 and year 5 did not significantly differ from one another (p > .05). Year 3 had significantly longer reaction times than year 5 (p = .003) (see Figure 21).

A significant main effect of test time was found in the slopes of the combined (front and back) condition, F(4, 136) = 12.67, p = .000, MSE = .000. Pairwise comparisons revealed that year 1 (M = .0061, SD = .0033) had significantly steeper slopes than all other test years (all p's < .05). Year 2 (M = .0046, SD = .0028) and year 3 (M = .0041, SD = .0030) did not significantly differ from one another (p > .05). Year 2 had significantly steeper slopes than year 4 (M = .0031, SD = .0026) and year 5 (M = .0026, SD = .0019) (all p's ≤ .011). Year 3 and year 4 did not significantly differ from one another (p > .05), and year 4 and year 5 did not significantly differ from one another (p > .05). Year 3 had significantly steeper slopes than year 5 (p = .004).

A significant main effect of test time was also found in the intercepts of the combined (front and back) condition, F(4, 136) = 11.31, p = .000, MSE = .007. Pairwise comparisons revealed that year 1 (M = 1.36, SD = .44), year 2 (M = 1.26, SD = .40), and

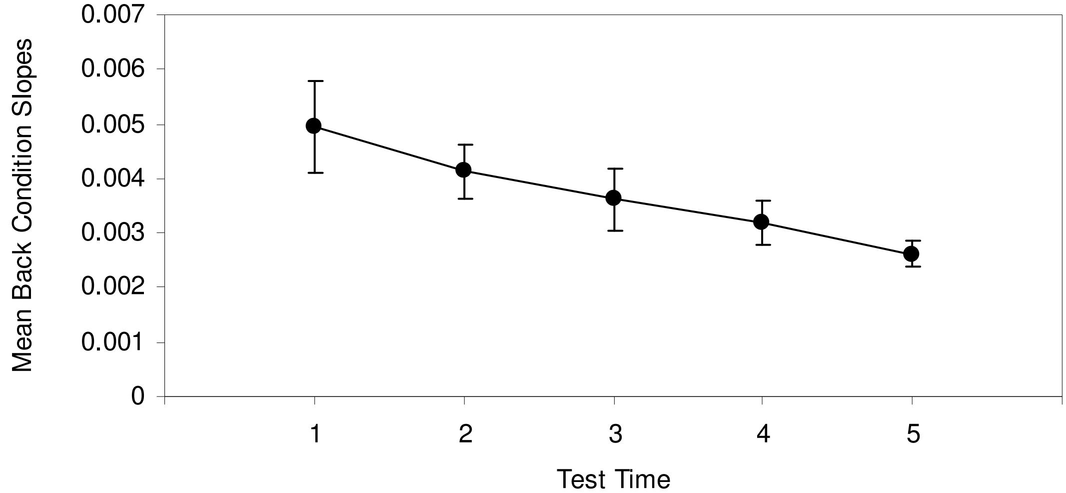


Figure 20. Mean Back (2-dimensional) Slopes on the Basketball Man Task by Test Time.

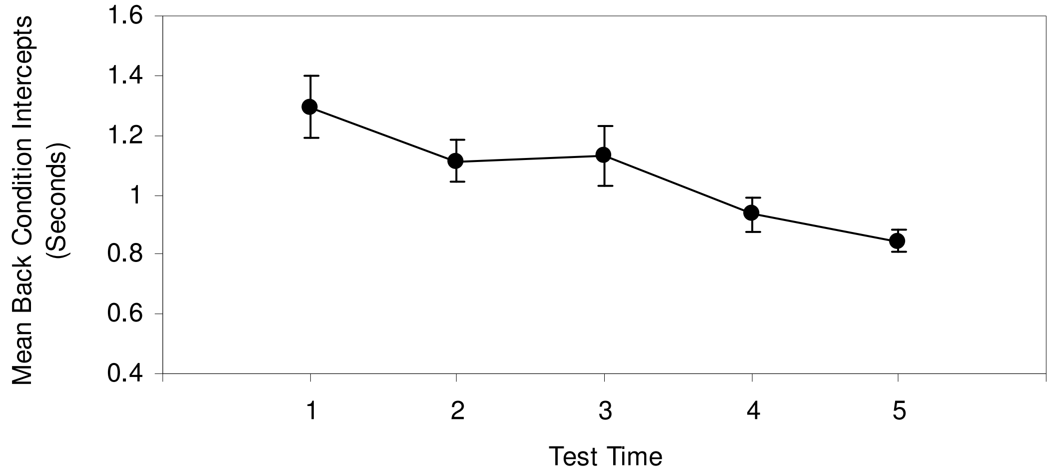


Figure 21. Mean Back (2-dimensional) Intercepts on the Basketball Man Task by Test Time.

year 3 (M = 1.21, SD = .37) did not differ significantly from one another (all p's > .05). Year 1, year 2, and year 3 had significantly longer reaction times than year 4 (M = 1.06, SD = .30) and year 5 (M = .98, SD = .31) (all p's ≤ .021). Year 4 and year 5 did not significantly differ from one another (p > .05).

Female-Sensitive Tasks

Purdue Pegboard. In each condition, a 2 (sex) x 5 (test time) mixed factorial ANOVA was conducted on the dependent variable of the number of correctly inserted pegs. Means and standard deviations for each condition are presented in Tables 13-16 by sex and test time.

The main effect of sex failed to reach statistical significance in the dominant hand condition, F(1, 62) = .44, p = .511, MSE = 1.10 (see Figure 22 for results by sex and test time). However, there was a significant main effect of test time in the dominant hand condition, F(4, 248) = 16.53, p = .000, MSE = 1.95. Pairwise comparisons revealed that year 1 (M = 13.64, SD = 1.79) inserted significantly less pegs than all other test years (all p's < .01). Year 2 (M = 14.58, SD = 1.64) and year 3 (M = 14.80, SD = 1.60) did not significantly differ from one another (p > .05) and inserted significantly less pegs than year 4 and year 5 (all p's ≤ .05). Year 4 (M = 15.47, SD = 1.37) and year 5 (M = 15.28, SD = 1.68) did not significantly differ from one another (p > .05) and inserted significantly more pegs than all test years (see Figure 23). The interaction between sex and test time in the dominant hand condition failed to reach statistical significance, F(4, 248) = .17, p = .952, MSE = 1.95.

The main effect of sex in the non-dominant hand condition failed to reach statistical significance, F(1, 62) = .67, p = .42, MSE = 1.33 (see Figure 24 for results by

Table 13

Mean Number of Correctly Inserted Pegs for the Dominant Hand Condition of the Purdue Pegboard Task by Sex and Test Time.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	13.83	1.87	29	13.49	1.74	35
Year 2	14.72	1.62	29	14.46	1.67	35
Year 3	14.86	1.75	29	14.74	1.48	35
Year 4	15.45	1.43	29	15.49	1.34	35
Year 5	15.38	1.66	29	15.20	1.71	35

Table 14

Mean Number of Correctly Inserted Pegs for the Non-Dominant Hand Condition of the Purdue Pegboard Task by Sex and Test Time.

Test Time	Males			Females		
	M	SD	N	M	SD	N
Year 1	13.24	1.27	29	13.29	1.49	35
Year 2	14.34	1.74	29	13.54	1.44	35
Year 3	13.97	1.74	29	14.23	1.59	35
Year 4	14.79	1.35	29	14.40	1.48	35
Year 5	15.07	2.87	29	14.77	1.40	35

Table 15

Mean Number of Correctly Inserted Pegs for the Both Hands Condition of the Purdue Pegboard Task by Sex and Test Time.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	21.48	3.30	29	22.46	2.75	35
Year 2	23.10	3.19	29	23.06	3.37	35
Year 3	23.97	3.12	29	23.91	2.09	35
Year 4	24.69	2.59	29	24.83	2.62	35
Year 5	24.48	3.45	29	24.34	2.75	35

Table 16

Mean Number of Correctly Inserted Pegs for the Assembly Condition of the Purdue Pegboard Task by Sex and Test Time.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	33.34	6.99	29	33.34	5.27	35
Year 2	35.69	5.76	29	34.63	5.29	35
Year 3	38.31	4.54	29	36.66	6.10	35
Year 4	38.34	5.11	29	39.89	5.13	35
Year 5	39.34	4.92	29	40.06	7.73	35

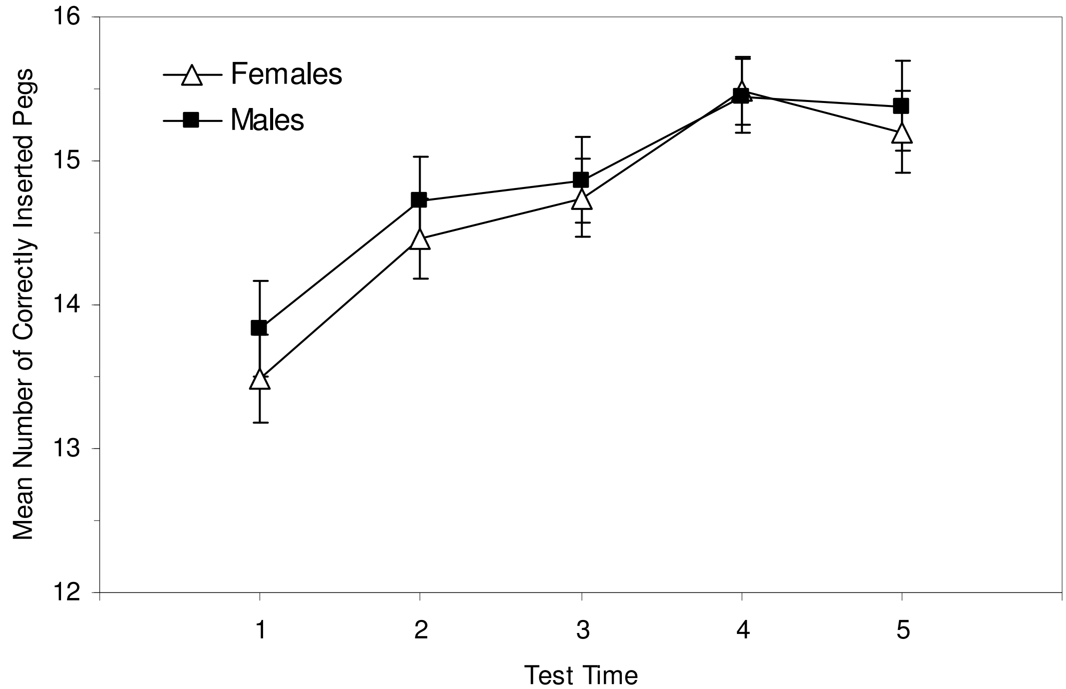


Figure 22. Mean Number of Correctly Inserted Pegs in the Dominant Hand Condition of the Purdue Pegboard Task by Sex and Test Time.

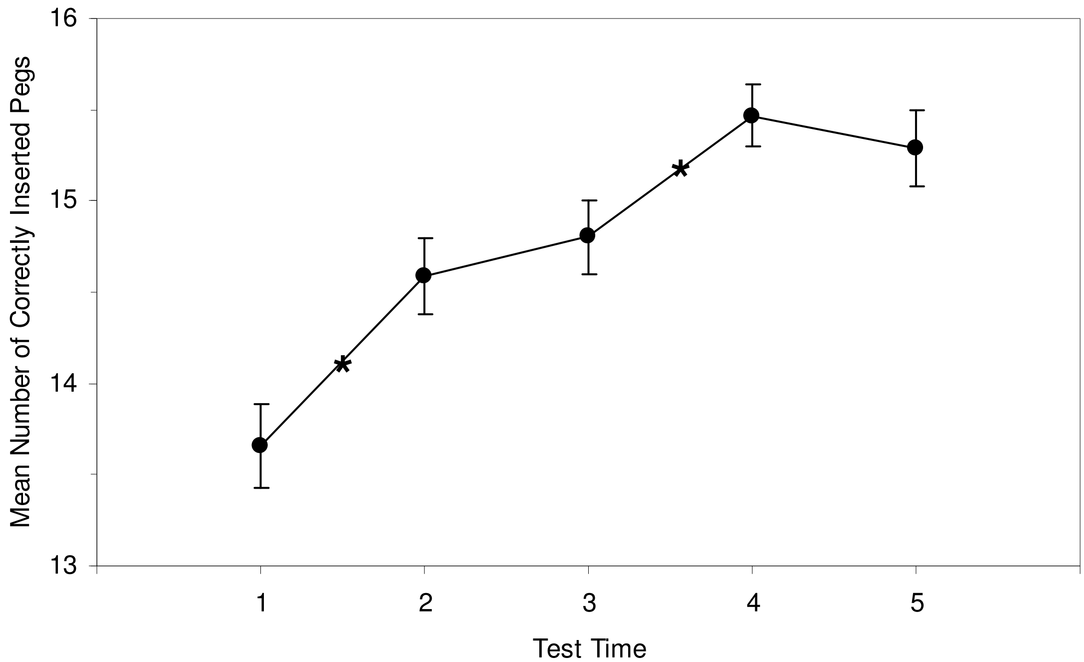


Figure 23. Mean Number of Correctly Inserted Pegs in the Dominant Hand Condition of the Purdue Pegboard Task by Test Time. Asterisks indicate significant differences between test times.

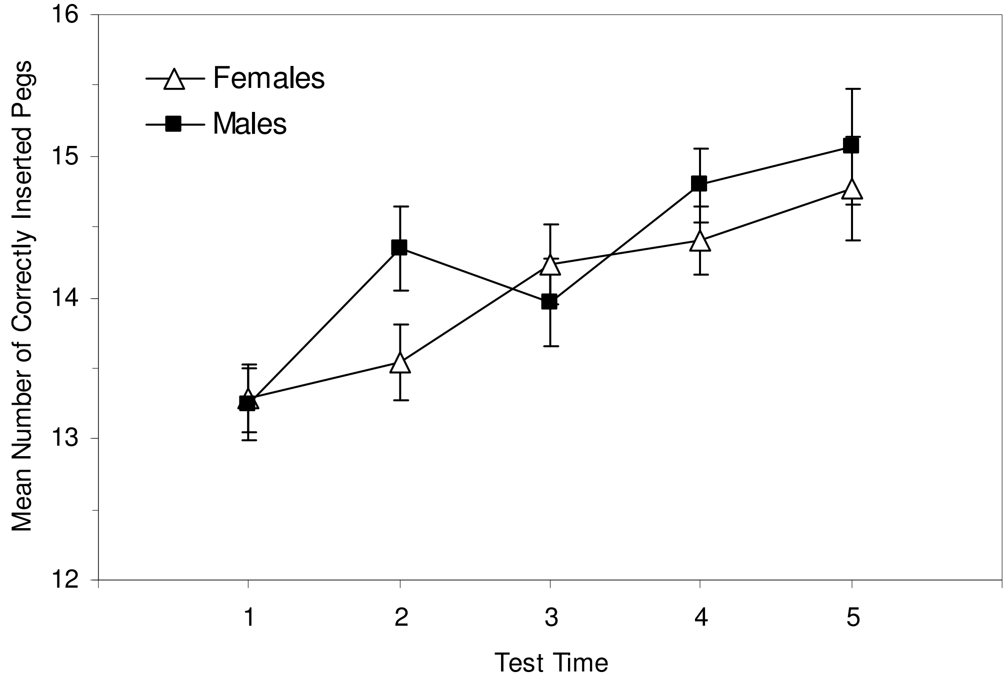


Figure 24. Mean Number of Correctly Inserted Pegs in the Non-Dominant Hand Condition of the Purdue Pegboard Task by Sex and Test Time.

sex and test time). However, a significant main effect of test time was found in the non-dominant hand condition, $\underline{F}(4, 248) = 14.03$, $\underline{p} = .000$, $\underline{MSE} = 1.84$. Pairwise comparisons revealed that year 1 ($\underline{M} = 13.27$, $\underline{SD} = 1.38$) inserted significantly less pegs than all other test years (all $\underline{p}$'s $< .01$). Year 2 ($\underline{M} = 13.91$, $\underline{SD} = 1.62$) and year 3 ($\underline{M} = 14.11$, $\underline{SD} = 1.65$) did not significantly differ from one another ($\underline{p} > .05$) and inserted significantly less pegs than year 4 and year 5 (all $\underline{p}$'s $< .05$). Year 4 ($\underline{M} = 14.58$, $\underline{SD} = 1.42$) and year 5 ($\underline{M} = 14.91$, $\underline{SD} = 2.17$) did not significantly differ from one another ($\underline{p} > .05$) and inserted significantly more pegs than all other test years (all $\underline{p}$'s $< .05$) (see Figure 25). The interaction between sex and test time failed to reach statistical significance, $\underline{F}(4, 248) = 1.46$, $\underline{p} = .214$, $\underline{MSE} = 1.84$.

The main effect of sex in the both hands condition failed to reach statistical significance, $\underline{F}(1, 62) = .12$, $\underline{p} = .73$, $\underline{MSE} = 19.79$ (see Figure 26 for results by sex and test time). However, a main effect of test time in the both hands condition was found to be statistically significant, $\underline{F}(4, 248) = 13.77$, $\underline{p} = .000$, $\underline{MSE} = 5.81$. Pairwise comparisons revealed that year 1 ($\underline{M} = 22.02$, $\underline{SD} = 3.03$) inserted significantly less pegs than all other test years (all $\underline{p}$'s $\leq .01$). Year 2 ($\underline{M} = 23.08$, $\underline{SD} = 3.26$) inserted significantly less pegs than year 3, year 4, and year 5 (all $\underline{p}$'s $< .05$). Year 3 ($\underline{M} = 23.94$, $\underline{SD} = 2.59$) and year 5 ($\underline{M} = 24.41$, $\underline{SD} = 3.06$) did not significantly differ from one another ($\underline{p} > .05$). Year 3 inserted significantly less pegs than year 4 ($\underline{p} < .05$). Year 4 ($\underline{M} = 24.77$, $\underline{SD} = 2.59$) and year 5 did not significantly differ from one another ($\underline{p} > .05$). Year 4 inserted significantly more pegs than year 1, year 2, and year 3 (all $\underline{p}$'s $< .05$) (see Figure 27). The interaction between sex and test time failed to reach statistical significance, $\underline{F}(4, 248) = .57$, $\underline{p} = .682$, $\underline{MSE} = 5.81$.

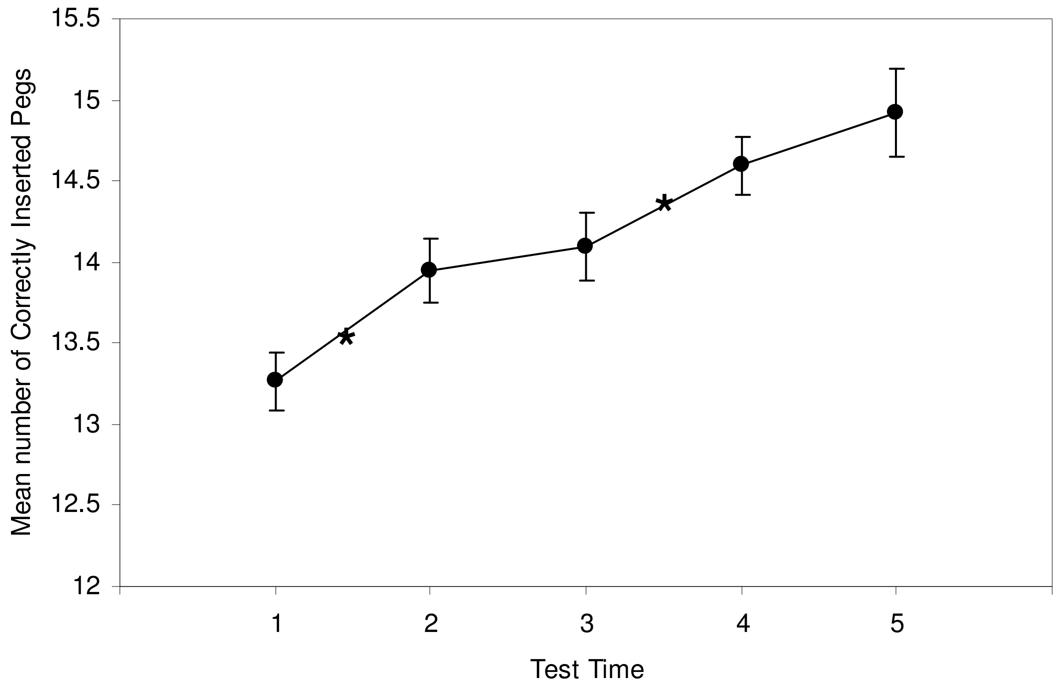


Figure 25. Mean Number of Correctly Inserted Pegs in the Non-Dominant Hand Condition of the Purdue Pegboard Task by Test Time. Asterisks indicate significant differences between test times.

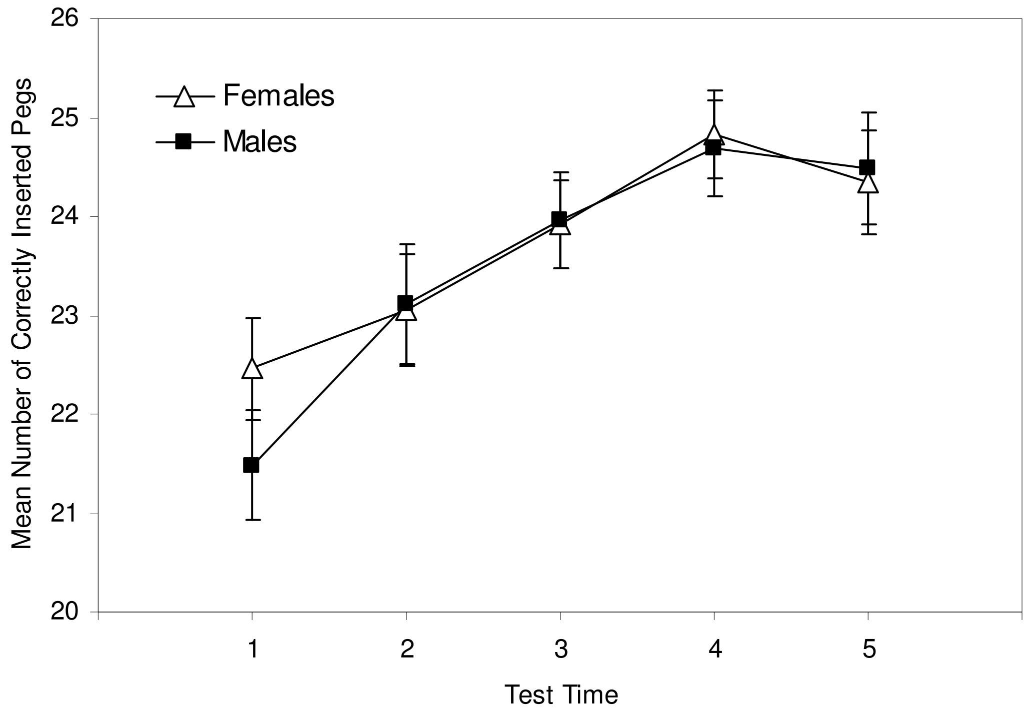


Figure 26. Mean Number of Correctly Inserted Pegs in the Both Hands Condition of the Purdue Pegboard Task by Sex and Test Time.

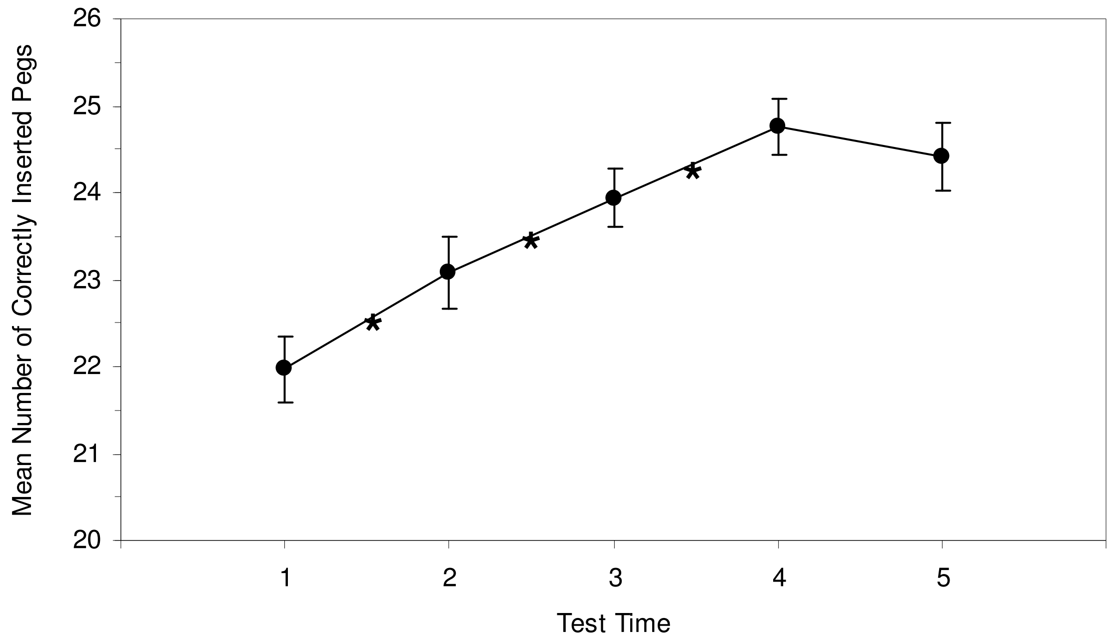


Figure 27. Mean Number of Correctly Inserted Pegs in the Both Hands Condition of the Purdue Pegboard Task by Test Time. Asterisks indicate significant differences between test times.

There was not a significant main effect of sex in the assembly condition, $\underline{F}(1, 62) = .01$, $\underline{p} = .94$, $\underline{MSE} = 103.20$ (see Figure 28 for results by sex and test time). However, a main effect of test time was found to be statistically significant, $\underline{F}(4, 248) = 28.48$, $\underline{p} = .000$, $\underline{MSE} = 16.01$. Pairwise comparisons revealed that year 1 ($\underline{M} = 33.34$, $\underline{SD} = 6.06$) inserted significantly less pegs than all other test years (all $\underline{p}$'s < .05). Year 2 ($\underline{M} = 35.11$, $\underline{SD} = 5.49$) inserted significantly less pegs than year 3 ($\underline{M} = 37.41$, $\underline{SD} = 5.47$), year 4 ($\underline{M} = 39.19$, $\underline{SD} = 5.14$), and year 5 ($\underline{M} = 39.73$, $\underline{SD} = 6.57$) (all $\underline{p}$'s < .05). Year 3 inserted significantly less pegs than year 4 and year 5 (all $\underline{p}$'s < .01). Year 4 and year 5 did not significantly differ from one another ($\underline{p} > .05$) and inserted significantly more pegs than all other test years (all $\underline{p}$'s < .01) (see Figure 29). The interaction between sex and test time failed to reach statistical significance, $\underline{F}(4, 248) = 1.66$, $\underline{p} = .160$, $\underline{MSE} = 16.01$.

Hypothesis II

Hypothesis II stated that there would be sex differences in each of the cognitive tasks in the adult sample. Hypothesis II also stated that performance on each of the cognitive tasks will be comparable for the year 5 males and the adult sample males. Likewise, performance on each of the cognitive tasks will be comparable for the year 5 females and the adult sample females. In order to evaluate Hypothesis II the following analyses were conducted.

Male-Sensitive Tasks

Water Level Task. For each participant, the scores from each angle of rotation were averaged to form a single score for the combined 45°/135° angle condition, the 90° angle condition, and the 180° angle condition. A 2 (sex) x 2 (sample: year 5 and adults) between-subjects factorial ANOVA was then conducted on the mean degrees from

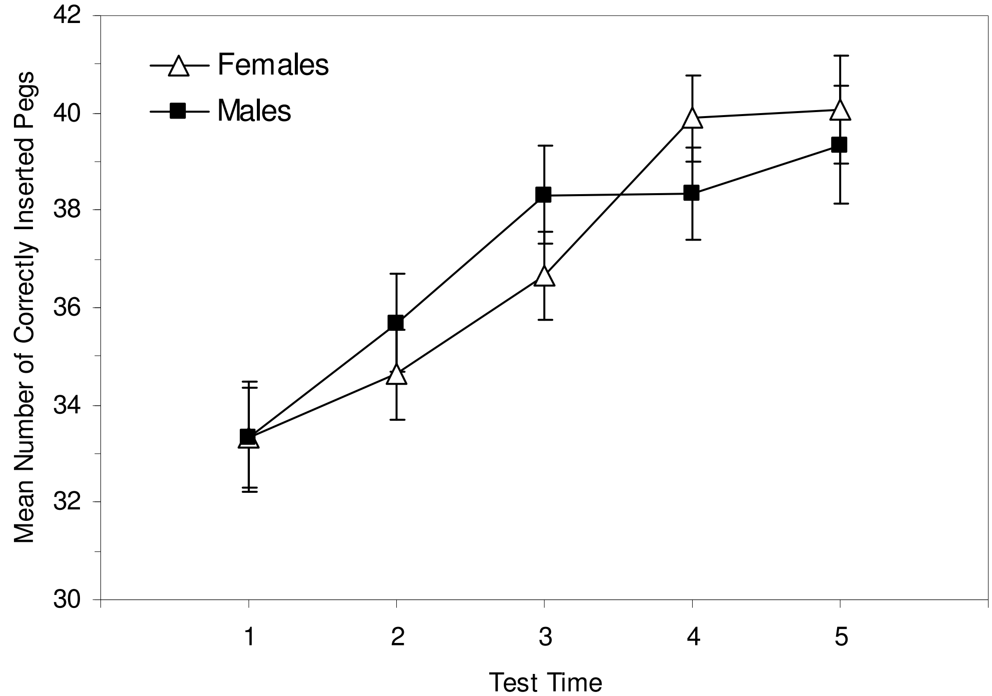


Figure 28. Mean Number of Correctly Inserted Pegs in the Assembly Condition of the Purdue Pegboard Task by Sex and Test Time.

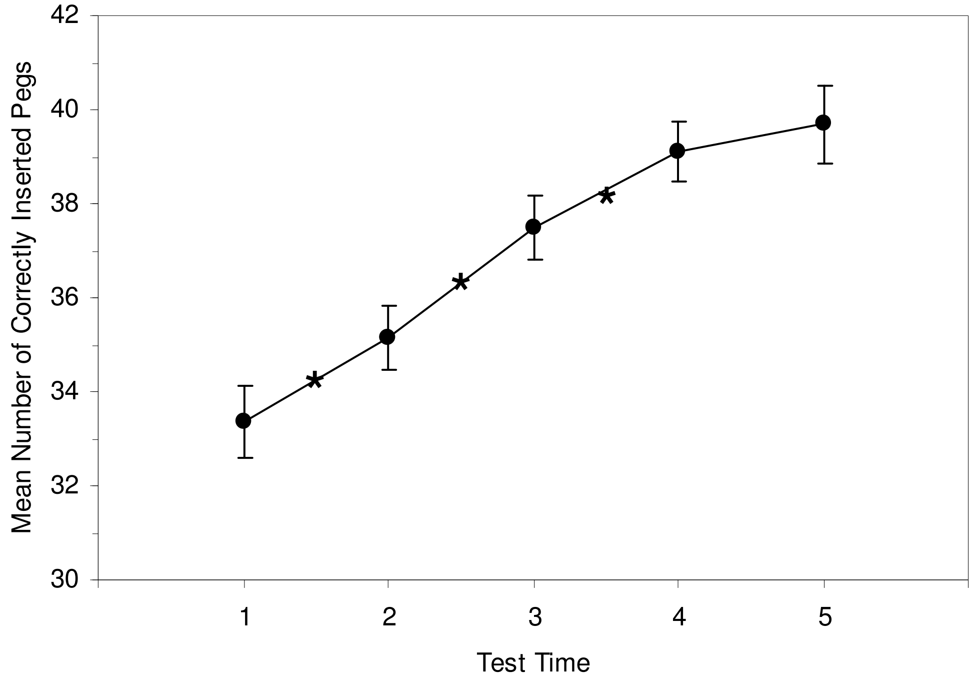


Figure 29. Mean Number of Correctly Inserted Pegs in the Assembly Condition of the Purdue Pegboard Task by Test Time. Asterisks indicate significant differences between test times.

horizontal for each of the three angle conditions.

Means and standard deviations for the combined 45°/135° angle condition of the Water Level Task are presented in Table 17 by sex, test time, and the adult sample. As predicted males (M = 6.55°, SD = 10.64) significantly outperformed females (M = 19.59°, SD = 21.20) in the combined 45°/135° angle condition, $F(1, 122) = 19.44$, p = .000, MSE = 280.63. The year 5 sample (M = 11.00°, SD = 15.65) and the adult sample (M = 15.69°, SD = 20.15) did not differ significantly from one another, $F(1, 122) = 2.89$, p = .092, MSE = 280.63. The interaction between sex and sample failed to reach statistical significance, $F(1, 122) = 2.72$, p = .09, MSE = 280.63. When the data were sorted by sex results indicated that there was a marginally significant effect of sample, such that the year 5 females (M = 14.74°, SD = 18.39) outperformed the adult females (M = 24.75°, SD = 19.59), $F(1, 64) = 3.84$, p = .054, MSE = 430.63. However, the year 5 males (M = 6.47°, SD = 10.08) and the adult males did not significantly differ from one another, $F(1, 58) = .003$, p = .957, MSE = 115.11. When the data were sorted by sample results indicated that the adult males significantly outperformed the adult females, $F(1, 62) = 16.03$, p = .000, MSE = 327.99 (see Figure 30).

The following analyses coded the responses for males and females according to the total percentage of external and internal responses for all four views. The total percentage of external responses by sex, test time, and the adult sample can be found in Table 18. The total percentage of internal responses by sex, test time, and the adult sample can be found in Table 19.

The following analyses coded the responses according to the percentage of male participants and the percentage of female participants that made all external and all

Table 17

Mean Degrees from Horizontal for the Combined 45°/135° Angle Condition of the Water Level Task by Sex, Test Time, and the Adult Sample.

Test Time	Males	Females	p-value
Year 1	22.64(20.09)	31.29(19.03)	0.087
Year 2	13.48(17.23)	22.94(19.11)	0.047
Year 3	10.76(13.90)	22.63(20.84)	0.012
Year 4	7.63(12.28)	15.80(18.00)	0.046
Year 5	6.47(10.08)	14.74(18.39)	0.038
Adult Sample	6.63(11.26)	24.75(23.00)	0.000

Values presented are means with standard deviations in parentheses.

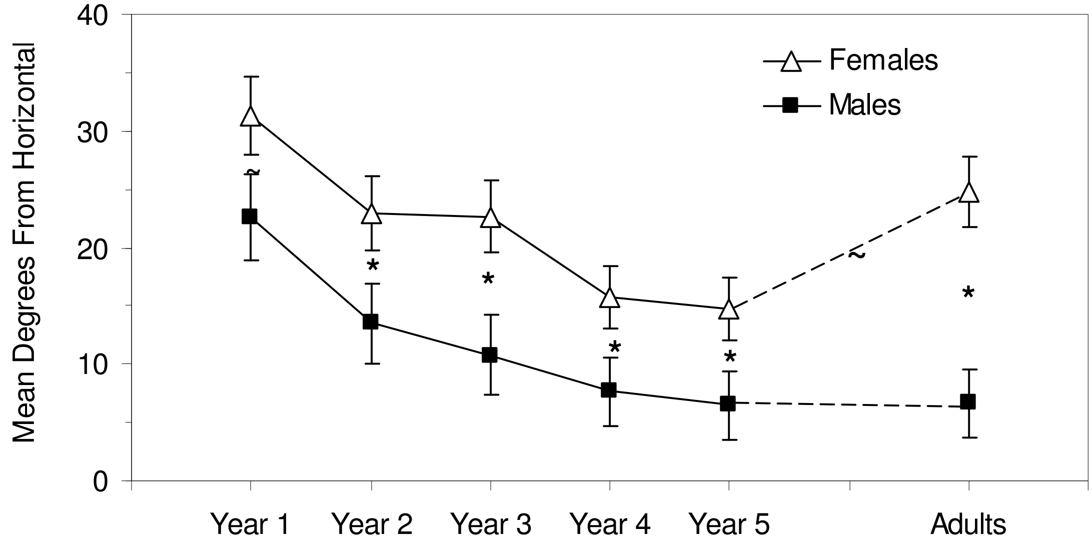


Figure 30. Mean Degrees from Horizontal for the Combined 45°/135° Angle Condition of the Water Level Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences and the ~ symbol indicates marginally significant differences.

Table 18

Total Percentage of External Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex, Test Time, and the Adult Sample.

Test Time	**Males**	**Females**
Year 1	61.0%	42.0%
Year 2	79.0%	59.5%
Year 3	88.0%	61.0%
Year 4	92.0%	73.0%
Year 5	94.0%	74.5%
Adult Sample	92.0%	61.0%

Table 19

Total Percentage of Internal Responses for the Combined 45°/135° Angle Condition of the Water Level Task by Sex, Test Time, and the Adult Sample

Test Time	**Males**	**Females**
Year 1	29.0%	51.0%
Year 2	18.0%	31.0%
Year 3	8.0%	27.0%
Year 4	7.0%	23.0%
Year 5	6.0%	24.0%
Adult Sample	7.0%	26.0%

internal responses for all four 45°/135° angle views. The percentage of participants that made all external responses can be found in Table 20 by sex, test time, and the adult sample. The percentage of participants that made all internal responses can be found in Table 21 by sex, test time, and the adult sample.

Means and standard deviations for the 90° angle condition of the Water Level Task are presented in Table 22 by sex, test time, and the adult sample. The 90° angle condition also confirmed predictions as males ($\underline{M}$ = 7.01°, $\underline{SD}$ = 23.01) significantly outperformed females ($\underline{M}$ = 24.64°, $\underline{SD}$ = 40.21), $\underline{F}$(1, 122) = 8.77, $\underline{p}$ = .004, $\underline{MSE}$ = 1117.15. The year 5 sample ($\underline{M}$ = 16.12°, $\underline{SD}$ = 34.55) and the adult sample ($\underline{M}$ = 16.37°, $\underline{SD}$ = 34.11) did not differ significantly from one another, $\underline{F}$(1, 122) = .03, $\underline{p}$ = .857, $\underline{MSE}$ = 1117.15. The interaction between sex and sample failed to reach statistical significance, $\underline{F}$(1, 122) = .01, $\underline{p}$ = .939, $\underline{MSE}$ = 1117.15. When the data were sorted by sex results indicated that the year 5 females ($\underline{M}$ = 23.90°, $\underline{SD}$ = 40.19) did not significantly differ from the adult females ($\underline{M}$ = 25.44°, $\underline{SD}$ = 40.86), $\underline{F}$(1, 64) = .02, $\underline{p}$ = .878, $\underline{MSE}$ = 1641.62. Also, as predicted the year 5 males ($\underline{M}$ = 6.68°, $\underline{SD}$ = 23.54) and the adult males ($\underline{M}$ = 7.30°, $\underline{SD}$ = 22.91) did not significantly differ from one another, $\underline{F}$(1, 58) = .011, $\underline{p}$ = .918, $\underline{MSE}$ = 538.42. When the data were sorted by sample results indicated that the adult males significantly outperformed the adult females, $\underline{F}$(1, 62) = 4.80, $\underline{p}$ = .032, $\underline{MSE}$ = 1097.18 (see Figure 31).

The main effect of sex in the 180° angle condition failed to reach statistical significance, $\underline{F}$(1, 122) = 1.56, $\underline{p}$ = .214, $\underline{MSE}$ = 32.60. The main effect of sample also failed to reach statistical significance, $\underline{F}$(1, 122) = 1.72, $\underline{p}$ = .193, $\underline{MSE}$ = 32.60. The interaction between sex and sample also failed to reach statistical significance,

Table 20

Percentage of Participants in the Combined 45°/135° Angle Condition of the Water Level Task Making All External Responses by Sex, Test Time, and the Adult Sample.

Test Time	**Males**	**Females**
Year 1	57.0%	32.0%
Year 2	75.0%	47.0%
Year 3	82.0%	44.0%
Year 4	89.0%	65.0%
Year 5	89.0%	71.%
Adult Sample	91.0%	50.0%

Table 21

Percentage of Participants in the Combined 45°/135° Angle Condition of the Water Level Task Making All Internal Responses by Sex, Test Time, and the Adult Sample.

Test Time	**Males**	**Females**
Year 1	29.0%	41.0%
Year 2	18.0%	29.0%
Year 3	7.0%	24.0%
Year 4	7.0%	23.0%
Year 5	4.0%	23.0%
Adult Sample	6.0%	22.0%

Table 22

Mean Degrees from Horizontal for the 90° Angle Condition of the Water Level Task by Sex, Test Time, and the Adult Sample.

Test Time	**Males**	**Females**	**p-value**
Year 1	25.50(40.25)	39.82(43.35)	0.186
Year 2	16.29(35.01)	25.28(40.15)	0.356
Year 3	8.14(24.63)	26.74(41.46)	0.041
Year 4	6.43(23.60)	21.19(38.74)	0.083
Year 5	6.68(23.54)	23.90(40.19)	0.050
Adult Sample	7.30(22.91)	25.44(40.86)	0.032

Values presented are means with standard deviations in parentheses.

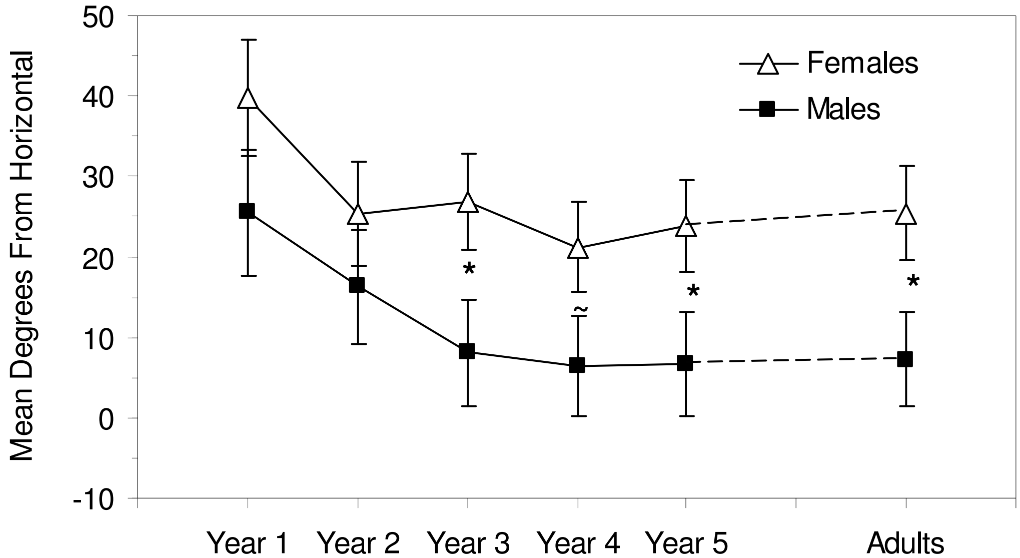


Figure 31. Mean Degrees from Horizontal for the 90° Angle Condition of the Water Level Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences and the ~ symbol indicates marginally significant differences.

F(1, 122) = .98, p = .325, MSE = 32.60. When the data were sorted by sex results indicated that the year 5 females (M = .24°, SD = .55) did not significantly differ from the adult females (M = .56°, SD = .99), F(1, 64) = 2.82, p = .098, MSE = .63 Also, as predicted the year 5 males (M = .50°, SD = 1.11) and the adult males (M = 2.84°, SD = 11.22) did not significantly differ from one another, F(1, 58) = 1.21, p = .276, MSE = 67.88. When the data were sorted by sample results indicated that the adult males did not significantly differ from the adult females, F(1, 62) = 1.31, p = .256, MSE = 63.45.

Flags Task. A 2 (sex) x 2 (sample: year 5 and adults) between-subjects factorial ANOVA was conducted on the mean number of correct responses out of 48 trials. Means and standard deviations for the Flags Task are presented in Table 23 by sex, test time, and the adult sample.

As predicted, a significant main effect of sex was found such that males (M = 44.80, SD = 3.76) correctly identified more trials than females (M = 41.38, SD = 6.72), F(1, 128) = 8.76, p = .004, MSE = 27.07. A significant main effect of sample was also found, such that the year 5 sample (M = 44.32, SD = 4.74) correctly identified more trials than the adult sample (M = 41.67, SD = 6.38), F(1, 128) = 15.46, p = .000, MSE = 27.07 (see Figure 32). A significant interaction was also found such that adult females (M = 38.66, SD = 7.11) scored significantly lower than all other groups, F(1, 128) = 7.39, p = .007, MSE = 27.07 (see Figure 33).

When the data were sorted by sex results indicated that the year 5 males (M = 44.91, SD = 3.91) and the adult males (M = 44.69, SD = 3.66) did not significantly differ from one another, F(1, 62) = .053, p = .818, MSE = 14.32. However, the year 5 females (M = 43.81, SD = 5.38) significantly outperformed the adult females (M = 38.66,

Table 23

Mean Number of Correct Responses on the Flags Task by Sex, Test Time, and the Adult Sample.

Test Time	Males			Females		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	38.73	9.87	30	35.06	10.16	36
Year 2	42.33	6.70	30	40.06	7.34	36
Year 3	43.20	6.35	30	40.94	7.44	36
Year 4	44.50	4.32	30	43.11	5.92	36
Year 5	44.93	4.03	30	43.81	5.38	36
Adult Sample	44.69	3.66	32	38.66	7.11	32

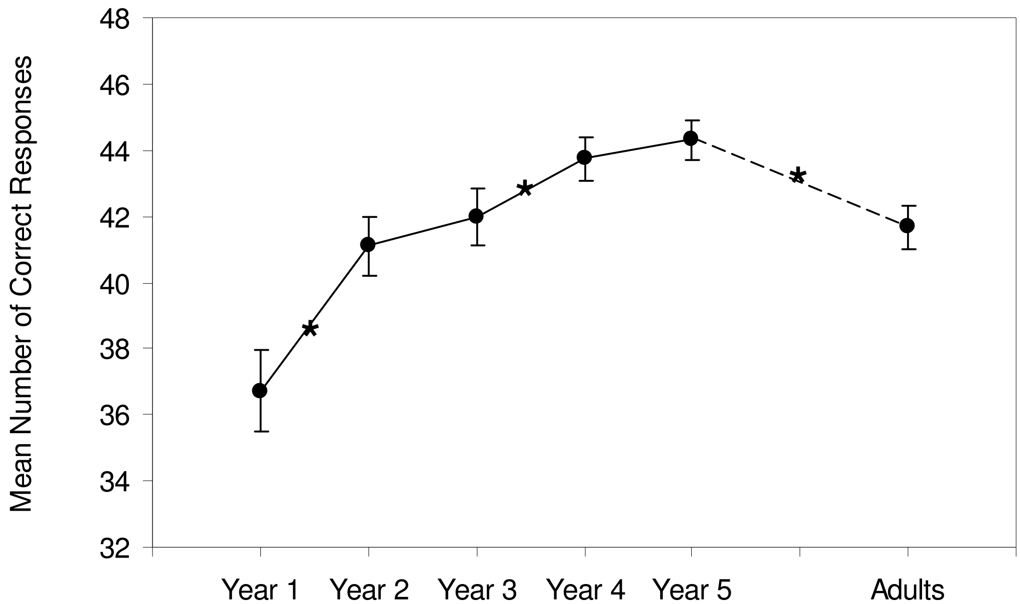


Figure 32. Mean Number of Correct Responses on the Flags Task by Test time and by the Year 5 Sample versus the Adult Sample. Asterisks indicate significant differences between groups. The dashed line represents statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

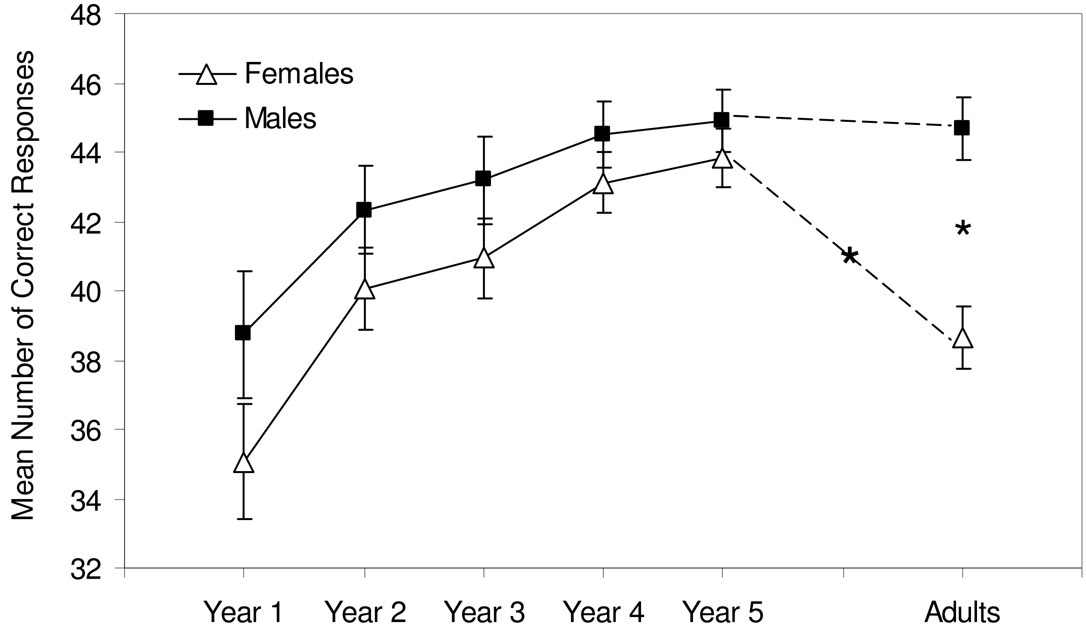


Figure 33. Mean Number of Correct Responses on the Flags Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

SD = 7.11), $F(1, 66) = 11.51$, $p = .001$, MSE = 39.04. Analyses comparing the adult sample revealed that the adult males (M = 44.69, SD = 3.66) significantly outperformed the adult females, $F(1, 62) = 18.22$, $p = .000$, MSE = 31.94. The analyses comparing the year 5 males with the adult females revealed that the year 5 males significantly outperformed the adult females, $F(1, 62) = 19.01$, $p = .000$, MSE = 32.87. The year 5 females and the adult males did not significantly differ from one another, $F(1, 66) = .61$, $p = .438$, MSE = 21.61.

Basketball Man Task. Separate regressions were performed for each participant in the adult sample and in the year 5 sample on the front (three-dimensional) and the back (two-dimensional) rotation conditions. The slopes, intercepts, R^2 values, and error rates from these regressions were also averaged for each participant in the adult sample and in the year 5 sample to form a combined front and back score.

Slopes obtained from these regression analyses were used to obtain estimates of the increase in reaction time as a function of rotation angle (0°, 45°, 135°, 180°). The slopes, intercepts, R^2 values, and error rates for the front, the back, and the combined conditions were then entered into a 2 (sex) x 2 (sample: year 5 and adults) between-subjects factorial ANOVA.

Participants with error rates greater than .2 were omitted from the slope, intercept, and R^2 value analyses. Mean scores for the participants that were included in the present analyses can be found in Table 24 (year 5 sample) and Table 25 (adult sample). Only significant findings are reported in the following analyses.

A main effect of sample was found in the front condition slopes, such that the adult sample (M = .005, SD = .007) had significantly steeper slopes than the year 5

Table 24

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for Year 5

Year 5	Females	Males

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.003	1.147	0.280	0.101		0.003	1.272	0.243	0.056
SD		0.003	0.600	0.204	0.175		0.003	0.577	0.257	0.068
N		19	19	19	28		18	18	18	24

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.003	0.811	0.296	0.110		0.003	0.877	0.305	0.038
SD		0.002	0.222	0.185	0.190		0.001	0.213	0.185	0.070
N		21	21	21	28		18	18	18	24

Note: The mean scores presented represent individuals that participated in all five years of the study. The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

Table 25

Mean Slopes, Intercepts, R^2 values, and Error Rates (%) in both the 2-D and 3-D Rotation Conditions for the Adult Sample

Adults	Females	Males

Front View (3-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.007	1.580	0.305	0.058		0.004	1.347	0.227	0.065
SD		0.009	0.761	0.237	0.070		0.004	0.692	0.194	0.840
N		25	25	25	26		30	30	30	32

Back View (2-D)

		Slope	Intercept	R^2	Errors		Slope	Intercept	R^2	Errors
Mean		0.006	1.234	0.284	0.533		0.004	1.162	0.254	0.278
SD		0.005	0.421	0.201	0.105		0.003	0.524	0.214	0.061
N		22	22	22	25		32	32	32	33

Note: The mean scores for the slopes, intercepts, and R^2 values represent individuals that had error rates less than .20.

sample (M = .003, SD = .004), $F(1, 115) = 5.61$, $p = .020$, MSE = .000. A significant interaction was also found, $F(1, 115) = 4.19$, $p = .043$, MSE = .000. When the data were sorted by sex, results indicated that the adult females (M = .007, SD = .009) had significantly steeper slopes than the year 5 females (M = .002, SD = .003), $F(1, 54) = 6.36$, $p = .015$, MSE = .000 (see Figure 34). The interaction also revealed that the adult males (M = .004, SD = .004) had significantly steeper slopes than the year 5 females, $F(1, 59) = 4.86$, $p = .031$, MSE = .000.

A main effect of sample was found in the front condition intercepts, such that the adult sample (M = 1.45, SD = .73) had significantly longer reaction times that the year 5 sample (M = 1.19, SD = .54), $F(1, 115) = 5.38$, $p = .022$, MSE = .404. When the data were sorted by sex, results indicated a significant effect of sample such that the adult females (M = 1.58, SD = .76) had longer reaction times than the year 5 females (M = 1.22, SD = .59), $F(1, 54) = 3.98$, $p = .050$, MSE = .450 (see Figure 35).

A main effect of sex was found in the back condition slopes such that the females (M = .005, SD = .004) had significantly steeper slopes than the males (M = .003, SD = .003), $F(1, 113) = 9.80$, $p = .002$, MSE = .000. A main effect of sample was also found such that the adult sample (M = .005, SD = .004) had significantly steeper slopes than the year 5 sample (M = .003, SD = .003), $F(1, 113) = 8.75$, $p = .004$, MSE = .000. When the data were sorted by sex, results revealed that the adult females (M = .006, SD = .005) had significantly steeper slopes than the year 5 females (M = .004, SD = .003), $F(1, 51) = 5.43$, $p = .024$, MSE = .000. When the data were sorted by sample, results revealed that the adult females also had significantly steeper slopes than the adult males (M = .004, SD = .003), $F(1, 52) = 5.44$, $p = .024$, MSE = .000 (see Figure 36).

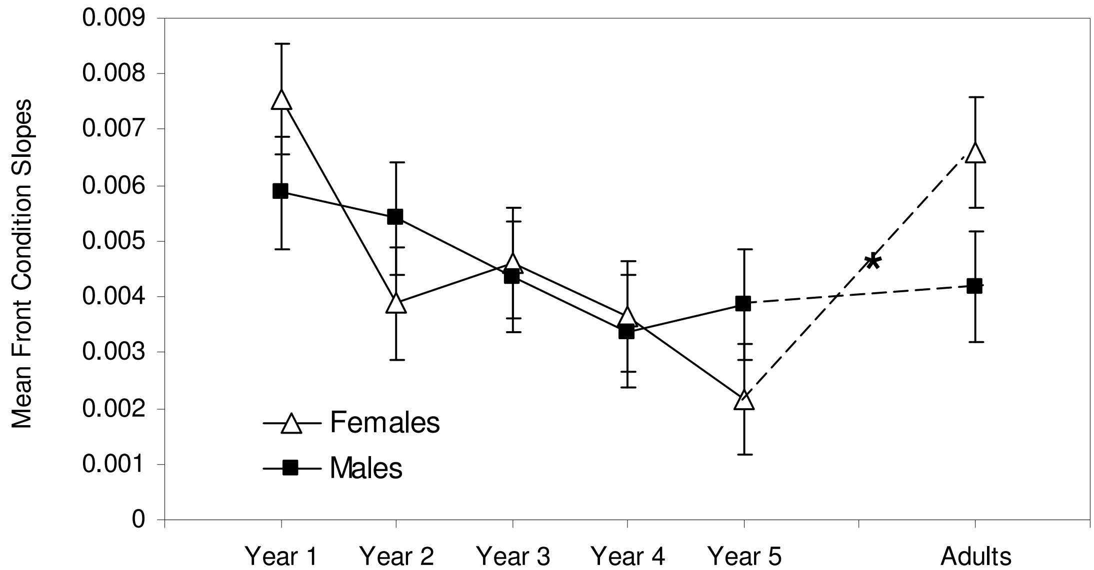


Figure 34. Mean Front (3-dimensional) Slopes on the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

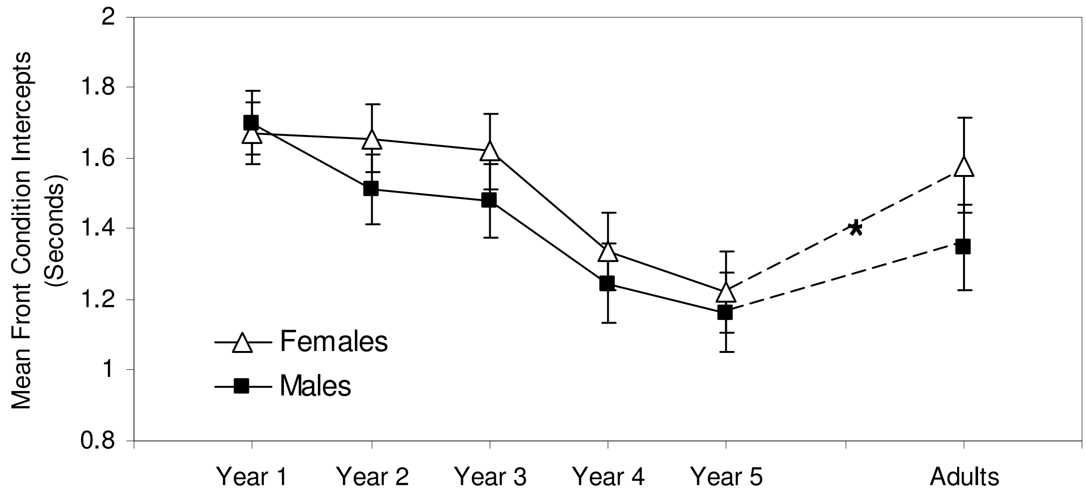


Figure 35. Mean Front (3-dimensional) Intercepts for the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

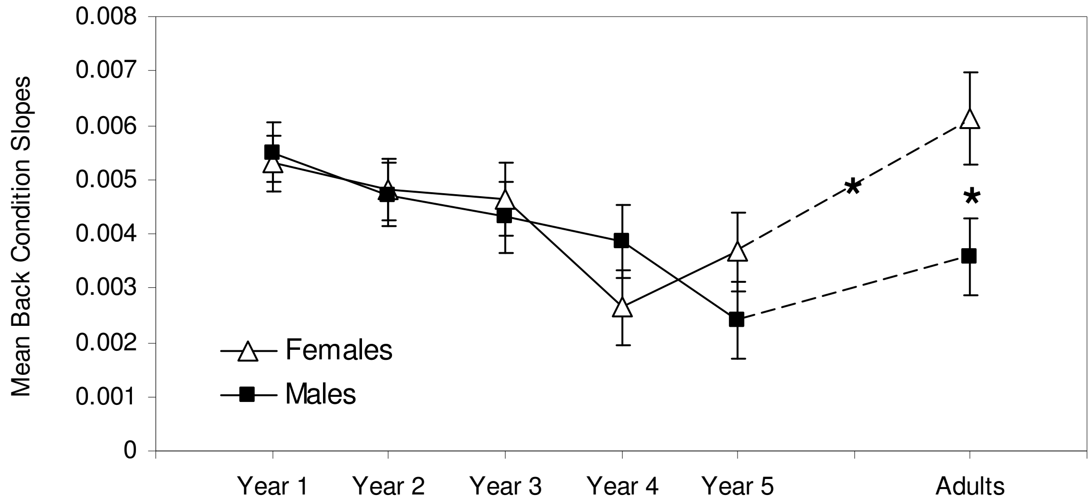


Figure 36. Mean Back (2-dimensional) Slopes for the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

A main effect of sample was found in the back condition intercepts such that the adult sample ($M = 1.19$, $SD = .48$) had significantly longer reaction times than the year 5 sample ($M = .88$, $SD = .27$), $F(1, 113) = 19.44$, $p = .000$, $MSE = .146$. When the data were sorted by sex, results indicated that the adult females ($M = 1.23$, $SD = .42$) had significantly longer reaction times than the year 5 females ($M = .82$, $SD = .23$), $F(1, 51) = 21.00$, $p = .000$, $MSE = .104$. The adult males ($M = 1.16$, $SD = .52$) also had significantly longer reaction times than the year 5 males ($M = .94$, $SD = .30$), $F(1, 62) = 4.27$, $p = .043$, $MSE = .182$ (see Figure 37).

The combined (front and back) slope condition revealed a main effect of sex such that the females ($M = .005$, $SD = .004$) had significantly steeper slopes than the males ($M = .003$, $SD = .003$), $F(1, 112) = 6.43$, $p = .013$, $MSE = .000$. A significant main effect of sample was also found such that the adult sample ($M = .005$, $SD = .004$) had significantly steeper slopes than the year 5 sample ($M = .003$, $SD = .003$), $F(1, 112) = 10.78$, $p = .001$, $MSE = .000$. A significant interaction was also found such that the adult females ($M = .007$, $SD = .005$) had significantly steeper slopes than all other groups, $F(1, 112) = 4.59$, $p = .034$, $MSE = .000$. When the data were sorted by sex results revealed that the adult females had significantly steeper slopes than the year 5 females ($M = .003$, $SD = .003$), $F(1, 52) = 9.73$, $p = .003$, $MSE = .000$. The adult analyses revealed that the adult females had significantly steeper slopes than the adult males ($M = .004$, $SD = .003$), $F(1, 51) = 6.90$, $p = .011$, $MSE = .000$ (see Figure 38). The interaction also revealed that the adult females had significantly steeper slopes than the year 5 males ($M = .003$, $SD = .003$), $F(1, 53) = 11.44$, $p = .001$, $MSE = .000$.

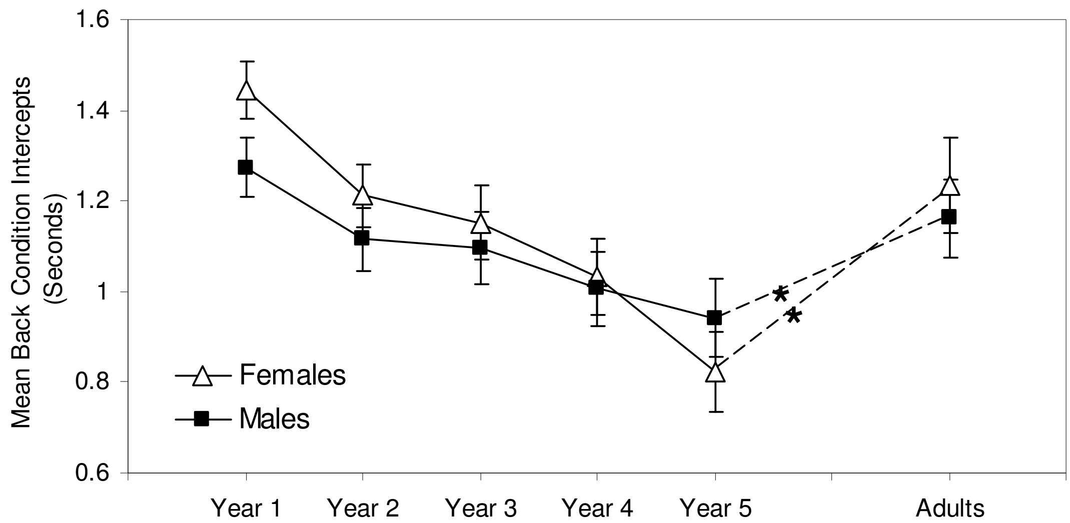


Figure 37. Mean Back (2-dimensional) Intercepts for the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

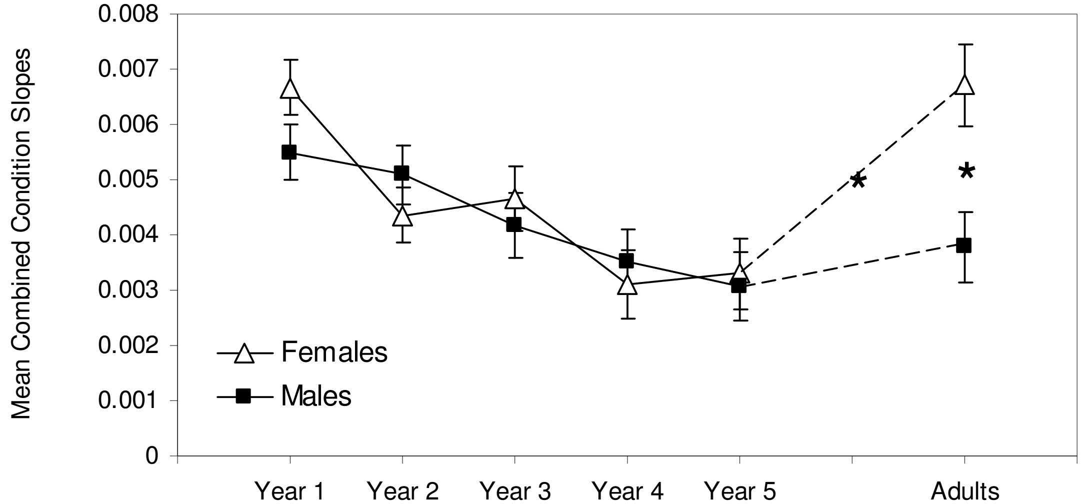


Figure 38. Mean Combined (Front and Back) Condition Slopes for the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

The combined (front and back) intercept condition revealed a significant main effect of sample such that the adult sample (M = 1.30, SD = .53) had significantly longer reaction times than the year 5 sample (M = 1.03, SD = .35), $F(1, 112) = 11.63$, $p = .001$, MSE = .196. When the data were sorted by sex results revealed that the adult females (M = 1.37, SD = .44) had significantly longer reaction times than the year 5 females (M = 1.00, SD = .36), $F(1, 52) = 11.54$, $p = .001$, MSE = .156 (see Figure 39).

Female-Sensitive Task

Purdue Pegboard. For each condition, a 2 (sex) x 2 (sample: year 5 and adults) between-subjects factorial ANOVA was conducted on the mean number of correctly inserted pegs. Means and standard deviations for each condition are presented in Tables 26-29 by sex, test time, and the adult sample.

The main effect of sex in the dominant hand condition failed to reach statistical significance, $F(1, 124) = 3.07$, $p = .08$, MSE = 3.13. However, a significant main effect of sample was found such that year 5 (M = 15.28, SD = 1.68) correctly inserted more pegs than the adult sample (M = 14.44, SD = 1.94), $F(1, 124) = 7.05$, $p = .009$, MSE = 3.13 (see Figure 40). A significant interaction was also found such that adult males (M = 13.82, SD = 1.76) scored significantly lower than all other groups, $F(1, 124) = 5.41$, $p = .022$, MSE = 3.13. When the data were sorted by sex results indicated that the year 5 females (M = 15.20, SD = 1.71) and the adult females (M = 15.10, SD = 1.94) did not significantly differ from one another, $F(1, 64) = .053$, $p = .819$, MSE = 3.32. However, the year 5 males (M = 15.38, SD = 1.66) significantly outperformed the adult males (M = 13.82, SD = 1.76), $F(1, 60) = 12.84$, $p = .001$, MSE = 2.93. The analyses from the adult sample revealed that the adult females significantly outperformed the adult males,

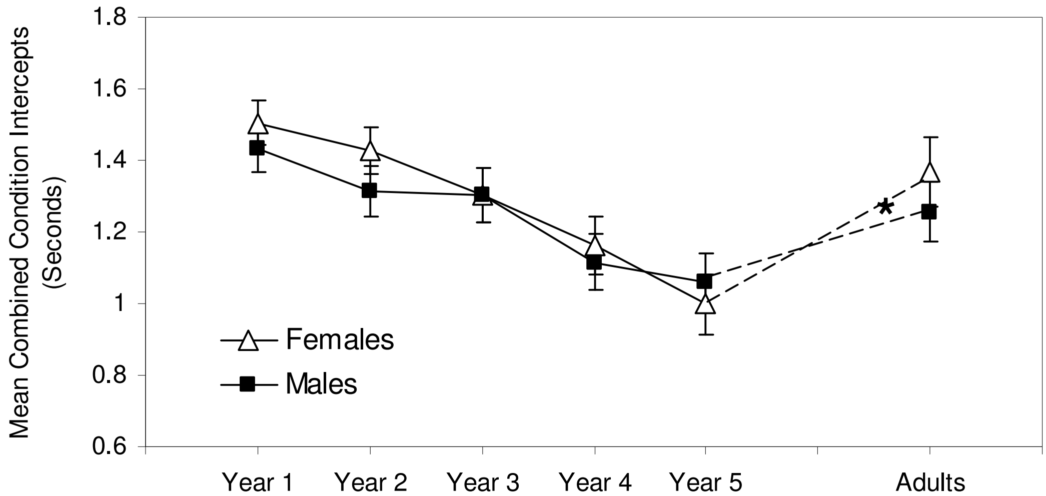


Figure 39. Mean Combined (Front and Back) Condition Intercepts for the Basketball Man Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

Table 26

Mean Number of Correctly Inserted Pegs for the Dominant Hand Condition of the Purdue Pegboard Task by Sex, Test Time, and the Adult Sample.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	13.83	1.87	29	13.49	1.74	35
Year 2	14.72	1.62	29	14.46	1.67	35
Year 3	14.86	1.75	29	14.74	1.48	35
Year 4	15.45	1.43	29	15.49	1.34	35
Year 5	15.38	1.66	29	15.20	1.71	35
Adult Sample	13.82	1.76	33	15.10	1.94	31

Table 27

Mean Number of Correctly Inserted Pegs for the Non-Dominant Hand Condition of the Purdue Pegboard Task by Sex, Test Time, and the Adult Sample.

Test Time	Males			Females		
	M	SD	N	M	SD	N
Year 1	13.24	1.27	29	13.29	1.49	35
Year 2	14.34	1.74	29	13.54	1.44	35
Year 3	13.97	1.74	29	14.23	1.59	35
Year 4	14.79	1.35	29	14.40	1.48	35
Year 5	15.07	2.87	29	14.77	1.40	35
Adult Sample	13.30	1.57	33	14.10	1.44	31

Table 28

Mean Number of Correctly Inserted Pegs for the Both Hands Condition of the Purdue Pegboard Task by Sex, Test Time, and the Adult Sample.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	21.48	3.30	29	22.46	2.75	35
Year 2	23.10	3.19	29	23.06	3.37	35
Year 3	23.97	3.12	29	23.91	2.09	35
Year 4	24.69	2.59	29	24.83	2.62	35
Year 5	24.48	3.45	29	24.34	2.75	35
Adult Sample	22.55	3.26	33	23.71	3.35	31

Table 29

Mean Number of Correctly Inserted Pegs for the Assembly Condition of the Purdue Pegboard Task by Sex, Test Time, and the Adult Sample.

Test Time	**Males**			**Females**		
	M	**SD**	**N**	**M**	**SD**	**N**
Year 1	33.34	6.99	29	33.34	5.27	35
Year 2	35.69	5.76	29	34.63	5.29	35
Year 3	38.31	4.54	29	36.66	6.10	35
Year 4	38.34	5.11	29	39.89	5.13	35
Year 5	39.34	4.92	29	40.06	7.73	35
Adult Sample	37.45	5.15	33	37.74	4.91	31

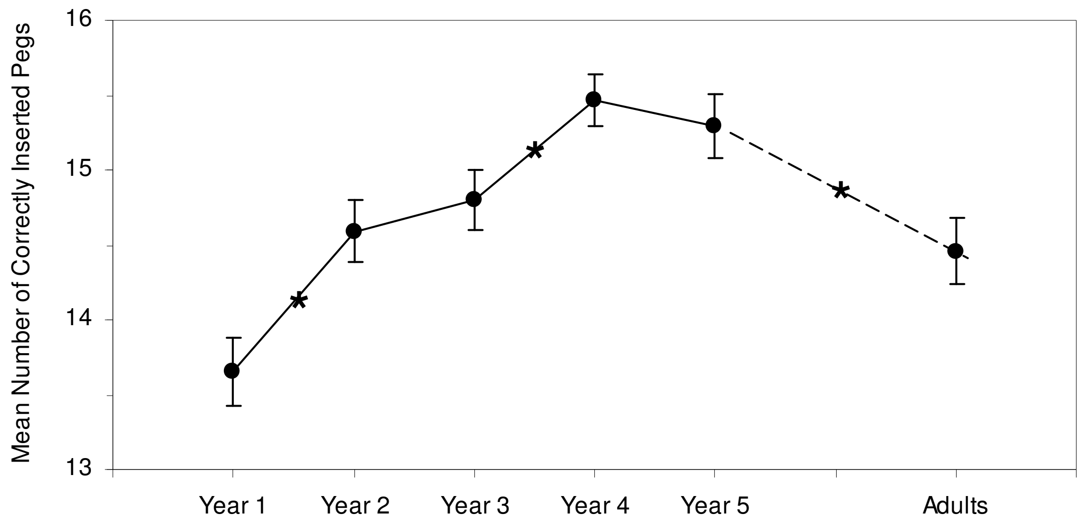


Figure 40. Mean Number of Correctly Inserted Pegs in the Dominant Hand Condition of the Purdue Pegboard Task by Test Time and by the Year 5 Sample versus the Adult Sample. Asterisks indicate significant differences between groups. The dashed line represents statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

F(1, 62) = 7.66, p = .007, MSE = 3.41 (see Figure 41). The year 5 females inserted significantly more pegs than the adult males, F(1, 66) = 10.78, p = .002, MSE = 3.01. The year 5 males and the adult females did not significantly differ from one another, F(1, 58) = .37, p = .548, MSE = 3.27.

There was not a significant main effect of sex in the non-dominant hand condition, F(1, 124) = .56, p = .46, MSE = 3.53. However, a significant main effect of sample was found such that year 5 (M = 14.91, SD = 2.17) correctly inserted more pegs than the adult sample (M = 13.69, SD = 1.55), F(1, 124) = 13.43, p = .000, MSE = 3.53 (see Figure 42). The interaction between sex and sample failed to reach statistical significance, F(1, 124) = 2.69, p = .104, MSE = 3.53. When the data were sorted by sex the results revealed a marginally significant effect of sample, such that the year 5 females (M = 14.77, SD = 1.40) outperformed the adult females (M = 14.10, SD = 1.45), F(1, 64) = 3.72, p = .058, MSE = 2.01. A significant effect of sample was found for the males, such that the year 5 males (M = 15.07, SD = 2.87) significantly outperformed the adult males (M = 13.30, SD = 1.57), F(1, 60) = 9.35, p = .003, MSE = 5.15. The analyses from the adult data revealed that the adult females significantly outperformed the adult males, F(1, 62) = 4.41, p = .040, MSE = 2.29 (see Figure 43).

The main effect of sex failed to reach statistical significance in the both hands condition, F(1, 124) = .82, p = .37, MSE = 10.22. However, a significant main effect of sample was found, such that the year 5 sample (M = 24.41, SD = 3.06) correctly inserted more pegs than the adult sample (M = 23.11, SD = 3.33), F(1, 124) = 5.15, p = .025, MSE = 10.22 (see Figure 44). The interaction between sex and sample failed to reach statistical significance, F(1, 124) = 1.32, p = .252, MSE = 10.22. When the data were

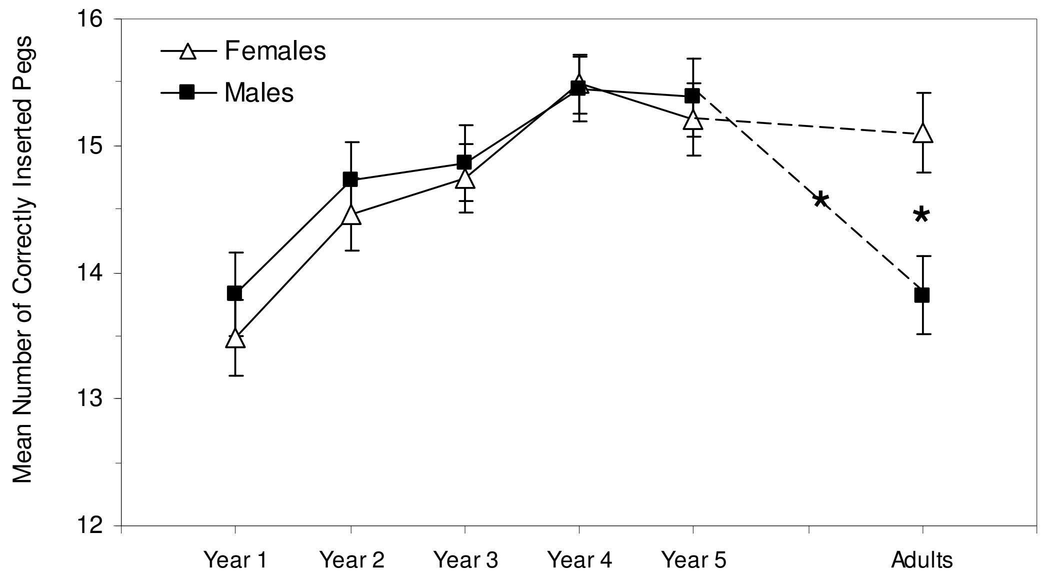


Figure 41. Mean Number of Correctly Inserted Pegs for the Dominant Hand Condition of the Purdue Pegboard Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

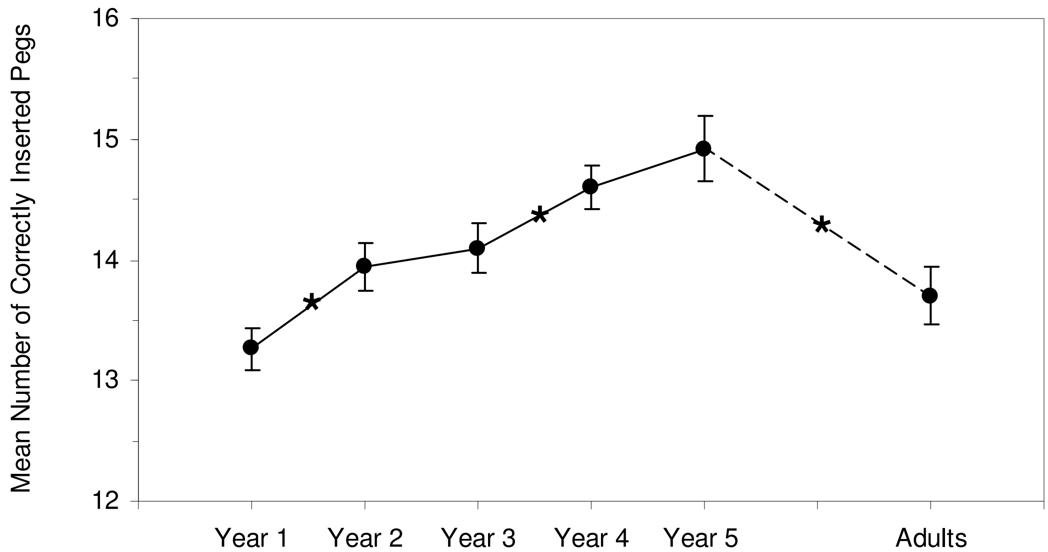


Figure 42. Mean Number of Correctly Inserted Pegs for the Non-Dominant Hand Condition of the Purdue Pegboard Task by Test Time and by the Year 5 Sample versus the Adult Sample. Asterisks indicate significant differences between groups. The dashed line represents statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

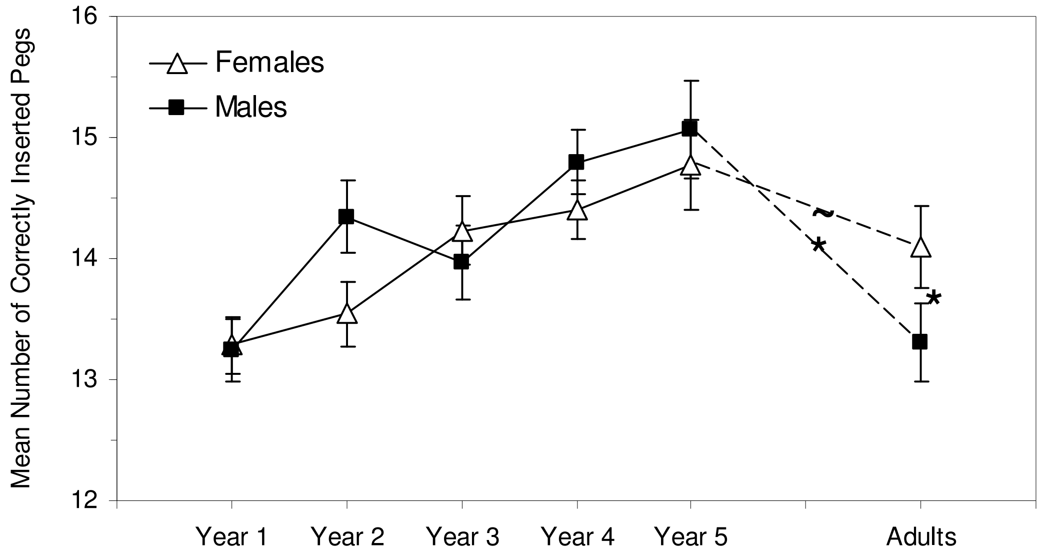


Figure 43. Mean Number of Correctly Inserted Pegs in the Non-Dominant Hand Condition of the Purdue Pegboard Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups and the ~ symbol indicates marginally significant differences between groups.

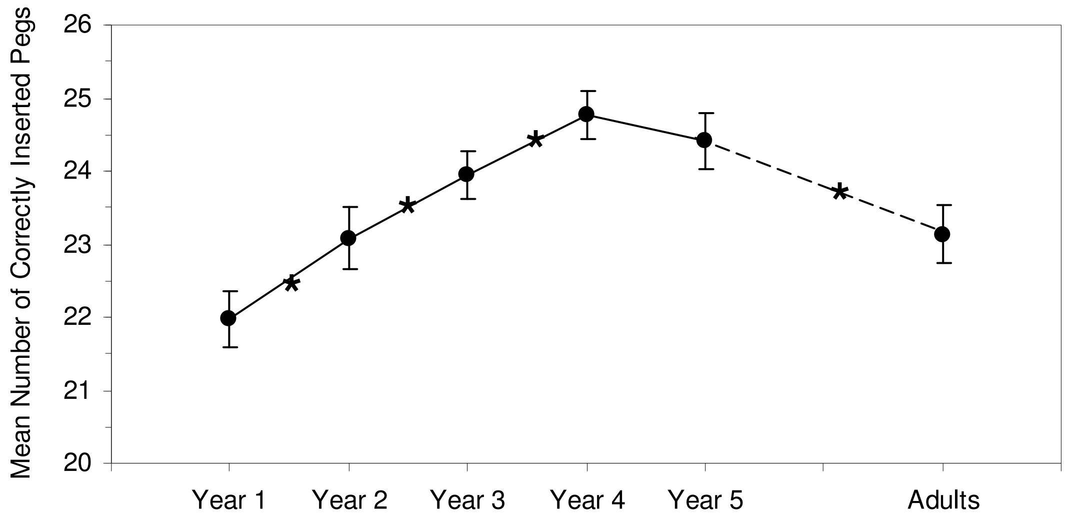


Figure 44. Mean Number of Correctly Inserted Pegs in the Both Hands Condition of the Purdue Pegboard Task by Test Time and by the Year 5 Sample versus the Adult Sample. Asterisks indicate significant differences between groups. The dashed line represents statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

sorted by sex results revealed that the year 5 females (M = 24.34, SD = 2.75) and the adult females (M = 23.71, SD = 3.35) did not significantly differ from one another, $F(1, 64) = .71$, $p = .403$, MSE = 9.29. However, the year 5 males (M = 24.48, SD = 3.45) did significantly outperform the adult males (M = 22.55, SD = 3.26), $F(1, 60) = 5.16$, $p = .027$, MSE = 11.22. The adult analyses revealed that the adult males and the adult females did not significantly differ from one another, $F(1, 62) = 1.99$, $p = .164$, MSE = 10.91 (see Figure 45).

The main effect of sex failed to reach statistical significance in the assembly condition, $F(1, 124) = .23$, $p = .630$, MSE = 34.54. However, a significant main effect of sample was found such that the year 5 sample (M = 39.73, SD = 6.57) correctly inserted more pegs than the adult sample (M = 37.59, SD = 5.00), $F(1, 124) = 4.08$, $p = .046$, MSE = 34.54 (see Figure 46). The interaction between sex and sample failed to reach statistical significance, $F(1, 124) = .04$, $p = .839$, MSE = 34.54. When the data were sorted by sex results revealed that the year 5 females (M = 40.06, SD = 7.73) and the adult females (M = 37.74, SD = 4.91) did not significantly differ from one another, $F(1, 64) = 2.05$, $p = .158$, MSE = 43.09. The year 5 males (M = 39.34, SD = 4.91) and the adult males (M = 37.45, SD = 5.14) also did not significantly differ from one another, $F(1, 60) = 2.17$, $p = .146$, MSE = 25.41. When the adult data were analyzed results revealed that the adult females and the adult males did not significantly differ from one another, $F(1, 62) = .05$, $p = .820$, MSE = 25.36 (see Figure 47).

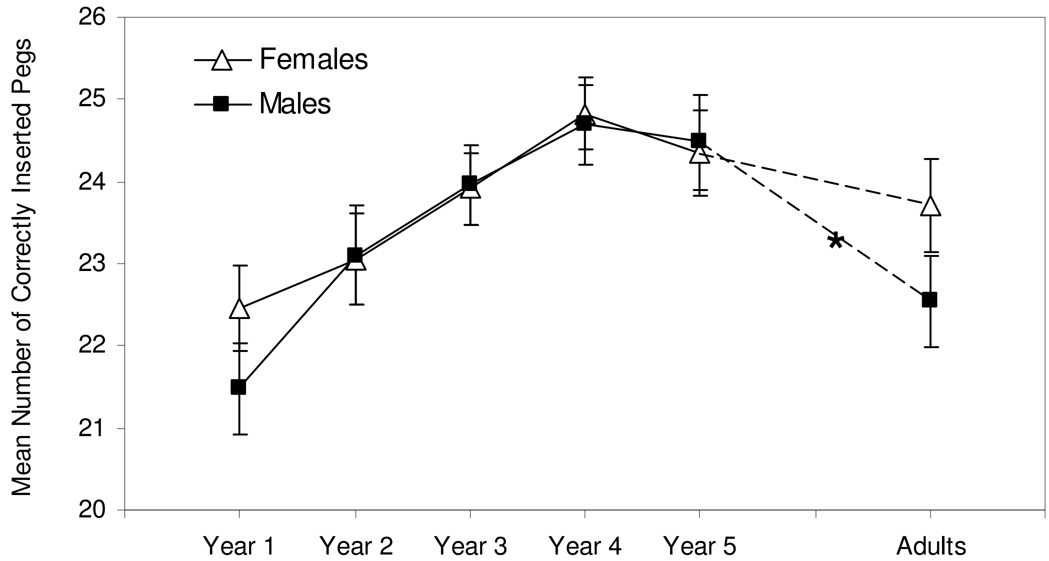


Figure 45. Mean Number of Correctly Inserted Pegs in the Both Hands Condition of the Purdue Pegboard Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons). Asterisks indicate significant differences between groups.

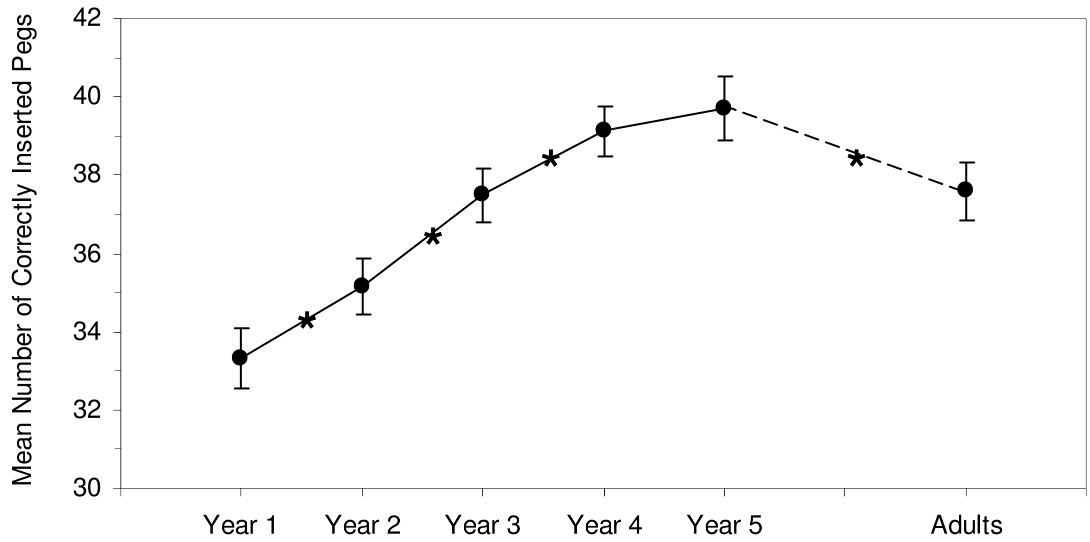


Figure 46. Mean Number of Correctly Inserted Pegs in the Assembly Condition of the Purdue Pegboard Task by Test Time and by the Year 5 Sample versus the Adult Sample. Asterisks indicate significant differences between groups. The dashed line represents statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

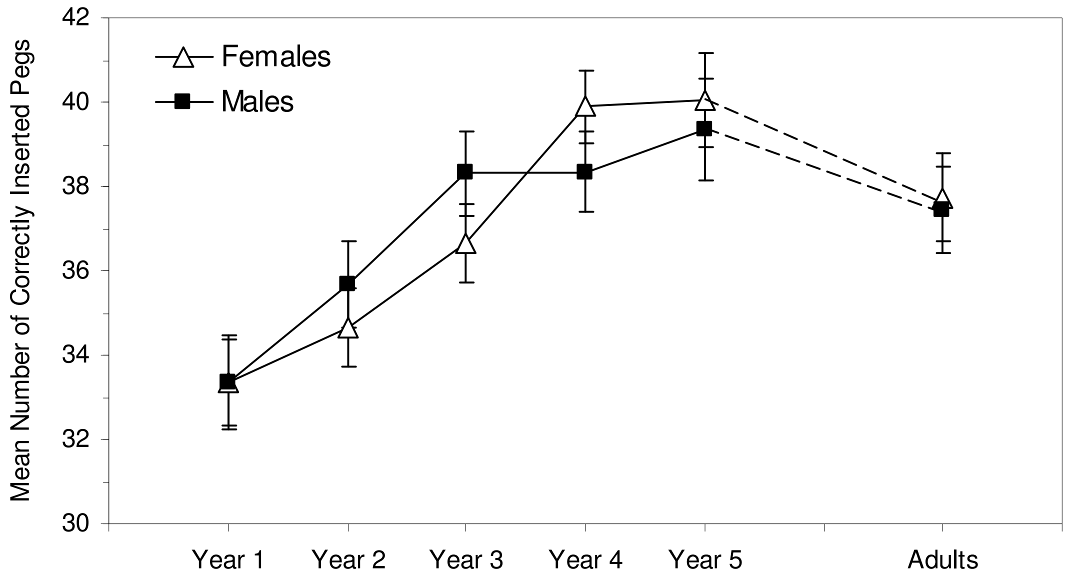


Figure 47. Mean Number of Correctly Inserted Pegs in the Assembly Condition of the Purdue Pegboard Task by Sex, Test Time, and the Year 5 Sample versus the Adult Sample. The dashed lines represent statistical comparisons of only the year 5 sample versus the adult sample (between-subjects comparisons).

DISCUSSION

It was hypothesized that sex differences would emerge and increase throughout the five years of this study, with males being superior on the Water Level Task, the Flags Task, and the Basketball Man Task and females being superior on the Purdue Pegboard Task. Sex differences were also hypothesized to exist in each of the cognitive tasks for the adult sample. Additionally, performance on each of the cognitive tasks was hypothesized to be comparable for the year 5 males and the adult sample males. Likewise, performance on each of the cognitive tasks was hypothesized to be comparable for the year 5 females and the adult sample females. Discussion of these hypotheses will be organized by task.

Male Sensitive Tasks

Water Level Task

The combined 45°/135° angle condition and the 90° angle condition revealed a main effect of test time such that increases in task performance were found across the five years of the study. These increases can be attributed to either cognitive maturation or practice effects. While this study cannot definitely rule out either hypothesis; practice effects are unlikely since participants were not provided with any feedback at any time regarding their performance. Without feedback there was no way for participants to know if they were improving each successive test year. Additionally, the one year interval between test sessions makes practice effects unlikely.

A marginally significant effect of sex emerged in year 1 in the combined 45°/135° condition, such that males outperformed females. This sex effect was significant across the remaining four test years in the combined 45°/135° condition. A significant effect of sex emerged in the 90° angle condition in year 3 and year 5 and was marginally significant in year 4. These findings are consistent with other studies that found consistent sex differences after adolescence for tasks measuring spatial perception (see Linn & Peterson, 1985; Maccoby and Jacklin, 1974; Voyer, Voyer, & Bryden, 1995 for review).

Voyer, Voyer, and Bryden’s (1995) meta-analyses suggested that sex differences in spatial perception increase with age. These findings are also consistent with the present analyses. As shown in Figure 30 and Figure 31, the sex differences found in the combined 45°/135° condition and in the 90° condition in the adult sample were larger than the differences found in the adolescent sample.

Perhaps the most interesting findings of the Water Level Task was the bimodal distribution that revealed one cluster of scores consisting of almost no errors (very close to horizontal) and another cluster of scores consisting of many errors with an average of 45° from horizontal in the combined 45°/135° angle condition. This distribution led to the conclusion that one group of participants relied on the correct external context (horizontal table) to make their decision on where to draw the water level line and another group of participants used the internal context (45° top and bottom of the glass jar) when deciding where to draw the water level line. These findings are consistent with another study (Thomas, Lohaus, & Kessler, 1999) that found that 3 different samples of children (ages 8-16 years old) in a three year longitudinal study performed the water level task according to 1 of 3 latent (strategy) states: (a) bottom-parallel responders, (b) random responders, or (c) accurate responders. Thomas et al. stated that bottom-parallel responders drew the water level line approximately parallel to the glass jar’s bottom, largely independent of the angle tilt of the glass jar. The present study termed these respondents as "internal responders" because it is believed that they either relied on the internal context to make their response or they ignored the tilt of the jar altogether. Thomas et al. also termed another group of respondents as "accurate responders," which would be synonymous with the group that was labeled "external responders" in the present study. These individuals, as previously stated, produce lines that are very close to horizontal. Finally, Thomas et al. termed another group of respondents as "random responders." According to Thomas et al. these respondents perform poorly on the Water Level Task and they show large variability in their responses. Analyses regarding random respondents will be reserved for a future paper.

Analyses discussing the total percentage of external and internal responses made and the analyses discussing the percentage of participants making all external (4/4) responses and all internal (4/4) responses revealed that both of these percentages were accurately predicted by test time, with the R^2 values ranging from 89-96%. These high R^2 values indicate that the power trendlines found in Figures 8-11 fit the data extremely well. The power trendlines found in Figure 8 and Figure 10 are important because they reveal that both sexes were increasing in their total percentage of external responses and the that the percentage of male and female participants making all external responses was also increasing across time. The power trendlines found in Figure 9 and in Figure 11 are also important because they reveal that both sexes were decreasing in their total percentage of internal responses and that the percentage of male and female participants making all internal responses was also decreasing across time.

Both analyses discussing the percentage of external responses revealed that both sexes increased at the same rate regarding these percentages. However, analyses comparing the intercepts of the regression lines revealed that the total percentage of external responses made by males and the percentage of male participants making all external responses was significantly greater than the percentage of external responses made by females and the percentage of female participants making all external responses, in the first test year. These results indicate that males started off making significantly more external responses and all external responses than the females, and this significant difference in external response making continued across the five years of the study.

Additionally, as can be seen in Figure 8 the trendline for the percentage of external responses made by males is approaching about 95-100%, whereas the female trendline is approaching about 75-80% of external responses. Since the slopes do not significantly differ from one another it can be inferred that males will remain in upper 95-100%, while the females will level off and stay in the 75-80% range. These ranges are mostly confirmed for males in the adult sample. Table 18 shows that 92% of adult male responses were external. The percentage of external responses made by adult females decreases from the expected ranges. Table 18 shows the adult females made 61% of external responses. In summary, adolescent males seem to increase in their "correct" responses across time, meaning that they are decreasing in their "incorrect" percentage of internal or other responses (only about 0-5% of their decisions). Adolescent females also increase their percentage of "correct" responses; however, they still seem to make about 20-25% "incorrect" internal or other responses.

The same pattern can be found in the percentage of participants making all external responses. Figure 10 shows that the percentage of male participants making all external responses is approaching about 90-95%, while the percentage of female participants making all external responses is approaching about 70-75%. Once again, the slopes did not differ so it can be inferred that the percentage of male participants making all external responses will remain in the upper 90-95% range, while the percentage of female participants making all external responses will level off and stay in the 70-75% range. These ranges are confirmed for males in the adult sample. Table 20 shows that the percentage of adult males making all external responses is 91%. The percentage of adult females making all external responses decreases from the expected ranges. Table 20 shows the percentage of adult females making all external responses is 50%. The combined results suggest that most males will probably master the "correct" response as they mature, while some females will probably continue to make the "incorrect" internal or other responses, even after allowing for maturation time.

With respect to the total percentage of internal responses, the analyses revealed that males decreased in their percentage of internal responses at a faster rate than the females. This is important because it shows that females are not learning that internal responses are not the correct response as fast as males. Analyses comparing the intercepts of the regression lines also revealed that females made significantly more total internal responses than males, in the first test year. This is important because it shows that females start off making significantly more internal responses than males and continue making significantly more internal responses than males across the five years of the study. Figure 9 shows that males are approaching about 0-5% of total internal the study. Figure 9 shows that males are approaching about 0-5% of total internal responses, while females still remain in the 20-25% range. These ranges are mostly confirmed in the adult sample. Table 19 shows that 7% of adult male responses were internal and 26% of adult female responses were internal.

When the analyses was analyzed according to the percentage of participants making all internal responses, results indicated that the percentage of male participants making all internal response did not significantly differ from the percentage of female participants in the first test year. However, analyses comparing the slopes revealed that the percentage of male participants making all internal responses decreased at a faster rate than the percentage of female participants making all internal responses. This analysis suggests that the percentage of male and female participants making all internal responses starts out similar, but then the percentage of male participants making all internal responses decreases at a faster rate then the females. As shown in Figure 11, the percentage of male participants making all internal responses is approaching 0-5%, while the percentage of females making all internal decisions still remains in the 20-25% range. Again, these ranges are mostly confirmed in the adult sample. Table 21 shows that percentage of adult male participants making all internal responses is 6%, whereas the percentage of adult female participants making all internal responses is 22%.

These findings suggest that the overall sex differences found in the Water Level Task may be mostly due to males and females using different contexts (external vs. internal) when making their responses. It is also important to note that task directions or instructions are an unlikely explanation of the lower performance of the females. For example, Thomas and Jamison (1975) provided detailed visual instructions to both males and females regarding the Water Level Task. Each participant watched as an experimenter rotated a real glass jar filled half way with colored water. This allowed the participants to visually see the water level remain horizontal while the bottle was tilted at various angles. Subsequent sex differences were found even after these detailed visual instructions were given. In that study, the male participants showed linear improvements and by the seventh and eighth grade the adolescent males were performing comparable to college men. The performance of the females did not show any clear pattern and the college age women as a group did no better on the task than many female groups much younger than them. Thomas and Jamison (1975) also tested the understanding of the task by examining written responses of college males and females. The experimenters asked the college males and females two important questions: (a) How did you know where to draw the water level line? and (b) What is the principle or the idea which determines where the water goes in the bottle? Results indicated that more females than males were judged to perform poorly on the task based on their written responses to the questions. Additionally, Thomas, Jamison, and Hummel (1973) performed two experiments on college women who were judged not to know the concept of horizontal water. These women were given tasks which were designed to elicit the concept of horizontal; however, they still did not learn the concept.

Flags Task

A main effect of test time was found such that increases in performance were found across test years. This finding can be attributed to cognitive maturation or practice effects. While this study cannot definitely rule out either hypothesis; practice effects are unlikely since participants were not provided with any feedback at any time regarding their performance. Without feedback there was no way for participants to know if they were improving each successive test year. Additionally, the one year interval between test sessions makes practice effects unlikely. The possibility of practice effects will be discussed in detail later in the discussion.

Although sex differences were not found in the Flag Task across the five years of the study, there was a trend in the expected direction with males averaging higher scores across all five test years. Interestingly, a previous study in our lab (Cummings, 2003), found significant effects of sex in each year when the data were analyzed as between subjects in years 1-3 (see Figure 48). In this study, each year was analyzed separately meaning that every participant that was tested each year was included in the analyses. Therefore, the previous analyses had an increase in the overall sample size. While only 66 participants were used in the current analysis, the previous analysis included 147 participants in year 1, 113 participants in year 2, and 97 participants in year 3. It is important to note that the previous analyses and the current analyses revealed highly similar means for males and females in each year (see Figure 49), therefore, the previous study's significant sex effect is attributed to the increase in sample size. Thus, it is believed that a significant sex differences would have been found in this study had the sample size been larger.

As predicted, the present analysis comparing the year 5 sample with the adult sample revealed that the year 5 males and the adult males did not differ significantly from one another. However, an interaction was revealed such that adult females were performing significantly lower than all other test conditions. This finding may reflect differences in task instructions given to the adolescent and the adult populations.

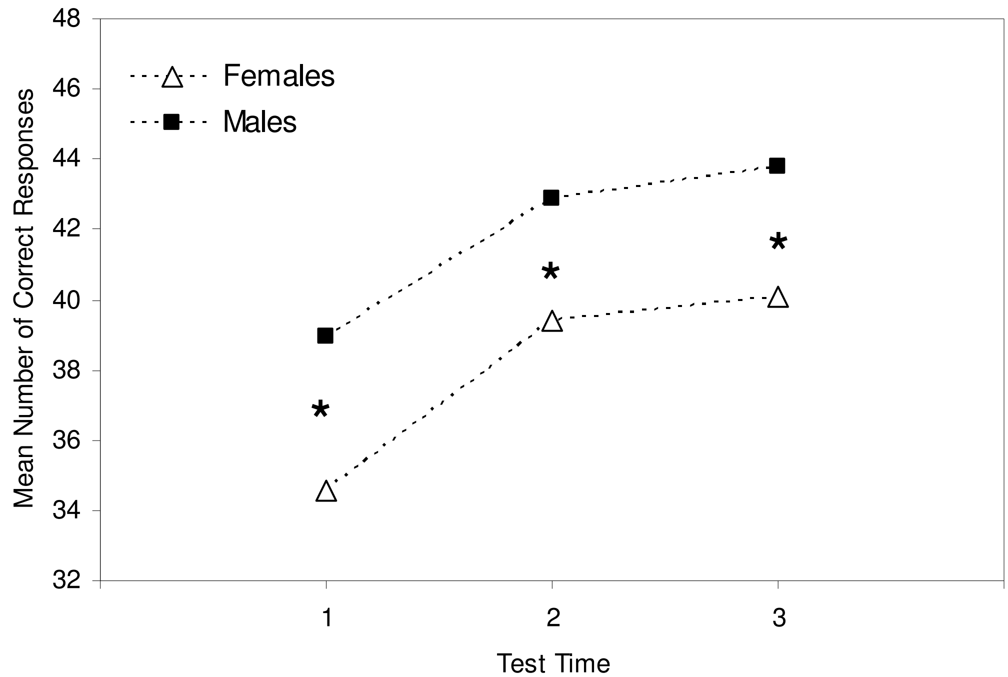


Figure 48. Mean Number of Correct Responses on the Flags Task by Sex and Test Time in a Previous Study (Cummings, 2003). Data was analyzed as between-subjects, meaning that every participant that was tested each year was included in the analyses (year 1, N = 147; year 2, N = 113; year 3, N = 97). Asterisks indicate significant differences between sexes.

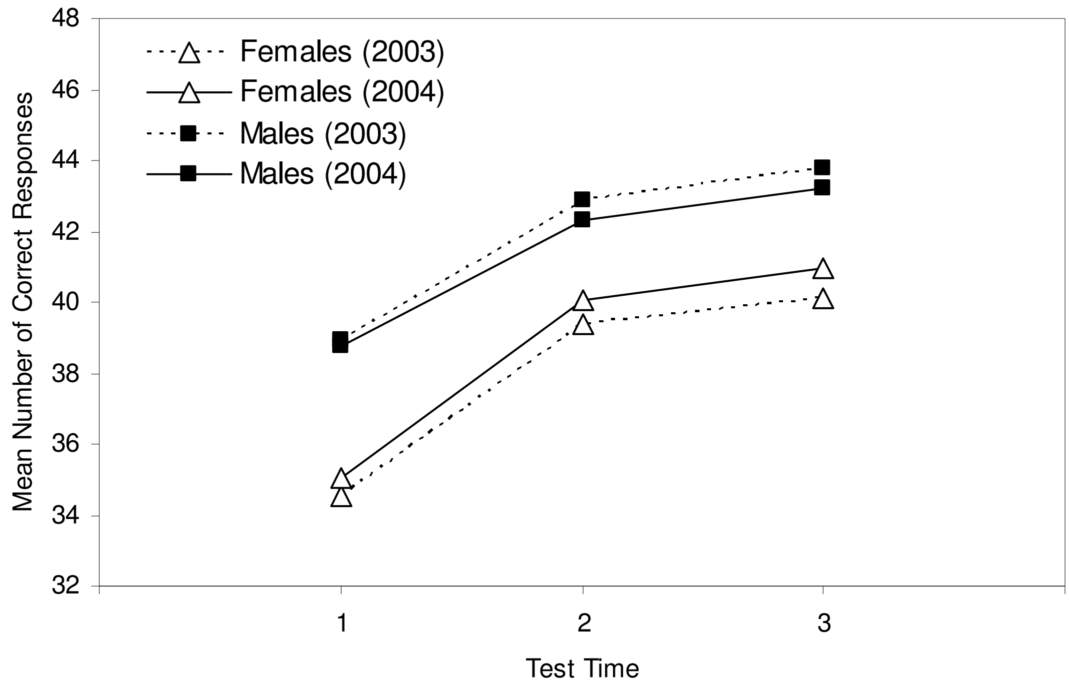


Figure 49. Mean Number of Correct Responses on the Flags Task by Sex and Test Time in a Previous Study (Cummings, 2003) and in the Present Study (Hill, 2004). Cummings (2003) year 1, N = 147; year 2, N = 113; year 3, N = 97. Hill (2004) N = 66 for year 1, year 2, and year 3.

It is important to note that the entire adult sample was tested before the adolescent sample. Therefore, the adult sample's performance provided a type of baseline for male and female performance. After the adult sample was tested, pilot testing began on the 7^{th} grade adolescents. Pilot results from the adolescent sample revealed a floor effect in the mean number of correct responses for this task. Experimenter observation and pilot results indicated that the adolescents did not fully understand the task. Therefore, the decision was made to provide more detailed instructions to the adolescent sample in order to correctly measure their mental rotation abilities. Paper cut-outs of the same sides of the flags and of the mirror images of the flags were constructed (see Figure 2). The adolescent participants were instructed to physically rotate these cutout flags, while the instructions were read to them. When the paper cut-outs were rotated the participants were able to visually see the difference between the same side flags and the mirror image flags. These detailed instructions eliminated the previous floor effect. Therefore, instructions using the paper cut-outs were incorporated into all five years of this study with the adolescent participants.

One possible explanation of the low performance of the adult females is that they may have not understood the task. Therefore, the more detailed instructions may have been beneficial to the adult females in the same way they were beneficial to the adolescent sample. These more detailed instructions could have been provided to the adults; however, at the time it was not apparent that the adult females did not understand the task. It was not until the females in year 3 of the adolescent sample began significantly outperforming the adult sample that this became apparent. Future research will test the same type of adult population using the more detailed instructions in order to test this hypothesis.

As previously mentioned, it is also possible that practice effects affected the results for the adolescents. The superior performance found in the year 5 adolescent sample when compared with the adult sample could also be attributed to some sort of practice effect. The results indicate that both male and female adolescent participants benefited from the brief (3 minute) exposure they received once a year on the Flags Task. It is not possible to determine exactly what aspect of behavior benefited from the exposure (i.e. motivation, task familiarity, attention, or perception). However, the data do indicate that the mean number of correct responses increased across the five years of the study for both males and females.

Exposure to the task also seems to differentially affect females' performance. While the year 5 males and the adult males did not significantly differ from one another, the adolescent females outperformed the adult females. Additionally, the adult males significantly outperformed the adult females, while sex differences were not found in the adolescent sample. Therefore, the results indicate that task exposure aided the adolescent females performance. If the year 5 females would have had only one test session, like the adult females, it can be reasoned that sex differences would have been found. In summary, task exposure seems to be benefiting females to a greater extent than males.

As previously stated, this study cannot fully explain why the adolescent females improved; however, these findings deserve further research attention as they have important implications.

Basketball Man Task

No statistically significant sex differences were found in the Basketball Man Task across the five years of this study. The lack of significant sex differences could be due to the small sample size that was the result of first excluding participants that did not participate in all five years of the study and the result of excluding the remaining participants with error rates greater than 20%. Participants with error rates greater than 20% were omitted from the analyses of slope, intercept, and R^2 data. This exclusion procedure is standard in modern mental rotation/reaction time research because it is not statistically appropriate to compare participants who not correctly mentally rotating (error rates greater than 20%), with the participants that are correctly mentally rotating (error rates less than 20%). This stringent exclusion criteria was not reported in the earlier traditional studies that showed significant sex differences in mental rotation tasks (see Linn & Peterson, 1985; Voyer, Voyer, & Bryden, 1995). Therefore, the exclusion criteria used in the present study may explain the lack of significant sex differences.

Analyses of the slopes and the intercepts revealed a main effect of test time in all three test conditions (i.e. front, back, combined front and back). Figures 18-21 show that rotation efficiency increased (slopes decreased) and that reaction times (intercepts) decreased across the five years of the study.

Again, interesting results were found when the year 5 sample and the adult sample were compared. Analyses of the slopes revealed an effect of group, such that the adults had significantly steeper slopes (less efficient) than the year 5 sample in all three test conditions. Additionally, the adult females had significantly steeper slopes (less efficient) than the year 5 females in all three test conditions. The same results emerged in the reaction time (intercept) analyses. Analyses on the intercepts revealed an effect of group, such that the adult sample had significantly slower reaction times than the year 5 sample in all three test conditions. Additionally, the adult females had significantly slower reaction times the year 5 females in all three test conditions.

The superior performance found in the year 5 adolescent sample when compared with the adult sample could not be attributed to cognitive maturation, because the mean age of the adolescent sample was 16.26 and the mean age of the adult sample was 20.23. Therefore, the adults should have had higher levels of brain (cognitive) maturation in the frontal, temporal, and parietal lobes (see Giedd et al., 1999) when compared with the year 5 adolescent sample.

The superior performance found in the year 5 adolescent sample could possibly be attributed to some sort of practice effect. The results indicate that both male and female adolescent participants benefited from the brief (5 minute) exposure they received once a year on the Basketball Man Task. It is not possible to determine exactly what aspect of behavior benefited from the exposure (i.e. motivation, task familiarity, attention, or perception). However, the data do indicate that speed of responding and speed of rotation increased across test years, as reaction time (intercept) decreased and slope decreased (efficiency increased).

Exposure to the task seems to differentially affect females' performance. While the adolescent males and the adult males only differed once (back slope condition), the adolescent females outperformed the adult females in all three slope conditions and all three reaction time conditions. Additionally, sex differences found in the back and combined (front and back) slope conditions of the adult sample were not found in the adolescent sample. Therefore, the results indicate that task exposure aided the adolescent females' performance (decreased slope). If the year 5 females would have had only one test session, like the adult females, it can be reasoned that sex differences would have been found in the slopes. In summary, task exposure seems to be benefiting females to a greater extent than males.

As previously stated, this study cannot fully explain why the females improved; however, these findings deserve further research attention as they have important implications.

Female-Sensitive Task

Purdue Pegboard

While adult females have been consistently found to outperform males on this manual dexterity task (Epting & Overman, 1998; Hampson & Kimura, 1992) there were no sex differences found among the adolescents in the four test conditions across the five years of this study. However, there were interesting findings in the between-subjects analysis comparing the year 5 adolescents with the adult sample. The year 5 sample consistently outperformed the adult sample in all four test conditions.

The year 5 adolescent sample was tested on the Purdue Pegboard in 2002, while the adult sample was tested on the Purdue Pegboard in 1999. These test dates may be critically important from the standpoint of opportunity to practice sophisticated manual dexterity video/computer games. The five year period preceding each of the test times was the 1998-2002 period for the year 5 adolescent sample versus the 1995-1999 period for the adult sample. These five year periods vastly differ in the advancements of technology. For example, the recent increase in the popularity, affordability, and technology of video/computer games and video/computer game controllers may be responsible for the year 5 sample consistently outperforming the adult sample.

One study (Drew & Waters, 1986) found that the videogame play significantly improved Purdue Pegboard scores in an elderly population. Purdue Pegboard scores were measured for a control group and an experimental group at the first testing session. Then the experimental group was trained (2 times/week for 1 ½ hours) for 2 months on an arcade-type videogame. After a two month interval, both groups were tested again. The experimental group demonstrated significant improvements on their Purdue Pegboard scores, while the control group showed no improvements. In the present study, if the year 5 sample did in fact have greater exposure to video/computer games and more advanced controllers, then this may have led to greater manual dexterity skills than the adult sample.

Although, this study cannot address this hypothesis or separate which variable is responsible for the increase in performance, this hypothesis is worth further investigation.

CONCLUSIONS

Performance improved for both sexes on all of the cognitive tasks across the five years of this study. These findings can be attributed to cognitive maturation and/or practice effects. While this study cannot rule out either hypothesis, the possible practice effects on the two mental rotation tasks (the Flags Task and the Basketball Man Task) seem to have important implications for females. Significant sex differences, with males outperforming females, were found in both mental rotation tasks in the adult sample; however, these differences were not found in the adolescent sample. These results indicate that task exposure somehow aided the adolescent females' performance. If these increases in performance were to generalize to other mental rotation tasks, then brief exposure/practice to these tasks may increase females' spatial skills, having important implications for the career fields of mathematics, science, and engineering because these fields require spatial skills and are all underrepresented by females (Halpern, 2000).

Predicted sex differences emerged and remained consistent across the five years of the study in the spatial perception Water Level Task. Perhaps the most interesting findings of the Water Level Task was the bimodal distribution that revealed one cluster of scores consisting of almost no errors (very close to horizontal) and another cluster of scores consisting of many errors with an average of 45° from horizontal in the combined 45°/135° angle condition. These findings suggest that the overall sex differences found in the Water Level Task may be mostly due to males and females using different reference contexts (external vs. internal) when making their responses.

Finally, significant differences were found in the year 5 sample and the adult sample in the manual dexterity Purdue Pegboard Task. As previously stated, these two samples may have differed in practice with sophisticated manual dexterity video/computer games. The recent increase in the popularity, affordability, and technology of video/computer games and video/computer game controllers may be responsible for the year 5 sample consistently outperforming the adult sample.

All of these findings have important implications and deserve further research attention. It is essential that we truly precisely the nature of these differences so that we can provide additional or different types of education that benefit both males and females of all ages.

REFERENCES

Adams, R. P., & Bradbard, M. R. (1985). Sex differences in preschool children's tactual Exploration of novel and familiar objects with parents and strangers. *The Journal of Genetic Psychology, 146(4),* 567-569.

Bayer, L. M., & Bayley, N. (1959). *Growth diagnosis.* Chicago: University of Chicago Press.

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

Benbow, C. P., & Lubinski, D. (1993). Consequences of gender differences in mathematical reasoning ability and some biological linkages. In M. Haug, R. E. Whalen, C. Aron, & K. L. Olsen (Eds.), *The Development of Sex Differences and Similarities in Behavior* (pp. 87-110). Boston: Kluwer Academic.

Benbow, C. P., & Stanley, J. C. (1980). Sex differences in mathematical ability: Fact or artifact? *Science, 210*, 1262-1264.

Berenbaum, S. A. (1999). Effects of androgens on sex-typed activities and interests In adolescents with congenital adrenal hyperplasia. *Hormones and Behavior, 35,* 102-110.

Berenbaum, S. A., & Hines, M. (1992). Early androgens are related to childhood sex-typed toy preferences. *Psychological Science, 3,* 203-206.

Casey, B. M., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31(4),* 697-705.

Collaer, M. & Hines, M. (1995). Human behavioral sex differences: A role for gonadal

Hormones during early development? *Psychological Bulletin, 1,* 55-107.

Cummings, D. H. (2003). *The role of gonadal hormones in the emergence and maintenance of cognitive sex differences.* Unpublished master's thesis, University of North Carolina at Wilmington.

Drew, B. & Waters, J. (1986). Video games: Utilization of a novel strategy to improve perceptual motor skills and cognitive functioning in the non-institutionalized elderly. *Cognitive-Rehabilitation, 4(2),* 26-31.

Eals, M., & Silverman, I. (1994). The hunter-gatherer theory of spatial sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociology, 15,* 95-105.

Engelhard, G, Jr. (1990). Gender differences in performance on mathematics items: Evidence from the United States and Thailand. *Contemporary Educational Psychology, 15, 13-26.*

Epting, L.K. & Overman, W.H. (1998). Sex-sensitive tasks in men and women: A search for performance fluctuations across the menstrual cycle. *Behavioral Neuroscience, 112(6),* 1304-1317.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., & Rapoport, J.L. (1999). Brain development during childhood and adolescence: A longitudinal MRI study. *Nature Neuroscience, 2(10),* 861-863.

Halpern, D. F. (2000). *Sex differences in cognitive abilities.* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Halpern, D. F., & Tan, U. (2001). Stereotypes and steroids: Using a psychobiological model to understand cognitive sex differences. *Brain and Cognition, 45,* 392-414.

Hampson, E., & Kimura, D. (1992). Sex differences and influences on cognitive function in humans. In J. D. Becker, S.M. Breedlove, & D. Crews (Eds.). *Behavioral Endocrinology.* Cambridge, MA: MIT Press.

Hier, D. B., & Crowley, W. F., Jr. (1982). Spatial ability in androgen-deficient men. *New England Journal of Medicine, 306,* 1202-1205.

Hyde, J. S. (1981). How large are cognitive gender differences? A meta-analysis using w^2 and d. *American Psychologist, 36(8),* 892-901.

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

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104(1),* 53-69.

Kimura, D. (1987). Are men's and women's brains really different? *Canadian Psychology, 28(2),* 133-147.

Kimura, D. (1992). Sex differences in the brain. *Scientific America, 214,* 84-90.

Levy, J. (1976). Cerebral lateralization and spatial ability. *Behavior Genetics,6,* 171-188.

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

Maccoby, E. E. (1966). Sex differences in intellectual functioning. In E. E. Maccoby (Ed.), *The development of sex differences* (pp. 25-55). Stanford, CA: Stanford University Press.

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

University Press: Stanford.

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

Noble, C. E., Baker, B. L., & Jones, T. A. (1964). Age and sex parameters in psycho-motor learning. *Perception and Motor Skills, 19,* 935-945.

Nyborg, H. (1983). Spatial ability in men and women: Review and new theory. *Advances in Behaviour Research and Therapy, 5,* 89-140.

Peterson, A. C. (1976). Physical androgyny and cognitive functioning in adolescence. *Developmental Psychology, 12,* 533-542.

Piaget, J. & Inhelder, B. (1956). *The child's conception of space*. New York: Norton.

Ratcliff, G. (1979). Spatial thought, mental rotation, and the right cerebral hemisphere. *Neuropsychologia, 17,* 49-53.

Resnick, S. M., Berenbaum, S. A., Gottesman, I. I., & Bouchard, T. J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Developmental Psychology,22,* 191-198.

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

Skinner, P. H., & Shelton, R. L. (1985). *Speech, language, and hearing: Normal Processes and disorders* (2nd ed., p. 308). New York: Wiley.

Stage, E. K., & Karplus, R. (1981). Mathematical ability: Is sex a factor? *Science, 212,* 114.

Stein, J. F. (1994). Developmental dyslexia, neural timing and hemispheric lateralisation.

International Journal of Psychophsiology, 18, 241-249.

Tanner, J.M. (1962). *Growth at adolescence.* London: Blackwell.

The College Board. (2002, August 27). *10-year trend in SAT scores indicates increased emphasis on math is yielding results; Reading and writing are causes for concern.* Retrieved June 16, 2003, from http://www.collegeboard.com/press/article/0,3183,11752,00.html

Thomas, H. & Jamison, W. (1975). On the acquisition of understanding that still water is horizontal. *Merrill-Palmer Quarterly, 21(1),* 30-44.

Thomas, H., Jamison, W., & Hummell, D.D. (1973). Observation is insufficient for discovering that the surface of still water is invariantly horizontal. *Science, 181,* 173-174.

Thomas, H., Lohaus, A., & Kessler, T. (1999). Stability and change in longitudinal water level task performance. *Developmental Psychology, 35(4),* 1024-1037.

Tiffin, J. (1968). Purdue Pegboard. Chicago: Science Research Associates.

U.S. Department of Education, Office of Educational Research and Improvement, & National Center for Educational Statistics. (1999, September). *The NAEP 1998 Writing report for the nation and the states.* Retrieved June 24, 2003 from http://nces.ed.gov/nationsreportcard//pubs/main1998/1999462.asp

Van Goozen, S. H. M., Cohen-Kettenis, P. T, Gooren, L. J. G., Frijda, N. H., & Van De Poll, N. E. (1995). Gender differences in behaviour: Activating effects of cross-sex hormones. *Psychoneuroendocrinology, 20(4),* 343-363.

Van Goozen, S. H. M., Slabbekoorn, D., Gooren, L. J. G., Sanders, G., & Cohen-Kettenis, P. T. (2002). Organizing and activating effects of sex hormones in homosexuals transsexuals. *Behavioral Neuroscience, 116(6),* 982-988.

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

Waber, D. P. (1976). Sex differences in cognition: A function of maturation rate? *Science, 192,* 572-574.

Waber D. P. (1977). Sex differences in mental abilities, hemispheric lateralization, and rate of physical growth at adolescence. *Developmental Psychology, 13,* 29-38.

Witelson, S. F. (1976). Sex and the single hemisphere: Specialization of the right hemisphere for spatial processing. *Science, 193,* 425-427.

Witkin, H. A., Goodenough, D. R., & Karp, S. A. (1967). Stability of cognitive style from childhood to young adulthood. *Journal of Personality and Social Psychology, 7(3),* 291-300.

APPENDIX

Appendix A. Informed Consent.

INFORMED CONSENT

The following study is being conducted to examine how hormones affect learning. You will complete six (6) short paper and pencil tasks.

In addition, an experienced nurse will be giving you a finger prick in order to obtain a few drops of blood to a lab in order to measure hormone levels. Finally, you will be asked to fill out a short questionnaire.

All of the information you provide us is strictly confidential. Code numbers, not names, will be used in the data analysis.

Again, all of the information is confidential. No names will be used in the data analyses or written reports.

Your participation is completely voluntary, and you may choose to leave at any time without penalty. Thank you for your participation in this important study.

This study may not have direct benefits and applications to the immediate participants, but we hope that, in the future, what we learn by your participation will help us understand how hormones affect learning and thinking.

I have read and understand the above information. By signing this consent form, I am indicating that I am voluntarily participating in this study. I understand that I may withdraw at any time without penalty.

____

Your signature Date Witness's signature Date

Appendix B. Parent Questionnaire

Confidential Information about parents

1. Mother's date of birth……………………………………..______ mo ______yr
2. Highest level of education completed by mother? __________ years
3. Occupation of mother (please be as specific as possible):

____

____

4. Father's date of birth………………………………………_______ mo ______yr
5. Highest level of education completed by father? _____________ years
6. Occupation of father (please be as specific as specific as possible):

____

____

7. With whom does the child live? ______Mother _______Father ______Both
8. Please list child's brothers and sisters (and ages) who are living at home:

____

9. Total family income last year:

_____ Below $12,000 _____ $12,000 – $16,000

_____ $ 16,001 - $20,000 _____ $20,001 - $30,000

_____ $30,001 - $40,000 _____ $40,001 - $50,000

_____ over $50,000

THANK YOU FOR RESPONDING TO THIS QUESTIONAIRE.

PLEASE RETURN IT IN THE SELF ADDRESSED ENVELOPE TO DR. WILLIAM OVERMAN, UNC-W PSYCHOLOGY DEPARTMENT

Appendix C. Adolescent Questionnaire

Survey

1. Middle School attended: ____

2. High School attending: ____

3. Birthday: Month__________ Day__________ Year___________

4. What is your race? Please Circle.

Caucasian African-American Hispanic Asian

Native-American Other__________

5. Do you take any medications on a regular basis? Is so, please indicate.

6. Did you take any medications today, that you usually do not use, such as cough syrup or decongestant? If so, please indicate.

7. Are you diabetic or do you have any other medical conditions?

8. Height__________ Weight__________

FEMALES ONLY

9. Have you started your period? Please circle. Yes No

- If yes, when did you experience your first period? If you are not sure, please try to make an accurate guess. Month__________ Year__________
- When was the first day of your last period? Month__________ Year________
- Do you have your period every month? Please circle. Yes No
- If yes, how many days per month are you on your period? 1 2 3 4 5 6 7

10. Do you use any of the following birth control methods? Please circle.

Birth control pills Depo-Provera Norplant

Other (please indicate)____