Distribution Agreement

Katherine Boice March 21, 2016

The Effect of Stereotype Threat on Spatial and Mathematical Performance in Young Girls

by

Katherine Boice

Stella F. Lourenco

Adviser

Psychology

Stella F. Lourenco

Adviser

Jessica Barber

Committee Member

Bradley Howard

Committee Member

2016

The Effect of Stereotype Threat on Spatial and Mathematical Performance in Young Girls

By

Katherine Boice

Stella F. Lourenco

Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2016

Abstract

The Effect of Stereotype Threat on Spatial and Mathematical Performance in Young Girls

By Katherine Boice

Recent concern about the underrepresentation of women in the fields of science, technology, engineering, and mathematics (STEM) has led to research demonstrating that the threat of confirming negative gender stereotypes, or "stereotype threat," may be a contributing factor to the sex differences in STEM performance and achievement, particularly in the domains of math and space. The current study examines the role of implicit gender stereotypes on mathematical and spatial performance in 6- and 7-year-old girls. Working memory and self-reported anxiety were measured to determine potential mechanisms of stereotype threat. Working memory ability moderated stereotype threat effects, with low-working memory girls performing worse than high-working memory girls when exposed to gender stereotype threat. Anxiety scores were not significantly correlated with performance. These findings offer important insight into potential mechanisms of stereotype threat and causes of the sex differences in spatial ability and the related deficit of women in STEM fields.

The Effect of Stereotype Threat on Spatial and Mathematical Performance in Young Girls

By

Katherine Boice

Stella F. Lourenco

Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Psychology

2016

Acknowledgements

I would like to thank the members of the Spatial Cognition Laboratory at Emory for their guidance and support, as well as my committee members for their feedback. Additionally, I would like to thank the families that participated in the study.

Table of Contents

The Effect of Stereotype Threat on Spatial and Mathematical Performance in Young Girls

Spatial thinking encapsulates a wide range of skills, including spatial navigation, visualization, and memory, that are essential to understanding the world around us. For instance, spatial thinking is involved in reading road maps, using an instruction manual to construct furniture, and imagining how an object would look if rotated. Spatial thinking is also necessary for academic and vocational success in a number of fields. Wai, Lubinski and Benbow (2009) found that spatial abilities assessed at age 13 predicted attaining an undergraduate and graduate degree in the fields of science, technology, engineering, and mathematics (STEM) and predicted a career in a STEM field 20 years after initial testing. These effects held even when controlling for verbal and mathematical aptitude at age 13, indicating the specific importance of spatial abilities for long- and short-term success in STEM fields. Additionally, it has been found that improvements in spatial thinking result in increased academic success in STEM fields, such as biology (Lennon, 2002), engineering (Sorby & Baartmans, 2000), and dentistry (Hegarty, Keehner, Khooshabeh, & Montello, 2009). The connection between spatial thinking and STEM success is well-documented, but much is still unknown about the specific mechanisms that might contribute to this relationship. The current study explores these mechanisms through an investigation of individual differences in spatial thinking, working memory, and the role of gender stereotypes in the early development of spatial thinking.

Recent concern about the underrepresentation of women in STEM fields, and the well-documented relationship between STEM success and spatial skills, has sparked interest in the corresponding sex differences in spatial abilities. There are robust sex differences in spatial cognition with men typically outperforming women on spatial reasoning tasks (for meta-analysis see Voyer, Voyer, & Bryden, 1995). Even before children begin elementary school there are sex differences in spatial tasks that require visually transforming shapes (Levine, Huttenlocher, Taylor, & Langrock, 1999) and navigation (Lourenco, Addy, Huttenlocher, & Fabien, 2011). While much of the research on the sex difference in spatial abilities acknowledges that there are likely many causes, there are two common perspectives on the origin of this difference: biological and sociocultural.

Biological arguments cite hormonal sex differences beginning prior to birth that contribute to the sex difference in spatial abilities later in life. A 1995 longitudinal study by Grimshaw, Sitarenios, and Finegan showed that prenatal testosterone levels were positively associated with girls' mental rotation abilities at 7 years of age, implicating the presence of hormones in the development of spatial abilities. Spatial abilities improve rapidly in the preschool years. However, even infants show notable differences between the sexes in the recognition of spatial features. One study showed that at 5 months of age, male infants were able to recognize a mirror reversal of an image whereas female infants could not (Moore & Johnson, 2008). Additional support for a biologically-based sex difference comes from studies demonstrating that sex differences in spatial abilities are consistent across many countries. In a recent cross-national study, men outperformed women on a test of mental rotation ability in all 52 countries in which data were collected (Lippa, Collaer, & Peters, 2010). The biological correlates of spatial ability early in development, as well as the seeming universality of the sex difference demonstrated by cross-national studies, offers compelling evidence for a biological component of the sex difference in spatial ability.

Gender differences in spatial aptitude have also been linked to psychosocial factors, including spatial anxiety (Ramirez, Gunderson, Levine, & Beilock, 2012) and gender stereotypes (McGlone & Aronson, 2006). Stereotypes that men excel at spatial reasoning compared to women may induce a threat in women of confirming these negative stereotypes, leading to poorer performance in women compared to men. This phenomenon, in which stereotypes negatively impact performance, is called “stereotype threat” (ST). The effect of ST on cognitive tasks has been demonstrated for a range of social identities in a variety of domains. Most commonly this has been shown in studies on race and academic achievement, in which stereotypes about a racial identity’s intelligence can impact task performance (Steele & Aronson, 1995). The same multi-national study conducted by Lippa and colleagues (2010) that revealed a consistent male advantage on spatial tasks across countries found, counterintuitively, that greater gender equality predicted larger sex differences in mental rotation ability. The authors suggested that women in societies with greater gender equality may have more awareness of the stereotype about sex differences in spatial abilities. Subsequently, they may have a greater interest in performing well on spatial tasks on which males stereotypically do better, making them more susceptible to the negative impact of ST.

In addition to the theory that ST is driven by concerns about reinforcing negative stereotypes, working memory (WM) has also been implicated in ST susceptibility. However, there are opposing perspectives on how WM, or the ability to hold and manipulate thoughts, may impact one’s susceptibility to performance-related concerns. One perspective argues that those with lower WM capacity may be poorly equipped to ignore the influence of distracting information in general and, more specifically, threat related to stereotypes (Régner et al., 2010; Schmader, Johns, & Forbes, 2008). This results in poorer performance among those lower in WM on a range of cognitive tasks compared to those higher in WM, particularly when concerns about salient stereotypes co-opt WM ability. The opposing view is that, compared to those with low WM, those with high WM capacity rely on these abilities to solve problems, and thus are more impacted by anxieties about their performance, which co-opt WM (Ashcraft & Kirk, 2001; Beilock, 2008).

The current study addresses the potential psychosocial causes of the sex difference in spatial abilities by investigating the role of gender ST as a function of WM and anxiety. Much of the work on gender ST and spatial ability is consistent with the findings in mathematics that women's performance is negatively impacted by making gender salient (McGlone & Aronson, 2006; Spencer, Steele, & Quinn, 1999). However, meta-analyses show less success inducing ST on spatial tasks in adult populations compared to studies inducing ST in mathematics. This may be because gender-math stereotypes are more widely accepted and more widely known than gender-spatial stereotypes (Doyle & Voyer, 2015). Another explanation offered by Flore and Wicherts (2015) regards the nature of the spatial measures used. Not only are there different kinds of spatial thinking that can be measured in relation to ST, there are also many assessments that can be used to measure individual differences in spatial thinking ability. This gives us reason to believe that successfully inducing ST may depend on the type of spatial ability being assessed. Because of this, the current study uses a variety of spatial tasks, including measures of spatial transformation, visualization, and visuospatial memory.

The effect of gender ST on math performance has been demonstrated in school-aged children, though less frequently than in adults. Ambady, Shih, Kim, and Pittinsky (2001) compared performance on a math task between a group of girls in which gender identity was implicitly activated and a group of girls in which it was not. In doing this, the authors demonstrated that when gender stereotypes are made salient they negatively impact young girls' math performance. Interestingly, this pattern of results did not hold for girls ages 8 through 10 years of age, indicating that children may not be susceptible to ST consistently throughout development. There is even less evidence of successfully inducing ST on spatial tasks in developmental populations. ST has been successfully induced in spatial tasks in older children but, to our knowledge, has never been tested in those younger than 9 years of age (Neurburger, Jansen, Heil, & Quaiser-Pohl, 2015). However, Ramirez and colleagues (2012) demonstrated that children as young as 5 years of age are already susceptible to anxiety regarding spatial tasks. Success in inducing ST effects on mathematical tasks in developmental populations and on spatial tasks in older children, coupled with research indicating the presence of spatial anxiety in young children, suggests that it is possible to induce ST on spatial tasks in early school-aged children.

Given the close relationship between spatial ability and success in STEM domains, it is critical to understand how stereotypes regarding girls' performance on spatial tasks affect performance. Using early school-aged children offers a population in which we can determine when in development ST effects begin and what degree of subjective anxiety about a stereotype is necessary to elicit ST effects. This will shed light on the developmental trajectory of sociocultural influences on performance and improve our understanding of how stereotype awareness, WM, and anxiety interact to impact girls' performance.

The current study aims to answer a number of questions related to ST. First, we attempt to replicate the effects of ST on young girls' math abilities demonstrated by Ambady and colleagues (2001) using implicit ST manipulation. To our knowledge, there have been no attempts to replicate this effect in school-aged children using the implicit manipulation. Second, we extend this work to the domain of spatial reasoning. Using the same implicit ST manipulation procedure, we assess young girls' spatial abilities using a range of spatial measures. In doing this, we will not only test the existence of ST in the spatial domain, but also determine the specificity of this effect. We hypothesized that young girls would perform more poorly on spatial and mathematical tasks when their gender was made salient, compared to a control group in which gender was not made salient. Moreover, in an effort to understand the mechanisms that may support ST, we investigate the effects of WM and anxiety on task performance. If WM is indeed a mechanism through which ST operates, then we should see significant interactions in our study. That is, whereas low-WM girls' performance may be negatively affected by ST, high WM girls' performance may be unaffected (Régner et al., 2010; Beilock, Rydell, & McConnell, 2007).

Method

Participants

Forty-eight 6- to 7-year-old (*M* = 86.35 months; *SD* = 7.34 months) girls from a metropolitan area participated in the current study. Parents provided written informed consent on behalf of their children. Children received a gift for participating. All procedures were approved by the local ethics committee.

Procedure

Children were randomly assigned to one of two conditions: an experimental condition, in which gender stereotype threat was induced, or a control condition, in which gender stereotype threat was not induced. Following Ambady and colleagues (2001), in the experimental condition (*n* = 24), girls chose to color one of two pictures intended to induce gender salience. One of these pictures depicted a young girl playing with a doll (*n* = 9) and the other depicted two girls dancing (*n* = 15). By depicting young girls engaged in typically feminine activities, these images made gender salient to participants. Girls in the control condition chose to color one of two neutral pictures, a mountain landscape (*n* = 15) or an underwater landscape (*n* = 9). These images contained no human figures or depictions of activities that could be categorized as either feminine or masculine and thus should not induce ST. There were two five-minute coloring sessions during the course of the experiment, one prior to the first block of tasks and another prior to the second block of tasks.

The first block contained four measures of spatial thinking (Thurstone test, Children's Mental Transformation Task, Porteus Maze Test, and NEPSY-Block Construction) and the second block contained one additional measure of spatial thinking (KABC-Spatial Short-Term Memory) and three non-spatial measures (WJ-Auditory Working Memory, WJ-Applied Problems, and WJ-Letter–Word Identification). The order of the blocks was held constant for all participants, but the order of tasks within each block was randomized across participants. After completing the second block, participants filled out a computerized questionnaire that measured domain-specific anxiety.

Testing was conducted by two experimenters. The primary experimenter administered all the tasks while a second experimenter supervised the coloring sessions. This was critical for ensuring the primary experimenter was blind to the participant's condition.

Spatial Measures

Thurstone Test. The Mental Rotation subtest of the Thurstone Primary Mental Abilities Test (Thurstone, 1974) was used as a measure of mental rotation ability. Children were administered one of three versions of the task that differed in the items presented and the order of presentation (version administered was randomized across participants and performance was comparable across versions). Children were shown a row of five shapes. On the left was the target shape, namely an incomplete square. On the right were four shapes. Children were asked to select the shape on the right that would complete the square. The shapes were presented in different orientations, requiring the use of mental rotation to correctly identifying the shape that completed the square. Participants completed all 16 items in the test. Scores were computed by summing the number of items answered correctly.

CMTT. Children were presented with 16 test trials of the Children's Mental Transformation Task (CMTT) developed by Levine and colleagues (1999). Each trial consisted of a shape divided along a vertical axis creating two target pieces. The two pieces were presented above four completed shapes that served as answer choices. Children were instructed to choose the shape among the four answer choices that would be formed if the two target pieces were moved together. Eight of the trials required mental translation (i.e., movement across the horizontal or vertical axis) whereas the other eight trials required mental rotation (i.e., movement across a diagonal axis). Scores were calculated by summing the number of correct responses.

Porteus Maze Test. The Vineland Revision of the Porteus Maze Test was used to assess visuospatial planning ability (Porteus, 1919). A set of nine mazes of increasing difficulty were administered. Children were directed to a starting place and were asked to draw a route out of the maze without crossing lines or going into blocked areas. The number of chances children received to correctly exit the maze varied across mazes. A score was given for each maze based on the number of attempts needed to correctly complete the maze. These scores were summed to provide the final numerical score for each participant. The task is internally consistent, with a Cronbach's alpha of .81 (Krikorian & Bartok, 1998).

NEPSY-Block Construction. The Block Construction subtest of the NEPSY-II is designed to measure spatial visualization and visuomotor development (Korkman, Kirk, & Kemp, 2007). Children were shown a two-dimensional image of a block formation and were asked to create a three-dimensional reconstruction of the image using a set of square blocks.

Children were given either 30 or 60 seconds to complete the structure, depending on the item's difficulty. Items became progressively more difficult and testing was discontinued after four consecutive incorrect responses. Standard scores, normed on a sample of children ages 3 through 16, were computed for each participant (M = 10; SD = 3). Reported reliability is estimated as r = .75 using a split-half procedure (Korkman et al., 2007).

KABC-Spatial STM. The Spatial Memory subtest of the Kaufman Assessment Battery for Children (KABC; Kaufman & Kaufman, 1983) was administered to assess children's short-term memory ability. Children were presented with a set of pictures on a page for five seconds. A grid was then overlaid on the array and children were asked to point to the boxes in which the pictures on the previous page were located. The number of images participants were required to recall increased across trials. Testing was discontinued after four incorrect responses. Scores were calculated by summing the number of correct responses. In children (5 years of age and older) reported reliability is high (split-half procedure, r = .80; Kaufman & Kaufman, 1983).

Non-Spatial Measures

WJ-Applied Problems. The Applied Problems subtest of the Woodcock-Johnson III Tests of Achievement (WJ; Woodcock, McGrew, & Mather, 2001) assesses mathematical ability. Questions include counting and basic arithmetic. Each item was read aloud to the participant and corresponded to visual images on a page. Testing was discontinued after six consecutive incorrect responses. Reported reliability is estimated as r = .92 using a split-half procedure (Woodcock et al., 2001).

WJ-Auditory Working Memory. The Auditory Working Memory subtest of the Woodcock-Johnson III Tests of Cognitive Abilities (WJ) was used to assess WM. Children were present aurally with a set of two to seven items containing numbers and nouns, referred to as "things" (e.g., the experimenter says, "1 – cat – milk"). They were then asked to repeat the set to the experimenter, beginning with the nouns and then the numbers (e.g., the participant responds correctly, "cat – milk – 1"). Testing was discontinued after three consecutive incorrect responses. Reported reliability is estimated as $r = .96$ using a split-half procedure (Woodcock et al., 2001).

WJ-Letter–Word Identification. The Letter–Word Identification subtest of the Woodcock-Johnson III Tests of Achievement (WJ) measures linguistic competence. Children were presented with a list of words on a page and asked to read each word aloud. Responses are considered correct if the word is read fluidly on the first attempt. Difficulty increases across items, as words listed become increasingly less common in written English. Testing was discontinued after six incorrect responses. Reported reliability is estimated as $r = .91$ using a split-half procedure (Woodcock et al., 2001).

Anxiety Questionnaires

Anxiety questionnaires were administered on a touchscreen computer. Questions were written at the top of the screen and were read aloud to the child by the experimenter. Children responded by touching one of five smiley faces at the bottom of the screen, indicating that the situation made them feel "very, very nervous," "a little nervous," "in the middle," "not very nervous, sort of calm" or "not nervous at all, very calm." Some items included an image to help guide the child's understanding of the question, such as a map or a maze.

Spatial Anxiety Questionnaire. Children completed the Child Spatial Anxiety Questionnaire (CSAQ), a self-report measure of spatial anxiety developed by Ramirez and colleagues (2012). All eight items from the CSAQ were administered. These items assessed children's anxiety about spatial tasks as they occur in academic or everyday settings (e.g., "How do you feel when you are asked to point to a certain place on a map, like this one?"). Responses were coded on a five-point scale and averaged, producing a mean spatial anxiety score for each participant.

Math Anxiety Questionnaire. A shortened version of the Child Math Anxiety Questionnaire was used to measure children's math anxiety in academic settings and in daily life (CMAQ; Ramirez, Gunderson, Levine, & Beilock, 2013). Four items from the original CMAQ were chosen to avoid repetition across items (e.g., "How do you feel when you have to solve 34 – 17?"). A math anxiety score was calculated by averaging participants' responses to each question.

Results

Task Performance

Task performance was analyzed across all participants to ensure that scores were normally distributed. Measures of skewness on all tasks ranged from -.297 to .752, which is within the accepted range (from negative one to positive one) of a normally distributed sample. Within condition we analyzed group differences between the two coloring-page options to ensure that performance was not significantly effected by coloring page instead of condition. In the control condition there was only one significant difference in performance between coloring pages. Girls who colored the underwater landscape picture performed significantly better on the NEPSY-Block Construction subtest than girls who colored the mountain landscape picture (p = .048). There was no significant difference between coloring pages on any other task in the control condition (ps > .059). Additionally, there were no significant group differences in task performance between girls who colored the picture of two girls dancing and the picture of a girl with a doll in the experimental condition (ps > .096).

Group Level Analyses

The primary aim of the current study was to test whether we could implicitly induce ST in young girls. We did this by comparing task performance between girls in the control group and girls in the experimental group (see *Table 1*). Analyses revealed significant group differences on the KABC-Spatial STM task and the Math Anxiety Questionnaire. Specifically, girls who were exposed to ST performed significantly worse on a test of spatial short-term memory compared to girls who were not exposed to ST, $t(46) = 2.080$, $p = .043$. Similarly, girls exposed to ST reported significantly more math anxiety than girls who were not exposed to ST, $t(46) = -2.545$, $p = .014$. There were no other significant effects (*p*s > .1). Although these findings reveal few differences when comparing girls' task performance across experimental and control groups, a possibility described in the Introduction is that the effect of ST is dependent on individual differences in WM (Régner et al., 2010; Beilock, Rydell, & McConnell, 2007). For example, Régner and colleagues (2010) found that ST effects held only for women with low WM, whereas women with high WM showed no effects of ST induction. Thus, we next considered the effect of ST in relation to children's WM ability.

Stereotype Threat in Relation to Working Memory

To understand the role of WM on young girls' susceptibility to ST, we divided participants into high- and low-WM groups, using a median split of scores on the WJ-Auditory Working Memory subtest. This created four separate groups: High-WM Control group ($n = 9$), Low-WM Control group ($n = 15$), High-WM Experimental group ($n = 14$), and Low-WM Experimental group ($n = 10$).

We then analyzed the effects of condition and WM on math task performance using a 2 (condition: Experiment, Control) x 2 (WM group: High, Low) between-subjects analysis of variance (ANOVA). This analysis revealed significant main effects of condition, $F(1, 44) =$

4.510, $p = .039$, $\eta^2 = .093$, and WM group, $F(1, 44) = 17.884$, $p = .000$, $\eta^2 = .289$. There was also a significant interaction between condition and WM group, $F(1, 44) = 13.739$, $p = .001$, $\eta^2 = .238$. Pairwise comparisons (Bonferroni corrected) of these results revealed a significant difference in math performance between WM groups in the experimental condition, with Low-WM girls performing significantly worse ($M = 96.40$, $SD = 6.75$) than High-WM girls ($M = 119.36$, $SD = 11.28$), $F(1, 44) = 32.069$, $p = .001$, $\eta^2 = .422$ (see *Figure 1*). There was no significant difference in the control condition, with High-WM girls and Low-WM girls performing comparably, $F(1, 44) = .134$, $p = .716$, $\eta^2 = .003$. Moreover, Low-WM girls in the experimental condition ($M = 96.40$, $SD = 6.75$) performed significantly worse than Low-WM girls in the control condition ($M = 113.27$, $SD = 10.17$), $F(1, 44) = 17.805$ $p = .001$, $\eta^2 = .288$.

We also found a significant effect of ST on a mental rotation task when accounting for WM. A 2 (condition: Experiment, Control) x 2 (WM group: High, Low) between-subjects ANOVA using performance on the Thurstone test as the dependent variable revealed a significant main effect of condition, $F(1, 44) = 5.577$, $p = .023$, $\eta^2 = .112$. There was a marginally significant condition-by-WM interaction, $F(1, 44) = 3.898$, $p = .055$, $\eta^2 = .081$, and pairwise comparisons (Bonferroni corrected) revealed a significant difference between WM groups in the experimental condition, with girls in the Low-WM group performing worse ($M = 6.40$, $SD = .70$) than girls in the High-WM group ($M = 9.07$, $SD = 2.56$), $F(1, 44) = 7.491$, $p = .007$, $\eta^2 = .153$ (see *Figure 2*). There was no significant difference between WM groups in the control condition, $F(1, 44) = .000$, $p = 1$, $\eta^2 = .000$. Moreover, girls in the Low-WM Experimental group ($M = 6.400$, $SD = .699$) performed significantly worse than girls in the Low-WM Control group ($M = 9.333$, $SD = 2.498$), $F(1, 44) = 9.848$, $p = .003$, $\eta^2 = .183$. The results of the pairwise comparisons should be taken with a degree of caution given that the ANOVA revealed a close-to- but not-quite-significant interaction of $p = .055$. However, if replicated, these findings indicate that gender ST can result in decreased performance on a spatial task.

We next examined these effects in relation to performance on the KABC-Spatial STM task. A 2 (condition: Experiment, Control) x 2 (WM group: High, Low) between-subjects ANOVA showed a significant main effect of condition, $F(1, 44) = 4.504$, $p = .039$, $\eta^2 = .093$, but no significant main effect of WM. Additionally, there was no significant interaction, $F(1, 44) = 1.377$, $p = .247$, $\eta^2 = .030$.

There were no significant main effects or interactions on the other spatial measures administered, namely the Porteus Maze Test, CMTT, and NEPSY Block Construction subtest (ps $> .09$; see *Table 2*). Importantly, there was no significant group difference on WJ-Letter–Word Identification scores, demonstrating that ST did not impact performance on a measure that is not linked to gender-based stereotypes. This, coupled with our findings that girls in the Low-WM Experimental group performed worse on a number of spatial and mathematical tasks compared to girls in the Low-WM Control group, offers support that group differences were due to induced salience of gender stereotypes. These findings extend the work on ST in math by demonstrating an effect of gender ST on multiple spatial tasks in young girls.

To ensure that the significant effects on the WJ-Applied Problems test, Thurstone test, and KABC-Spatial STM test were specifically moderated by WM, we conducted additional analyses in which girls were divided into groups based on high- and low-linguistic competence, which was determined using a median split of scores on the WJ-Letter–Word Identification test. If ST functions as a mechanism of WM, then we should not see significant interactions in another domain, such as language.

Using math performance as the dependent variable, a 2 (condition: Experiment, Control) x 2 (linguistic competence group: High, Low) between-subjects ANOVA revealed a significant main effect of linguistic competence, $F(1, 44) = 13.203$, $p = .001$, $\eta^2 = .231$, and an interaction, $F(1, 44) = 4.448$, $p = .041$, $\eta^2 = .092$. Pairwise comparisons (Bonferroni corrected) of these results revealed a significant difference in math task performance between linguistic competence groups in the experimental condition, with girls in the Low-linguistic competence group (M = 98.22, SD = 9.080) performing significantly worse than those in the High-linguistic competence group (M = 116.73, SD = 13.504), $F(1, 44) = 16.194$, $p = .000$, $\eta^2 = .269$ (see *Figure 3*). There was no significant difference in the control condition between Low-linguistic competence girls and High-linguistic competence girls, $F(1, 44) = 1.184$, $p = .263$, $\eta^2 = .026$. Among girls with Low-linguistic competence, those in the experimental condition (M = 98.22, SD = 9.080) performed significantly worse than those in the control condition (M = 111.79, SD = 9.713), $F(1, 44) = 8.468$, $p = .006$, $\eta^2 = .161$.

A 2 (condition: Experiment, Control) x 2 (linguistic competence group: High, Low) between-subjects ANOVA using performance on the Thurstone test as the dependent variable revealed a significant main effect of condition, $F(1, 44) = 4.174$, $p = .047$, $\eta^2 = .087$, but no significant main effect of linguistic competence, $F(1, 44) = .176$, $p = .677$, $\eta^2 = .004$, or significant interaction, $F(1, 44) = 1.468$, $p = .232$, $\eta^2 = .032$. The lack of a significant interaction between condition and linguistic competence on the Thurstone test indicates that WM is specifically responsible for the negative impact of ST on girls in the Low-WM Experimental group.

Using performance on the KABC-Spatial STM test as the dependent variable, a 2 (condition: Experiment, Control) x 2 (linguistic competence group: High, Low) between-subjects

ANOVA revealed a significant main effect of condition, $F(1, 44) = 5.490$, $p = .024$, $\eta^2 = .111$. Again, there was no significant main effect of linguistic competence, $F(1, 44) = 1.536$, $p = .222$, $\eta^2 = .034$, or significant interaction, $F(1, 44) = 1.012$, $p = .320$, $\eta^2 = .022$, offering support that our ST induction on the KABC Spatial STM test was moderated specifically by WM.

Controlling for Age and General Intelligence

To ensure that our effects were not due to age or individual differences in general intelligence, we performed a 2 (condition: Experiment, Control) x 2 (WM group: High, Low) between-subjects analysis of covariance (ANCOVA) on the math and spatial tasks that showed significant main effects above, with chronological age and linguistic competence (WJ-Letter–Word Identification score) as our covariates. There were significant main effects of condition, $F(1, 42) = 8.239$, $p = .006$, $\eta^2 = .164$, and WM group, $F(1, 42) = 8.340$, $p = .006$, $\eta^2 = .166$, on math task performance. There was also a significant condition-by-WM interaction, $F(1, 44) = 16.679$, $p = .000$, $\eta^2 = .284$. This provides additional evidence that other variables (i.e., age and linguistic competence) do not account for our effects. Analysis of the Thurstone test also revealed significant main effects of condition, $F(1, 42) = 5.271$, $p = .027$, $\eta^2 = .112$, and WM group, $F(1, 42) = 6.119$, $p = .018$, $\eta^2 = .127$. There was no significant condition-by-WM interaction on the Thurstone test, $F(1, 42) = 2.827$, $p = .100$, $\eta^2 = .063$. However, this is not surprising as there was only a marginally significant interaction ($p = .055$) prior to controlling for the effects of age and linguistic competence. There was a significant main effect of condition, $F(1, 42) = 5.178$, $p = .028$, $\eta^2 = .110$, but not WM, $F(1, 42) = .139$, $p = .711$, $\eta^2 = .003$, on the KABC-Spatial STM task. As with prior analyses of the KABC-Spatial STM task, there was no significant interaction, $F(1, 42) = .577$, $p = .452$, $\eta^2 = .014$. The results described above replicate the findings produced by our ANOVAs, indicating that our effects are not due to age or general intelligence, but rather to WM ability.

Anxiety-Related Effects

In addition to measuring group differences in task performance, we also measured self-reported spatial and mathematical anxiety. There were significant negative correlations between age and spatial anxiety, $r(46) = -.444$, $p = .002$, as well as between age and math anxiety, $r(46) = -.411$, $p = .004$, with older girls reporting less anxiety than younger girls. Thus, we controlled for the effects of age on the relationship between anxiety and task performance. First, we addressed the significant group difference in math anxiety between conditions, in which girls in the experimental condition reported more math anxiety than those in the control condition (see *Table 1*). However, we found that math anxiety was not significantly related to task performance in either the control or experimental group (ps > .124). Similarly, when controlling for age, spatial anxiety was not significantly related to performance in the control or experimental group (ps > .215). Additionally, there were no significant differences in spatial or mathematical anxiety between Low-WM and High-WM girls in the control and experimental groups (see *Table 2*). These results indicate that, although ST resulted in lower performance among girls with low-WM ability on both a math test and a test of mental rotation, reported anxiety was not related to performance.

Discussion

The current study demonstrated that early school-aged girls are susceptible to ST on both spatial and mathematical tasks and that the effect of ST is modulated by WM ability. When analyzing groups by WM ability, we successfully replicated the use of implicit ST manipulation first used by Ambady and colleagues (2001) with relation to math performance. Our results also extend this work by demonstrating these effects in the spatial domain. Of the five spatial measures administered, ST was successfully induced in girls with low-WM in two of these, the Thurstone test and the KABC-Spatial STM task. The Thurstone Mental Rotation test is a measure of one's ability to mentally rotate images in space. Mental rotation shows the largest sex difference of any cognitive task so it is not surprising that gender ST was successfully induced on this task (Linn & Peterson, 1985).

Inducing ST was unsuccessful on three measures of spatial thinking: the CMTT, NEPSY-Block Construction subtest, and Porteus Maze Test. Children's scores were normally distributed on all measures, giving us reason to believe that these results were not due to a lack of variability in performance. Instead, the lack of significant findings on these measures may be because the types of spatial thinking assessed are not as strongly associated with gender stereotypes and thus are not easily influenced by gender ST. More work should be done to understand whether spatial tasks other than mental rotation and short-term memory are susceptible to gender ST. Additionally, task difficulty has been shown to be an important factor in ST (Flore & Wicherts, 2015). Perhaps tasks in which ST was not successfully induced were not challenging enough to show an effect of ST.

The significant condition-by-working memory interactions shown in our results indicate that girls with low-WM ability are uniquely affected by ST, performing worse on spatial and mathematical tasks when their gender identity is activated. (It should be noted that, while there was a moderating effect of linguistic competence on children's susceptibility to ST on the WJ-Applied Problems task, the significant condition-by-WM interaction effect remained significant when controlling for age and linguistic competence. Taken as a whole, these data support the notion that WM is a more likely mechanism of ST than linguistic competence.) Our findings replicate the work by Régner and colleagues (2010) who suggest that those with low-WM are not equipped to handle the additional WM load caused by the stress of ST and thus are highly susceptible to its negative effects.

Interestingly, the results of the current study differ from the work of Ramirez and colleagues (2012, 2013), and showed the opposite effect of WM in relation to performance anxiety. Ramirez and colleagues (2013) measured individual differences in math ability and math anxiety in children and found that poor math performance was associated with high math anxiety in those with high-WM ability. The same authors showed a similar pattern of effects in the domain of space (Ramirez et al., 2012). High spatial anxiety in children was associated with poor performance on a measure of mental rotation ability. However, girls with high-WM were uniquely affected by spatial anxiety, while boys demonstrated no significant spatial anxiety-by-WM interaction. The authors suggest that those with high-WM ability rely on WM-intensive strategies to solve problems and thus perform more poorly on tasks when faced with performance-related anxieties that co-opt WM.

When interpreting these findings in relation to our own it is important to note that, though the terms are colloquially used interchangeably, ST and performance anxiety are different concepts that may tap WM in different ways. Performance deficits seen in conjunction with ST are related to the expectation of failure due to one's association with a particular identity. This differs from performance deficits stemming from anxieties about the pressure to succeed at a particular task. While WM is implicated in both phenomena, Régner and colleagues (2010) suggest that WM plays a different role in the experiences of ST and performance anxiety and that the differing performance focus in each produces differential WM interactions across studies.

More specifically, high-WM individuals are protected from the effects of ST but may suffer when experiencing performance anxiety.

Our results showed no significant differences on anxiety scores between groups when divided by condition and WM. Additionally, neither spatial nor math anxiety were related to task performance on any of our measures. Past work has demonstrated significant relationships between task performance and the anxiety measures used in the current study (CSAQ, CMAQ; Ramirez et al., 2012, 2013, respectively). However, as previously mentioned, these studies simply measured performance anxiety and did not induce ST. Therefore, it is quite possible that anxiety is not an explicit component of ST. An alternative possibility is that, though inducing ST has been shown to cause physiological stress arousal (Schmader, Johns, & Forbes, 2008), an individual facing ST might remain consciously unaware of this stress and feel well-equipped to handle a problem. To reiterate, though the individual is experiencing an elevated stress response to ST strong enough to negatively impact performance, they are not aware of anxiety on a conscious level. In contrast, an individual who suffers from high math or spatial anxiety experiences this anxiety on a conscious level with feelings of nervousness or apprehension (Ashcraft & Kirk, 2001). These performance-related worries co-opt WM ability and lead to subsequent performance deficits in those who already rely on WM-intensive problem-solving strategies. In this way, our data provide evidence that ST may occur on a less conscious level and, thus, may not be associated with the explicit experience of anxiety. These results have important implications for the understanding of ST by demonstrating that children are still heavily impacted by ST even though they may not explicitly express anxiety related to task performance.

The current study replicated the relationship between math and gender ST and expanded the literature by successfully inducing ST on a spatial task in young girls. This adds to the growing body of research demonstrating that gender stereotypes affect performance in a number of domains. It also offers insight into the conflicting reports of the moderating effect of WM on task performance when task-related anxieties are present. More work is needed to enhance our understanding of how WM contributes to ST and performance pressure in unique ways. Additionally, the current study contributes to our knowledge of the role of anxiety on ST. Future studies should address participants' stress response, in addition to explicit anxieties, in order to understand how stress manifests itself in relation to ST. Gender ST should not be deemed unimportant simply because girls are not reporting anxiety about the effect of gender stereotypes on their performance. This study offers hope that preventing the negative impact of ST is possible through further investigation of stress and ST, as well as through improved WM ability. With increased knowledge of the mechanisms behind gender ST, we can address the performance deficits that women face on spatial and mathematical skills. Understanding the relationship between spatial thinking and gender ST is an essential step in addressing the persistent sex difference in achievement on spatial tasks and in STEM fields.

References

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

Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, *130*(2), 224.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*(5), 339-343.

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology*, *136*(2), 256-276.

Doyle, R. A. & Voyer, D. (2015). Stereotype threat manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences*, *47*, 103-116.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology*, *53*(1), 25-44.

Grimshaw, G. M., Sitarenios, G., & Finegan, J. A. K. (1995). Mental rotation at 7 years-relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition*, *29*(1), 85-100.

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, *19*(1), 61-70.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children*. John Wiley & Sons, Inc.

Korkman, M., Kirk, U., & Kemp, S. (2007). *NEPSY-II: Administration Manual*. San Antonio, TX: The Psychological Corporation.

Krikorian, R. & Bartok, J. A. (1998). Developmental data for the Porteus Maze Test. *The Clinical Neuropsychologist*, *12*(3), 305-310.

Lennon, P. A. (2000). Improving students' flexibility of closure while presenting biology content. *The American Biology Teacher*, *62*(3), 177-180.

Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, *35*, 940–949.

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

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, *39*, 990-997.

Lourenco, S. F., Addy, D., Huttenlocher, J., & Fabian, L. (2011). Early sex differences in weighting geometric cues. *Developmental Psychology*, *14*, 1365-1378.

McGlone, M. S. & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, *27*, 486–493.

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants a sex difference. *Psychological Science*, *19*(11), 1063-1066.

Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2015). A threat in the classroom. *Zeitschrift für Psychologie*, *220*(20), 61-69.

Porteus, S. D. (1919). *Porteus tests–the vineland revision*. Vineland, N. J.: Research Dept. Publications.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. *Quarterly Journal of Experimental Psychology*, *65*, 474-487.

Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2013). Math anxiety, working memory, and math achievement in early elementary school. *Journal of Cognition and Development*, *14*(2), 187-202.

Régner, I., Smeding, A., Gimmig, D., Thinus-Blanc, C., Monteil, J., Huguet, P. (2010). Individual differences in working memory moderate stereotype-threat effects. *Psychological Science*, *21*(11), 1646-1648.

Schmader, T., Johns, M., Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, *115*(2), 336–356.

Sorby, S. A., & Baartmans, B. J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students. *Journal of Engineering Education*, *89*(3), 301-307.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*, 4-28.

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797.

Thurstone, T. G. (1974). *PMA readiness level*. Chicago, IL: Science Research Associates.

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.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817-835.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III*. Rolling Meadows, IL: Riverside Publishing.

Table 1.

T-test Comparisons of Mean Task Performance for Control and Experimental Groups.

	Control Group (N = 24) *Mean (SD)*	*Experimental Group (N = 24)* *Mean (SD)*
Spatial Measures		
Thurstone Test	9.33 (2.48)	7.96 (2.39)
CMTT	10.46 (2.43)	9.17 (2.91)
Porteus Maze Test	11.40 (1.90)	11.08 (2.28)
NEPSY-Block Construction	9.00 (3.23)	9.54 (3.53)
KABC-Spatial STM	**13.21 (2.19)***	**11.83 (2.39)***
Non-spatial Measures		
WJ-Applied Problems	113.83 (9.70)	109.79 (14.95)
WJ-Auditory WM	107.38 (19.17)	114.58 (16.34)
WJ-Letter–Word Identification	116.33 (9.43)	121.42 (13.78)
Anxiety Questionnaires		
Spatial	2.61 (.70)	2.82 (.58)
Mathematical	**2.39 (.88)***	**3.03 (.88)***

Note. WJ tasks include Scaled Scores, Anxiety measured on a 5-point scale
*Group differences present at $p < .05$

Table 2.

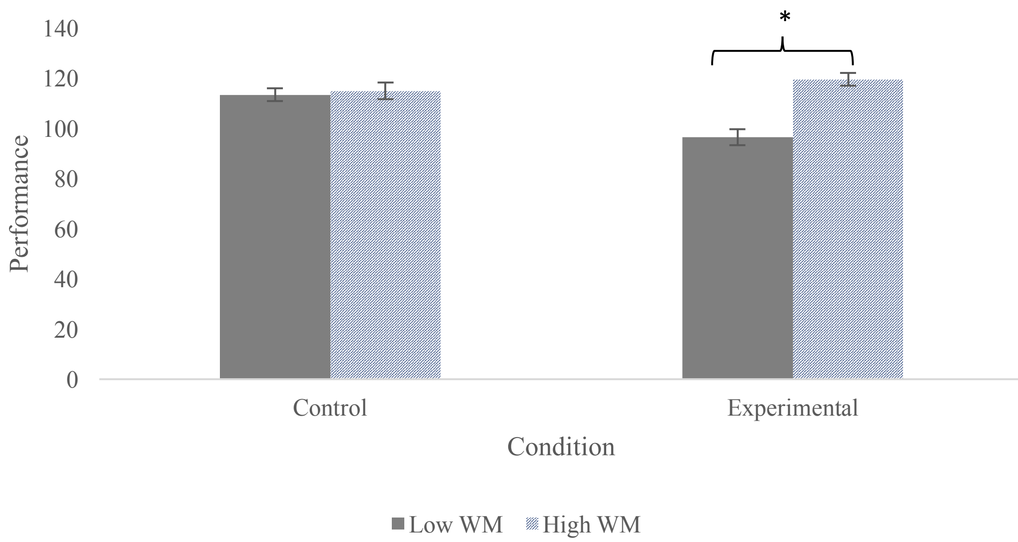
T-test Comparisons of Mean Task Performance for Condition and Working Memory Groups.

	Low WM		High WM	
	Control Group (N = 15) *Mean (SD)*	*Experimental Group (N = 10)* *Mean (SD)*	*Control Group (N = 9)* *Mean (SD)*	*Experimental Group (N = 14)* *Mean (SD)*
Spatial Measures				
Thurstone Test	**9.33 (2.94)****	**6.40 (0.70)****	9.33 (2.60)	9.07 (2.56)
CMTT	10.47 (2.56)	8.60 (2.59)	10.44 (2.35)	9.57 (3.20)
Porteus Maze Test	11.70 (2.14)	10.35 (1.94)	10.89 (1.39)	11.61 (2.42)
NEPSY-Block Construction	9.73 (3.15)	8.30 (3.20)	7.78 (3.15)	10.14 (3.59)
KABC Spatial STM	**13.33 (2.47)***	**11.10 (1.66)***	13.00 (1.73)	12.36 (2.74)
Non-spatial Measures				
WJ-Applied Problems	**113.27 (10.17)****	**96.40 (6.75)****	114.78 (9.39)	119.36 (11.28)
WJ-Auditory WM	95.67 (12.57)	99.30 (9.80)	126.89 (9.71)	125.50 (9.88)
WJ-Letter–Word Identification	113.13 (8.00)	115.20 (13.13)	121.67 (9.67)	125.86 (12.87)
Anxiety Questionnaires				
Spatial	2.73 (0.81)	2.89 (0.54)	2.40 (0.43)	2.78 (0.62)
Mathematical	2.47 (0.92)	3.10 (0.86)	2.25 (0.86)	2.98 (0.92)

Note. WJ tasks include Scaled Scores, Anxiety measured on a 5-point scale
*Group differences present at $p < .05$, ** $p < .01$

Figure 1.

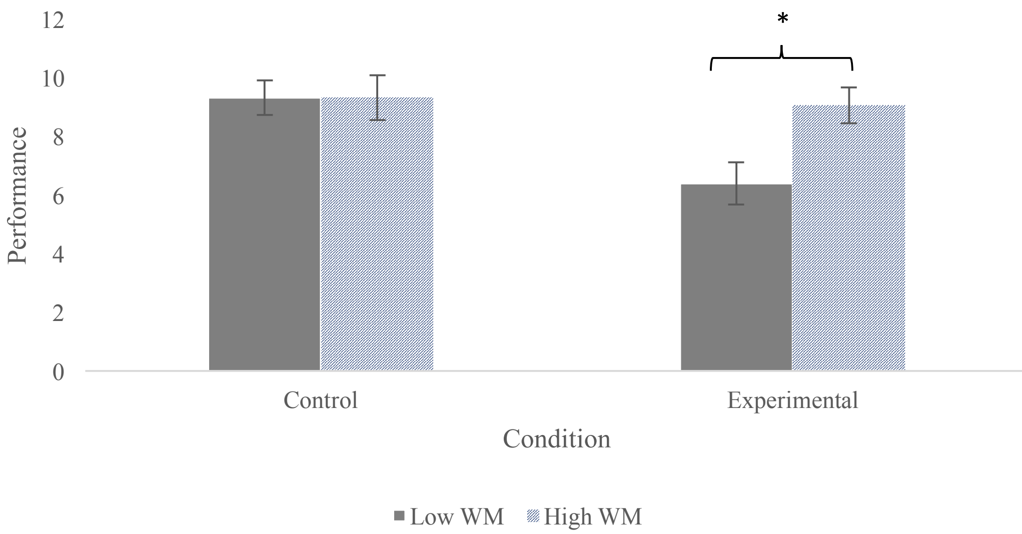
Mean Score Differences in Math Performance in Relation to WM



Note. Mean scores on the WJ-Applied Problems test for condition and WM groups. Girls in the Low-WM Experimental group performed significantly worse than those in the High-WM Experimental group, $F(1, 44) = 32.069$, $p = .001$, $\eta^2 = .422$.
*Group differences present at $p < .01$

Figure 2.

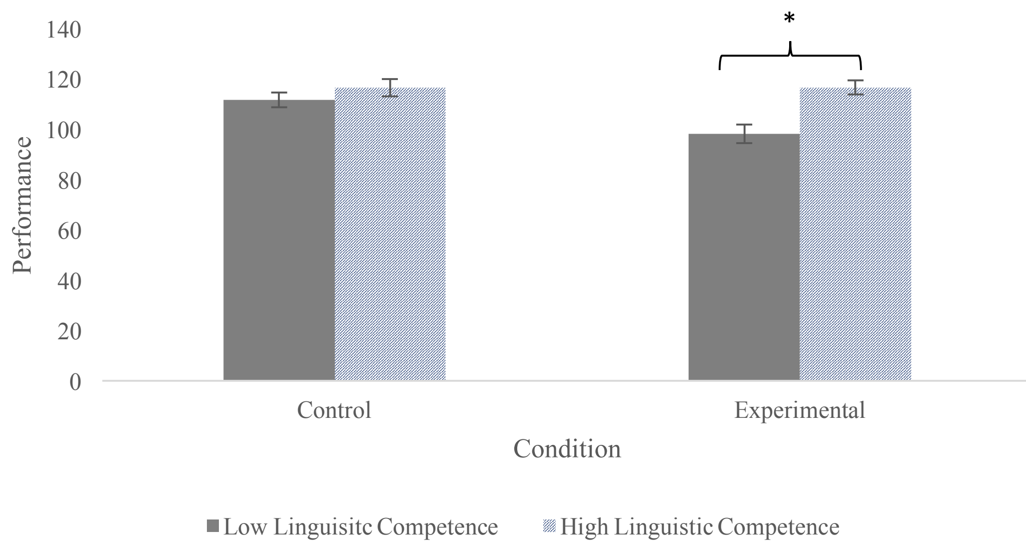
Mean Score Differences in Spatial Performance in Relation to WM



Note. Mean scores on the Thurstone test for condition and WM groups. Girls in the Low-WM Experimental group performed significantly worse than those in the High-WM Experimental group, $F(1, 44) = 7.491$, $p = .007$, $\eta^2 = .153$.
*Group differences present at $p < .01$

Figure 3.

Mean Score Differences in Math Performance in Relation to Linguistic Competence



Note. Mean scores on the WJ-Applied Problems test for condition and linguistic competence groups. Girls in the Low-WM Experimental group performed significantly worse than those in the High-WM Experimental group, $F(1, 44) = 16.194$, $p = .000$, $\eta^2 = .269$.
*Group differences present at $p < .01$