Journal of Educational Psychology
2014, Vol. 106, No. 1, 105–120

0022-0663/14/$12.00 DOI: 10.1037/a0034099

The Role of Anxiety and Working Memory in Gender Differences in Mathematics

Colleen M. Ganley
Boston College and University of Illinois at Urbana–Champaign

Marina Vasilyeva
Boston College

This research examined a potential mechanism underlying gender differences in math performance by testing a mediation model in which women's higher anxiety taxes their working memory resources, leading to underperformance on a mathematics test. Participants for the 2 studies were college students ($N = 87$, $N = 118$) who completed an anxiety measure, 2 working memory tasks (verbal and visuospatial), and a challenging math test including both geometry and algebra items. Findings showed a significant gender difference in math performance, anxiety, and visuospatial working memory. Further, there was a mediating chain from gender to the worry component of anxiety to visuospatial working memory to math performance. The results suggest that women's heightened worry may have utilized their visuospatial working memory resources, and the resulting gender differences in working memory were associated with gender differences on a math test. The present research contributes to our understanding of affective and cognitive factors underlying gender differences in mathematics. The findings are discussed in terms of their implications for interventions aimed at reducing anxiety and improving working memory skills.

Keywords: gender differences, mathematics performance, anxiety, working memory, college students

There is a large body of literature investigating gender differences in math performance, and multiple explanations for these differences have been proposed. Some studies have emphasized the role of affective processes, such as anxiety, whereas others have focused on cognitive differences that may be associated with a male advantage in mathematics. The present investigation seeks to examine the joint role of affective and cognitive factors underlying gender differences in math performance. Specifically, we test a mediational model, whereby females have higher anxiety about a math test, which subsequently utilizes working memory resources needed for the task, potentially leading to poorer test performance. Although there is much research supporting the possibility of this model, this is the first empirical investigation of the full mediation model that incorporates both cognitive and affective factors.

Gender Differences in Math Performance

Research shows that gender differences in math performance are generally small or nonexistent early in development but become larger and more consistent later on (Hyde, Fennema, & Lamon, 1990; though see Robinson & Lubienski, 2011). In particular, a recent meta-analysis has indicated that differences favoring males appear in high school students and persist into college (Lindberg, Hyde, Petersen, & Linn, 2010). High-school males outscore females on national achievement tests including the quantitative portion of the Scholastic Assessment Test (SAT) and the mathematics Advanced Placement (AP) exams, both of which have implications for students' academic futures (College Board, 2009, 2010). These differences emerge even though female students take advanced mathematics courses at a similar rate to male students in high school (College Board, 2010). As females progress through college and graduate school, they tend to drop out of the mathematics pipeline at higher rates. Although in the last decade the majority (57%) of college students have been female, they made up 44% of college math majors and only 31% of recipients of doctoral degrees in mathematics (National Science Foundation, National Center for Science and Engineering Statistics, 2011). Even larger gender gaps are seen later in math-related career choices in Science, Technology, Engineering, and Mathematics (STEM) careers (Snyder & Dillow, 2011).

This article was published Online First August 19, 2013.

Colleen M. Ganley, Department of Counseling, Developmental, and Educational Psychology, Boston College, and Departments of Curriculum and Instruction and Educational Psychology, University of Illinois at Urbana–Champaign; Marina Vasilyeva, Department of Counseling, Developmental, and Educational Psychology, Boston College.

This research was supported by Boston College through a dissertation fellowship from the Lynch School of Education and an individual research grant from the Graduate Student Association, both to Colleen M. Ganley, and by the Institute of Education Sciences, U.S. Department of Education, through Grant R305B100017 to the University of Illinois at Urbana–Champaign. The opinions expressed are those of the authors and do not represent views of the Institute of Education Sciences or the U.S. Department of Education. We thank Martha Makowski, Paul Mathewson, and Katie-Beth Price for assistance with data collection as well as Beth Casey, Lillie Albert, Elida Laski, Michelle Perry, Alana Dulaney, Sarah Lubienski, and Joe Robinson for helpful feedback on this project.

Correspondence concerning this article should be addressed to Colleen M. Ganley, who is now at the Department of Psychology and the Learning Systems Institute, Florida State University, 1107 West Call Street, Tallahassee, FL 32306. E-mail: ganley@psy.fsu.edu

In attempting to understand the age-related increase in the observed gender gap, it is important to consider the types of mathematical tasks that tend to reveal gender differences. Recent studies have shown that a male advantage is only apparent in certain areas of math, whereas females maintain an advantage in other areas. Females perform better than males on computation tasks, as well as other tasks that rely heavily on recall of procedures and information (Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Vasilyeva, Casey, Dearing, & Ganley, 2009). However, males tend to outperform females in more complex math that involves problem solving, particularly in areas that rely on spatial thinking (Becker, 1990; Gibbs, 2010; Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010; Vasilyeva et al., 2009).

In the context of these findings, Gibbs (2010) suggested that the fact that gender differences favoring boys appear to emerge with development is due, in part, to the nature of skills required on math assessments at different ages. Specifically, as students progress through their schooling, a more pronounced male advantage is observed in association with increased task complexity. Gibbs's research suggests that we may underestimate existing gender differences by aggregating performance across assessments that require a wide range of skills. Many curriculum-based standardized assessments do not include items at a higher level of complexity, which could explain why we do not see gender differences on these types of assessments (Hyde, Lindberg, Linn, Ellis, & Williams, 2008). Thus, despite the fact that some researchers emphasize that gender differences in math are small—and perhaps disappearing—it appears that if more complex items are included in math assessments, the male advantage may be more robust across ages.

Explaining Gender Differences in Math Performance

Potential Role of Anxiety

Researchers consistently find that math anxiety is negatively related to math performance. This relation has been found both with measures of trait anxiety, which reflect a general tendency to feel anxious about mathematics (Ganley & Vasilyeva, 2011; Ma, 1999; Miller & Bichsel, 2004), and measures of state anxiety, which capture anxious feelings experienced in the context of a math testing situation (Beilock, Rydell, & McConnell, 2007; Brodish & Devine, 2009; Cadinu, Maass, Rosabianca, & Kiesner, 2005; Osborne, 2001). Although most of these studies are correlational in nature, there is also some research that has experimentally tested the effect of interventions designed to alleviate anxiety on later performance. These studies have generally found that behavioral interventions that reduce anxiety also improve performance (in math and other academic areas), suggesting a causal relation between anxiety and performance (Hembree, 1988; Ramirez & Beilock, 2011; Wood, 2006).

This relation is particularly important in the context of gender, as women are consistently found to be more anxious about math tests than men (e.g., Hembree, 1988, 1990; Hyde, Fennema, Ryan, Frost, & Hopp, 1990; Miller & Bichsel, 2004). It has been suggested that heightened anxiety may reflect the process of gender socialization (Eccles & Jacobs, 1986). As women are socialized as members of their gender, they may become familiar with the stereotype that women are not as good as men at mathematics, which may lead them to develop higher levels of anxiety about their own mathematical abilities.

Evidence of a negative relation between anxiety and math performance combined with evidence of heightened math anxiety in women raise the possibility that gender differences in anxiety play a role in gender differences in math achievement. To address this issue, two past studies have tested anxiety as a potential mediator of the relation between gender and math performance with mixed results. In considering the discrepancy in these findings, it is important to note differences in the studies' methodologies that could be relevant. In particular, Osborne (2001), who found evidence for mediation, used a measure of state anxiety during math testing, whereas Casey, Nuttall, and Pezaris (1997), who did not find such evidence, examined a measure of trait math anxiety. It has been argued that the relation with test performance is stronger for state, compared to trait, anxiety because the former directly captures the affective state during testing (e.g., O'Neil & Fukumura, 1992). In the present studies, we chose to use a measure of state anxiety, as it reflects feelings associated with a particular testing situation.

Relation Between Anxiety and Working Memory

Given the evidence of the link between anxiety and math performance, a further question concerns the psychological mechanism underlying this relation. Past research suggests that anxiety may impact math performance through its relation with working memory. Working memory is the part of the memory system that involves temporarily storing information while simultaneously manipulating it. According to traditional models, there are two distinct types of working memory, verbal and visuospatial, which are coordinated and controlled by the central executive component of working memory (Baddeley & Hitch, 1974). Additionally, more recent models incorporate the concept of the episodic buffer, where information is integrated across multiple dimensions (i.e., verbal and visual; Baddeley, 2000, 2012). Together, the various components of working memory provide a mental space for problem solving.

There is extensive research showing that working memory is related both to anxiety and math performance. First, several studies have documented a relation between anxiety and working memory functioning (e.g., Beilock, 2008; Eysenck & Calvo, 1992). In particular, in the context of the processing efficiency theory, it has been theorized that individuals with high anxiety would perform less efficiently on tasks requiring working memory resources because their worrisome thoughts interfere with working memory, making them unable to fully utilize their working memory capacity for task performance (Eysenck & Calvo, 1992; Eysenck, Derakshan, Santos, & Calvo, 2007). The processing efficiency theory posits that anxiety could affect both verbal and visuospatial components through its effect on the central executive, which is utilized by both types of working memory. Existing research, however, shows inconsistent results in regard to whether anxiety impacts the verbal or visuospatial component, or both. Some researchers have found that anxiety exclusively has its impact on verbal working memory—presumably because intrusive worry-related thoughts may require processing of verbal information

(Ikeda, Iwanaga, & Seiwa, 1996; Lee, 1999; Owens, Stevenson, Norgate, & Hadwin, 2008; Rapee, 1993)—whereas other researchers have found that anxiety leads to lower performance on visuospatial working memory tasks but not on verbal tasks (Crowe, Matthews, & Walkenhorst, 2007; Shackman et al., 2006). Due to the inconsistencies in past research, in the current studies we examine both types of working memory.

In addition to establishing the link between anxiety and working memory, researchers have also examined the link between working memory and academic performance. It has been shown that working memory plays a critical role in problem solving across academic domains. This can be expected as most cognitive tasks require holding multiple pieces of information in one's mind while also processing information (Raghubar, Barnes, & Hecht, 2010). With respect to mathematics, some studies find evidence that verbal working memory (Adams & Hitch, 1997; Gathercole, Pickering, Knight, & Stegmann, 2004; Owens et al., 2008), visuospatial working memory (Holmes, Adams, & Hamilton, 2008; Kyttälä & Lehto, 2008), or both (Berg, 2008; Holmes & Adams, 2006; Jarvis & Gathercole, 2003; Miller & Bichsel, 2004) are related to performance. Given the integrative functions of the central executive and episodic buffer, it is not surprising that both visuospatial and verbal working memory have been implicated in solving math problems. The extent to which visuospatial versus verbal processing is involved in problem solving may vary—it may be dependent on the particular features of the mathematical task.

At present, there has been only one study that examined working memory as a mediator of the relation between anxiety and math performance (Owens et al., 2008). These researchers found that working memory indeed partially explained the relation between anxiety and math performance in fifth-grade children. However, gender was not examined as part of this model, and therefore it is not clear if the same explanatory mechanism can be applied to understanding the nature of gender differences in math performance.

Anxiety and Working Memory as Mediators of Gender Differences in Math Performance

The integration of the bodies of research reviewed above suggests the possibility that anxiety and working memory together play a role in gender differences in math performance. First, there is substantial research demonstrating gender differences in anxiety, in its many manifestations (Hembree, 1988, 1990; Miller & Bichsel, 2004). Second, there are studies showing that anxiety is related to working memory (Owens et al., 2008; Rapee, 1993; Shackman et al., 2006). Finally, there is evidence that anxiety and working memory are both related to math performance (Hembree, 1988; Holmes & Adams, 2006; Jarvis & Gathercole, 2003; Miller & Bichsel, 2004). Combining these findings suggests that female students may be more anxious about a math testing situation, that their heightened anxiety may interfere with working memory, and this interference may be related to poor performance in math, as shown in Figure 1. To our knowledge, the present studies are the first to empirically test this entire model.

The Present Research

This research expands on previous work that has examined individual pieces of the model depicted in Figure 1. We conduct two studies testing this model in two different samples to better assess the generalizability of findings. In both studies, we chose college students as participants, because gender differences in math have been found to be particularly robust at this age, and we presented these students with challenging math problems (Lindberg et al., 2010). We first examined whether the math test revealed gender differences in each of the samples. Then we tested the model in which anxiety and working memory create a mediating chain between gender and math performance.

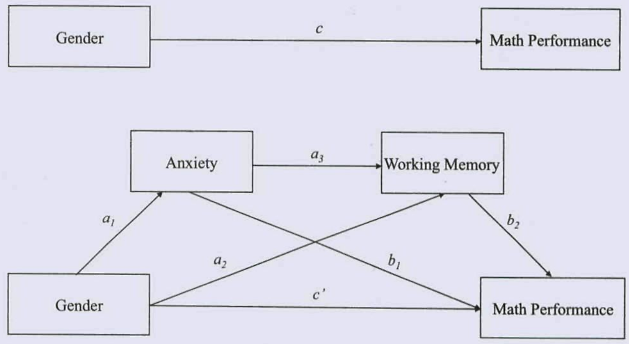


Figure 1. Proposed mediation model for the relation between gender and math test performance.

In considering potential mediators of the gender difference in mathematics, we examined *both types of working memory (verbal and visuospatial)*, so that the component(s) of working memory that are impacted by anxiety and related to math performance could be better understood in the context of gender differences. The type of anxiety examined in the present studies was the *worry component of state anxiety*. As noted earlier, state anxiety reflects the individual's experience during a testing situation and thus may be a stronger predictor of performance than trait anxiety (O'Neil & Fukumura, 1992). The worry component refers to anxious thoughts, such as concerns related to performing poorly, as opposed to the emotionality component, which refers to the physiological arousal response (Deffenbacher, 1977, 1978; Liebert & Morris, 1967). We focus on the worry component of anxiety because some research suggests that it is more strongly related to working memory and test performance than is emotionality (Deffenbacher, 1980; Hembree, 1988; Kim & Rocklin, 1994; Zeidner & Nevo, 1992).

Study 1

The first study was conducted as part of a larger investigation, which also examined effects of stereotype threat on the math performance of female students. Thus, half of the students in the sample were tested under stereotype threat conditions, whereas the other half were not. We found that performance did not vary as a function of the stereotype threat manipulation—this variable did not produce either main effect or interactions. Therefore, in the present study we included students from both conditions. Because the main focus of our study was on the relation between working memory, anxiety, and math performance, and stereotype threat condition did not interact with any of these variables, we did not include it in the analyses reported here. It should be noted that when we tested our models using stereotype threat condition as a covariate, we found the same pattern of results.

Method

Participants. Participants were 87 undergraduate students from a private university in the Northeast region of the United States. The university has been categorized as "most selective" (U.S. News & World Report, 2012). Students were recruited from introductory level psychology courses by advertising the study in their classes. There were no particular selection restrictions; everyone who signed up for the study was tested. Students received extra credit in their course for participation. The sample included 63 female students and 24 male students. This gender distribution was approximately proportional to that of the students in the courses from which participants were recruited. On average, participants were 19 years, 11 months of age. Sixty-six percent of participants took calculus in high school (62% of female students and 75% of male students), and, on average, students had taken between one and two math courses in college (M = 1.15; female students M = 0.97, male students M = 1.63).[1]

Materials. Participants completed four measures during the testing session: a measure of the worry component of state anxiety, two working memory assessments (verbal and visuospatial), and a mathematics test.

Worry. Worry was measured with a four-item self-report scale, adapted from Morris, Davis, and Hutchings (1981). Participants rated their feelings on a scale ranging from 1 (*strongly disagree*) to 4 (*strongly agree*) for statements about their worry (e.g., "I am worried that I may not do well on this test"), two of which were reverse coded (e.g., "I feel very confident about my performance on the test I'm about to take"). A worry score was calculated by summing the responses on the four items after appropriate reverse coding. The internal consistency for the worry scale as measured by Cronbach's alpha was .78.

Working memory. Two dual-task working memory measures were administered to the participants on a computer: a verbal working memory measure (a word recall task) and a visuospatial working memory measure (a spatial recall task).

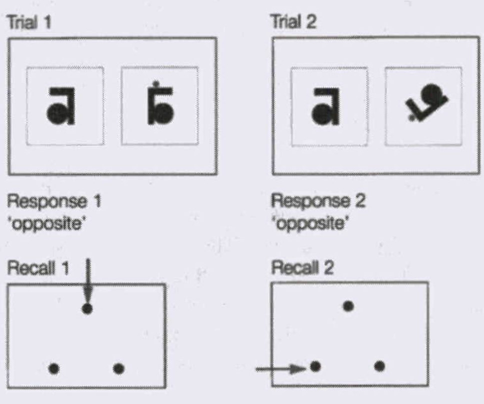


Figure 2. Sample visuospatial working memory task trial. From *Automated Working Memory Assessment* by T. P. Alloway, S. E. Gathercole, and S. Pickering, 2004. Reprinted with permission. Copyright by Pearson Education.

Verbal working memory. To assess verbal working memory, we used a word recall task (listening recall) from the Automated Working Memory Assessment (AWMA; Alloway, Gathercole, & Pickering, 2004). In each trial, the students heard a series of sentences and judged whether they were true or false. After they heard 1–6 sentences in a row that they judged true or false, they were asked to recall the last word of each sentence in the order in which the sentences were presented. For example, in a trial with two sentences, the participants heard "Magazines have pages" and said true or false, then heard "Apples play football" and said true or false, and then were asked to say the last word of each of the sentences in order ("pages, football"). The task was divided into six blocks: The first block consisted of trials that included only one sentence, the second had trials with two sentences, and so forth, up to six sentences. Each block had a maximum of six trials with a particular number of sentences. Due to the ease of trials within the first three blocks (i.e., with 1, 2, or 3 sentences), students completed only one trial from each of these blocks. For the remaining blocks (4, 5, or 6 sentences), the participants completed trials within the block until they either (a) got 4 trials correct, in which case they moved on to the next block without completing the last two trials in the current block, or (b) got 3 trials incorrect, in which case the participants were finished with the task. Thus, the verbal working memory task measured the ability to remember words from several sentences while making true/false judgments about the meaning of each sentence. The outcome measure was the recall score: the number of times a participant accurately recalled the words in the correct order. The test–retest reliability for this measure is .81 (Alloway, Gathercole, & Pickering, 2006).

Visuospatial working memory. To assess visuospatial working memory, we used the spatial recall task, also from the AWMA (Alloway et al., 2004). The task followed a parallel structure to the word recall task but with visual, instead of verbal, stimuli. On each trial, two shapes were presented next to each other (see Figure 2 for sample). The shape on the right was either identical to or a mirror image (opposite) of the shape on the left. Importantly, the shape on the right was rotated relative to the shape on the left by 0, 120, or 240 degrees; a red dot marked the top of the shape to indicate the degree of rotation. The task was divided into seven blocks: The first block consisted of trials that included only one pair of shapes; the second had trials with two pairs of shapes, and so forth, up to seven pairs of shapes. Within a trial, participants had to determine if the shape on the right was the same or opposite of the shape on the left. After making the same–opposite judgments for all pairs of shapes within a trial, participants had to point to one of three dots (at 0, 120, and 240 degrees) that matched the dot location for the shape on the right in the order in which the pairs of shapes were presented. Thus, the visuospatial working memory task measured the participants' ability to remember the location of a dot across trials while making same/opposite judgments about shapes. The administration procedure for the trial blocks was parallel to that in the verbal working memory task (i.e., students completed one trial for 1, 2, and 3 pairs of shapes and up to 4 trials for 4, 5, 6, and 7 pairs). Participants' recall scores, which were based on their correct recall of the dot locations, were used as the

[1] Analyses were also run with the number of college courses that students had taken as a covariate, which did not change the results.

Table 1
Study 1: Descriptive Statistics: Means, Standard Deviations, and Effect Sizes for Gender Differences

Measure	Female students		Male students		Cohen's *d*
	M	*SD*	*M*	*SD*	
Worry scale	10.10	2.09	7.63	1.88	-1.25^{**}
Verbal working memory	22.63	4.16	23.21	3.99	0.14
Visuospatial working memory	25.83	6.39	32.79	5.41	1.18^{**}
Math score (%)	40.67	18.99	56.71	21.62	0.79^{**}

Note. The scale for worry is from 4 to 16; the scale for verbal working memory is from 0 to 36; the scale for visuospatial working memory is from 0 to 42; the scale for math score is percent accuracy, ranging from 0 to 100.
$^{**}p < .01$.

outcome measure for this task. The test–retest reliability for this measure is .82 (Alloway et al., 2006).

Math performance. Participants solved 12 math problems (see the Appendix) presented in a multiple-choice format (six algebra items, six geometry/measurement items). These test items were selected from several widely-used standardized tests: National Assessment of Educational Progress (NAEP) for 12th grade, the Massachusetts Comprehensive Assessment System (MCAS) for 10th grade, and the SAT. The items were selected by a panel of experts in math and math education based on the coverage of mathematical content. Another criterion for item selection was difficulty level. Because gender differences are often found on more difficult problems (Gibbs, 2010; Lindberg et al., 2010), we chose items that showed 30%–40% accuracy rates in national samples of high school seniors. In a separate study, we obtained evidence for the criterion-related validity of the math test by showing a high correlation between accuracy scores obtained on this test and the scores on the math portion of the state-administered standardized test, MCAS ($r = .63$, $p < .001$). The internal consistency for the mathematics test as measured by Cronbach's alpha was .61. Participants completed the six algebra problems in one block and the six geometry/measurement problems in another, and they were given 6 min for each block. Math scores were calculated as the percent of items answered correctly (out of the total number of items attempted).[2]

Procedure. Participants were tested individually; each student received a booklet that included the worry scale and the math test. Prior to conducting the study, the purpose was explained to the students as follows, "The goal of this study is to understand the relationship between what students think while doing math, how they do in math, and their memory skills." At the beginning of testing, students were told, "The main thing we are going to do is a math test. The math test is made up of 12 word problems that are multiple-choice. Also you will be answering some questions about your thoughts and feelings and doing some memory tasks." Prior to completing test measures, participants were given a practice math problem, which included a stereotype threat manipulation (the content of the problem either activated the gender stereotype or did not). Immediately after the practice math problem, they completed the worry scale. The memory tasks were administered next, with the order of the verbal and visuospatial measures counterbalanced. Finally, students completed the math test, with the order of the algebra and geometry/measurement blocks counterbalanced.

Results

Preliminary and descriptive analyses. Prior to conducting the main analysis of interest, we ran multiple sets of preliminary analyses. In one set we examined whether the order of task administration affected performance. We found that the order of administering the two types of working memory tasks and the order of administering the two types of math problems did not affect the outcome variable, students' math scores ($p = .47$ and .94, respectively), and therefore the data were collapsed across the different orders. Then we examined whether performance varied across the two types of math items (algebra and geometry/measurement). The algebra items ($M = 42\%$) were slightly more difficult than the geometry/measurement items ($M = 47\%$), but this difference was not statistically significant ($p = .11$). Therefore, in the main analysis we relied on the overall math test score across content areas. Finally, as indicated earlier, a preliminary analysis showed that the stereotype threat manipulation did not produce either a main effect or any interactions, and thus the data from both conditions were combined in further analysis.

It should be noted that when conducting analyses of variance (ANOVAs), unequal sample size can be problematic (as was the case for the gender distribution in the present study). However, it is only an issue if there is heterogeneity of variance (Keppel & Wickens, 2004). Thus, we conducted Levene's test of homogeneity of variance, which was nonsignificant across all outcomes ($.28 < ps < .93$). This indicated that the unequal sample size does not present a serious problem in our subsequent analyses because we found no evidence for heterogeneity of variance.

Descriptive statistics by gender are presented in Table 1. As shown in the table, significant gender differences were found in worry, visuospatial working memory, and math performance ($.79 < |ds| < 1.25$) but not in verbal working memory. Table 2 presents correlations between the variables. These correlations show that worry, visuospatial working memory, and math performance were all significantly correlated with one another. Verbal working memory was related to visuospatial working memory and worry but not to math performance. This nonsignificant relation, in combination with the lack of a gender difference in verbal working memory, informed the decision to exclude verbal working memory

[2] The same pattern of results was obtained when analyses were run with the percent correct out of all items as well as when the scores were corrected for guessing (Frary, 1988).

Table 2
Study 1: Correlations Among Measures

Measure	Gender	Worry scale	Verbal working memory	Visuospatial working memory
Worry scale	$-.48^{**}$			
Verbal working memory	.06	$-.25^{*}$		
Visuospatial working memory	$.46^{**}$	$-.41^{**}$	$.33^{*}$	
Math score	$.35^{**}$	$-.47^{**}$	.19	$.51^{**}$

$^{*}p < .05$. $^{**}p < .01$.

as a potential mediator in the model. Thus, only visuospatial working memory was used in the subsequent mediation analysis.

Analytic strategy for mediation analysis. After we established gender difference in students' math test scores, the next step was to test whether worry and working memory served as a potential mechanism through which gender was linked to math test performance. In statistical terms, this pattern is captured in a model in which worry and working memory mediate the relation between gender and math performance (Hayes, Preacher, & Myers, 2011). To test for mediation, we ran a series of regression models.[3] The first two regression analyses tested whether the predictor, gender (females = 0, males = 1), was related to the outcome measure, math test score (Equation 1), and the first potential mediator, worry (Equation 2). Next, we ran a regression model that included gender and worry as predictors of visuospatial working memory (Equation 3). The last regression analysis tested whether math test performance was predicted from gender, worry, and visuospatial working memory (Equation 4).

$$\text{Math Test Score} = \beta_{01} + c(\text{Gender}) + e_1 \qquad (1)$$

$$\text{Worry} = \beta_{02} + a_1(\text{Gender}) + e_2 \qquad (2)$$

$$\text{Visuospatial Working Memory} = \beta_{03} + a_2(\text{Gender}) + a_3(\text{Worry}) + e_3 \qquad (3)$$

$$\text{Math Test Score} = \beta_{04} + c'(\text{Gender}) + b_1(\text{Worry}) + b_2(\text{Visuospatial Working Memory}) + e_4 \qquad (4)$$

Figure 1 depicts the model and the coefficients that represent each effect from these four analyses. Critical to the present analysis is the combined path from gender to worry (a_1), worry to working memory (a_3), and working memory to math test score (b_2). Mediation is likely evident if the relation between gender and math test score (c) disappears or significantly decreases after the addition of worry and visuospatial working memory into the model (c').

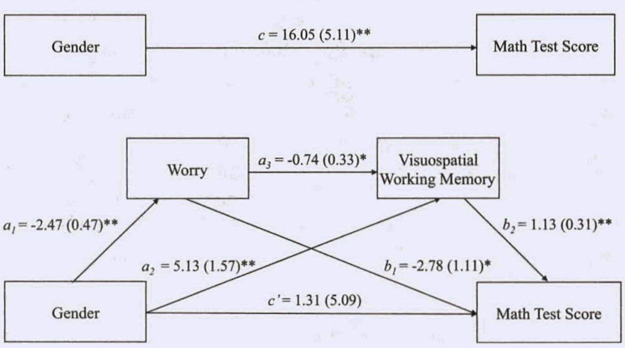


Figure 3. Study 1 mediation analysis results. $^{*}p < .05$. $^{**}p < .01$.

The indirect effects were tested by calculating the product of the path coefficients (henceforth referred to as a point estimate) between the predictor and mediator(s) and between the mediator(s) and the outcome. The three specific indirect effects were (1) the indirect effect of gender on math test score through worry only ($a_1 \times b_1$), (2) the indirect effect of gender on math test score through visuospatial working memory only ($a_2 \times b_2$), and (3) the indirect effect of gender on math test score through worry and then visuospatial working memory ($a_1 \times a_3 \times b_2$). The test of the total indirect effect is the sum of the three specific indirect effects ($a_1 \times b_1 + a_2 \times b_2 + a_1 \times a_3 \times b_2$). This is equivalent to testing the difference between the total effect of gender on math performance (c) and the direct effect of gender on math test performance (c'). Note that the total effect, c, captures the relation between gender and math test performance without taking any other variables into account, whereas the direct effect, c', captures the remaining relation between gender and math test score after taking into account the contributions of worry and working memory to math test performance. The main interest of the present study is the third indirect effect: the path from gender, to worry, to visuospatial working memory, to math test score.

To estimate confidence intervals for the point estimate of the mediated effect, we used bias-corrected bootstrapping (Dearing & Hamilton, 2006; Hayes et al., 2011; Preacher & Hayes, 2008). This method was recommended by Taylor, MacKinnon, and Tein (2008), who conducted a Monte Carlo study comparing the Type I error, power, and coverage of six different methods for estimating this type of mediation model (i.e., a three-path model). Bootstrapping has an advantage over other tests of the indirect effect (e.g., the Sobel test) in that it does not assume that the sampling distribution of the indirect effect is normal (Hayes et al., 2011). The procedure, instead, involves resampling to estimate the characteristics of the sampling distribution and uses this information to create confidence intervals around the point estimate. Specifically, the bootstrapping strategy used in the present study involved selecting 50,000 bootstrap samples with replacement from the

[3] Another way to test the paths that we were interested in is to use structural equation modeling with maximum likelihood estimation. It should be noted that this approach has some advantages, but ordinary least squares (OLS) regression is better here because it is more difficult to establish whether one has satisfied the strict assumptions necessary for using maximum likelihood estimation when the sample size is small (Ding, Velicer, & Harlow, 1995).

Table 3
Study 1: Magnitude of Indirect Effects

Effect type	Formula	Point estimate	95% bias-corrected confidence interval	
			Lower limit	Upper limit
Total indirect effect	$a_1 \times b_1 + a_2 \times b_2 + a_1 \times a_3 \times b_2$	14.74	9.40	21.87
Indirect effect through worry only	$a_1 \times b_1$	6.86	2.34	13.22
Indirect effect through visuospatial working memory only	$a_2 \times b_2$	5.81	1.82	12.45
Indirect effect through worry, then visuospatial working memory	$a_1 \times a_3 \times b_2$	2.07	0.51	5.26

Note. If the confidence interval does not overlap with 0, it indicates a significant indirect effect.

current sample and then calculating the point estimates for the indirect effects within each of these samples (Preacher & Hayes, 2008).[4] Then, the distribution of the point estimates obtained in these samples was used to create 95% confidence intervals based on the sampling distribution of these estimates (Hayes et al., 2011). The mediation analysis was run with the PROCESS SPSS macro created by Hayes et al. (2011), which uses bootstrapping and reports bias-corrected confidence intervals. Using this method, the total indirect effect as well as the three specific indirect effects were examined. An indirect effect was considered significant if its 95% confidence interval did not overlap with zero (Preacher & Hayes, 2004).

Estimation of the mediation model. Results of the mediation analysis are presented in Figure 3. In the first step of the analysis, we ran the regression models described in Equations 1 and 2. The results showed that gender was significantly related to both the outcome measure—math test score (Equation 1: $c = 16.05$, $s_c = 5.11$, $p = .002$)—and the potential first mediator—worry (Equation 2: $a_1 = -2.47$, $s_{a1} = 0.47$, $p < .0001$). The next regression analysis (Equation 3) included gender and worry as predictors of visuospatial working memory performance. The findings revealed that both gender ($a_2 = 5.13$, $s_{a2} = 1.57$, $p = .002$) and worry ($a_3 = -0.74$, $s_{a3} = 0.33$, $p = .03$) were significant predictors of visuospatial working memory. The final regression analysis predicted math performance from gender, worry, and visuospatial working memory performance. The results showed that gender was no longer a significant predictor of math test performance ($c' = 1.31$, $s_{c'} = 5.09$, $p = .80$), but both worry ($b_1 = -2.78$, $s_{b1} = 1.11$, $p = .014$) and visuospatial working memory ($b_2 = 1.13$, $s_{b2} = 0.31$, $p = .0005$) were significant predictors.

As a final step in the analysis, we estimated confidence intervals for the different mediation effects. Specifically, we used bias-corrected bootstrapping to calculate confidence intervals—both for the entire model and for each of the three specific indirect effects (through worry only, visuospatial working memory only, and both worry and visuospatial working memory). As seen in Table 3, the total indirect effect was significant. Critical to our research question, the indirect effect of gender on math test score through both worry and visuospatial working memory ($a_1 \times a_3 \times b_2$) was significant with a point estimate of 2.07 (95% CI [0.51, 5.26]). Thus, there was evidence of a mediating chain from gender to math test performance through worry and then visuospatial working memory.[5] The indirect effect through worry and then visuospatial working memory (2.07% on the math test) is 12.92% of the total effect (16.05% on the math test).

Study 2

Study 2 addressed some of the potential weaknesses of Study 1. First, we recruited a larger sample of students (118 vs. 87) that had a more balanced distribution of males and females (56 male students, 62 female students). Second, we recruited students from a university that varied in a number of ways from that in Study 1: It was a less selective public university with a more diverse student body and was located in a different region of the United States. Therefore, we were able to determine whether the results obtained in Study 1 would generalize to another group of college students. Third, we did not use a stereotype threat manipulation in this study, and thus any potential influence of stereotype threat was eliminated.

Method

Participants. Participants were 118 undergraduate students from a public university in the Midwest region of the United States. The university has been categorized as one of the "more selective" national universities (U.S. News & World Report, 2012); its acceptance rate is about 2.5 times higher than the university where participants were recruited for Study 1. Participants were recruited from introductory level psychology courses by advertising the study in their classes. There were no particular selection restrictions; everyone who signed up for the study was tested. Students received either subject pool credit, extra credit in their course, or $5 for participation. The sample included 62 female students and 56 male students. Participants were on average 20 years, 7 months of age. Sixty-two percent of participants took a calculus course in high school (57% of female students and 68% of male students), and, on average, students had taken between one and two math courses during college ($M = 1.79$; female students $M = 1.52$, male students $M = 2.09$).

[4] It is important to note that the bootstrapping algorithm will give somewhat different values for the confidence intervals around the point estimate each time the bootstrapping is done because it is using different samples within the larger sample to compute the point estimates. Despite this issue, this method still offers multiple advantages over other existing methods.

[5] We compared the current model that includes both mediators with models that include each mediator separately and found that the model with both mediators explained a larger proportion of the total effect.

Table 4
Study 2: Descriptive Statistics: Means, Standard Deviations, and Effect Sizes for Gender Differences

	Female students		Male students		
Measure	*M*	*SD*	*M*	*SD*	Cohen's *d*
Worry scale	9.04	1.98	7.39	2.21	−0.79**
Verbal working memory	22.42	7.38	22.20	4.38	−0.04
Visuospatial working memory	26.53	6.40	31.66	6.59	0.79**
Math score (%)	36.29	20.62	56.76	22.69	0.94**

Note. The scale for worry is from 4 to 16; the scale for verbal working memory is from 0 to 36; the scale for visuospatial working memory is from 0 to 42; the scale for math score is percent accuracy, ranging from 0 to 100.
** $p < .01$.

Materials. Participants completed the same four measures used in Study 1: a worry scale ($\alpha = .84$), two working memory tasks (verbal and visuospatial), and a mathematics test ($\alpha = .70$).

Procedure. The procedure was identical to that in Study 1 except that this study did not include a stereotype threat manipulation. All other measures were administered in the same way.

Results

Preliminary and descriptive analyses. Preliminary analyses showed that the order of administering the two types of working memory tasks and the order of administering the two types of math problems had no effect on students' math test scores ($p = .94$ and .53, respectively). Thus, the data were collapsed across different orders. We also found that the algebra items ($M = 44\%$) were slightly more difficult than the geometry/measurement items ($M = 48\%$), but this difference was not statistically significant ($p = .08$), and thus, parallel to Study 1, we used the overall math test score as the dependent variable in main analyses.

Descriptive statistics by gender are presented in Table 4. As shown in the table, significant gender differences were found in worry, visuospatial working memory, and math performance ($0.79 < |ds| < 0.94$), but not in verbal working memory. Correlations among the variables are presented in Table 5. These correlations show that worry, visuospatial working memory, and math performance were all significantly related to one another. Verbal working memory was related to visuospatial working memory, but it was not correlated with either worry or math performance and was therefore excluded from mediation analysis. In sum, descriptive analyses—both in terms of gender differences on the variables of interest and correlations between these variables—showed similar patterns to those in Study 1.

Estimation of the mediation model. We conducted a mediation analysis parallel to that in Study 1—testing the indirect path from gender to worry to visuospatial working memory to math test performance (see Figure 4). In the first step of the analysis, we ran the regression models described in Equations 1 and 2 above. The results showed that gender was significantly related to both the outcome measure—math test score (Equation 1: $c = 20.47$, $s_c = 4.04$, $p < .0001$)—and the potential first mediator—worry (Equation 2: $a_1 = -1.65$, $s_{a1} = 0.39$, $p = .0001$). The next regression analysis (Equation 3) included gender and worry as predictors of visuospatial working memory performance. The findings revealed that both gender ($a_2 = 3.64$, $s_{a2} = 1.33$, $p = .007$) and worry ($a_3 = -0.90$, $s_{a3} = 0.27$, $p = .001$) were significant predictors of visuospatial working memory performance. The final regression analysis predicted math test scores from gender, worry, and visuospatial working memory performance. Results showed that although the coefficient for gender decreased from 20.47 to 13.67, it was still a significant predictor of math test performance ($c' = 13.67$, $s_{c'} = 4.36$, $p = .002$), as was visuospatial working memory ($b_2 = 0.86$, $s_{b2} = 0.31$, $p = .007$), but worry ($b_1 = -1.46$, $s_{b1} = 0.88$, $p = .10$) was not a significant predictor.

As a final step in the analysis, we estimated bias-corrected bootstrap confidence intervals for the mediation effects. The results are presented in Table 6, and as can be seen, the total indirect effect was significant. The key finding is that the indirect effect of gender on math test score through both worry and visuospatial working memory ($a_1 \times a_3 \times b_2$) was significant with a point estimate of 1.28 (95% CI [0.35, 3.33]). Thus, there was again evidence of a mediating chain from gender to math test performance through worry and then visuospatial working memory. The indirect effect through worry and then visuospatial working mem-

Table 5
Study 2: Correlations Among Measures

Measure	Gender	Worry scale	Verbal working memory	Visuospatial working memory
Worry scale	−.37**			
Verbal working memory	−.02	.01		
Visuospatial working memory	.37**	−.39**	.19*	
Math score	.43**	−.34**	−.06	.41**

* $p < .05$. ** $p < .01$.

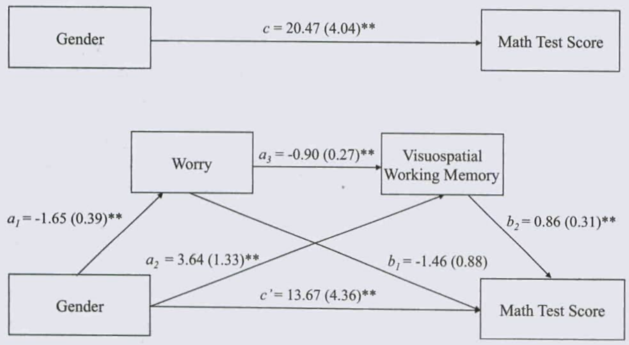


Figure 4. Study 2 mediation analysis results. ** $p < .01$.

ory (1.28% on the math test) is 6.24% of the total effect (20.47% on the math test).

Discussion

The main goal of this work was to examine the cognitive and affective factors related to gender differences in math performance. This investigation grew from research on gender socialization, which suggests that as women are socialized, they are exposed to gender stereotypes, and this may lead them to be anxious about their ability to do well in mathematics (Eccles & Jacobs, 1986). Past research suggested that heightened anxiety may have negative consequences for cognitive processing, which in turn, could lead to poor performance on math tests. Both lower levels of performance and the anxiety itself may deter female students from choosing educational paths leading to STEM careers. To develop educational programs aimed at reducing the gender gap in math-intensive fields, it is critical to better understand whether the affective and cognitive factors suggested in the literature as potentially underlying gender differences in math achievement form a mediating chain from gender to math performance. The present research provided the first empirical investigation of which we are aware testing the entire mediational relation from gender differences in math anxiety to working memory to math performance.

Gender Differences in Math Performance

Recently, there has been a discussion in the literature concerning the magnitude and even the very existence of gender differences in math achievement (Hyde et al., 2008; Lindberg et al., 2010). Some evidence suggests that these differences have been decreasing over time. In particular, when looking at grades or standardized test performance in school students, researchers often find comparable levels of performance for males and females, and sometimes a female advantage, especially when measured in terms of grades (e.g., Pomerantz, Altermatt, & Saxon, 2002). The findings of a male advantage appear to be more persistent at the college level, where the evidence consistently shows a small but significant gender difference (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010).

Further, it has been argued that math problems that are more complex tend to show larger differences favoring males (Gibbs, 2010). Thus, in the present studies we utilized a test instrument that included items proven to be challenging in earlier large-scale assessments, to maximize our chances of finding a gender difference. Our results showed that the mean scores for both genders were quite low, indicating that this was indeed a difficult assessment for our participants (on average, female students = 38% correct, male students = 57% correct). More importantly, there was a robust gender difference (Study 1 $d = 0.74$; Study 2 $d = 0.94$). The magnitude of this difference was substantially larger than in research using less difficult assessments (Hyde, Fennema, & Lamon, 1990; Lindberg et al., 2010). Note that the ability to solve challenging math problems may be particularly predictive of an individual's desire to persist in seeking a career in a math-intensive field. The present findings, consistent with Gibbs's (2010) arguments, highlight the importance of considering the nature of the math task, and particularly, the level of difficulty, in discussing the existence and magnitude of gender differences.

Mediators of the Relation Between Gender and Math Performance

The key novel finding of the present work is the mediating pathway from gender to math performance created by anxiety and working memory. These results suggest that one possible reason gender differences in math performance might exist is because of increased worry in female students, which taxes their visuospatial working memory, which, in turn, leads them to perform more poorly on difficult math assessments. We should be clear: The finding of a mediated relation between variables does not provide evidence of a causal link between them. Yet, it provides a critical

Table 6
Study 2: Magnitude of Indirect Effects

Effect type	Formula	Point estimate	95% bias-corrected confidence interval: Lower limit	Upper limit
Total indirect effect	$a_1 \times b_1 + a_2 \times b_2 + a_1 \times a_3 \times b_2$	6.80	3.27	11.82
Indirect effect through worry only	$a_1 \times b_1$	2.40	−0.13	5.71
Indirect effect through visuospatial working memory only	$a_2 \times b_2$	3.13	0.75	7.25
Indirect effect through worry, then visuospatial working memory	$a_1 \times a_3 \times b_2$	1.28	0.35	3.33

Note. If the confidence interval does not overlap with 0, it indicates a significant indirect effect.

piece of evidence for the possibility of such a link, which can be further addressed in experimental work.

The present results generally fit with prior work that has shown evidence for the individual pathways involved in the model. Here, we combined pathways associated with cognitive and affective predictors and found that together they explain a significant part of the relation between gender and math performance. In fact, the model that includes both mediators is stronger (i.e., explains a larger portion of the total effect) than models that include each of the mediators separately. Further, our studies allowed us to take a more in-depth look into specific affective and cognitive factors that are associated with gender differences in math performance.

The role of worry. In these studies, the worry component of anxiety, as measured before a math test, was found to be involved in the mediational path from gender to math performance. The present analyses confirmed the bivariate relations documented in prior research, namely, relations between gender and anxiety (Hembree, 1988, 1990; Miller & Bichsel, 2004), anxiety and working memory (Ikeda et al., 1996; Lee, 1999; Owens et al., 2008; Shackman et al., 2006), and anxiety and math performance (Ganley & Vasilyeva, 2011; Miller & Bichsel, 2004; Osborne, 2001). Two previous studies examined math anxiety in relation to gender differences in math performance with mixed results. Osborne's (2001) findings supported the mediational role of anxiety, whereas Casey et al.'s (1997) findings did not. To understand potential sources of inconsistency between the findings, we take a closer look at the specific dimensions of anxiety measured in these studies.

The first potentially relevant distinction is between trait and state anxiety. The study by Casey et al. (1997) utilized a measure of trait math anxiety (i.e., a general tendency to feel anxious about math tests), whereas the present studies as well as Osborne's (2001) study explicitly tested state anxiety experienced during the current testing situation. Of course, the two types of anxiety are not completely independent: Individuals with higher trait anxiety are more likely to have higher state anxiety (Hembree, 1990; Miller & Bichsel, 2004). At the same time, these two types of anxiety are viewed as separate constructs. It has been suggested that due to its proximity to the testing situation, state anxiety may show a stronger relation to performance than trait anxiety (e.g., O'Neil & Fukumura, 1992). This may, in part, explain why studies looking at trait anxiety arrive at different findings than the studies testing state anxiety. In the current studies, state anxiety in some students may have been particularly heightened by their knowledge that they would have a rather limited amount of time to take the math test.

Another potentially relevant distinction is between the two components of anxiety—worry (anxious thoughts) and emotionality (affective arousal; e.g., feeling tense or nervous/jittery). These components of anxiety have been found to differentially relate to performance (Deffenbacher, 1980; Hembree, 1988; Kim & Rocklin, 1994; Zeidner & Nevo, 1992). Worry, in particular, appears to be strongly linked to working memory, as monitoring anxious thoughts utilizes a substantial amount of working memory resources. In the study by Casey et al. (1997), the anxiety scale combined worry and emotionality items. It is possible that the relation between worry and math performance was diluted by combining it with emotionality. Thus, the type of anxiety (trait vs. state) and the component of anxiety measured (worry vs. emotionality) are critical dimensions to consider in examining the relation between anxiety and math performance.

Working memory and its relation to worry and math performance. Prior research has allowed us to identify working memory as a potential mediator of the relation between anxiety and math performance. There are multiple findings pointing to the link between anxiety and working memory (e.g., Ikeda et al., 1996; Lee, 1999; Owens et al., 2008; Shackman et al., 2006), and between working memory and academic performance (e.g., Adams & Hitch, 1997; Berg, 2008; Holmes et al., 2008; Owens et al., 2008). The findings, however, have been somewhat inconsistent in implicating verbal versus visuospatial working memory, or both, in math performance. The present research contributed to the literature by directly examining both types of working memory in relation to gender, anxiety, and math performance. Consistent with previous research, we found that visuospatial and verbal components of working memory were moderately correlated with each other (Gilhooly, Wyn, Phillips, Logie, & Della Sala, 2002; St Clair-Thompson & Gathercole, 2006). We found mixed results in terms of the relations between the two types of working memory and worry. In both studies, visuospatial working memory was related to worry; however, verbal working memory was correlated with worry in Study 1 but not in Study 2. The results of Study 1 fit with the processing efficiency theory (Eysenck & Calvo, 1992), which suggests that anxiety affects the central executive component of working memory that is responsible for processing information stored in verbal and visuospatial memory. However, the results of Study 2 suggest that perhaps there is something more specific about visuospatial working memory that worry interferes with. Future research should further examine these relations.

The pattern of relations between working memory and the other two variables examined in the present studies—gender and math performance—was different for visuospatial and verbal measures. In particular, visuospatial working memory was strongly related to both gender and math performance, whereas verbal working memory was not significantly related to either gender or math test performance. Thus, verbal working memory could not act as a mediator, given a nonsignificant relation with both the predictor and outcome.

The lack of involvement in the mediation model for verbal working memory is somewhat surprising given other studies that have implicated this component of working memory as a mediator of the relation between anxiety and math performance (Owens et al., 2008). Based on current conceptualizations of working memory, which emphasize the integrative role of the episodic buffer, one could expect a mixture of visuospatial and verbal working memory effects on math performance. Yet, it is important to note that the present studies, in contrast to the study done by Owens et al. (2008), focused specifically on the role of working memory in the context of gender differences. There have been numerous studies revealing a male advantage in visuospatial abilities (e.g., Voyer, Voyer, & Bryden, 1995), thus the inclusion of gender in the model may explain the strength of visuospatial working memory as a predictor of the observed differences in math performance.

Limitations and Future Directions

One limitation of the present studies pertains to the measures used. With regard to the math assessment, the test consisted of only 12 items (6 in each content area) due to constraints on how much time we could request from undergraduate students for test taking. We did not find differences between algebra and geometry/measurement items, and it is possible that the small number of problems within each content area made it difficult to find such differences. A more extensive test with a more comprehensive representation of each domain would be useful in comparing gender effects and the mediational model across mathematical content.

With regard to the measure of anxiety, we used a worry scale that was based on self-report. Although this is a common approach to assessing affective states, it may have introduced a confound—it is possible that male students were less likely to report worry as it would not be seen as acceptable. However, there is evidence suggesting that the gender difference in worry is unlikely to be due to males' general tendency to downplay their worry. In particular, in other subjects (e.g., English) male students report lower levels of confidence than do female students (Correll, 2001). This may reflect cultural expectations about the strengths and weaknesses of certain groups in certain academic areas. Using multiple measures of anxiety and not just self-report may allow one to get a more complete picture of a person's emotional state during problem solving.

With respect to the overall method, as we mentioned earlier, the mediation analysis is correlational in nature, and thus causal relations cannot be established definitively. Most likely, some relations reported here are bidirectional in nature (e.g., the relation between worry and math performance). Conceptually, mediation offers a potential mechanism underlying the relation between the predictor and outcome (Dearing & Hamilton, 2006). Establishing this type of relation is critical to the investigation of gender differences in math performance. Yet, the strongest test of the relations uncovered in the present studies would be provided by an experimental investigation of whether changes in anxiety and working memory could lead to changes in math outcomes.

We cannot make causal claims based on the present studies; however, there have been a few studies exploring the causal relation between test anxiety and performance, although not focusing specifically on mathematics (Hembree, 1988; Wood, 2006). In a meta-analysis of intervention studies, Hembree (1988) found that anxiety during testing can be reduced through behavioral and cognitive-behavioral interventions and that this diminished anxiety is followed by an improvement in test performance. Further, a number of studies have provided experimental evidence that cognitive interventions can lead to improvements in working memory (Klingberg, Forssberg, & Westerberg, 2002; Thorell, Lindquist, Nutley, Gunilla, & Klingberg, 2009) and that there are subsequent gains in math performance (Holmes, Gathercole, & Dunning, 2009). Although these studies provide important evidence of malleability of the factors related to gender differences in mathematics, none of them have yet looked at the joint role of anxiety and working memory. In future work, it would be useful to design intervention studies aimed at examining whether the mediational chain uncovered in the present research can be reproduced through experimental manipulations. That is, would the interventions designed to alleviate math anxiety increase the working memory of female students during testing, and would this increase lead to an improvement in math performance? Researchers' ability to answer such questions could have important consequences for women's achievement in math-intensive fields.

Finally, it would be important to explore the mechanisms addressed in the present studies not only with the goal of explaining between-group gender differences but also to identify potential sources of individual difference within groups. After all, there are substantial individual differences within each gender that need to be better understood. Furthermore it may be useful to examine the strength of the predictive relations for males compared to females. For example, the relation between worry and math performance and the extent to which this relation is mediated by working memory may be different for the two gender groups.

Educational Implications

The findings of the present studies not only lead to important suggestions for future research but also have potential implications for educational practice. One of the key findings in this respect is the evidence of large gender differences on a difficult mathematics test among college students. This finding adds to the recent evidence (Gibbs, 2010), suggesting that, despite improvements in female students' performance on standardized test and classroom-based assessments, the issue of gender differences in mathematics has not been completely resolved. Thus, it is important for policy makers and educators to continue their efforts to improve the math achievement of female students and incorporate research evidence concerning factors related to gender differences in mathematics achievement.

In this context, the other key finding of the present studies—concerning the role of anxiety and working memory as mediators of gender difference in math performance—appears to be particularly relevant. The results raise the possibility that efforts to ameliorate the problem of gender differences in math-related fields may not be adequate unless they target specific factors, such as worry about math, in girls and women. In fact, recent findings suggest that heightened math anxiety in female teachers at the beginning of the school year is associated with lower math performance over a school year in their female students (Beilock, Gunderson, Ramirez, & Levine, 2010). Thus, addressing the problem of math anxiety in females may have far-reaching implications. Other recent findings indicate that female students' performance can be improved through interventions targeting specific cognitive skills, including visuospatial reasoning (Terlecki, Newcombe, & Little, 2008; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008). The converging evidence from the present studies and other recent research highlights the importance of considering both cognitive and affective processes in targeting gender differences in educational settings.

References

Adams, J., & Hitch, G. (1997). Working memory and children's addition. *Journal of Experimental Child Psychology, 67,* 21–38. doi:10.1006/jecp.1997.2397

Alloway, T. P., Gathercole, S. E., & Pickering, S. (2004). *Automated Working Memory Assessment.* Test battery available from authors at http://www.pearsonclinical.co.uk/Psychology/ChildCognitionNeuropsychologyandLanguage/ChildMemory/AutomatedWorkingMemoryAssessment(AWMA)/AutomatedWorkingMemoryAssessment(AWMA).aspx

Alloway, T. P., Gathercole, S. E., & Pickering, S. J. (2006). Verbal and visuo-spatial short-term and working memory in children: Are they separable? *Child Development, 77,* 1698–1716. doi:10.1111/j.1467-8624.2006.00968.x

Baddeley, A. D. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4,* 417–423. doi:10.1016/S1364-6613(00)01538-2

Baddeley, A. (2012). Working memory: Theories, models, and controversies. *Annual Review of Psychology, 63,* 1–29. doi:10.1146/annurev-psych-120710-100422

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.

Becker, B. (1990). Item characteristics and gender differences on the SAT-M for mathematically able youths. *American Educational Research Journal, 27,* 65–87. doi:10.3102/00028312027001065

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science, 17,* 339–343. doi:10.1111/j.1467-8721.2008.00602.x

Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, USA, 107,* 1860–1863. doi:10.1073/pnas.0910967107

Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General, 136,* 256–276. doi:10.1037/0096-3445.136.2.256

Berg, D. H. (2008). Working memory and arithmetic calculation in children: The contributory roles of processing speed, short-term memory, and reading. *Journal of Experimental Child Psychology, 99,* 288–308. doi:10.1016/j.jecp.2007.12.002

Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology, 45,* 180–185. doi:10.1016/j.jesp.2008.08.005

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidences for the role of negative thinking. *Psychological Science, 16,* 572–578. doi:10.1111/j.0956-7976.2005.01577.x

Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology, 33,* 669–680. doi:10.1037/0012-1649.33.4.669

College Board. (2009). *SAT percentile ranks for males, females, and total group*. Retrieved from http://professionals.collegeboard.com/data-reports-research/sat/data-tables

College Board. (2010). *AP summary reports: National report.* Retrieved from http://www.collegeboard.com/student/testing/ap/exgrd_sum/2010.html

Correll, S. J. (2001). Gender and the career choice process: The role of biased self-assessments. *American Journal of Sociology, 106,* 1691–1730. doi:10.1086/321299

Crowe, S. F., Matthews, C., & Walkenhorst, E. (2007). Relationship between worry, anxiety and thought suppression and the components of working memory in a non-clinical sample. *Australian Psychologist, 42,* 170–177. doi:10.1080/00050060601089462

Dearing, E., & Hamilton, L. C. (2006). Contemporary approaches and classic advice for analyzing mediating and moderating variables. *Monographs of the Society for Research in Child Development, 71,* 88–104. doi:10.1111/j.1540-5834.2006.00406.x

Deffenbacher, J. L. (1977). Relationship of worry and emotionality to performance on the Miller Analogies Test. *Journal of Educational Psychology, 69,* 191–195. doi:10.1037/0022-0663.69.2.191

Deffenbacher, J. L. (1978). Worry, emotionality, and task-generated interference in test anxiety: An empirical test of attentional theory. *Journal of Educational Psychology, 70,* 248–254. doi:10.1037/0022-0663.70.2.248

Deffenbacher, J. L. (1980). Worry and emotionality in test anxiety. In I. Sarason (Ed.), *Test anxiety: Theory, research, and application* (pp. 111–128). Hillsdale, NJ: Erlbaum.

Ding, L., Velicer, W. F., & Harlow, L. L. (1995). Effects of estimation methods, number of indicators per factor, and improper solutions on structural equation modeling fit indices. *Structural Equation Modeling: A Multidisciplinary Journal, 2,* 119–143. doi:10.1080/10705519509540000

Eccles, J. S., & Jacobs, J. E. (1986). Social forces shape math attitudes and performance. *Signs, 11,* 367–380. doi:10.1086/494229

Eysenck, M. W., & Calvo, M. G. (1992). Anxiety and performance: The processing efficiency theory. *Cognition & Emotion, 6,* 409–434. doi:10.1080/02699939208409696

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion, 7,* 336–353. doi:10.1037/1528-3542.7.2.336

Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing). *Educational Measurement: Issues and Practice, 7,* 33–38. doi:10.1111/j.1745-3992.1988.tb00434.x

Ganley, C. M., & Vasilyeva, M. (2011). Sex differences in the relation between math performance, spatial skills, and attitudes. *Journal of Applied Developmental Psychology, 32,* 235–242. doi:10.1016/j.appdev.2011.04.001

Gathercole, S. E., Pickering, S. J., Knight, C., & Stegmann, Z. (2004). Working memory skills and educational attainment: Evidence from national curriculum assessments at 7 and 14 years of age. *Applied Cognitive Psychology, 18,* 1–16. doi:10.1002/acp.934

Gibbs, B. G. (2010). Reversing fortunes or content change? Gender gaps in math-related skill throughout childhood. *Social Science Research, 39,* 540–569. doi:10.1016/j.ssresearch.2010.02.005

Gilhooly, K. J., Wyn, V., Phillips, L. H., Logie, R. H., & Della Sala, S. (2002). Visuo-spatial and verbal working memory in the five-disc Tower of London task: An individual differences approach. *Thinking & Reasoning, 8,* 165–178. doi:10.1080/13546780244000006

Hayes, A. F., Preacher, K. J., & Myers, T. A. (2011). Mediation and the estimation of indirect effects in political communication research. In E. P. Bucy & R. Lance Holbert (Eds.), *Sourcebook for political communication research: Methods, measures, and analytical techniques* (pp. 434–465). New York, NY: Routledge.

Hembree, R. (1988). Correlates, causes, effects and treatment of test anxiety. *Review of Educational Research, 58,* 47–77. doi:10.3102/00346543058001047

Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education, 21,* 33–46. doi:10.2307/749455

Holmes, J., & Adams, J. W. (2006). Working memory and children's mathematical skills: Implications for mathematical development and mathematical curricula. *Educational Psychology, 26,* 339–366. doi:10.1080/01443410500341056

Holmes, J., Adams, J. W., & Hamilton, C. J. (2008). The relationship between visuospatial sketchpad capacity and children's mathematical skills. *European Journal of Cognitive Psychology, 20,* 272–289. doi:10.1080/09541440701612702

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12,* F9–F15. doi:10.1111/j.1467-7687.2009.00848.x

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107,* 139–155. doi:10.1037/0033-2909.107.2.139

Hyde, J. S., Fennema, E., Ryan, M., Frost, L. A., & Hopp, C. (1990). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly, 14,* 299–324. doi:10.1111/j.1471-6402.1990.tb00022.x

Hyde, J. S., Lindberg, S. M., Linn, M. C., Ellis, A. B., & Williams, C. C. (2008, July 25). Gender similarities characterize math performance. *Science, 321,* 494–495. doi:10.1126/science.1160364

Ikeda, M., Iwanaga, M., & Seiwa, H. (1996). Test anxiety and working memory system. *Perceptual and Motor Skills, 82,* 1223–1231. doi:10.2466/pms.1996.82.3c.1223

Jarvis, H. L., & Gathercole, S. E. (2003). Verbal and non-verbal working memory and achievements on national curriculum tests at 11 and 14 years of age. *Educational and Child Psychology, 20,* 123–140.

Keppel, G., & Wickens, T. D. (2004). *Design and analysis: A researcher's handbook* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Kim, S. H., & Rocklin, T. (1994). The temporal patterns of worry and emotionality and their differential effects on test performance. *Anxiety, Stress & Coping: An International Journal, 7,* 117–130. doi:10.1080/10615809408249339

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory in children with ADHD. *Journal of Clinical and Experimental Neuropsychology, 24,* 781–791. doi:10.1076/jcen.24.6.781.8395

Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education, 23,* 77–94. doi:10.1007/BF03173141

Lee, J. H. (1999). Test anxiety and working memory. *Journal of Experimental Education, 67,* 218–240. doi:10.1080/00220979909598354

Liebert, R. M., & Morris, L. W. (1967). Cognitive and emotional components in test anxiety: A distinction and some initial data. *Psychological Reports, 20,* 975–978. doi:10.2466/pr0.1967.20.3.975

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136,* 1123–1135. doi:10.1037/a0021276

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education, 30,* 520–540. doi:10.2307/749772

Miller, H., & Bichsel, J. (2004). Anxiety, working memory, gender, and math performance. *Personality and Individual Differences, 37,* 591–606. doi:10.1016/j.paid.2003.09.029

Morris, L. W., Davis, M. A., & Hutchings, C. H. (1981). Cognitive and emotional components of anxiety: Literature review and a revised worry–emotionality scale. *Journal of Educational Psychology, 73,* 541–555. doi:10.1037/0022-0663.73.4.541

National Science Foundation, National Center for Science and Engineering Statistics. (2011). *Women, minorities, and persons with disabilities in science and engineering (NSF 11-309).* Retrieved from http://www.nsf.gov/statistics/wmpd

O'Neil, H. F., & Fukumura, T. (1992). Relationship of worry and emotionality to test performance in the Juku environment. *Anxiety, Stress & Coping: An International Journal, 5,* 241–251. doi:10.1080/10615809208249525

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement? *Contemporary Educational Psychology, 26,* 291–310. doi:10.1006/ceps.2000.1052

Owens, M., Stevenson, J., Norgate, R., & Hadwin, J. A. (2008). Processing efficiency theory in children: Working memory as a mediator between trait anxiety and academic performance. *Anxiety, Stress & Coping: An International Journal, 21,* 417–430. doi:10.1080/10615800701847823

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance. *Journal of Educational Psychology, 94,* 396–404. doi:10.1037/0022-0663.94.2.396

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers, 36,* 717–731. doi:10.3758/BF03206553

Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods, 40,* 879–891. doi:10.3758/BRM.40.3.879

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and Individual Differences, 20,* 110–122. doi:10.1016/j.lindif.2009.10.005

Ramirez, G., & Beilock, S. L. (2011, January 14). Writing about testing worries boosts exam performance in the classroom. *Science, 331,* 211–213. doi:10.1126/science.1199427

Rapee, R. M. (1993). The utilisation of working memory by worry. *Behaviour Research and Therapy, 31,* 617–620. doi:10.1016/0005-7967(93)90114-A

Robinson, J. P., & Lubienski, S. T. (2011). The development of gender achievement gaps in mathematics and reading during elementary and middle school: Examining direct cognitive assessments and teacher ratings. *American Educational Research Journal, 48,* 268–302. doi:10.3102/0002831210372249

Shackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D., Lavric, A., & Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion, 6,* 40–61. doi:10.1037/1528-3542.6.1.40

Snyder, T. D., & Dillow, S. A. (2011). *Digest of education statistics 2010* (NCES 2011-015). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *The Quarterly Journal of Experimental Psychology, 59,* 745–759. doi:10.1080/17470210500162854

Taylor, A. B., MacKinnon, D. P., & Tein, J. (2008). Tests of the three-path mediated effect. *Organizational Research Methods, 11,* 241–269. doi:10.1177/1094428107300344

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22,* 996–1013. doi:10.1002/acp.1420

Thorell, L. B., Lindquist, S., Nutley, S. B., Gunilla, B., & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12,* 106–113. doi:10.1111/j.1467-7687.2008.00745.x

U.S. News & World Report. (2012). *College search.* Retrieved from http://colleges.usnews.rankingsandreviews.com/best-colleges

Vasilyeva, M., Casey, B. M., Dearing, E., & Ganley, C. M. (2009). Measurement skills in low-income elementary school students: Exploring the nature of gender differences. *Cognition and Instruction, 27,* 401–428. doi:10.1080/07370000903221809

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. doi:10.1037/0033-2909.117.2.250

Wood, I. (2006). Effect of anxiety reduction on children's school performance and social adjustment. *Developmental Psychology, 42,* 345–349. doi:10.1037/0012-1649.42.2.345

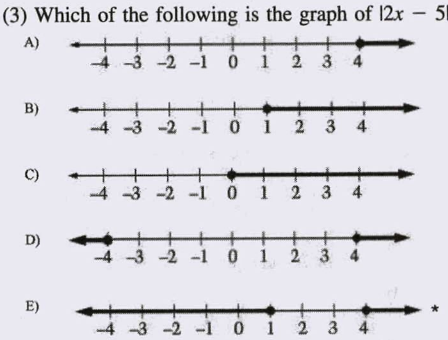
Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15,* 763–771. doi:10.3758/PBR.15.4.763

Zeidner, M., & Nevo, B. (1992). Test anxiety in examinees in a college admission testing situation: Incidence, dimensionality, and cognitive correlates. In H. M. van der Ploeg, R. Schwarzer, & C. D. Spielberger (Series Eds.) and K. A. Hagtvet & T. B. Johnsen (Vol. Eds.), *Advances in test anxiety research* (Vol. 7, pp. 288–303). Amsterdam/Lisse, the Netherlands: Swets & Zeitlinger.

(Appendix follows)

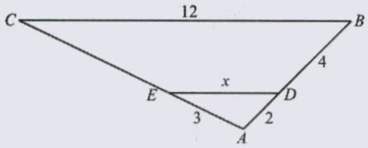
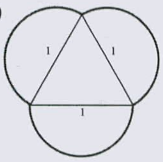
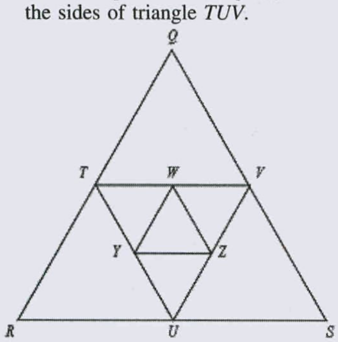
Appendix

Mathematics Test Items

<table>
<tr><th>Test item</th><th>Female percent correct[a]</th><th>Male percent correct[a]</th></tr>
<tr><td>(1) Yvonne has studied the change in cost of tickets over time for her favorite sports team. She has created a model to predict the cost of a ticket in the future. Let C represent the cost of a ticket in dollars and y represent the number of years in the future. Her model is as follows:
$$C = 2.50y + 13$$
Based on this model, how much will the cost of a ticket increase in two years?
(A) $5* (B) $8 (C) $13 (D) $18 (E) $26</td><td>49%
36%</td><td>63%
52%</td></tr>
<tr><td>(2) The length of a rectangle is 3 more than its width. If L represents the length, what is an expression for the width?
(A) $3 \div L$ (B) $L \div 3$ (C) $L \times 3$ (D) $L + 3$ (E) $L - 3$*</td><td>40%
42%</td><td>71%
57%</td></tr>
<tr><td>(3) Which of the following is the graph of $|2x - 5| \geq 3$?

</td><td>25%
29%</td><td>38%
48%</td></tr>
<tr><td>(4) For what value of x is $8^{12} = 16^{x}$?
(A) 3 (B) 4 (C) 8 (D) 9* (E) 12</td><td>16%
23%</td><td>30%
36%</td></tr>
<tr><td>(5) If $3x + 2y = 11$ and $2x + 3y = 17$, what is the average (arithmetic mean) of x and y?
(A) 2.5 (B) 2.8* (C) 5.6 (D) 5.8 (E) 14</td><td>26%
34%</td><td>57%
55%</td></tr>
<tr><td>(6) The table below shows a linear relationship between x and y.
<table><tr><th>x</th><th>y</th></tr><tr><td>−7</td><td>a</td></tr><tr><td>−3</td><td>10</td></tr><tr><td>−1</td><td>6</td></tr><tr><td>0</td><td>4</td></tr><tr><td>5</td><td>−6</td></tr></table>
What is the value of a?
(A) −18 (B) −14 (C) 14 (D) 18* (E) 36</td><td>66%
38%</td><td>67%
80%</td></tr>
</table>

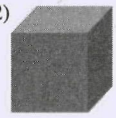
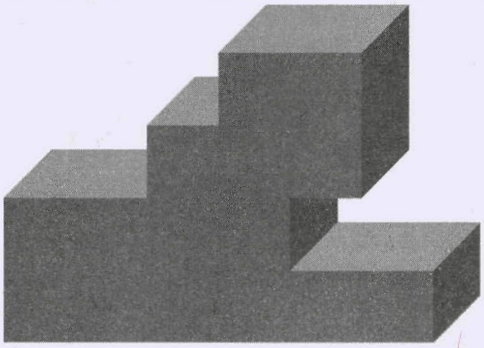
(Appendix continues)

Appendix (*continued*)

Test item	Female percent correct[a]	Male percent correct[a]
(7) 	48%	67%
If triangles *ADE* and *ABC* shown in the figure above are similar, what is the value of x? (A) 4* (B) 5 (C) 6 (D) 8 (E) 10	44%	59%
(8) A circle with diameter 10 centimeters is to be cut from a square of paper 10 centimeters on a side. Of the following, which is closest to the amount of paper left over after the circle is cut out?	32%	29%
(A) 9 square centimeters (B) 21 square centimeters* (C) 24 square centimeters (D) 69 square centimeters (E) 84 square centimeters	33%	45%
(9) 	42%	74%
Semicircles are constructed on the sides of an equilateral triangle, as shown in the figure above. Of the following, which best approximates the sum of the lengths of the three darkened arcs? (A) 4.4258 (B) 4.7124* (C) 6.0000 (D) 6.7124 (E) 9.4258	39%	59%
(10) In the *xy*-plane, a line parallel to the x-axis intersects the y-axis at the point (0, 4). This line also intersects a circle in two points. The circle has a radius of 5 and its center is at the origin. What are the coordinates of the two points of intersection?	36%	78%
(A) (2, 1) and (2, −1) (B) (3, 4) and (3, −4) (C) (3, 4) and (−3, 4)* (D) (5, 4) and (−5, 4) (E) (5, 0) and (−5, 0)	38%	66%
(11) In the figure shown below, triangle *TUV* is formed by joining the midpoints of the sides of equilateral triangle *QRS*. Triangle *WYZ* is formed by joining the midpoints of the sides of triangle *TUV*.	63% 56%	86% 75%
		
If the area of triangle *QRS* is 64 square inches, what is the area of triangle *WYZ*? (A) 1 square inch (B) 4 square inches* (C) 8 square inches (D) 16 square inches (E) 64 square inches		

(*Appendix continues*)

Appendix (*continued*)

Test item	Female percent correct[a]	Male percent correct[a]
(12) 	41%	30%
How many of the unit cubes above would it take to make the object below?  (A) 15 (B) 16 (C) 30* (D) 32 (E) 45	25%	47%

Note. An asterisk indicates the correct answer.
[a] Study 1 means are presented first with Study 2 means below.

Received January 31, 2012
Revision received July 1, 2013
Accepted July 5, 2013 ■