A CROSS-CULTURAL COMPARISON OF GENDER AND MATH PERFORMANCE

AMONG COLLEGE STUDENTS IN CHINA AND THE UNITED STATES

A Report of a Senior Study

by

Yilong Zheng

Major: Psychology

Maryville College

Fall, 2010

Date Approved__________________, by ______________________

Faculty Supervisor

Date Approved__________________, by ______________________

Editor
ABSTRACT

Cross-cultural studies of math performance have shown that biological factors alone cannot fully explain the differences in math performance. Therefore, social factors, such as stereotypes and self-esteem, were considered in the present study with undergraduate students from both China ($N = 153$) and the United States ($N = 87$). Participants from both countries were asked to answer a math-related self-esteem scale and a gender stereotype scale. They also provided demographic information and their math entrance examination score. Students in the United States also completed an abridged sample SAT-math test after the surveys. Both analyses of variance and linear regression analyses were performed. Results showed that there was not a significant difference between males and females on math performance, but rather significant differences on math-related self-esteem and math-related stereotypes. Math-related self-esteem, age, and major had strong influence on changing one’s math performance. Cross-cultural comparisons between Chinese students and American students showed significant differences on math-related self-esteem and stereotypes: Chinese students on average had lower math-related self-esteem, but a higher degree of math-related stereotypes than American students. Chinese females and American males had greater amount of math-related stereotypes within their own cultures. More future studies on the influence of social factors need to be conducted.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>26</td>
</tr>
<tr>
<td></td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>30</td>
</tr>
<tr>
<td>III</td>
<td>33</td>
</tr>
<tr>
<td>IV</td>
<td>42</td>
</tr>
<tr>
<td>Appendices</td>
<td>53</td>
</tr>
<tr>
<td>References</td>
<td>79</td>
</tr>
</tbody>
</table>
CHAPTER I

INTRODUCTION

Charles Darwin (1897) once said: “Man is more courageous, pugnacious and energetic than woman, and has a more inventive genius” (p. 557). Apparently, a hundred years ago, women in general were in an absolutely disadvantaged situation in term of being treated unequally. Thankfully, after decades of women's rights movements, women are able to see the trend towards greater gender equality in this society. However, in the field of science and mathematics, male superiority has still been assumed by most people. What is even worse is that statistically, males comprise the majority of individuals in the science of mathematics population.

In 2003, the specific shortage of women in science fields was announced by the National Science Board, along with the report of needing a workforce with higher level math and science skills in the U.S. (Halpern et al., 2007). This means that women are badly needed in the field of science as well as mathematics, especially in the advanced level. Another group of data from National Science Foundation in 2002 illustrates this phenomenon in a more obvious way by saying that almost one-fourth of the labor force in science and mathematics are woman, while close to half of the overall work force in the U.S. is made up by women. Also, high school and college-age women have shown
less likelihood of choosing mathematics as their major or future career (Kieger & Sekaquaptewa, 2007). Therefore, many researchers have said that the assumption of male dominance in science and mathematics has once again ‘been proven’ in these silent ways.

It seems like females are losing their position in regard to the proportion of overall population in math field. However, does that mean males are dominant? Or do these facts mean that females in general are not good at math? Does the quantity of female underperformance in math indicate its lower quality? Debates about gender equality in mathematics have simply started with the basic argument of math performances between males and females. Starting from this discussion, the facts of women’s underrepresentation in mathematic research and women’s limited success in academic careers in mathematics have also been mentioned in order to counter the disagreements of male superiority (Chipman, 2005).

Looking at the facts that women are still underrepresented in the field of science and mathematics, could there be other possible explanations and outcomes of gender inequality from the past decades? Ceci and William (2007) give additional possible reasons to explain the underperformance by saying that women might be less interested in a professional career in math than men due to the decreased encouragement and social support. In their study on the evaluation of how well people were doing in a job, they mentioned that women will be less favored by others if they are good in a male-dominant field. What is more interesting was that both men and women in the study rated men as more likeable in male-dominant fields throughout the evaluation process. However, these facts might not be convincing enough to explain
why females in general underperform in mathematics. Therefore, the questions have been raised: What do we know about women’s performance and participation in mathematics field in this 21st century? What factors can be the possible reasons of understanding the underrepresentation of women in the world of mathematics? In this modern society, if the gender equality is well assumed and practiced, is there still a difference on math performance and ability between males and females from both biological and social perspectives? Put another way, in the past 30 years, the emphasis of gender equality has been raised and is reaching its highest point ever, which means that despite the unequal education between men and women, there is no reason to hypothesize that men and women should perform differently than one another. However, if the recent studies still show that there is a gender difference of overall math ability, we should then take a closer look at the issue from biological and sociopsychological points of view.

To further understand the issue of gender differences on mathematics, we need to look at the past 20 years’ studies beginning in the 1990s. This was the time when more psychologists started to find out that gender differences in mathematics had begun to decline (Ding, Song, & Richardson, 2006). Some researchers attributed this change to the changing of social gender role climate. This might also indicate that biological factors were not the only cause of gender differences in math performance. Also, it was mentioned that in the ‘90’s studies, the timing of gender differences in math was generally varied. It could be started in elementary school or even as late as in high school. These possibilities have all been mentioned and supported in different studies (Penner, 2003). Due to the fact that it was hard to determine a specific developmental
stage that influenced gender difference the most, it is better to assume that the changing of gender difference happened throughout developmental stages.

As to the results of the studies, most researchers tended to hold the traditional belief that males were in a more advantaged position even though the gap between male and female on math performance was very small (Brown & Joseph, 1999; Ding et al., 2006; Halpern et al., 2007; Kiefer & Sekaquaptewa, 2007). Ding et al. (2006) indicated that the male superiority in mathematics performance is found even in elementary school, but it tended to decrease as both males and females grow older. Kurdek and Sinclair (2001) mentioned that the difference did not occur until the early fourth grade of elementary school due to the different teaching content in mathematics, in which teachers taught more geometry in later grades than algebra in early elementary school years. Ding et al. (2006) also included several past research studies by Amrein and Berliner that showed females absolutely outperformed males in general starting from middle school, while falling behind males on math tests. On the contrary, these statements did not seem to be very convincing when other researchers found that females have higher math grades in school in their studies. Leahey and Guo (2001) found that females actually had higher scores between the years of grade seven and ten, but the improving rate of males was higher. Ryan and Ryan (2005) also mentioned that females generally have higher math grades when compared to males in school, while males have higher math scores on tests. The debate of which gender performs better comes to a point where studies suggest both males and females have their advantages in mathematics. However, when deciding which one has a better score in general, males are being favored in terms of having more evidence and data support.
Despite the results saying males are performing better, we should not forget the fact that the gap is small, and the difference is declining.

Sex and Gender

For those who do not clearly understand the ‘difference’ between sex and gender, it seems there is no difference in use of either these two terms to describe the population. For some people, they are just two different expressions but for the same purpose. However, they are different. The term ‘sex’ generally refers to the biological sex and differences between two sexes, whereas ‘gender’ is the culturally-defined term for two sexes and the differences that arise in a social environment. In Halpern et al.’s (2007) article, a much more precise definition has been provided; it is said that more authors prefer ‘sex’ to discuss about innate differences, such as chromosomes and hormones, and ‘gender’ is the role that comes from a environmental and social perspective.

Many researchers believe the differences between ‘sex’ and ‘gender’ are meaningless due to the intertwined relationship between one another: it is hard to identify biological differences from cultural influences (Severin & Wyer, 2000). The most recent edition of the APA *Publication Manual* (American Psychological Association, 2009) has further suggested the change due to language inaccuracy about gender and race. In the manual, more usages such as “the other sex” and “transsexual” are adopted to replace “the opposite sex” and “transgender” (pp. 73-74). These small but clear changes have indicated that the correct terms are very important in producing better research when considering gender and sex. The usage of term is clear and will
be used for different purposes in this study, especially for the debate of ‘nature or nurture’ dilemmas.

Ability and Achievement

When reading research studies related to tests, ‘ability’ and ‘achievement’ are usually quite confusing. An achievement test is used to test one’s ability, but does it really tell one’s ‘ability’ or one’s test performance? Although they are quite different in meaning, they are widely used interchangeably and referring to similar things in everyday life. If one’s performance is examined, does it count as more influence on ability or achievement? We may easily understand that two individuals can have same achievement test scores but not really have the same ability, because these two people may or may not have the same score again in the next achievement test. Halpern et al. (2007) counted the subtle difference between these two as the result of whether they reflect individual’s curriculum or the likelihood to benefit from instruction. For ‘achievement,’ it is more about curriculum, whereas ‘ability’ is more about likelihood to succeed from instruction. Even though it is still ambiguous, one may see the difference as having internal consistency: for achievement, it has less internal consistency than ability. The debate between ‘ability’ and ‘achievement’ leads to the concern of whether tests or grades show more individual ability. Many researchers, such as Caplan and Caplan (2005), point out the interesting fact that female middle and high school students generally have higher grades in school, including subjects like math since two decades ago. However, when it comes to tests such as SAT or GRE tests, females score lower on average than males in science and mathematics (Caplan & Caplan, 2005; Ceci & Williams, 2007; Halpern et al., 2007; Penner, 2008). Since females’ performance is
usually underpredicted by the standardized test, questions have been raised about whether grades or standardized tests are more suitable for showing one’s ability. Due to the fact that most studies in the past were conducted based on comparing the test performance between two different genders, the present study will continue the focus on test performance rather than grades in school. Also, as the gap of gender performance in standardized math tests is getting smaller, it is more interesting to study the determining factors behind this issue.

The Effect Size

The effect size is generally used in statistics and psychology in meta-analysis studies. It is usually used to compare two groups of sample data by comparing their sizes, and to see if there is any relationship between the two variables. Meanwhile, the result will also show whether the sample size can predict each population mean. There are many well known effect sizes being mentioned in psychology and statistics textbooks, such as Pearson’s correlation and its square, \( t \)-test, ANOVA tests, ANCOVA tests, etc. Halpem et al. (2007) explained one kind of effect size called “Cohen d” by giving an example of comparing males and females math standardized test score: If the sample mean of a group of male students is subtracted by the sample mean of a group of female students in the model, the positive value will lead to a “male advantage”, and the negative value will lead to a “female advantage” (p. 4). The effect size ranges from -3 to +3, and larger or smaller than ±.8 is usually considered as a large difference. There are usually three outcomes determine by this model: small, medium, and large. More effect size tests will be introduced later in this chapter.
Biological Perspective

The debate of nature and nurture continues, and apparently one should not prefer any side and say one is more influential than another. When studying human intelligence development, most researchers found their way in the middle, which is studying from both nature and nurture views. The studies of math performance between sexes can be generalized into several categories: hormones and evolution, brain structure, brain function, as well as cognitive skills that are integrated within the former two categories.

Among most psychologists, the brain was the first and most significant part that made the differences between men and women, even between individuals. Kimura (1993) indicated that not only cognitive variations but also hormones functions are the major focus of brain development. Cognitive variations suggest differences in preferring words or objects. Interestingly, men perform better on spatial tasks than women on average, and women perform better on identifying and matching, as well as verbal fluency. One of the old examples for explaining the differences is driving behaviors: females tend to remember the landmarks close the destination whereas males tend to use compass routes more. Cognitive skills are easily improved through teaching methods; using software and models are very helpful tools in geometric math teaching. Does the cognitive difference still exist between men and women after decades? Halpern (2007) referenced one study from Spelke conducted in 2005 that shows boys and girls develop their early cognitive skills equally well. However, many researchers still believe the difference is there throughout developmental stages. The answer from most researchers is positive, that is, the differences in cognitive characteristics for
males and females still exist. Therefore, it is easily seen how the gender stereotypes about males and females developed in the past decades: the idea of ‘males are good at math, females are good at verbal skills’ indeed comes from some scientific evidence. However, how people interpret these results can make a big difference.

Another interesting fact that Kimura (1993) pointed out is that all human organisms have the potential to be male or female. However, if a Y chromosome is present, a series of male hormone and organs will be formed. If not, then a series of female hormone and organs will be formed. The researcher emphasized that the early exposure of the brain to the resulting sex hormones will change brain function and behaviors. Kimura noted several past research studies on androgens, and suggested that women who have high testosterone (an androgen) generally perform better in spatial and mathematical reasoning tests, while men who have lower testosterone perform much better than other males in spatial, mathematical, and perceptual speed tests. More studies about hormones were done and similar results supported in several Halpern’s studies after Kimura.

Kimura (1993) also showed brain structure can explain the difference in performance. The splenium, part of corpus callosum, is different in males and females. Women have a larger splenium whereas men have a significantly smaller size. The size may indicate the amount of fibers connecting two hemispheres in the brain, which further shows that women have more information communication between two hemispheres—the symmetry of female brain. The asymmetry of male brain relates to male’s hormone growth in the left cortex, which is the critical part for developing spatial and cognitive skills (Kimura).
If those differences do exist, how effective are they on influencing one’s math performance? Do males always have the advantage on math tests? One way to compare the difference is to look at the college entrance examination. Casey, Nuttal, Pezaris, and Benbow (1995) focused their study on spatial ability on gender, and they hypothesized that spatial ability was the major factor why men have led in math test scores in past decades. They also examined the correlation between mental rotation ability and math ability. A sample of 274 undergraduate students (195 women and 79 men) at two liberal art colleges participated in the study and both SAT math score and mental rotation score were collected. The sample was divided into ‘high/low ability’ and ‘talented’ groups. The SAT-M results showed that males performed better among high ability students and talented students. For high ability students, males on average scored 5.7 points higher than females. For talented students, males scored 7.3 points higher than their counter partners. The overall Pearson’s correlation showed a positive relationship between SAT score and mental rotation score. Due to small sample size and other limitations, one cannot say that males perform better than females overall. In Halpern et al.’s (2007) study, the SAT-Math average test scores between 1967 and 2004 were examined, showing that males on average did perform better, as shown in Figure 1. However, the gender gap is quite consistent over the years.

From Casey et al.’s (1995) study, there is small evidence of the relationship between visuospatial ability and math performance. Spatial ability can be one of the factors to explain the difference. One result that came from the distribution analysis was that the differences mainly exist mainly at the higher level math population; that is,
differences between males and females tend to occur among the top students. Similar results about visuospatial abilities were found in Halpern et al.’s (2007) review article: the results of these tasks still favor men over women. However, it should be noticed that the scores are not much different, and the effect size of some studies is small.

Halpern, Wai, and Saw (2005) examined sex differences from a hormonal perspective. Geschwind’s theory of prenatal hormonal effects was adopted to explain the hand preference. It is said that high levels of testosterone slow down the development in left side of the brain, which leads to right brain dominance. Also, it is believed that reasoning and spatial skills are under greater control of right hemisphere of our brain. Halpern et al. mentioned that many research studies have confirmed that there are more males are left handed. Therefore, if both Geschwind’s theory and findings about having more males being left-handed are scientifically supported, there should be more males who are good at reasoning and spatial tasks than females. The
theory and statement could be the crucial part in explaining how effective sex hormones are and finally lead to the conclusion of ‘male born advantage.’ Being a controversial theory, Geschwind’s theory has many ideas that lack scientific evidence, and the relation of motor control and sex hormone is hard to prove.

Another way to look at the influence of hormones is to look at abnormalities. This is a very important and theoretically preferred practice due to ethical limitations of conducting experimental research. Researchers look at the populations that have hormone problems to learn about hormone’s role in cognitive differences and thus further explain the difference in math performance. Halpern et al. (2005) looked at the past studies of low hormone level patients and found out that for females, their visuospatial ability was affected by the low level of hormones, whereas nothing changed in verbal ability and IQ. Similar research outcomes have been found in Kimura’s (1993) study.

One of the major limitations of hormone studies is due to an interesting fact—biologically, there are more than two sexes. However, this does not strongly affect the fact that males and females each follow their own characteristics from the influence of sex chromosomes. Therefore, there is a clear idea of how hormones are altering spatial ability and math performance on a regular basis.

Evolutionists, like Darwin, attributed sex differences to the theory of sexual selection. Most evolutionists think of the cognitive differences between males and females as the result of traditional male-male competition, female choice and survival rules of hunting. These factors gradually require male to be better at cognitive processes through the history of human development (Halpern et al., 2005). Why was it
so? According to Halpern et al., males who survived were the ones that were better at maintaining group’s territorial, long distance hunting, and tribal warfare. All of these visuospatial-related activities required advanced development in brain. However, other evolutionary psychologists did not think so. Halpern et al. (2007) referred to Murdock’s argument, who said that females were also required to develop their cognitive skills due to their roles. They usually needed to travel a long distance to collect food. Moreover, they were the ones who needed to conduct household works that require good spatial skills, such as weaving and tools making. Overall, people accepted that both roles were demanding on developing their spatial skills.

The differences between males and females in terms of brain structure and brain function are probably the most popular concepts. People are convinced when they see differences between males and females from all kinds of brain images. However, whether there is a strong correlation between math performance and differences in brain still remain questionable. As the neuroimaging techniques have become better developed, the differences in cerebral volume and corpus callosum between sexes have been shown. Halpern et al. (2007) mentioned that the differences in cerebral volume are well documented, with the different proportion of gray matter in men and women’s brains. Grey matter shows where the major cell bodies are located in the cerebrum. Researchers have collected enough information to determine that “males have a higher percentage of gray matter in the left hemisphere, but females have a higher portion of gray matter in both hemispheres,” yet another way females’ brains are generally more symmetric (p. 25). The correlation between volume and the cognitive
tasks has not yet been found, but one should not ignore that males show preference in using spatial ability in the brain than females, as evidenced by brain imaging.

Studies that examined the differences in brain function between sexes can usually be divided into three major areas: cerebral blood flow, cerebral glucose metabolism, and neurotransmitter function (Halpern et al., 2007). The main reason that researchers study these areas is because they all show the brain usage and the activity in each hemisphere. This may or may not suggest how important brain is in relation to cognitive ability. Females generally had more activation in parietal and prefrontal regions, but males had more activity in the hippocampus. This suggests that males have more advantages in solving geometric problems. Males also have higher metabolic glucose in the cerebellum, which is another area that may affect cognitive ability. Also, for neurotransmitter function, different levels of dopamine were found that suggested males perform better on visuospatial tests, whereas females are outstanding in verbal tests.

Other biological factors were also considered in determining math performance. A study of math performance as a function of sex, laterality (especially handedness), and age of puberty was conducted. Sappington and Topolski (2005) questioned the effect of laterality and ‘late blooming’ in puberty in math tests. Sex, laterality, and age of puberty were collected along with SAT-M scores. An ANOVA test was used and the results showed the significant effects of sex and pubertal onset. The combination of all the factors was associated with high SAT-M scores. To their surprise, early maturation seemed to be beneficial in the study.
Overall, it seems like there are many studies that suggest that biological differences in sexes that may influence math test scores. However, the limitations of these studies are fairly clear—all the studies have not found any strong scientific correlation with cognitive performance. Therefore, it will take more steps to show that these factors can influence math performance. Also, none of these biological findings was found longitudinally: longitudinal studies are crucial in helping researchers find a better understanding between brain and math performance. After reviewing those studies, one may wonder: how important are cognitive skills in affecting math tests? Are there any other factors that may also influence math performance? To find out the answers, a close look at the issue is given from a biopsychosocial aspect.

Biopsychosocial perspective

Caplan and Caplan (2005) illustrate how one should take biological perspective of view seriously, but not as the most important aspect in explaining the difference in math performance between males and females:

Today, the search for sex differences in neuronal activities or glucose metabolism in particular brain areas may sound highly scientific and rigorous…However, none of these approaches make sense because they represent a search for anatomical or physiological correlates of a behavioral effect that has not been reliably demonstrated to exist, certainly not so reliably and immutably that it makes sense to posit physiological correlates. (p. 25)

Even though most biological studies have their limitations, researchers have never rejected how critical visuospatial skills are in predicting math test scores. One of the best ways to indicate this is to look at the biopsychosocial model, because one
cannot simply look at how cognitive skills can alter math performance without knowing the powerful influence of ‘nurture.’ Figure 2 shows a model that illustrates the connection between biology and social environment. Halpern et al. (2007) point out that each piece in the model can help people understand how all the elements work as a whole to influence one’s cognitive ability. However, where did the other elements come from? Why are they equally important as cognitive skills in altering math performance?

Social perspectives

It seems that the differences of math performance between males and females are well explained by biological and cognitive findings. However, that is not always the case. Humans are not only biological but also social beings. Assuming there are fewer biological differences within the same sex category than between sexes, the test performance should not be much different. However, there exists that diversity,
including math performance. Not every single male or female is good at math, and not every offspring of mathematicians is excellent at mathematics. Even within the same person at different developmental stages, his/her math performance may be different. This is due to the development of his/her cognitive skills. For different reasons, there is a possibility that one can have different levels of math performances during the same development stage, too.

Ding et al. (2006) conducted a longitudinal study of gender differences in math in the U.S. Biological differences were mentioned, but social factors that affect the differences were mainly concerned. Ding et al. compared students’ math performance in terms of test scores over time from elementary school to high school. SAT-M scores, and GPA were collected by gender. Results showed that both males and females had improvement in math, but the growth rate was not significantly different. Females tend to have a higher GPA in math in both middle and high school. If most of the past cognitive studies can explain why favoring males over females in math performance, then what biological factors can be used to explain why females are good at class work globally? It is true that studies may show females are good at verbal skills and cognitive skill like memorization (Halpern et al., 2007). However, Halpern et al. also point out that as schools start to teach geometry and calculus, there is a major change for male students in terms of having more advantages. The conflicting results have shown that there was probably more than one factor that made this situation complicated.

Looking at the interaction of gender and math from a cross-national perspective could probably show how strongly biological differences could affect math performance around the world, and most importantly, the influence of social factors besides the
biological factors. If there are significant differences cross-nationally in terms of the math scores, which probably means there is more influence of social factors. Therefore, logically, the opposite view should be: if there is no international variation, then a stronger biological influence is indicated. Penner (2008) found that males have significantly higher math scores in almost all the countries examined, but the differences between countries vary greatly. Both of the findings raised doubts of how important the biological influences are.

Penner (2003) compared the interaction between gender and item difficulty in math tests in 10 countries. The interaction could explain whether test scores favor men or women as the difficulty level increases. The data were collected from Third International Mathematics and Science Survey (TIMSS), and the multi-linear regression test showed there was a certain degree of male privilege for all ten countries. The interaction indicated that as the difficulty level increased, the females perceived the test as more difficult than males did.

Examining the importance of social factors, Else-Quest, Hyde, and Linn (2010) examined the two largest international data sets of math scores. The results show that gender differences are changing at an international level. Gender-typed behavior, self-efficacy, stereotypes were mentioned in attributing the past changes in math performance between males and females. The effect sizes were also tested. Females outperformed males in algebra but still reported having less motivation and more negative attitudes toward mathematics. Males were little slightly better at other domains, including geometry, but they showed a relatively large effect size for self-confidence level and valuing math. Similar results had been observed in another data sets, in
which the effect size of anxiety for females is .28, whereas the self-concept and self-efficacy for males were both .33 (Else-Quest et al.). The interesting effect that males are generally more confident in math even though they are not necessarily better was found in both Penner’s (2003) and Else-Quest et al.’s (2010) studies.

Looking at both longitudinal and cross-national studies can give a general idea of how important social factors are in this study. Longitudinal studies showed that both males and females were improving over time, but the changes of performance were not equivalent to its biological assumption of sex differences. The international studies indicate the universal trend of math performance between genders, while also showing in some countries that females could surely be better in certain math areas than males. These powerful statistical results give the clear idea that not only biological, but also social perspectives have the potential to affect math performance, and sometimes social factors are even more important.

The major drawback to these cross-cultural studies is that different countries have different levels in educating and emphasizing the importance of mathematics. Moreover, the social factors and gender equity levels were fairly different across countries.

According to the studies discussed above, there is an understanding that social forces are important in determining the relationship between gender and math performance. Halpern et al. (2007) mentioned many social factors that are influential: Family, peers, schools, stereotype threats, training studies, cross-cultural analyses, and trends over time.
Halpern et al. (2007) found parental involvement and expectations for their children’s school are correlated with children’s performance in mathematics—the more parents are involved with their children’s school work, the higher level of performance their children have in math tests. People mentioned that parents who encourage sex-typed behavior may also make a difference (Entwisle & Alexander, 1996). Sex-typed behaviors, like girls playing with dolls indoors and boys playing outdoor sports, are considered to have potential influence on math abilities. However, it is not known that to what extent this sex-typed behavior has affected math performance.

Children are influenced by their peers. It is interesting that one’s math grade is related to how supportive their peers are (Halpern et al., 2007). Starting from middle school, females have less peer support from their peers for math activities than males in general. Besides the stereotype of math and gender, females also have less confidence rating themselves on math performance than males. This phenomenon is not shocking from a stereotype theory perspective but from the finding that females are increasingly not disadvantaged compared to males in math tests, and their grades are usually higher.

What is the secret of mathematical tests that then prevents female students from performing well? Biological factors cannot explain the reason well, so psychologists started to pay attention to stereotype threat, which is another important factor that explains this interesting phenomenon. Stereotype threat has a great influence on one’s test performance (Brown & Josephs, 1990; Ryan & Ryan, 2005); this significant influence has been found both in experimental studies and in daily life. Halpern et al. (2007) give several examples about people having stereotypes regarding one’s ethnicity
and gender that really made differences in math performance. Different racial groups, such as Caucasians, African Americans, and Asian Americans, were compared in one study about academic performance. Stereotypes about intellectual inferiority of Caucasian and African Americans compared with Asian Americans in mathematics, and African Americans’ inferiority when compared with Caucasian in academics in general were illustrated. Another classic stereotype example is believing that males perform better than females in math tests. Both of these examples give a clear idea of how the majority in this society judges performance based on their old stereotypes. These stereotypes generally create stereotype threat effect, which can possibly harm one’s performance by altering cognitive abilities like memory and sympathetic nervous system activity. People who were influenced by stereotype threat will consciously or unconsciously confirm stereotype expectations; this kind of stereotype behavior is called ‘Self-fulfilling stereotypes’ (Ryan & Ryan, 2005). Self-fulfilling stereotype can sometimes be positive, such as confirming the stereotypes that all Asians or males are generally good at math, and this will to some degree encourage Asians or males to do better on math tests.

In one of the early studies of stereotype threat, Brown and Josephs (1999) pointed out that stereotypes of gender differences in mathematics are less severe, but related stereotypes still exist. An ANOVA test was conducted to find the relationship between gender and SAT-M score. The interaction was highly significant showing that gender-specific stereotypes can affect differential performance between different genders. The researchers designed a ‘weak-ability test,’ in which the questions are similar to the regular test, but participants who were not good at mathematics were
given external excuses for weak performance, and the results were quite surprising: females who were given the external attribution as ‘handicapped’ performed much better in the test compared to both females who did not get any attribution and males who were given or not given external reason (Brown & Josephs). Researchers believe this can explain how much influence self-handicapping stereotype has in real life situations among females. The reason that males’ average score did not change by being given an external reason was because males generally do not experience this stereotype—males are more confident in their math performance even though some of them are not individually as superior as females.

Ryan and Ryan (2005) studied how stereotype threat moderated one’s motivation, cognitive ability, and affect during a math test, and their relationships to gender performance. Researchers believed stereotypes of women have haunted female students in school and their lives. They also worried that one’s achievement and self-efficacy can be greatly harmed once a stereotype is accidently confirmed. For example, a “performance-avoid goal” like ‘we females are not good at math, so I am not going to try hard’ has been linked to low performance among female students (p. 55). This performance-avoid goal usually goes with low self-efficacy.

Rivardo, Rhodes, and Klein (2008) studied physiological responses to stereotype threat at a liberal arts college. Researchers examined differences between males and females on math performance by measuring salivary adrenal cortex hormone—cortisol (CORT). Pre-test CORT, post-test CORT, GRE, and SAT-M scores were collected from over 100 students. The results showed that not only the interaction of sex and condition failed to reach statistical significance, but also CORT levels between
conditions are not significantly different. The results showed no evidence that there is a difference on math performance due to stereotype threat in this sample.

This kind of stereotype threat has not only been found in the U.S., but in other countries as well. In Wilkin’s (2004) study, TIMSS data and self-concept of students were used in a linear regression model, which was tested to show whether there is a relationship between achievement and self-concept. A sample of 290,000 students’ math performance of 41 countries were studied. The results showed that there is generally a positive view of self about mathematics performance in the world, but it declines differently for males and females over the years. Similar results were also found in Penner's (2003) study.

Some researchers have wondered whether stereotype threat is only a laboratory artifact. Kiefer and Sekaquaptewa (2007) further suggested self identification should be considered in studying stereotype threat, because generally talking about stereotypes of females’ performance in math still cannot explain why females have different responses to stereotypes by having significant different levels of scores. Females who have both low levels of self-identification and stereotypes can succeed and perform better on math tests, and thus may have the potential to pursue math-related careers. The results showed the effect of gender identification and implicit stereotyping for affecting exam score, but not for career goals. Other effects were not significant. To researchers’ surprise, explicit stereotyping was not significant. An example of explicit stereotyping is female participants expecting males to be better than they are in math. However, most females in the study did not agree with the idea that males are better at math. This finding shows that stereotypes are perhaps starting to change among females.
Ceci and William (2007) examine stereotypes about women being underrepresented in mathematics from a cultural perspective. According to Ceci and William, gender schemas and stereotypes have been exaggerated in past decades based on small differences; if all of these significant social factors have influences on individuals, then it is not hard to understand why females’ average math performance has improved over the past years as more equal classroom environment in math course is encouraged. Also, females’ decreased interest in pursuing professional math careers can also be one of the explanations for underrepresentation.

Sex discrimination is mentioned in Ceci and William (2007)’s book and Halpern et al. (2007) ’s study. Females have reported being sexually discriminated against in both job searching and graduate school experiences in traditional ‘masculine’ areas; these unfair and unequal experiences have greatly influenced their decisions to continue careers in math-related areas. Also, comparing female experiences in mathematics and other science fields with the experiences in female-dominated fields, females reported significant higher rates of sex discrimination (Halpern et al.). This result did not even include subtle forms of sexism. Sex bias treatment and gender stereotypes were most commonly mentioned in all the related studies. However, there were not many studies about sex discrimination in studying gender performance in math.

Summary

The purpose of the present study is to compare college math performance between males and females both in the United States and China. Also, the math performance of males and females who are pursuing a future profession in mathematics field is going to
compared with non-majors. Since the psychological perspective is considered to be the most influential and well studied aspect of the problem, a social perspective of the study will be used. Another goal of this study is to determine whether math-related stereotypes exist among students both in the United States and China. A third goal relates to stereotype threat in the United States: after being primed on math-gender stereotypes using an attitude survey, it is also important to see whether this stereotype threat may affect students’ performance significantly on a mock standardized math test.
CHAPTER II

METHOD

Participants

In the study in China, 153 participants (102 males, 49 females, 2 unidentified) were recruited from a comprehensive university, East China Jiao Tong University (ECJTU). These participants were chosen by cluster sampling in the Department of Basic Sciences. The sample included students from Freshman to Junior rank, with ages ranging from 17 to 22 years ($M_{age} = 19.9$ years, $SD = 1.05$). Most of them had taken college algebra and college statistics. More sensitive background information and socioeconomic status of each participant in this study were not collected. Their voluntary participation and confidentiality in the study were respected and protected, and no compensation was paid.

In the United States, the sample consisted of 87 participants (51 men and 36 women) at a southern liberal art college, Maryville College, in Tennessee. Participants were mostly recruited in three general education Statistics classes (STA120) and an Introductory Psychology class (PSY101). Other mathematics major students were recruited from several upper-level math courses, including Calculus, Linear algebra, Inferential Statistics, and Advanced Calculus. The participants in this second sample
were not randomly selected but rather were a convenience sample. Most of the participants were freshmen. Students also had different levels of mathematics backgrounds, with an average of SAT-M score of 599, and an average of ACT math score of 24.4. The majors among participants varied, including Biochemistry, Business, Computer Science, English, Mathematics, Physical Education, Psychology, Writing Communication, etc. The age of the undergraduates was from 18 to 26 years ($M = 19.1$, $SD = 1.4$). As with the Chinese sample, sensitive background information and socioeconomic status of each participant in this study were not collected. In most of the classes, extra credit was provided by the course instructors; however, no compensation was paid.

**Materials**

All the participants in the two studies were required to complete three basic questionnaires, but each in relevant languages (Chinese or English). Two of the three questionnaires used in both countries were exactly the same. In addition to those questionnaires, participants in the US were asked to take an abridged SAT-M practice test which contained ten math-related questions. Informed consent and debriefing statements were given prior to and after the study, respectively.

*Informed Consent (American and Chinese version):* Participants were informed in the introduction that consent to participate in the experiment would be implied by their completion of the measures. No names were collected and confidentiality of individual results was maintained. It indicated that results were going to be reported in group format only (see Appendix A1 and A2).
Math self-esteem scale (American and Chinese version): This scale was adapted from the Rosenberg's Self-Esteem Scale (Rosenberg, 1989), which has been widely used to measure self-esteem. It also has been used in several research studies to measure the self-concept in one's ability to function in various domains. The 10-item, four-point Likert scale with '1' being 'Strongly Agree' to '4' being 'Strongly Disagree,' was modified to assess math self-esteem (e.g., "All in all, I am inclined to feel that I am a failure in math.") (see Appendix B1). The Chinese version was a translation of the scale into Mandarin Chinese according to the meaning of each item. It was modified to match the Chinese concept of self-esteem, so that participants could understand the choices more easily (see Appendix B2). The Chinese version was reviewed by another Chinese native speaker to help corroborate meaning.

Math-Gender Stereotype Survey (American and Chinese version): This six-item survey assesses math and gender stereotyping. This questionnaire is supposed to indicate whether a participant has certain kinds of stereotypical thoughts or opinions related to gender. This stereotyping includes both implicit gender stereotyping and explicit gender stereotyping. The survey contains six statements with a five-point Likert response scale about math-related gender stereotypes. In the survey, a response of '1' indicates 'Disagree/ Not important/ Not common,' while '5' indicates 'Agree/ Very important/ Very common,' with '3' is considered as 'Neutral.' The same format was used for the Chinese version. (see Appendices C1 and C2).

Personal information survey (American and Chinese version): In the survey for American students, the survey includes items for high school ACT and/or SAT-M
scores, age, sex, year in college, high school GPA, working status, and major. Both ACT and SAT-M scores were included due to the fact that most high school students in Tennessee take the ACT test rather than SAT test. However, an ACT math score will give an approximate SAT-M estimate for the performance. A general ACT math and SAT-M convertor scale was used (Grover, 2010). The reasons for using SAT-M scores as the standard were because: (1) More students take SAT tests than ACT tests in the nation; and (2) Most previous research (e.g., Brown & Josephs, 1990; Casey et al., 1995; Ding, Song, & Richardson, 2006; Halpern et al., 2007; Ryan & Ryan, 2005; Sappington & Topolski, 2005) used SAT-M as their measures. Sex is an independent variable that is a main focus in this research. Age, year, high school GPA, working status were the variables that were going to be used in linear regression model. Major is another variable that is critical here, because of the comparison of males and females for both non-math majors and mathematics major, and even possible comparisons between non-math majors and math majors.

The Chinese version of this survey is the same except instead of ACT/SAT-M math scores, the score for the Chinese general entrance examination test was requested. An equivalent high school GPA is not available because Chinese high schools use a different grading system; it is also difficult to obtain GPA because GPA is not very significant to Chinese high school students (see Appendices D1 and D2).

**Brief SAT-M test:** This is an abridged version of a practice SAT-Math test, available from an online SAT-M sample test (SAT math problem solving, 2010). It consists of 10 items from a practice SAT-M test, suitable for administration in group format. The time allowed for students to take the SAT-M test is 15 minutes. This brief SAT-M test
includes Algebra and Geometry problems that are equivalent to general mathematics requirement for high school level students. Therefore, it is considered to be a way of testing college students’ math ability (see Appendix E).

*Debriefing statement (American and Chinese version):* This is a statement that explains the purpose of the study and provides the researcher’s contact information for the results of the study (see Appendix F). The Chinese version of debriefing was a translated statement of the English version (see Appendix G).

**Procedure**

Before conducting the study, permission was attained from the Maryville College Institutional Review Board (IRB) (see Appendix H). Permission from the appropriate authorities at ECJTU was also obtained before collecting data in China (see Appendix I). ECJTU does not have formal IRB procedures, so the chairperson of the Math Department at ECJTU gave permission to recruit participants in the department and followed the procedures outlined in the IRB proposal approved at Maryville College. The informed consent and the various surveys were translated into Mandarin Chinese and checked for accuracy of translation by an independent rater.

Students were notified and encouraged to participate in the study through the department email a week before the study was conducted. Additionally, students were asked in their classes to participate in the study. Questions and concerns were answered by emails and phone calls. Students who were willing to participate in this study were required to reply by email to either the chair of their math department or the researcher. By doing so, numbers of participants were confirmed and the same number
of surveys were prepared. A specific time and location were selected for participants to complete the study. Each individual survey was assigned a number code starting from 001 to 200. Informed consent was given before they started to complete the surveys. Students were told not to write their names or any information that might indicate their identities on the survey. Participants were asked to complete the translated surveys in order. The entire process took around 15 minutes. The Brief SAT-M test was not given to the Chinese sample due to translation issues.

The study in the United States was done after the study in China. All the participants at Maryville College were recruited from three Statistics classes, one Introductory Psychology class, and several advanced Mathematics classes. The surveys and Brief SAT-M test were administered during their regular class periods in the regular classrooms, with permission by all course instructors. The surveys were collected at different days and class periods. After the researcher met all the participants in each classroom, the researcher explained the purpose and the topic of the study. Questions were asked about the study before the surveys were completed. The informed consent was the cover for the survey packet, which included the Personal Information survey, Math Self-Esteem scale, the Math-Gender Stereotype scale, and the Brief SAT-M test. The abridged SAT-M test was the last page of the survey packet. For each packet, no names were collected. After participants finished with the scales, the Brief SAT-M test was administered. The entire session lasted 25-30 minutes. Extra credit was given by STA 120 instructors and PSY101 instructor to those who participated in the study. When participants completed the Brief SAT-M test, they were given the debriefing statement. The math responses were then graded, and the
ACT math scores were converted to SAT-M scores. Then a numeric code from 1 to 87 was given to each of all the survey packets.
CHAPTER III

RESULTS

Since the Chinese samples were not used to examine the research question regarding stereotype threat, they were used as a comparison with the American sample to determine whether there is a gender difference on math performance cross-culturally. If results in China and the United States show similar gender differences, then it supports the notion of genuine gender differences in mathematical performance.

Scores for math self-esteem and stereotyping were computed for all participants. The range of scores of math self-esteem varies from 10 to 40, and the lower the score, the greater is a person's self-esteem. The total score for stereotype score was summed from the six survey questions, ranging from 6 to 30 points. The higher score a participant obtained, the greater gender stereotype he/she had.

*Chinese sample*

Descriptive statistics were used first to describe the data, then three single factorial analyses of variance (ANOVA) between gender and math score, gender and math self-esteem score, gender and stereotype score were conducted. In addition, coefficient correlations were used to determine the relationship between gender, math score, math self-esteem score, and stereotype score. A linear regression model was also applied to see which independent variable had the strongest effect on math score,
and whether there are other factors influencing math score besides gender. All analyses were performed with a significance level of $\alpha = .05$.

Although the mean of the math score among males $(M = 96.89, SD = 14.04)$ is little higher than that of females $(M = 93.96, SD = 13.49)$, there were six outliers at the higher level. The result from the first one-way ANOVA showed no significant difference between male and female on math score, $F(1, 150) = 1.48, p = .226$. However, a highly significant relationship between gender and math-related self-esteem score was found, $F(1, 150) = 8.07, p = .005$. Female participants on average had a higher self-esteem score $(M = 23.96, SD = 2.65)$ than that of male participants $(M = 22.31, SD = 3.17)$, which means that female students generally had lower self-esteem than male students. On the other hand, no significant interaction between gender and stereotype score was found in Chinese sample: both males and females scored similarly on the survey (see Figure 3).
As to the correlations between age, major, math score, self-esteem score and stereotype score, there were several interesting findings. First, there was a significant positive correlation between age and math score with an $r = .21$, $p = .01$. This means that the older a participant is, the more likely he/she had a higher math score. Also, there was a highly significant positive correlation ($r = .23$, $p = .005$) between gender and self-esteem score, which matches the result from the ANOVA test between gender and self-esteem. Moreover, there is an $r = -.20$ ($p = .013$) correlation between math score and self-esteem scores, indicating that the higher self-esteem score one has, the lower math score he/she had. Remember in this case that higher self-esteem score actually represents a lower self-esteem. The particular item, “Overall, males are better than females in math” from the stereotype score survey was also examined and showed a negative correlation with gender, ($r = -.20$, $p = .014$). This showed that females were much more likely to score low on this question, whereas males tended to score higher on the scale.

A linear regression model was used to analyze the data for Chinese sample. Multiple factors including age, major, class, job (working or not), self-esteem math score, and stereotype math score were used to build the model. Among these factors, major, class, and job information were considered as the non-numerical dummy variables. After eliminating major as the least significant factor, the second least significant factor among all was gender. The regression model was statistically significant, $F (5, 147) = 2.89$, $p = .02$. The $R^2$ (adj) of this model was 5.9%, indicating that only 6% of the data from the Chinese sample was explained by the linear regression model. As shown in Table 1, the only two statistically significant explanatory
factors were age ($p = .058$) and self-esteem score ($p = .056$), with a coefficient of 2.28 and -0.66, respectively. (See Table 1)

Table 1. Regression equation for predicting Chinese sample math scores

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>$T$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>73.02</td>
<td>2.68</td>
<td>.008*</td>
</tr>
<tr>
<td>Gender (M=1, F=0)</td>
<td>1.42</td>
<td>0.59</td>
<td>.555*</td>
</tr>
<tr>
<td>Age</td>
<td>2.28</td>
<td>1.91</td>
<td>.058*</td>
</tr>
<tr>
<td>Class (R=1, O=0)</td>
<td>-2.16</td>
<td>-0.81</td>
<td>.428*</td>
</tr>
<tr>
<td>SE Score</td>
<td>-0.66</td>
<td>-1.93</td>
<td>.056*</td>
</tr>
<tr>
<td>Stereotype Score</td>
<td>-0.41</td>
<td>-1.04</td>
<td>.298*</td>
</tr>
</tbody>
</table>

$R^2$(%) = 6

$F (5, 147) = 2.89^*$

*Note. $N = 153$; M = Male, F = Female; R = Freshman, O = Other classes; SE = Self-Esteem

$^* p < .05.$

The United States sample

For the sample in the United States, more factors and statistical analyses were necessary. Besides the same coefficient correlations, three one-way ANOVAs, and the linear regression models, and a 2 (Gender: Male, Female) × 2 (Major: Math, Non-Math) mixed factorial ANOVA were also conducted. All analyses were performed at a significance level of $\alpha = .05$.

Among all the correlations, math-related self-esteem score and math score, as well as math related self-esteem score and stereotype score, had highly statistically
significant coefficient correlations. The correlation between math score and self-esteem score was $r = -0.29$, $p = 0.006$. This correlation indicates a negative relationship between math score and math-related self-esteem score. This means that participants with greater math self-esteem tended to have higher math scores (a low score on the self-esteem scale indicates greater self-esteem). The positive correlation between math-related self-esteem score and stereotype score was $r = 0.327$, $p = 0.002$, indicating that among US students, a higher self-esteem score would lead to a higher stereotype score. In other words, lower self-esteem students had higher stereotype scores, and thus have more stereotypical thoughts.

The $t$-test between gender and math score did not show statistical significance. However, the average math score for males was $M = 3.92$ ($SD = 2.14$) and $M = 4.58$ ($SD = 2.44$) for females. The results from the $t$-test for gender and self-esteem score were also not significant at all, which showed that there was really not much difference between males and females in math-related self-esteem. However, there was a highly statistical significant relationship between gender and stereotype score in the third $t$-test, $t(86) = 9.85$, $p = 0.002$. Females ($M = 13.72$, $SD = 3.02$) scored significantly lower than the males ($M = 15.65$, $SD = 2.67$) in holding gender-based stereotypes.

Similar to the Chinese sample, a linear regression model was used to show whether other factors might also have influence on math scores. Gender, age, major, class, job (working or not), self-esteem score, stereotype score, converted SAT math score, and ACT math score were variables considered in building this model. GPA was not included in the model due to different scales being used in different high schools. The entire model was highly statistical significant, $F (1, 86) = 3.43$, $p = 0.002$. However,
the percentage of $R^2$ (adj) was only 27%, which showed that not much of the regression model has been explained by these factors. The model showed major and ACT score were the only marginally significant factors that needed to be included (see Table 2).

Table 2. Regression equation for predicting American sample math scores

<table>
<thead>
<tr>
<th>Variables</th>
<th>Coefficient</th>
<th>T</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-4.21</td>
<td>-0.45</td>
<td>.653*</td>
</tr>
<tr>
<td>Gender (M=1, F=0)</td>
<td>-0.18</td>
<td>-0.31</td>
<td>.758*</td>
</tr>
<tr>
<td>Age</td>
<td>-0.07</td>
<td>-0.16</td>
<td>.870*</td>
</tr>
<tr>
<td>Major (Math=1, O=0)</td>
<td>2.19</td>
<td>2.57</td>
<td>.013*</td>
</tr>
<tr>
<td>Class (R=1, O=0)</td>
<td>0.17</td>
<td>0.20</td>
<td>.841*</td>
</tr>
<tr>
<td>Job</td>
<td>0.49</td>
<td>0.85</td>
<td>.402*</td>
</tr>
<tr>
<td>SE Score</td>
<td>0.07</td>
<td>1.00</td>
<td>.323*</td>
</tr>
<tr>
<td>Stereotype Score</td>
<td>-0.04</td>
<td>-0.36</td>
<td>.722*</td>
</tr>
<tr>
<td>SAT-M Score (converted)</td>
<td>.002</td>
<td>0.30</td>
<td>.768*</td>
</tr>
<tr>
<td>ACT-M Score</td>
<td>0.28</td>
<td>1.64</td>
<td>.108*</td>
</tr>
</tbody>
</table>

$R^2(\%) = 38.2$

$F(9, 59) = 3.43**$

*Note. N= 155. M = Male, F = Female; R = Freshman, O = Other classes; SE = Self-Esteem

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

In addition, the results from the 2 (Gender: Male, Female) × 2 (Major: Math, Non-Math) mixed factorial ANOVA showed a significant Major main effect on math score, $F(1, 86) = 8.19, p = .005$. Math major students ($M = 6.091, SD = 1.87$) scored 2 points on average higher than non-math major students ($M = 4.02, SD = 2.28$). The main effect of
gender on math score difference among math majors was not significant. There was no statistically significant gender main effect and no significant interaction between gender and major.

Lastly, two more linear regression models were built to examine other variables. Both math-related self-esteem score and math-related self-stereotype score were used as dependent variables. Factors such as gender, age, major, class, job (working or not), stereotype score, converted SAT math score, and ACT math score were considered in building the regression model for math-related self-esteem score, whereas gender, age, major, class, job (working or not), self-esteem score, converted SAT math score, and ACT math score were selected for the stereotype score model. Both models went through modification and selection based on the rule of “whether each has significant p-value.” The results for both models were worth considering. Surprisingly, ACT math score had highly significant positive relationship with personal self-esteem, but not converted SAT math score, \( p < .001 \). Stereotype score had a significant relationship with math-related self-esteem score, indicating a negative relationship, \( p < .001 \). Gender and major each was only marginally significant in determining a person’s math-related self-esteem. The regression equation of math-related stereotype score showed that gender \( (p = .003) \), major \( (p = .002) \) and math-related self-esteem score \( (p < .001) \) were the critical factors in determining a person’s stereotype. According to the \( R^2 \) (adj)s for both these linear regression models, there was about 30% to 35% of the data that had been explained through the linear regression model.
Chinese vs. the United States

After showing whether there are gender differences in these two countries, a cross-cultural comparison was also done. All analyses were performed with a significance level of $\alpha = .05$. Two 2(Gender: Male, Female) $\times$ 2(Culture: Chinese, American) factorial ANOVAs were conducted on math related self-esteem score and stereotype score.

The results were quite surprising (see Table 3). For self-esteem scores, there was a highly significant main effect for Culture in this study, $F(1, 234) = 8.73, p = .003$. Chinese college students ($M = 22.89, SD = 3.4$) on average had a higher self-esteem score, in other words, lower actual math-related self-esteem than American college students ($M = 21.31, SD = 5.1$). There was no significant gender main effect in these two countries. There was also no significant interaction between gender and culture on self-esteem score.

Table 3.
ANOVA on SE Score and Stereotype Score among Chinese and American Students

<table>
<thead>
<tr>
<th></th>
<th>Self-Esteem Score</th>
<th>Stereotype Score</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F(1, 237)$</td>
<td>$p$</td>
</tr>
<tr>
<td>Gender (M, F)</td>
<td>3.63</td>
<td>.058*</td>
</tr>
<tr>
<td>Culture (C, A)</td>
<td>8.73</td>
<td>.003**</td>
</tr>
<tr>
<td>Gender $\times$ Culture</td>
<td>1.77</td>
<td>.185*</td>
</tr>
</tbody>
</table>

Note: M = Male, F = Female; C = Chinese, A = American; * $p < .05$. ** $p < .01$. *** $p < .001$. 


The results of the two-way ANOVA test on stereotype score are also presented in Table 3. There was not a significant gender main effect difference for males ($M = 16.5$, $SD = 2.8$) and females ($M = 15.9$, $SD = 3.6$) on stereotype score. However, there was again a highly significant main effect for culture, $F(1, 234) = 32.9$, $p < .001$. The average stereotype score among Chinese students ($M = 17.2$, $SD = 2.9$) was significantly higher than the average score among American students ($M = 14.9$, $SD = 3.0$). There was also an interaction between gender and culture on stereotype score, $F = .10.07$, $p = .002$. Male students ($M = 16.9$, $SD = 2.7$) had an average stereotype score that was lower than the female students ($M = 17.5$, $SD = 3.2$) in China, but male students ($M = 15.6$, $SD = 2.7$) in the U.S. had a higher average on stereotype score than females students ($M = 13.7$, $SD = 3.0$).
CHAPTER IV

DISCUSSION

The main purpose of the study was to examine whether gender differences in math performance and math-related stereotype thoughts exist among college students both in the United States and China. Could a cross-cultural comparison indicate a universal trend of math performance between genders, while also showing that not only biological, but also social perspectives have the potential to affect math performance? Another goal was to see whether stereotype threat might have a significant effect on a standardized math test after students being primed on math-gender stereotypes using an attitude survey. The hypotheses were that: 1) there is no gender difference between males and females on math performances; 2) the math-related gender stereotype threat may significantly influence one’s math performance. Surveys on math-related self-esteem and math-related gender stereotypes were collected from both China and the United States at two different periods. Additionally, a math test was completed by the U.S. sample.

The present study showed various significant results, but the most interesting result was that there were no significant differences in math performance between males and females both in China and the U.S. In the sample in China, the average math
score of college entrance examination for males was just three points higher than females on a 150-point scale. However, the similar performance between males and females did not also translate to differences in math self-esteem among the Chinese sample. The results for gender and math-related self-esteem showed that females have lower self-esteem. Interestingly, both males and females groups in China believed that males have better math ability and thus should perform better in math tests. This was also indicated from the mean stereotype score of 17, nearly three points higher than the average of stereotype score among students in the United States.

After examining the results of both samples, it is evident that self-esteem and stereotypical thoughts might be the two greatest factors that have influence on one’s math performance. The comparison between the Chinese and American samples showed that Chinese college students on average had lower self-esteem in math than American college students, regardless of actual math results. Meanwhile, a similar pattern of math-related self-esteem scores between males and females in these two countries led to a insignificant $p$-value for the gender main effect, even though there was a significant gender main effect on math-related self-esteem for Chinese participants. No significant interaction between gender and culture on self-esteem was found in the present study.

Another two-way ANOVA test showed that Chinese students had stronger math-related, stereotypical thoughts than American students in general. What is more interesting is that males had lower stereotype scores than the female students in China, but male students in the U.S. had higher stereotype scores than female students. This means that male students in China had weaker or fewer math-related, stereotypical
ideas than female students, but this was not the same case in the U.S. samples.

Chinese females and American males both believed that males could and should perform better on math tests in general. This particular result leads to an interesting question: Why do Chinese female college students think they would underperform their male counterparts and American male college students think they would perform better than their female counterparts, while in fact there was no significant difference between genders on performance or past math achievement scores?

Chinese students had lower self-esteem in mathematics along with stronger stereotypical thoughts, especially among female Chinese students. The fact that Chinese students had lower math-related, self-esteem could possibly be explained by the influence of Chinese culture. It is believed that no matter how accomplished a person might be at math, he/she still needs to show politeness and modesty through admitting ‘he/she is still not good enough.’ Also, in contemporary Chinese society, academics are highly competitive. Students in school believe there are always people who are better than them in mathematics. The fact that most people still believe males are performing better than females on the math exams could possibly make female students to think that males are in dominant roles, and thus females are less confident in their overall math ability.

Interestingly, the lower confidence level of math among females did not decrease female’s math performance. In fact, in the present study, Chinese female students who had stronger stereotypical thoughts and lower self-esteem in general performed as well as their male counterparts. The existence of stereotypical thoughts was definitely affected by the male dominant society and the continued male dominance in advanced
mathematics and science fields in China. It is possible that females who were not willing to confirm the stereotypes thus tried their best in the exam to break this stereotype even they had less self-esteem in their math ability.

On the other hand, American students had higher math-related, self-esteem and held fewer stereotypes. This could be explained by the culture as well. The fact that American culture encourages one to show his/her confidence and pride perhaps boosts one’s self-esteem. It is interesting to see that there was no statistical difference in math-related self-esteem between genders, yet still there were different levels of math-related, stereotypical thoughts among male and female students.

In the present study, the second hypothesis that stereotype threat might affect one’s math performance was not shown in the subsample from American students. Male students who had the stereotype that males outperform females in math tests did not actually perform better in the subsequent math test. Stereotype threat was not necessarily related to the math performance of female participants either. The linear regression model indicated that one’s self-esteem contributes more to predict one’s math performance, rather than stereotypes in this case. However, it is still surprising to find that believing that ‘males can perform better than females on math’ did not necessarily increase male’s performance in the math test. Age and major were the more influential factors in the model, which means that the more one studies mathematics and the more effort or interest one puts on studying mathematics, the more likely it is that one can get a better math performance. Other factors, such as ACT scores, also had the potential to influence one’s math performance among American students as well.
In reviewing the study results, it is easy to see that males and females have similar math performance in term of math scores, but these results are not consistent with former studies. Most past studies showed there was a difference between males and females in math tests, but the results from present study were not strong enough to give its support. Penner (2008) found that males have significantly higher math scores in almost all the countries examined, but the differences between countries vary greatly. Penner (2003) suggested that males had a certain degree of male privilege as difficulty level of a math test increased in the study in all ten countries. However, in this study, there are no other levels of math achievement for comparison.

In the present study, gender stereotype threat was not the factor that affected one’s math performance. However, results from both Brown and Josephs (1999) and Ryan and Ryan (2005) showed the significant effects of gender stereotypes on math performance. In the study by Brown and Josephs (1999), researchers pointed out that stereotypes of gender differences were less severe now but still exist and are related to math performance. The interaction was highly significant, showing that gender-specific stereotypes might affect differential performance between genders. However, such an interaction between genders and cultures on stereotype threat was not found in the present study, very likely due to several limitations of the study.

The significant finding from the present study that might support results from other former studies was that social factors have an influence on math performance. Else-Quest et al. (2010) found that other factors such as self-confidence level, motivation, self-efficacy, stereotypes were contributing the changes in math performance. Both Penner (2003) and Else-Quest et al. (2010) say that males are
generally more confident in math even though they are not necessarily better. These findings are supported by the result in the present study that college males in the U.S. sample had, on average, a similar math-related self-esteem level, and had similar math scores to females. In the present study, the difference of self-esteem and stereotype scores between Chinese and American students might be explained by factors, such as family, peers, and schools (Halpern et al., 2007). Kiefer and Sekaquaptewa (2007) found the majority of women in their study did not agree with the idea that men have superior math ability, and surprisingly similar results were found in the present study.

Even though there were interesting findings and meaningful results shown in the study, there are surely many limitations that might affect the interpretation of results. Although there was a good sample size of students both in the United States ($N = 87$) and China ($N = 153$), gender balance was skewed within the two subsamples. In both studies, female subsamples were much smaller than male subsamples, especially in the Chinese study. The subsample size that the sample size of male students in China was more than twice the female students. The big differences in unbalanced gender sample size might lead to very different statistical results. Sadly, not much effort was done when considering about equalizing the sampling size before starting the research. Also, there was a small subsample of math major students in the American sample, for which only 4 males and 7 females were recruited. Comparing the pool size in this study with other former similar studies, a shortage of participants was quite significant. Another limitation of this pool was the generalizability of the participants representing the entire student population in China and the U.S. In the study in China, all the students were recruited from East China Jiao Tong University (ECJTU). In the study in the United
States, a similar approach was adapted: all the students were from Maryville College. In all, students were not randomly selected in most cases, and a better design on sampling method should be used in the future.

In addition, a standardized SAT math sample test that contained only 10 questions could not necessarily and accurately measured one’s math performance. A formal SAT test lasts more than 60 minutes, including more than 40 questions. Even at that level of intensity, one could not guarantee the consistency and reliability of the test. In this case, a sample test of 10 questions that lasted less than 20 minutes was definitely not a relatively satisfying math task. However, because of recruiting challenges, the length of the test was limited. In the future study, a more representative and longer length sample math test is strongly encouraged.

Besides these two basic limitations of the study, more factors in the study could possibly be modified and improved as well. Several variables, such as work, converted SAT-M score, and high school GPA in regression analysis could not contribute expected explanations for the model. Whether one had a job was simply used as a dummy variable rather than a numerical variable which may contribute more in the model. Converted SAT math scores from ACT math scores could not truly present the level of differences between students not only because these were two different results from two similar but indeed different standardized tests. A conversion scale that helps to convert from ACT to SAT could be reliable in terms of anchoring a similar score for college applications, but the actual performance and position in one of these tests from a student could not be estimated. Ideally, this conversion could work, but it seems like there are other potential factors such as home states, the difficulty level between ACT
and SAT exams, and the overall performance of the students in both tests that made the converted SAT math scores just an approximation. However, SAT math score was crucial in the regression model in terms of finding whether students had improved their math ability over time and whether stereotype threat might have the influence on math performance.

High school GPA could be another contributing factor. However, the variety of point scales in school systems and level of rigor from school to school made it impossible for any valid comparisons. Therefore, it is really hard to believe that high school GPA from the samples might make a difference in the present study. In a future study, a consistency between a standardized sample test used in the study and the results from a past standardized math test collected from participants should be maintained. Also, other possible factors such as ethnicity group, family background, and amount of time spent on math-related studying could be collected in the survey.

It is a pity that neither biological factors nor coping strategies were able to be explored in the study. More psychologists have started to focus on training studies that may suggest potential new techniques for teaching and improving skills in mathematics at school. Halpern et al. (2007) looked at several past training studies and found some encouraging techniques and outcomes. Other similar studies suggested that both female and male spatial ability can be improved with training, but the gap between the two genders remains the same. Another software training study was mentioned by Halpern et al., in which students who had software training was shown to be better at the end of the first year of college. Geist and King (2008) believe there is a difference between females and males in math performance, but rather consider the difference as
simply ‘different,’ not ‘better.’ From a social perspective, researchers introduced 10 classroom supporting suggestions for teachers that really make a difference in improving the math class environment and equalizing students’ chances to learn:

- Avoiding labeling.
- Get to know your student’s learning styles.
- Get to know the developmental differences of your children.
- Allow children to solve problems in many different ways.
- Using active and exploratory methods of teaching.
- Visual spatial versus language based approaches to mathematics.
- Developing activities based on different attention levels of boys and girls.
- Competition verses cooperation.
- Individual versus group.
- Inductive versus deductive reasoning. (pp. 44-49)

In the present study, there was no focus on coping techniques and the improvement of cognitive behaviors. However, these methods that are mentioned could be used in future studies.

Although there were several major limitations that had an impact on the results, there are some new findings that are quite important and meaningful. The translated version of Chinese survey made it possible to demonstrate that Chinese students have lower self-esteem in mathematics and higher degree of math stereotypes than American students. Previous studies have not compared the math-related self-esteem and stereotype between these two countries and thus provide little explanation. There are several reasons that might possibly explain the situation. First, because there are
different ways of perceiving one’s self-esteem in Chinese culture, the overall self-esteem is hard to measure. Also, in Chinese culture, etiquette subtly dictates that one should be modest and polite; people would rather underestimate their own abilities than the opposite in order to show respect to others. Another explanation could be that Chinese people identify themselves as members of a collective society and thus value self-esteem less as an individualistic construct. Therefore, it may not be valid to translate just the words of a self-esteem scale into Chinese and neglect the underlying cultural meanings of the construct. Translating and modifying the questions of the scale to their similar meanings was a difficult process, and that is the reason it required multiple people to proofread the Chinese version scale after the translation work was done.

The study not only used traditional inferential statistics test, such as $t$-tests and ANOVA tests, but also linear regression analysis to see whether factors other than gender, math-related self-esteem, and math-related stereotype threat have their effect on math performance. The results showed that age, major and ACT math scores might also have the potential to influence one’s math performance. It showed that linear regression was able to predict one’s math performance based on different factors. This also suggested that multilevel regression, which is a more sophisticated regression model, might be helpful and adapted to predicting one’s math performance in the future.

The discussion of using multilevel regression model leads to a future possible study or goal of determining whether the behaviors from professors and teachers, family factors and peers, and students all have influence on the math performance of a particular group of students.
Overall, the study showed that there was not a significant difference between males and females on math performance, but rather significant differences on math-related self-esteem and math-related stereotypes. The differences between Chinese and American students led to several interesting explanations that could possibly account for the cross-cultural difference. The first hypothesis that there would be no gender difference between male and female students was supported by the results. All three comparisons on math performance showed no difference between genders. It was expected that there might be a gender difference on the math result due to factors, such as math-related self esteem, and gender stereotype threat. The effect of math-related self-esteem, age, major was also demonstrated by this study, showing that confidence level and the time and effort one spends on studying mathematics can affect math performance. The effect of stereotype on math performance was, however, not evident, due perhaps to the limitations of the study. Therefore, the results on effect of stereotype threat failed to support the second hypothesis in the study. Future study is needed to take a closer look at the relationship among stereotypes, self-esteem and math performance in a larger and gender-balanced sample to see whether stereotype threat really exists. Research should also examine ways to reduce the harmful effects of math-self-esteem, as well as techniques to reduce math stereotypes.
APPENDICES
Hello. My name is Yilong Zheng. I am a senior psychology major at Maryville College. I am conducting research for my senior study. First, I would like to thank you for coming to participate in my senior research. This research is about factors that are related to math performance and is supervised by Dr. Lori Schmied in the Department of Psychology.

This study will take approximately 35-45 minutes. Your name will not be collected and data will be reported only as group result. Your individual data will be kept confidential.

Although your participation in this endeavor is voluntary, your sincere input is extremely important. Your completion of the surveys constitutes informed consent to participate in the study. At any time during the study, you may refuse to answer questions and you may stop participation at any time without penalty.

For more information, please contact me at yilong.zheng@my.maryvillecollege.edu or 865-293-7041; my thesis supervisor Dr. Schmied at lori.schmied@maryvillecollege.edu.

Thank you for your assistance!

Sincerely,

Yilong Zheng
心理学期课题研究实验告知

2010 年 6 月到 12 月

同学你好！我的名字叫郑毅龙。我是在美国一所大学学习的大四心理学系学生，我的毕业论文的中文课题是“探讨和研究中美大学生差别和数学考试发挥的关系”，英文课题是“Gender and Mathematics performance among college students in China and the United States”。

首先，非常感谢您参加我的研究实验调查。这个实验调查总共会花 15-20 分钟，短小精炼，而且非常直接简单，没有深层次意思，所以请放松回答。我会询问一些对您来说可能比较敏感的问题，希望您如实回答。请务必留意，不要担心焦虑这些问题会被除了您以外的人知道，您的所有个人信息和所有回答都不会泄露于他人，而且，所有的信息的汇报都是以整体出现，不会有分析个例的现象。所有统计资料在毕业论文完结后，都会被销毁，请再次放心。

您的参加代表您会认真如实回答我的问题，您对这些问题回答的准确性深深影响着我的实验调查统计结果和整个毕业论文的完整性，所以我希望您如实回答一下的问题。如果您觉得这项调查真的让您感到不适或者反感，您可以在任何时候退出这项调查。

我在这里真心感谢您的帮助，谢谢。

如有相关问题，请来信或致电：

Yilong Zheng

502 E Lamar Alexander Pkwy,

Maryville, TN 37804

01-865-293-7041

申请人：郑毅龙
Appendix B1: Self-concept--Rosenberg's Self-Esteem Scale (modified version):

Below is a list of statements dealing with your general feelings about yourself. If you strongly agree, circle 1. If you agree with the statement, circle 2. If you disagree, circle 3. If you strongly disagree, circle 4. Everything that you rate will be remained confidential. Thank you.

<table>
<thead>
<tr>
<th>Strongly Agree</th>
<th>Agree</th>
<th>Disagree</th>
<th>Strongly Disagree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. On the whole, I am satisfied with my math performance.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>2. At times, I think I am no good at all in math.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3. I feel that I am pretty good at math.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>4. I am able to do math as well as most other people.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>5. I feel I do not have much to be proud of in math.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>6. I occasionally feel uncomfortable and very distressed during math tests.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>7. I feel that I’m a person of worth, at least on an equal plane with others regarding math performance.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8. I wish I could have more respect for my math performance.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>9. All in all, I am inclined to feel that I am a failure in math.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>10. I take a positive attitude toward myself when it comes to doing math.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
### Appendix B2: Self-concept 自我认知和概念—Rosenberg's Self-Esteem Scale (modified version):

下面有一系列的陈述，需要您从这些陈述中选择最适合您观点的。下面左边是陈述栏，下面右边有四个选项：“1—非常同意”、“2—同意”、“3—不同意”、“4—非常不同意”，请对每个问题在相应选项打勾。下面有十个问题，每一个问题只能选一个选项，请选择最符合您观点的选项。您的所有答案都会被保密，请放心作答。

<table>
<thead>
<tr>
<th>序号</th>
<th>项 目</th>
<th>非常同意</th>
<th>同意</th>
<th>不同意</th>
<th>非常不同意</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>总体来说，我对自己很满意。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>很多时候，我觉得我自己什么都不是，什么事情都做不好。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>我觉得我的数学其实还不错。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>数学课堂和考试中，我觉得可以我可以做的和别人一样好。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>我觉得我的数学不是太好，没什么让我感到骄傲和自在的。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>6</td>
<td>我很确定我自己在几乎每次数学考试中都非常紧张，很不自在，但这和考试难易没太大关系。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>7</td>
<td>我觉得我做任何事都可以和别人一样好。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>8</td>
<td>我很想我自己获得自己和别人的赞扬。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>9</td>
<td>总的来说，我认为在学习上，尤其是数学，‘我有点失败’的趋向。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>我喜欢鼓励自己，积极地看待自己。</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
APPENDIX C
Appendix C1: Gender Attitudes and Math

Below is a list of statements concerned with your general feelings about gender and mathematics. Please circle the response that best reflects your personal feeling about the statement.

1. How important is your gender for your identity?

Not important at all 1 Neutral 2 Very important 4

2. How common is it for men to be good at mathematics?

Not common at all 1 Neutral 2 Very common 4

3. How common is it for women to be good at mathematics?

Not common at all 1 Neutral 2 Very common 4

4. There is no difference between male and female students in math performance.

Strongly Disagree 1 Neutral 2 Strongly Agree 4

5. A person’s math ability has nothing to do with one’s sex.

Strongly Disagree 1 Neutral 2 Strongly Agree 4

6. Overall, males are better than females in math.

Strongly Disagree 1 Neutral 2 Strongly Agree 4
Appendix C2: 对性别和对数学的态度

下面有一系列的陈述，需要您从这些述中选择最适合您观点的。下面左边是陈述栏，下面右边有五个选项："1 - 非常重要/同意"、"2 - 重要/同意"、"3 - 一般"、"4 - 不重要/不同意"、"5 - 非常不重要/不同意"。下面有 6 个问题，每一个问题只能选一个选项，请选择最符合您观点的选项。您的所有答案都会被保密，请放心作答。请不要选择不代表你真正内心观点的答案，谢谢。

1. 您认为您的性别对于您来说有多重要？

<table>
<thead>
<tr>
<th>非常重要</th>
<th>重要</th>
<th>一般</th>
<th>不重要</th>
<th>非常不重要</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

2. 您觉得男生数学好是一件很有代表性、很重要的事吗？

<table>
<thead>
<tr>
<th>非常重要</th>
<th>重要</th>
<th>一般</th>
<th>不重要</th>
<th>非常不重要</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

3. 您觉得女生数学好是一件很有代表性、很重要的事吗？

<table>
<thead>
<tr>
<th>非常重要</th>
<th>重要</th>
<th>一般</th>
<th>不重要</th>
<th>非常不重要</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

4. 总体来说，男生女生在数学考试中的发挥没有任何不同。

<table>
<thead>
<tr>
<th>非常同意</th>
<th>同意</th>
<th>一般</th>
<th>不同意</th>
<th>非常不同意</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

5. 认为一个人的数学能力和男女生理上的不同没什么关系。

<table>
<thead>
<tr>
<th>非常同意</th>
<th>同意</th>
<th>一般</th>
<th>不同意</th>
<th>非常不同意</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

6. 您同意总体来说男生比女生在数学考试中发挥更好的观点吗？

<table>
<thead>
<tr>
<th>非常同意</th>
<th>同意</th>
<th>一般</th>
<th>不同意</th>
<th>非常不同意</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>
Appendix D1

This is the last survey that I am going to collect from you. Thank you for your patience and participation so far. In this survey, some information about you will be collected.

1. SAT-Math scores (only if you took SAT; if you took it multiple times, you may indicate all scores):

   __________________________

2. ACT Math test scores (only if you took ACT; if you took it multiple times, you may indicate all scores):

   __________________________

3. High school GPA: ________________________

4. Sex: _____ Male _______ Female

5. Age: ______

6. Year in School: ___Freshman _____ Sophomore _____ Junior _____ Senior

7. Major (declared or intended): ________________________________

8. Currently Working or not? ________________________________
Appendix D2

这是最后一张调查卷，谢谢您到目前为止的的耐心和参加。最后这部分，我会问您您的一些个人过去信息。请记住，所有您的答案都会被保密，毕业论文完成后都会被销毁，请放心作答。您的回答和真实性至关重要。谢谢！

1. 请务必提供您的高考数学分数(Math Test Score)；

2. 您现在是否在工作(Job Info)?
   是    否

3. 您现在大学年级(Class/Year)(请圈示)：大一    大二    大三    大四

4. 您的性别(Sex/Gender)是：______________

5. 您的年龄(Age)是：__________________

6. 您的专业(Major)是：__________________
1. 6 pints of a 20 percent solution of alcohol in water are mixed with 4 pints of a 10 percent alcohol in water solution. The percentage alcohol in the new solution is

A. 16  B. 15  C. 14  D. 13  E. 12

2. ASB is a quarter circle. PQRS is a rectangle with sides PQ = 8 and PS = 6. What is the length of the arc AQB?

A. $5\pi$  B. $10\pi$  C. 25  D. 14  E. 28

3. If $x \circ y = (x + y)^2 - (x - y)^2$. Then $\sqrt{5} \circ \sqrt{5} =$

A. 0  B. 5  C. 10  D. 15  E. 20

4. The total weight of a tin and the cookies it contains is 2 pounds. After $\frac{3}{4}$ of the cookies are eaten, the tin and the remaining cookies weigh 0.8 pounds. What is the weight of the empty tin in pounds?

A. 0.2  B. 0.3  C. 0.4  D. 0.5  E. 0.6

5. Which of the following pairs of angles must be equal?

A. a and e only  B. a and e, and c and d only  C. c and d only  D. d and e only  E. c and d and d and e only

6. The number 0.127 is how much greater than $\frac{1}{8}$?

A. $\frac{1}{2}$  B. 2/10  C. 1/50  D. 1/500  E. 2/500

7. Which of the following could not be the lengths of the sides of a right angled triangle?

A. 3, 4, 5  B. 5, 12, 13  C. 8, 15, 17  D. 12, 15, 18  E. 9, 12, 15

8. Two equal circles are cut out of a rectangle of card of dimensions 16 by 8. The circles have the maximum diameter possible. What is the approximate area of the paper remaining after the circles have been cut out?

A. 104  B. 78  C. 54  D. 27  E. 13

$$a^2 - b^2 = \frac{a^2 + b^2}{a + b}$$

9. A. $a^2 + b^2 + 1$  B. $a + b$  C. $a - b$  D. ab  E. it cannot be simplified further

10. Which of the following could be a solution of the equation $|x| = |4x - 3|$?

A. -1  B. -0.6  C. 0  D. 0.6  E. 1.5
APPENDIX F
Appendix F: Debriefing Statement

Dear Participant,

Thank you so much for your participation in my senior thesis. I wanted to tell you a little about my project so that you would know what I am researching. I hypothesized that there is no gender difference among Maryville College students in their math performance. I was also trying to see whether gender-related stereotypes about math affect one’s math performance. Thank you for your time and effort. If you would like to know about the results of my project you can leave me a message in my campus mailbox (3253) or email me at

vilong.zheng@my.maryvillecollege.edu

Or please contact my thesis advisor Dr. Lori Schmied at

lori.schmied@maryvillecollege.edu. Once again, thank you for your time and participation.

Yilong Zheng
APPENDIX G
Appendix G: Debriefing Statement (调查结果报告)

亲爱的同学们，

感谢你们参加今天的调查。有了您的帮助，我的毕业论文才能顺利完成，而且定会更加优秀。下面我来告诉您一些关于我的实验调查的背后故事：我打算调查男女性别和数学考试成绩的关系。我是想通过对男生和女生对自己和数学态度的调查来分析是否个人自信程度和性别相关

的数学观念（比如：男生数学比女生好）会成男生和女生在考试中有和原有水平不吻合的发挥。尤其是女生，是否会因为这些传统观念对自己的能力产生质疑，从而没有发挥出应有的水平。在中国的调查因为没有一个统一的标准的测试，我只能通过高考成绩来做比较，其实这并不具备代表性，但我的目的在于发现中国的学生和美国的学生一样，男生女生会因为一些传统的观念在数学考试发挥中有差异。而这些差异，只有通过对分数来体现。如果您有兴趣了解更多

的信息，在毕业论文结题后，请联系我的邮箱 yilong.zheng@my.maryvillecollege.edu, 我会提供更多信息，最后，再次感谢您的鼎力帮助。

郑毅龙
APPENDIX H
The purpose of this study is to examine whether there are gender differences in math performance among undeclared freshmen and math major students in the United States and China. Using standardized test scores (SAT-M in the U.S. and the national mathematics entrance examination in China), math performance will be related to self-esteem in math ability. Another goal of this study is to determine whether math-related stereotypes exist among first-year students and students who are pursuing a future profession in mathematics field, both in the United States and China. A third goal relates to stereotype threat in the United States: after being primed on math-gender stereotypes using an attitude survey, it is also important to see whether this stereotype threat may affect students’ performance significantly on a standardized math test.
Participants:

The sample for the study will consist of a minimum of 50 participants in each country. In the United States, freshmen will be recruited in the general education statistics class. There will be no further selection for the participants. Students from the U.S. are preferred due to their similar mathematic background from high school. Math majors will be recruited from upper-level math courses. Confidentiality is very important in this study because their SAT-M score will be collected. Participant names will not be used.

In the study in China, students will be recruited from a comprehensive university, “East China Jiao Tong University (ECJTU). The cross-cultural comparison is the math performance between genders; therefore, the same math measure for the comparison of math performance between the two countries is not necessary. The difference between first year versus upper-level math major students will also be studied by asking their class/year in the last survey.

Methods and procedures: (attach instruments, interview guides, information letters etc.)

Informed Consent (& its Chinese version): Participants will be informed in the introduction that consent to participate in the experiment will be implied by their completion of the measures. No names will be collected and confidentiality of individual results will be maintained. Results will be reported in group format only (see Appendix A).

Math self-esteem scale (American and Chinese version): This is adapted from the Rosenberg's Self-Esteem Scale*, which has been widely used to measure one’s self-esteem, and it also has been used in several journal studies to measure the self-concept in one’s ability to function in various domains. The 10-item, four-point Likert scale has been modified to assess math self-esteem (See Appendix B1). The Chinese version is a translation of the above into Mandarin
Chinese according to the meaning of each item. It is modified to match the Chinese concept of self-esteem, so that participants can understand the choices more easily (See Appendix B2).


Math-Gender Stereotype Survey (American and Chinese version): This six-item survey assesses math and gender stereotyping, based on findings from the literature. The survey contains several statements with a five-point Likert response scale about math-related gender stereotypes. Also, the Chinese version is provided. See Appendices C1 and C2.

Personal information survey: This includes items for high school ACT and/or SAT-M scores, age, sex, year in college, high school GPA, working status, and major. The Chinese version of this last survey is the same except instead of ACT/SAT-M scores, the score for the Chinese general entrance examination test is requested. An equivalent high school GPA is not available because Chinese high schools use a different grading system; it is also difficult to obtain GPA due to GPA is not very significant to Chinese high school students. See Appendices D1 and D2.

Brief SAT-M test: This is an abridged version of the SAT-M, available from the Maryville College Math department. It consists of 15-20 items from a practice SAT-M test and will be administered in group format.

Debriefing statement: This is a statement that explains the purpose of the study and provides my contact information for the results of the study (see Appendix E).
Debriefing statement Chinese version: This is a statement that explains the purpose of the study and provides my contact information for the results of the study for the participants in Chinese University ECJTU (see Appendix F).

Permission for conducting study in a Chinese university:

Permission from the appropriate authorities at ECJTU will be obtained before collecting data in China (See Appendix G). ECJTU does not have formal IRB procedures. At this time, the chairperson of the math department at ECJTU has given permission to recruit participants in his department and will follow the procedure outlined in this IRB proposal. The informed consent and the various surveys have been translated into Mandarin Chinese and checked for accuracy of translation by an independent rater.

Principal ________________________________ Faculty ________________________________
Researcher ________________________________ Signature ________________________________
Supervisor ________________________________ Signature ________________________________

Committee Approval ________________________________ Date ________________________________
Signature
APPENDIX I
Appendix I: Study Permission in China

Approval and Permission for Research Study

心理调查申请书

My name is Yilong Zheng. I am a senior psychology major at Maryville College. I am conducting research for my senior study. This research is about factors that are related to math performance and is supervised by Dr. Lori Schmied in the Department of Psychology. I am planning to conduct a research study. The study will take approximately 35-45 minutes. Students' names will not be collected and data will be reported only in group form. Participation in this endeavor is voluntary. Students' completion of the surveys will constitute informed consent to participate in the study. At any time during the study, a student may choose not to answer questions and may stop participation at any time without penalty. I hypothesize that there is no gender difference on math performance, and I am also interested in examining whether gender stereotype affect one’s math performance. Please allow me to conduct the study, to recruit participants and collect data in your class at ECJTU. The time for conducting the study is between June and December, 2010. Thank you very much! For more information, please contact me at yilong.zheng@my.maryvillecollege.edu or 01-865-293-7041; or my supervisor at lori.schmied@maryvillecollege.edu.

您好！我的名字叫郑毅龙，我是就读于美国 Maryville College 的一名大四心理学学生。我有兴趣在贵校（院）开展一项小型心理学实验调查，请您给予相应指导并予以批准同意我在贵校的调查研究。我的调查将会进行 35-45 分钟，学生可以在任何时候参加和退出这项调查。我的调查时间是今年 6 月到 12 月。我的谢谢！我的邮箱是 yilong.zheng@my.maryvillecollege.edu; 电话是 01-865-293-7041; 我的导师 Dr. Lori Schmied 的邮件是 lori.schmied@maryvillecollege.edu.

Researcher Signature: ______________________  实验人签名     Date: ______________

School Representative: ______________________ 校方代表公章     Date: ______________

School Representative: ______________________ (签名 print)
REFERENCES


