5-2017

The Role of Stereotype Threat in Mental Rotation

Bridget A. Parler
Northern Michigan University, bparler@nmu.edu

THE ROLE OF STEREOTYPE THREAT IN MENTAL ROTATION

BY

BRDIGET A. PARLER

THESIS

Submitted to
Northern Michigan University
In partial fulfillment of the requirements
For the degree of

MASTERS OF SCIENCE DEGREE

Office of Graduate Education and Research

April 2017

SIGNATURE APPROVAL FORM

The Role of Stereotype Threat in Mental Rotation

This thesis by Bridget A. Parler is recommended for approval by the student's Thesis Committee and Department Head in the Department of Psychology and by the Assistant Provost of Graduate Education and Research.

_____

Committee Chair: Dr. Sheila Burns Date

_____

First Reader: Dr. Mary Pelton-Cooper Date

_____

Second Reader (if required): Dr. Judith Puncochar Date

_____

Department Head: Dr. Adam Prus Date

_____

Dr. Lisa S. Eckert Date
Interim Director of Graduate Education

Abstract

THE ROLE OF MENTAL ROTATION IN STEREOTYPE THREAT

By

Bridget A. Parler

Numerous studies have been conducted on mental rotation ability in both males and females. A lot of the literature discusses male superiority in mental rotation task as opposed to females. This paper examined the role of stereotype threat in a mental rotation task. Particularly, the research study focused on gender stereotype threat in female college students. The study was completed in a two part design with students completing two mental rotation task in a group setting. The low numbers in the sample make it impossible to reliably say that the Vandenberg Mental Rotation Task (VMRT) and the Purdue Mental Rotation Task (Purdue) were measuring the exact same thing. There was a confound in part 2 of the study and we were unable to draw conclusions on the role of gender stereotype threat. However, the findings do suggest more support for the research on video game experience and mental rotation performance. The implications are discussed.

Dedication

I would like to dedicate this work to my mother Brenda K. Wheeler and my godfather Patrick Cobb. Both passed away last year months apart before they could see this milestone be accomplished and this work come to fruition. Thank you for all your sacrifices that led me to this moment. Thank you for the selfless unconditional love you poured into me. *Dreams do come true*.

ACKNOWLEDGEMENTS

First, I would like to acknowledge and give all thanks to God. For without my faith and faithfulness I would not have made it this far. I would like to thank all the participants who gave their time to help fulfill my passion through this thesis work. Thank you to my thesis committee Dr. Sheila Burns (chair), Dr. Mary Pelton-Cooper and Dr. Judith Puncochar for their help and feedback on the project. A special thanks to Grace Albert for her help with so many small details of the project. Thank you to Professor Mike Strahan for his unlimited help not only with this project but with every step of my program. Thank you for the encouragement, the feedback, and for viewing me "as a whole person" and allowing me to grow as one. Thank you to Tyler Harris and Kelly Morrow for their help and feedback on various aspects of my project. Special thanks to Kristen Mills (Millz) for unmatched support, guidance and love throughout the process of me working on my Master's degree. It's a process I could not have gotten through without you. A special thank you to the entire DSO family for becoming my second family here during my time in Marquette, for providing love and support I did not even know I would need and for always being a safe space. Finally, thank you to my friends and family for your support in this long process.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The Role of Stereotype Threat in Mental Rotation Ability

Setting the Stage: Early Studies in Mental Rotation Ability

Shepard and Metzler (1971) published the paper which has motivated the current studies of mental rotation. Mental rotation is defined as a cognitive process in which an individual mentally rotates an object in the mind (Shepard & Metzler, 1971). The process is analogous to being able to physically touch the object and rotate it. Shepard and Metzler examined the reaction time of participants to observe the orientation of different drawings in three dimensional space and decide if the two were (Shepard & Metzler, 1971). The original study involved only eight participants given 1600 different pairs of line drawings to make judgments on. Subjects were measured over the course of eight to 10 sessions of one hour each. The subjects were timed on their ability to distinguish whether two shapes were congruent or mirror images. The major finding was that there was a strong linear relationship between reaction time and angular difference of the two pictures (Shepard & Metzler, 1971). This study is important because it was the seminal research for modern studies of mental rotation and reaction time. Since 1971 the literature surrounding mental rotation has become quite expansive. This paper will explore some of the relevant literature on the subject and examine the studies. Further, this paper will propose a new study to contribute to the body of literature.

A few years after Shepard and Metzler's study, Vandenberg and Kuse published their 1978 paper that introduced a different way to measure spatial ability. They developed a paper and pencil test using figures similar to those of the Shepard and Metzler. This test is now widely known as the Vandenberg Mental Rotation Test or the VMRT for short (Vandenberg & Kuse, 1978). The advantage of the VMRT is that it can be administered to a group. The VMRT is a five page document with a total of 20 items and is scored in terms of number of items correct

(Vandenberg & Kuse, 1978). This study was important because it established the first paper and pencil test to model after the original Shepard and Metzler task and is widely used today. While the early studies of computerized mental rotation make no reference to sex differences, sex differences became an issue in the literature after the introduction of the VMRT.

Collins and Kimura (1997) pointed out that the Vandenberg and Kuse paper and pencil test has consistently showed sex differences in mental rotation performance. It's not clear why the test has yielded such a strong sex difference. In their study, a new apparatus was developed to test the mental rotation ability in subjects. Students were administered a paper and pencil task consisting of both easy and hard items to complete. Unlike most VMRT studies, the subjects in this study were tested individually as opposed to the common group setting. Still, the results of the study revealed large sex differences on the test.

A meta-analysis by Voyer, Voyer, and Bryden (1995) found sex to be quite pronounced in certain mental rotation type tasks and smaller in others. The criteria for studies being included in the meta-analysis was quite stringent among the qualifications being studies used techniques well documented in the literature. A careful look at the studies included suggest those sex differences are large on the VMRT but smaller on computer based tasks.

Cherney (2008) found that computer game practice generally improved mental rotation scores. Women's improvement on the computer game was significantly greater than men's. In terms of method of delivery for computer video game practice they found that the most significant improvement when practice was massed. This finding contradicted both their hypothesis predicting that there would be greater improvement when the skill practice was distributed. Also, the trend in literature tends to favor more distributed practice of a skill.

This paper took the view that in general gender differences in spatial cognition exist. The study wanted to investigate what they felt was pretty understudied basic capacities that support spatial cognition. At the time of the current study only one study done in 2007 had looked at something similar (Kaufman, 2007). They had participants complete a pretest which include the useful-field-of-view (UFOV) task and a MRT task. Next, they did 10 hours of individual training on a video game that was supervised. Finally, they completed a post test which included the exact same measures as the pretest. They found that playing a action video game can enhance males' and females' performance on a spatial task. Females in the experimental group showed greater improvement than men (Feng, Spence, & Pratt, 2007).

Stereotype Threat

What is Stereotype Threat?

According to Steele and Aronson (1995) stereotype threat is the idea that one may confirm negative stereotypes about a group to which they belong. For example, if women are told that they significantly underperform males on a mental rotation task because men are better at the task, then women stereotype threat theory states this idea alone can hinder a woman's performance on the task because she's confronted with a stereotype about her group. Knowledge of the stereotype increases anxiety which inadvertently decreases performance especially on a different task. Steele & Aronson (1995) provides foundation for the theory of stereotype threat, as well as its importance. The theory was developed in the context of racial differences on standardize test. The paper includes four studies that examine various ways the stereotype can be shown in different circumstances. For instance, in study one of the experiment students, African American and White, were given a task that involved questions from the GRE verbal section and which they had 30 minutes to complete. The participants were given different directions for the

test depending on the condition they were placed in; stereotype threat condition or control group. In the control group students were told it was a general task. However, the stereotype threat group was told the test was linked to measures of intelligence. No difference in the control condition, but that African Americans' performance diminished on the test under the threat condition. As a whole the entire study showed the effects of stereotype beliefs and how they can function in some cases to enhance performance affirming belief and in others to diminish performance. Further, the study showcase the variance in the performance of African American students versus white students when confronted with negative group threats. Steele and others have replicated the initial findings many times (Steele, 1997).

Studies examining the theory of Stereotype Threat

Research has even shown that a person doesn't necessarily have to belong to a minority group to be affected by the stereotype but "…care enough about performing well to be bothered by a stereotype implication that they may lack the ability to do so" (Aronson et al., 1999, p. 40). Aronson and colleagues found that in a sample of white males with relatively high performing math abilities their ability to perform was diminished when given stereotype threat instructions.

Rodriguez (2014) shows that stereotype threat can be found when presenting negative stereotypes to Hispanics students. He pointed out there is lack of research on Hispanics students in regards to stereotype threat. Also, a predominant amount of the research has looked at White vs. African American students. Therefore, his study was one a few to begin to fill in an important literature gap missing in the theory. Rodriguez was able to replicate with Hispanic students the findings that are found with African American students. This reinforces the importance of the theory in potentially helping to explain students' decreased academic achievement when a stereotype can be elicited. In the Rodriguez study students (N=62) were

randomly placed in either a control condition or threat condition. All students regardless to condition were given an exam to complete that sampled questions from the SAT exam. The questions included various components of the exam (i.e. reading or grammar). Students in the threat condition read literature concerning the achievement gap that exist in education while those in the control condition did not. The results showed those students exposed to the negative threat showed less achievement than the control group (Rodriguez, 2014).

Research has also shown a difference in performance based on the source given for poor performance in a given area. Dar-Nimrod & Heine (2006) investigated how women's math performance is affected when manipulating the reasons given for women's underachievement in areas such as mathematics. In particular the researchers looked at genetic and experiential accounts for explanations on women's math performance. Genetic sources are seen as internal and something that cannot be changed while experiential sources are seen as changeable. Women took a Graduate Records Exam (GRE) like test where they completed two separate math sections that were separated by a verbal section. The verbal section contained the manipulation used in the study regarding math related sex differences. What they found is that women in the experiential and the no difference condition significantly outperformed those women in the genetic and stereotype threat conditions (Dar-Nimrod & Heine, 2006). They concluded that the study provides evidence that women's stereotype threat in mathematics can be reduced when women are presented with experiential reasoning for the stereotype (Dar-Nimod & Heine, 2006).

Moe (2009) looked at the role of the gender belief explanation in stereotype threat for men and women. This study looked at two kinds of motivational beliefs to examine the effects on participant's performance on a mental rotation task when confronted with a stereotype threat. In particular, the researcher looked at both the perceived task difficulty (i.e. representing the task

as easy or difficult) and the belief that one's own gender is better or worse than the other gender in the task (researchers called this 'able in the task stereotype'). The findings indicate that female participants were not affected by the potential difficulty of the task while male participants were. Additionally, females increased performance when instructed they typically perform better than males on the task (Moe, 2009).

Similarly, Moe (2012) found that participant's performance was negatively affected by instructions that stressed genetic factors as the reasoning for gender differences. On the other hand they found a positive relationship for instructions that stressed external sources such as time limit. Unlike most of the researcher looking at college level men and women this study looked at high school students (age range 14-18 years old; mean age 15.50 years old). Therefore, this study helped in broadening the scope of the literature.

McGlone & Aronson (2008) explored the role of identity salience in stereotype threat and spatial reasoning abilities in men and women. The researchers found that women who were primed to consider their identity as a member of a "selective private college" performed better than those women who were primed to consider their identity as women. The researchers concluded that reminding women of important achieved identities can be important in buffering stereotype threat (McGlone & Aronson, 2008).

Greater Implications: The Learning Process

An implication from the stereotype threat theory is whether or not the idea itself can hinder the process of learning itself. Taylor and Walton (2011) examined whether stereotype threat hinders the process of learning. In the first experiment, black and white students were assigned to either a threatening condition or a non-threatening condition where they would

complete two sessions looking at learning and word recall. Students were asked to learn the definitions for rare words and study them. However, the threat was present in the instructions these students received. Some were told the study wanted to understand how they learned; others were told the study looked at their ability to learn the material (threat condition) (Taylor & Walton, 2011). After about a week students were brought back in to test their recall. This study found that Black students in the threatening conditioning consistently underperformed their White counterparts when confront with negative group stereotypes. Interestingly, white students in this study were not found to be affected when compared to the control group.

The second study was conducted to look at whether or not Black students in the experiment could self-affirm, in a way that would moderate the negative stereotypes presented to them. Most of the same procedures were the same as described in their first experiment. However, noted that prior to the students in experiment two doing any "learning" with the words they were given another task at the beginning. In this study, students either wrote about a value that was important to them or they wrote about a value that wasn't important to them but could matter to someone else (Taylor & Walton, 2011). This exercise was used as a moderator to see what effect it would have throughout the study and this method is one that was previously shown to have positive effects (Sherman & Cohen, 2006). The results suggest that self-affirmation can moderate the stereotype threat effects.

Current Study

Purpose and Hypothesis

From the research cited in the literature review above there are circumstances where being confronted with a negative stereotype can impair performance. The current research study used stereotype threat theory to examine the gender differences in mental rotation. The purpose of the study will be to examine the ability of participants to mentally rotate objects even when faced with gender threats. Additionally, to observe how gender threats can impact a female college student's ability on a given task, to contribute to the literature on mental rotation and stereotype threat. The main hypothesis of the current study is that female college students who are confronted with a negative stereotype before beginning their task will perform worse than those students who are not confronted with a threat (control) before beginning their task. The second hypothesis of the study is that female college students who are explicitly told that the task they are about to perform shows strong gender differences favoring women will perform better.

Methods

IRB Approval

This research project was approved by Northern Michigan University Institutional Review Board (IRB). The research included into two smaller parts and separate IRB applications were completed. The first study was approved under the expedited review, proposal number: HS-787 on September 6, 2016. The second study was approved under the expedited review, proposal number: HS16-799 on September 30, 2016.

Participants

Participants in the study were recruited from psychology courses at Northern Michigan University. The sample size of study one included approximately N= 31 college students (n= 10 male students; n= 21 female students). The sample size of study two included approximately N=37 female college students. The ages ranged between 18-53 years for both of the samples. Students participating in the study may have received course credit for their time in the study. No other compensation was provided for study participants.

Measures

Two measures of mental rotation were used in the study. The first was the Vandenberg Mental Rotation Test (VMRT; Vandenberg and Kuse, 1978). The VMRT is a 20 item paper and pencil mental rotation task. The second was the Purdue Visualization of Rotations Task (ROT; Bodner and Guay, 1997) which is a condensed version of the Spatial Visualization Test Battery. The Purdue Visualization of Rotation Task is also a 20 item paper and pencil mental rotation task. Both measures were be used to assess participants' ability to mentally rotate objects.

General Procedure

Students were run through the experiment in groups of up to twenty students though the majority of groups included less than 10 participants. Upon entering the lab students were assigned to a seat by a researcher after verifying their names on the sign-up sheets. Each seat included an experiment booklet, consent sheet and pencil on the desk for the participant all face down so no information showed. Each booklet contained instructions for beginning each section of the experiment, the Vandenberg and Kuse Mental Rotation paper and pencil task, the Purdue Visualization of Rotations test, a demographic questionnaire, and a passage to read with questions to answers (study two only). Participants were instructed to sit but asked not to touch or turn over any of the papers on the desk until asked to do so. The researcher ensured all cellphones were turned off before officially beginning the experiment. Next, each student was instructed to read the informed consent and asked if they have any questions. If they agreed to participate in the study each student was asked to sign an informed consent form for the study.

Study One

Study one included a mixed group of male and female college students. The participants completed both the Vandenberg Mental Rotation Task and the Purdue Visualization Task in a counterbalanced order to control for order effects. Before beginning the first mental rotation task participants completed practice problems. They had exactly five minutes to work on the example problems for the first mental rotation task. If participants were still working when time was up the experimenter simply stated “please stop”. The timer was reset and the instructions were read to participants to start the first mental rotation actual set of problems. The mental rotation task was timed for exactly eight minutes and participants were told when to begin and end by the researcher. If participants were still working when time was up on the mental rotation task the experimenter simply stated “please stop”. This same sequence was followed for the second

mental rotation task. Afterwards, all participants in study one completed a demographic questionnaire at the end of the study (please refer to appendix for an example of the experiment booklets given to participants). Primarily, study one was looking at whether the sex differences are similar on both mental rotation tasks.

Study Two

Study two was comprised of only female college students. Again, all participants were run in groups using the same parameters as study one of up to twenty students. The general procedure applies to study two as well. The participants in study two completed both the Vandenberg Mental Rotation Task and the Purdue Visualization Task in a counterbalanced order. However, before beginning any mental rotation task participants were instructed to read a brief passage and answer the questions that followed. Research has demonstrated that a commonly used way to elicit stereotype threat is by instructing that gender differences are present on the task you are asking participants to perform (Moe, 2009) or by priming gender (Steel & Aronson, 1995; Moe & Pazzaglia, 2006). There were three identical possible passages participants could have read with the exception of the last two lines. The last two lines of each passage contained the stereotype threat manipulation using either female threat, male threat or neutral condition. Female threat stated "….Interestingly, the tasks we are asking you to perform are from a set which showed strong gender differences, specifically females outperform males on these tasks. Part of our interest is in whether these differences still exist." Male threat stated "….Interestingly, the tasks we are asking you to perform are from a set which showed strong gender differences, specifically males outperform females on these tasks. Part of our interest is in whether these differences still exist." Finally, the neutral condition stated "….Interestingly, the tasks we are asking you to perform are from a set which showed no gender differences,

specifically males and females perform the same on these tasks. Part of our interest is in whether these similarities still exist." Please see appendix for the complete passage used for all three conditions.

Before beginning the first mental rotation task participants completed practice problems and had exactly five minutes to work on them. If participants were still working when time was up the experimenter simply stated "please stop". The timer was reset and the instructions were read to participants to start the first mental rotation actual set of problems. The mental rotation task was timed for exactly eight minutes and participants were told when to begin and end by the researcher. If participants were still working when time was up on the mental rotation task the experimenter simply stated "please stop". Again, this same sequence was followed for the second mental rotation task. Afterwards, all participants in study two completed the same demographic questionnaire at the end of the study (please refer to appendix for an example of the experiment booklets given to participants in study two).

Results

For purposes of comparison between the VMRT and the Purdue test, this discussion uses two percent correct measures for the VMRT and one for the Purdue. The standard scoring of the VMRT has a correction for guessing, we call this the VMRTstnd. One can also look at the VMRT as having a total of 40 potentially correct responses. The number correct of these forty (without a guessing correction) is called the VMRTlt. The Purdue standard scoring has no guessing correction and includes 20 possible correct responses. So that a comparison between the VMRT and the Purdue can be made all scores have been converted to percent correct. The VMRT% score = VMRTstnd/40*100; the VMRTlt% = VMRTlt/40*100; and the Purdue% = Purdue/20*100.

Part I

For part 1 of this study there were 31 participants (F = 21, M = 10). The small sample size, especially for males make it difficult to make the intended comparison. The mean age for the participants was 19 years old, females = 18.57 years and males = 19.10 years. The age .251 difference was not significant, t (29) = 1.17, p = .251.

Table 1 shows the descriptive statistics on each of the five variables which were run to examine sex differences on the VMRT and Purdue tests: VMRTstnd = standard scoring, with a correction for guessing; VMRT40 = VMRT scoring without a correction; VMRTatt = number of problems attempted of 20; Purdstnd = the number correct on the Purdue test of 20; and Purdatt = the number of Purdue items attempted of 20. This table shows the means, number of cases, standard deviations, and standard errors of the mean for the primary variables in Part 1.

Table 1
Group Descriptive Statistics

Variable	Gender	N	Mean	Std. Deviation	Std. Error Mean
VMRT stnd	Female	21	17.43	9.87	2.15
	Male	10	20.8	13.38	4.23
VMRT 40	Female	21	23.71	8.66	1.89
	Male	10	26.6	10.54	3.33
VMRT att	Female	21	16.57	4.4	0.96
	Male	10	18	3.37	1.06
Purd stnd	Female	21	10.62	3.5	0.76
	Male	9	11.33	4.36	1.45
Purd att	Female	21	16.86	3.31	0.72
	Male	9	16	4.09	1.36

Males outperformed females, in absolute numbers, on all measures except the number of Purdue items attempted. However, none of these differences were significant, as will be shown below. Note that one male did not complete the Purdue test.

Table 2 shows independent groups t-test for each of the variables. Because the sample sizes were so different, both the t assuming equal variances and the t not assuming equal variances are displayed. There were no significant differences between females and males on any of these measures, all $t(28 \text{ or } 29) < 1.00$, $p > .05$.

Table 2. Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
VMRTstnd	Equal var. assumed	2.36	.14	-.79	29	.44	-3.37	4.26	-12.08	5.34
	Equal var. not assumed			-.71	13.85	.49	-3.37	4.75	-13.57	6.82
VMRT40	Equal var. assumed	.12	.73	-.81	29	.43	-2.89	3.57	-10.18	4.41
	Equal var. not assumed			-.75	15.01	.46	-2.89	3.83	-11.05	5.28
VMRTatt	Equal var. assumed	1.93	.18	-.91	29	.37	-1.43	1.58	-4.66	1.80

Mental Rotation

	Equal var. not assumed			-1.00	22.81	.33	-1.43	1.43	-4.40	1.54
Purdstnd	Equal var. assumed	1.33	.26	-.48	28	.64	-.71	1.50	-3.79	2.36
	Equal var. not assumed			-.44	12.65	.67	-.71	1.64	-4.27	2.84
Purdatt	Equal var. assumed	.62	.44	.61	28	.55	.86	1.41	-2.04	3.75
	Equal var. not assumed			.56	12.70	.59	.86	1.54	-2.48	4.20

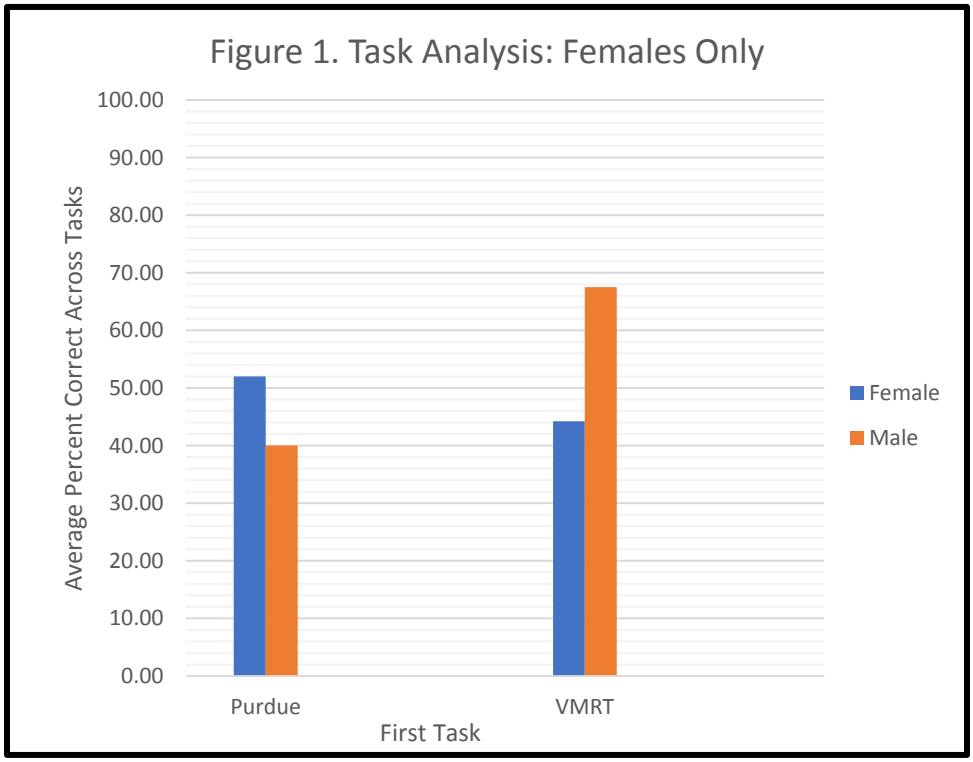
Table 3

Participant Means on Primary Measures

Gender	Measure	VMRT stnd	VMRT 40	VMRT att	Purdue stnd	Purdue att
Female	Mean	17.43	23.71	16.57	10.62	16.86
Male	Mean	20.8	26.6	18	11.33	16
Total	Overall Mean	18.52	24.65	17.03	10.83	16.6
Total	N	31	31	31	30	30
Total	Pooled Stnd Dev	11.01	9.23	4.09	3.71	3.51
Total	g	0.31	0.31	0.35	0.19	-0.24

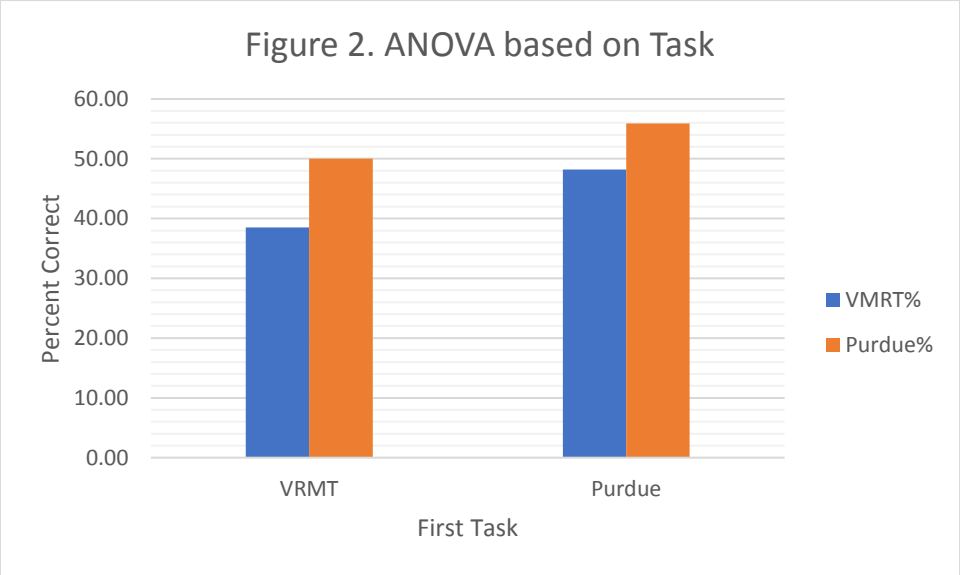
Table 3 shows the means for female and male participants on the primary measures, along with the pooled standard deviation and g. Note that for the VMRT, the size of effect, g, was consistently just above .30, which is actually low for the paper-and-pencil mental rotation test Voyer et al., 1995). If this is an accurate assessment of the size of effect it would take a much larger sample to detect the effect as a significant difference. The sex difference effect on the Purdue test may be smaller (g = .19), but it is difficult with the small number of males to draw a strong conclusion.

Analyzes of order effects. A repeated measures mixed model, task (repeated measure) by sex by first task, analysis of variance was done to evaluate potential order effects. The VMRT and Purdue scores were converted to percent correct (of 40 for the VMRT and of 20 for the Purdue) scores to allow a comparison of performances. The only significant effect was the first task x sex interaction, $F(1, 26) = 4.88$, $p = .038$. Figure 1 below shows this interaction. When the Purdue test was first, females performed better than males; when the VMRT was first, males performed better than females. However, with the small number of males in this study, it is difficult to assess the reliability of this result. The analysis is in Appendix xxx.



In part 1 number of males and females very different and males performances were more variable (larger standard deviations), analysis of order effects and task performance was done for females only.

Table 2 shows the mean percent correct on each task as a function of whether it was a first or second task. A two-way, mixed model ANOVA (task x first task) was completed and the only significant effect was the difference between the two tasks. The Purdue test was easier, in terms of percent correct than the VMRT. This can be seen in Figure 2.



Part 2 Summary

For part 2 of this study there were 37 participants who were all female (F =37). The mean age for the participants was XX years old. Design was a three (threat condition) by two (tasks) factorial.

Design note. While the original design included a counterbalanced presentation of the two tasks, that was not accomplished in Part 2. As can be seen in Table 1, overall, the first task was balanced, but this was not the within each threat condition. In the condition indicating that mental rotation was a "female" task, all 15 participants received the VMRT first. In the condition indicating that mental rotation was a "male" task, 12 of 13 participants had the Purdue first.

Table 4
Breakdown of First Task Completed by Condition

		Female Task	**Neutral Task**	**Male Task**	**Total**
First Task	**Purdue**	0	4	12	16
	VMRT	15	5	1	21
Total		15	9	13	37

Figure 1 shows performance as measured by the three percent correct scores as a function of the script instruction. For the VMRT the performance is actually best when participants are told it is a male task. The differences on the Purdue are minimal (see table 2 for means and standard deviations). It cannot be determined whether the apparently superior performance under the male script condition is because this was the second task or because of the instruction.

Mental Rotation

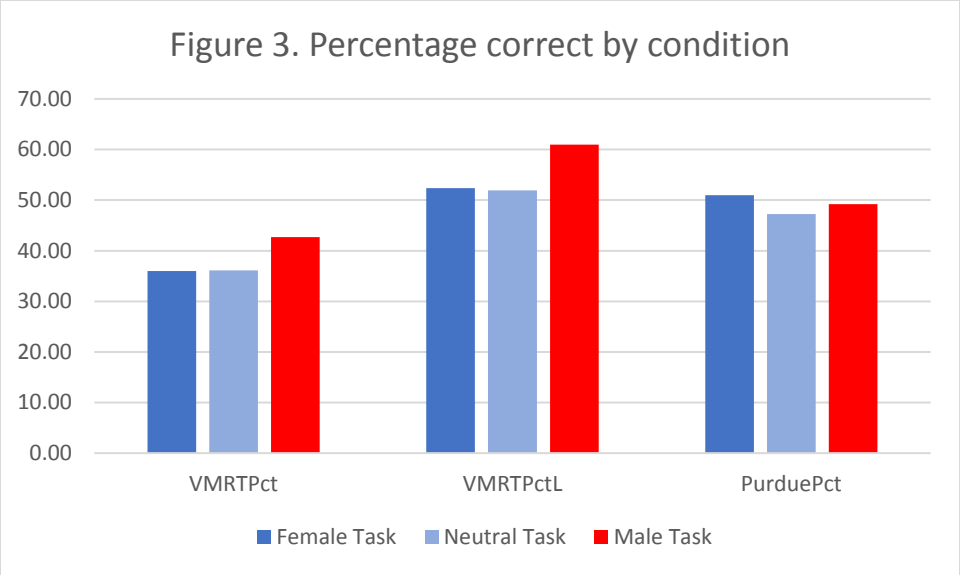


Table 5
Percent Correct on the Mental Rotation Tasks

Script		VMRTPCT	VMRTPCTL	PurduePCT
Female Task	Mean	36	52.33	51
	N	15	15	15
	St Dev	21.65	15.99	16.92
Neutral Task	Mean	36.11	51.94	47.22
	N	9	9	9
	St Dev	22.29	20.8	20.63
Male Task	Mean	42.69	60.96	49.23
	N	13	13	13
	St Dev	26.07	18.27	24.57
Total	Mean	38.38	55.27	49.46
	N	37	37	37
	St Dev	23.01	18.02	20.23

Comparison of VMRT and Purdue Test Redux. Understanding that the tests are different, the three percent correct measures, VMRT%, VMRTL%(no guessing correction), and Purdue% were compared. With all the data pooled (59 females and 9 males), a mixed model (tasks-core by sex) analysis of variance does not show any sex difference, $F(1, 65) = 2.177$, $p = .145$, $eta^2_{partial} = .032$; there were no test score differences.; and no interaction. Using the pooled data, the size of effect, g, for these comparisons is .46, .48, .30 respectively, possibly indicating that the sex effect in the Purdue mental rotation test is smaller than in the

VMRT. However, these are again all smaller than one would expect from a paper-and-pencil test according to the Bryden et al (1995) review.

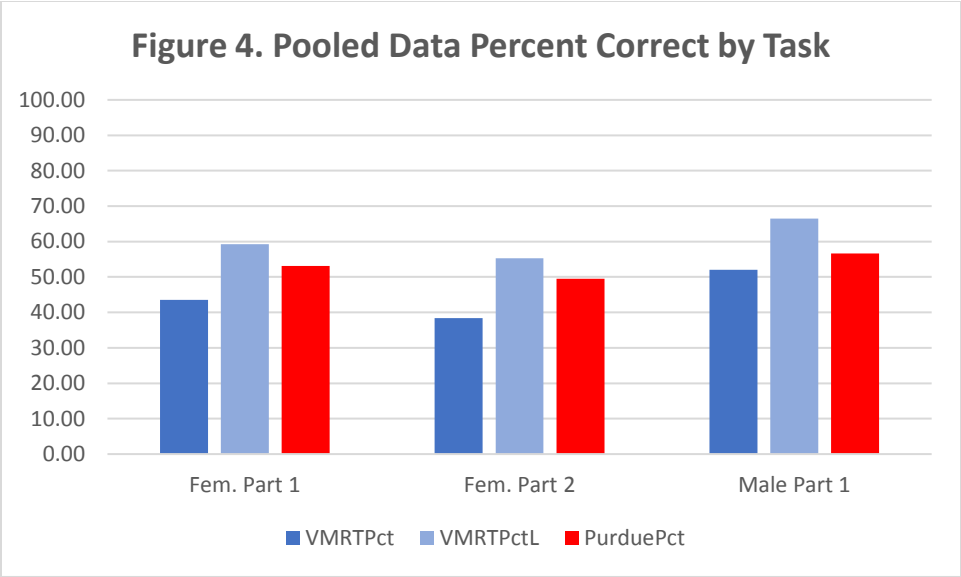


Table 6
Pooled Data for the Participants in both Parts

Sex		**VMRT % Ct**	**VMRT %ctL**	**Purdue %ct**
Female	Mean	40.26	56.72	50.78
	N	58	58	58
	Std. Deviation	23.55	19.32	19.21
Male	Mean	52.00	66.50	56.67
	N	10	10	9
	Std. Deviation	33.45	26.36	21.79
Total	Mean	41.99	58.16	51.57
	N	68	68	67
	Std. Deviation	25.29	20.57	19.51

A secondary interest in this research was the comparison of two paper-and-pencil tests of mental rotation. The VMRT has been used extensively to demonstrate sex differences in spatial abilities. The Purdue test, is relatively newer and has not been as extensively used. One way of examining whether the two tests measure the same thing is to look at correlations between performances on the tests. In personality research this usually falls under the establishment of construct validity. If Test A is an established measure of a concept and is highly correlated with a new instrument, Test B, we say that there is a clear overlap in what the tests are measuring and that Test B also seems to measure the Test A concept. In this case, to assess whether the Purdue Test was measuring the same information as the VMRT, correlations were run between the two measures. Pooling all subjects together (N = 67). The overall correlation $r(65) = .476$, $p < .001$, was significant suggesting that the percent overlap in variance was roughly 23%, a moderate correlation. Table 4 shows the correlations figured on subgroups as well as the total group of participants. The correlations are all in the greater than .40 and less than .60 range except for the

females in part 2. These correlations suggest that the two tests have a moderate, but clearly not perfect correlation.

Table 7
Correlations between VMRT AND Purdue Task

	All cases combined	All Females	Females Part 1	Females Part 2	Males Part 1
N	67	58	21	37	9
df	65	56	19	35	7
r r^2	.476 .23	.440 .21	.595 .35	.355 .13	.598 .36
	P < .001	p = .001	p = .004	p = .031	p = .089

Finally, it is interesting to note that for the females (in both parts) when the correlation is calculated separately for those participants who had the VMRT first the relationship between the two performances is lower, $r(29) = .226$, $p = .222$, $r^2 = .05$, than when calculated for those who had the Purdue first, $r(25) = .593$, $p = .001$, $r^2 = .35$. The relationship appears to be stronger when starting with the Purdue. However, there was no significant difference between VMRT or Purdue performances when compared as a first or second task. This is speculation since the research was not designed to look at transfer between the two tasks.

Discussion

Part 1 Summary

The purpose of part 1 was to compare the performance of males and females on the Vandenberg Mental Rotation paper and pencil test with the Purdue paper and pencil mental rotation test. We were interested in whether or not the Purdue test would also yield high gender differences as the VMRT has done for years. Low number of male subjects and uneven distribution of males and females made this impossible to reliably establish that the VMRT and the Purdue show the same gender differences. For sample, there were no sex differences shown but it was also a small uneven sample so that also has to be taken into account. The size of effect, g=.3 which is lower than expected for a paper and pencil mental rotation test. However, there was some indication that the Purdue test was easier for the females than the VMRT. It must also be note that cross test comparisons are hampered by both the fact that these are different tests, and hence different dependent variables, and that the scoring of the VMRT includes a guessing correction and that is not true for the Purdue.

Part 2 Summary

The purpose of part 2 was to add to existing literature by examining stereotype threat on two mental rotation tasks. A confound in part 2 of the study in which the test were not administered to participants in a counterbalanced order. For this reason we are unable to draw conclusions on stereotype threat or task transfer outcomes. When all of the data was pooled together for both part 1 and part 2 a moderate correlation for both of the mental rotation test was found. There was a significant correlation between the video game performance and mental

rotation performance for both part 1 and part 2 which is consistent with existing literature (Cherney, 2008; Spence & Feng, 2010). This suggest that experience with computers and video-game experience may account for some of the differences between males and females on mental rotation tasks. Further research is needed to determine the specifics and limits of this relationship. This suggest that the Purdue test does overlap but not perfectly with the VMRT in what they are measuring. It is not clear from our research that the two test are equivalent. For future researchers this is an area that would need to be looked at and expanded upon more in the literature.

In conclusion, the experiment did not fulfill the original purpose as stated in the beginning of the discussion section. However, the data could be looked at as supporting the literature on the relationship between video games and mental rotation performance. While we did observe some overlap in the two mental rotation task chosen it was a moderate relationship at best and needs to be looked at by future research.

References

Aronson, J., Lustina, M. J., Good, C., Keough, K., Steele, C. M., & Brown, J. (1999). When white men can't do math: Necessary and sufficient factors in stereotype threat. *Journal of Experimental Social Psychology, 35*(1), 29–46. http://dx.doi.org/10.1006/jesp.1998.1371

Collins, D. W., & Kimura, D. (1997). A large sex difference on a two-dimensional mental rotation task. *Behavioral Neuroscience*, *111*(4), 845-849. http://dx.doi.org/10.1037/0735-7044.111.4.845

Dar-Nimrod, L., & Heine, S. J. (2006). Exposure to scientific theories affects women's math performance. Science, 314, 435.

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

Moe, A., (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. Learning and Individual Differences, 19 (1), 21-27.

Moe, A., (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual differences, 22* (1), 20-24.

Moe, A., & Pazzaglia, F., (2006). Following the instructions! Effects of gender beliefs in mental rotation. *Learning and Individual differences, 16,* 369-377.

Rodriguez, B. A. (2014). The threat of living up to expectations: Analyzing the performance of Hispanic students on standardized exams. *Journal of Hispanic Higher Education*, *13*(3), 191–205. http://dx.doi.org/10.1177/1538192714531292

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*(3972), 701–703. http://dx.doi.org/10.1126/science.171.3972.701

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, *38*, 183–242. http://dx.doi.org/10.1016/S0065-2601(06)38004-5

Spence, I., & Feng, J. (2010). Video games and spatial cognition. *Review of General Psychology, 14*(2), 92-104. doi:http://dx.doi.org/10.1037/a0019491

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*(6), 613–629. http://dx.doi.org/10.1037/0003-066X.52.6.613

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*(5), 797-811. http://dx.doi.org/10.1037/0022-3514.69.5.797

Taylor, V. J., & Walton, G. M. (2011). Stereotype threat undermines academic learning. *Personality and Social Psychology Bulletin*, *37*(8), 1055–1067. http://doi.org/http://dx.doi.org/10.1177/0146167211406506

Vandenberg, S. G., & Kuse, A. R., (1978). Mental rotation: A group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47* (2), 599-604. http://dx.doi.org/10.2466/pms.1978.47.2.599

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270. http://doi.org/http://dx.doi.org/10.1037/0033-2909.117.2.250

Appendix 1

Research Participant Informed Consent

Purpose of the study

You are being asked to voluntarily participate in a research study interested in mental rotation ability as well as factors that can influence performance in the task. I am interested in the mechanisms that reinforce gender stereotypes and thus, affect performance in female college students. Before we can begin, we need you to carefully read over this sheet and decide if you would like to continue participation in the study. You should feel free to ask the researchers any questions you have.

Nature of the task

In this session, you will be asked to complete two separate mental rotation tasks. Each task will take 10-15 minutes. Both tasks will be timed by a researcher. You will be instructed when to start and stop each mental rotation task by a researcher. You will also be asked to complete a demographics questionnaire at the end of the session including questions such as your major, your minor and age. The entire experiment should be about 45 minutes to complete from beginning to end.

Potential Risk and Benefits

Risk in this research study is low. The mental rotation tasks you will complete in this study are commonly used in psychological research. Some participants may find the mental rotation tasks boring while others may find the tasks interesting. Sometimes performing in a new situation such as this can feel stressful

You will not directly benefit from participation in this study. The primary benefit will be a greater understanding of mental rotation and factors influencing performance on mental rotation task. Finally, participations will have the option to obtain research credit if allowed in their psychology course.

Cost

There are no cost to participating in the current research study beyond the time you invest to participate.

Confidentiality

Your confidentiality will be protected because no names will be associated with any data collected by researchers. All data will be coded by a numbering system that cannot be connected to your name. Finally, data will be reported only in terms of group averages. We collect names only to check whether a person has already participated. We only report average performance, never individual performance.

Right to withdraw

Your participation in the current study is voluntary. You may discontinue at any time without penalty. You have the right to say no and not participate in the study. You may choose not to answer specific questions without penalty. If you decide not to participate there will be no negative consequences. You may discontinue at any time without penalty.

Questions

If you have any further questions, comments or concerns please contact either the primary investigator Bridget Parler, at bparler@nmu.edu or the faculty research advisor Dr. Shelia Burns, at sburns@nmu.edu or the Dean of Research, Dr. Robert Winn at rwinn@nmu.edu.

Signature of Participant Date

Participant Name (please print)

Appendix 2

Manipulation Check Questions (Part II only)

Please read each question and select the answer that best answers the question.

1. In the instructions, were you told that in these tasks:
   (a) Women outperform men.
   (b) Men outperform women.
   (c) Men and women perform equally.

2. Based on your answer given for question 1 do you feel this is true?
   (a) Yes
   (b) No

3. If the statement is true, why do you feel it is true?

Appendix 3

Achievement Gap Article

The term “achievement gap” in education refers to the disparity in performance between groups of students. The achievement gaps often show up from early childhood into adulthood and is in grades, standardized-test scores, course selection, dropout rates, and college-completion rates, among other success measures. It is often used to describe the troubling performance gaps between female and male students. In the past decade, though, scholars and policymakers have begun to focus increasing attention on achievement gaps, such as those based on sex, English-language proficiency and learning disabilities.

In recent years, closing achievement gaps between various student groups has become a focus of federal education accountability. This created greater awareness of gender disparities and contributed to rising concern about kinds of achievement gaps. The attention led to more targeted interventions for different groups of students, but had not closed most achievement gaps to an appreciable degree a decade after the law was passed. Interestingly, the tasks we are asking you to perform are from a set which showed strong gender differences, specifically females outperform males on these tasks. Part of our interest is in whether these differences still exist.

Appendix 4

Achievement Gap Article

The term “achievement gap” in education refers to the disparity in performance between groups of students. The achievement gaps often show up from early childhood into adulthood and is in grades, standardized-test scores, course selection, dropout rates, and college-completion rates, among other success measures. It is often used to describe the troubling performance gaps between female and male students. In the past decade, though, scholars and policymakers have begun to focus increasing attention on achievement gaps, such as those based on sex, English-language proficiency and learning disabilities.

In recent years, closing achievement gaps between various student groups has become a focus of federal education accountability. This created greater awareness of gender disparities and contributed to rising concern about kinds of achievement gaps. The attention led to more targeted interventions for different groups of students, but had not closed most achievement gaps to an appreciable degree a decade after the law was passed. Interestingly, the tasks we are asking you to perform are from a set which showed strong gender differences, specifically males outperform females on these tasks. Part of our interest is in whether these differences still exist.

Appendix 5

Achievement Gap Article

The term “achievement gap” in education refers to the disparity in performance between groups of students. The achievement gaps often show up from early childhood into adulthood and is in grades, standardized-test scores, course selection, dropout rates, and college-completion rates, among other success measures. It is often used to describe the troubling performance gaps between female and male students. In the past decade, though, scholars and policymakers have begun to focus increasing attention on achievement gaps, such as those based on sex, English-language proficiency and learning disabilities.

In recent years, closing achievement gaps between various student groups has become a focus of federal education accountability. This created greater awareness of gender disparities and contributed to rising concern about kinds of achievement gaps. The attention led to more targeted interventions for different groups of students, but had not closed most achievement gaps to an appreciable degree a decade after the law was passed. Interestingly, the tasks we are asking you to perform are from a set which showed no gender differences, specifically males and females perform the same on these tasks. Part of our interest is in whether these similarities still exist.

Appendix 6

Research Participant Information Sheet

1. Can you tell me how you did the mental rotation task?

2. How often do you play video games?
   Not 0------1-------2------3------4------5------6------7------8------9------10 All the
   At all Time

3. In the mental rotation task, did you feel as though you were mentally rotating the image? If so, in what direction did you rotate the image?

4. How difficult did you find it to mentally rotate the image? (Scale 0-10 with 0 being NOT AT ALL and 10 being EXTREME DIFFCULTY).
   Not 0------1-------2------3------4------5------6------7------8------9------10 Extremely
   At all Difficult

5. How confident were you in completing the mental rotation task? Why?

6. Did your confidence change throughout the experiment? If so, please explain.

7. Did you feel pressure to do well? Why or why not?

8. Are you aware of negative stereotypes against your group? If so, what negative stereotypes?

9. Are you aware of negative stereotypes against women? If so, what negative stereotypes?

10. What do you think this research study is about and why?

11. Sex/Gender (Circle): Female Male
12. Age in years: _____
13. Handedness (Circle): Right Left Ambidexterity
14. Eyesight (Circle): Normal Corrected Other _____
15. Did you have any problems reading the instructions or seeing the figures?

16. What is your race/ethnicity? (Circle one)
    - White/Caucasian
    - Black/African American
    - Hispanic
    - American Indian/Alaska Native
    - Asian/Pacific Islander
    - Bi-racial
    - Other _____

Appendix 7

IRB Approval Part I

TO: Bridget Parler
Psychology Department

CC: Sheila Burns
Psychology Department

DATE: September 6, 2016

FROM: Rob Winn, Ph.D.
Interim Assistant Provost/IRB Administrator

SUBJECT: IRB Proposal HS16-787
IRB Approval Dates: 9/6/2016 - 9/6/2017
Proposed Project Dates: 8/22/2016 - 8/22/2017
"Stereotype Threat in Mental Rotation Performance"

The Institutional Review Board (IRB) has reviewed your proposal and has given it final approval. To maintain permission from the Federal government to use human subjects in research, certain reporting processes are required.

A. You must include the statement "Approved by IRB: Project # HS16-787" on all research materials you distribute, as well as on any correspondence concerning this project.

B. If a subject suffers an injury during research, or if there is an incident of non-compliance with IRB policies and procedures, you must take immediate action to assist the subject and notify the IRB chair (dereande@nmu.edu) and NMU's IRB administrator (rwinn@nmu.edu) within 48 hours. Additionally, you must complete an Unanticipated Problem or Adverse Event Form for Research Involving Human Subjects

C. Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding. Informed consent must continue throughout the project via a dialogue between the researcher and research participant.

D. If you find that modifications of methods or procedures are necessary, you must submit a Project Modification Form for Research Involving Human Subjects before collecting data.

E. If you complete your project within 12 months from the date of your approval notification, you must submit a Project Completion Form for Research Involving Human Subjects. If you do not complete your project within 12 months from the date of your approval

notification, you must submit a Project Renewal Form for Research Involving Human Subjects. You may apply for a one-year project renewal up to four times.

NOTE: Failure to submit a Project Completion Form or Project Renewal Form within 12 months from the date of your approval notification will result in a suspension of Human Subjects Research privileges for all investigators listed on the application until the form is submitted and approved.

All forms can be found at the NMU Grants and Research website:
http://www.nmu.edu/grantsandresearch/node/102

Appendix 8

IRB Approval Part II

TO: Bridget Parler
Psychology Department

CC: Sheila Burns
Psychology Department

DATE: September 30, 2016

FROM: Rob Winn, Ph.D.
Assistant Provost/IRB Administrator

SUBJECT: IRB Proposal HS16-799
IRB Approval Dates: 9/30/2016 - 9/30/2017
Proposed Project Dates: 10/1/2016 - 10/1/2017
"Stereotype Threat in Mental Rotation Performance Part II"

The Institutional Review Board (IRB) has reviewed your proposal and has given it final approval. To maintain permission from the Federal government to use human subjects in research, certain reporting processes are required.

A. You must include the statement "Approved by IRB: Project # HS16-799" on all research materials you distribute, as well as on any correspondence concerning this project.

B. If a subject suffers an injury during research, or if there is an incident of non-compliance with IRB policies and procedures, you must take immediate action to assist the subject and notify the IRB chair (dereande@nmu.edu) and NMU's IRB administrator (rwinn@nmu.edu) within 48 hours. Additionally, you must complete an Unanticipated Problem or Adverse Event Form for Research Involving Human Subjects

C. Please remember that informed consent is a process beginning with a description of the project and insurance of participant understanding. Informed consent must continue throughout the project via a dialogue between the researcher and research participant.

D. If you find that modifications of methods or procedures are necessary, you must submit a Project Modification Form for Research Involving Human Subjects before collecting data.

E. If you complete your project within 12 months from the date of your approval notification, you must submit a Project Completion Form for Research Involving Human Subjects. If you do not complete your project within 12 months from the date of your approval

notification, you must submit a Project Renewal Form for Research Involving Human Subjects. You may apply for a one-year project renewal up to four times.

NOTE: Failure to submit a Project Completion Form or Project Renewal Form within 12 months from the date of your approval notification will result in a suspension of Human Subjects Research privileges for all investigators listed on the application until the form is submitted and approved.

All forms can be found at the NMU Grants and Research website:
http://www.nmu.edu/grantsandresearch/node/102