ERASMUS UNIVERSITY ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

MSc Thesis Behavioural Economics (Marketing track)

# Stereotype threat in Greek universities

An experimental study on the effect of stereotype threat on math performance of female STEM students in Greece.

Ourania Ntinou
(581389)

Supervisor: Francesco Capozza
Second assessor: Georg Granic
April 4, 2022

# Abstract

Stereotypes play a significant role in society and can affect human behaviour in a number of different ways. When people are afraid of confirming a specific negative stereotype related to them, they experience the so-called “stereotype threat” (Steele and Aronson, 1995). This situational fear can negatively affect the target’s behaviour, and even their (academic and not only) performance. It is believed that stereotype threat accounts for the underrepresentation of women in STEM education and careers. This study aims to explore the effect of stereotype threat on the math performance of female STEM students, which is a sample strongly related to math and math-related sciences. This is done by means of an online experiment, in the form of a math test. The participants (both male and female students) were randomly assigned into a control and a treatment group. Prior to the math test, the treatment group received information regarding the underperformance of women in similar math tests, in the past. This is a cue that based on the literature causes stereotype threat to women about their math ability (e.g. Spencer et al., 1999). In contrast with the theory, women performed better in the stereotype threat condition, compared to the non-threatened female participants. In addition, men’s performance remained constant in both conditions. Finally, although it is believed that anxiety is involved in the relationship between stereotype threat and performance, the current study does not provide evidence for this hypothesis. The findings of the paper, as well as, the limitations and suggestions for future research are discussed.

# Acknowledgments

The completion of this study could not have been possible without the valuable support of my supervisor, Francesco Capozza. His guidance, advice and positive mindset helped and motivated me through all the stages of the thesis. I would also like to acknowledge all my professors of the Behavioural Economics Masters program, of the Erasmus University Rotterdam, for sharing their knowledge and expertise.

In addition, I would like to thank my colleagues, Dimitrios Sipitanos and Stefanos Stamatakis for their constant support since the first day of the program. This experience would not be the same without them.

Finally, I would also like to give special thanks to my family, my friends and my partner for giving me the encouragement I needed throughout the process of my Masters thesis.

# Table of Contents

# 1. Introduction

A stereotype is a fixed, generalised, and simplified belief about the attributes and behaviours of a certain group of people, as it characterises the members of the group, ignoring their individual personal qualities and characteristics. Stereotypes are everywhere. They can be gender (e.g. “Girls like dolls, boys like cars”), racial (e.g. “Asians are good at math, Blacks are good at sports”), cultural (e.g. “Greeks are lazy”), social (e.g. “Poor are less educated compared to rich people”), religious (e.g. “Muslims are aggressive”). As standardised beliefs, stereotypes can be classified based on their source. Stereotypes, as mental representations of real group differences, enable us to process information more efficiently about others and usually lead to realistic perceptions (Hilton and Von Hippel, 1996). However, stereotypes that originate from one’s enduring characteristics (e.g. gender) can lead to major potential errors with negative and unfair consequences to stereotyped people, such as prejudice (Hilton and Von Hippel, 1996).

A phenomenon that stems from being the target of discrimination and stereotypes is called *“stereotype threat”*. Stereotype threat is the experience of being in a situation, where one confronts a negative stereotype associated with one’s group (Spencer et al., 1999). Essentially, the person is “threatened” by confirming the stereotype and the perception that (s)he will be judged based on that. Ironically, stereotype threat does not discriminate, since it is not a unique experience associated with a certain group or stereotype. Taking into consideration the multitude of -negative- stereotypes related to numerous groups of the society, we can argue that stereotype threat is a widely applicable theory. However, one of the most focused areas in the stereotype threat research is academic performance, with their negative relationship gaining great experimental support (e.g. Bell et al., 2003; Harrison et al., 2006; O'Brien and Crandall, 2003).

One of the first papers that stereotype threat is discussed is related to women’s math performance, aiming to explain the gender gap in this domain (Spencer et al., 1999). Since then, a number of studies have found analogous results, i.e. that stereotype threat negatively influences women’s math performance (e.g. O’Brien and Crandall, 2003; Schmader, 2002; Schmader and Johns, 2003). However, stereotype threat can have chronic consequences on women’s decisions regarding their career and educational opportunities. According to Steele (1997), when the threat is a long-term situation, it can cause *disidentification.* More specifically, women, who are identified with a male-oriented math environment or domain, can subtract the domain from their self-identity with a negative impact on their motivation. In addition, recent experimental findings suggest that undergraduate female students, with high math SAT (i.e. Scholastic Aptitude Test) scores, underperformed in a math test in the presence of stereotype threat, compared to equally qualified males (Lesko and Corpus, 2006). Analogous results have been found regarding engineering students' performance in math-engineering exams (Bell et al., 2003). Overall, we conclude that even for women who have achieved identification with their -male-dominated- domain, stereotype threat is still there.

Still, nowadays, STEM (science, technology, engineering, and mathematics) is one of the most male-dominated fields. As of 2021, only eleven women, out of 405 awarded scientists, have won the Nobel Prizes in Physics and Chemistry (The Nobel Prize, n.d.). Although the past decades have been preeminent efforts to mitigate women's underrepresentation in STEM education and careers (e.g. UNESCO, 2017; European Institute for Gender Equality, 2018), gender differences in STEM-related fields still exist. For instance, based on a report from the European Institute about Gender Equality (i.e. EIGE) (2018) regarding gender segregation and equality across the European Union, women's proportion among STEM graduates in the EU was only 22%. Thus, it would be interesting to explore the stereotype threat phenomenon in the European Union, as the gender gap in STEM-related higher education is significant. Among the EU countries, we can focalise even more on Greece specifically, since the country has ranked last in the EU since 2010, on the Gender Equality Index (European Institute for Gender Equality, 2021). Furthermore, despite the considerable amount of literature that focuses on stereotype threat and women math performance, most studies use samples from undergraduate students from numerous fields without strong relation with math or math-related sciences (e.g. Ben-Zeev et al., 2005; Cadinu et al., 2005; Pronin et al., 2004; Spencer et al., 1999).

In relation to all these insides, I decided on the following research question: *Does a stereotype threat activation influence women's math performance in Greek STEM higher education?*

In order to test the research question above, I have conducted an online experiment. The findings of the experiment reveal that stereotype threat, indeed, affects math performance of female STEM students, however in a positive way. In addition, there is no evidence that self-reported anxiety plays a role in this effect. Finally, while women's performance shows to be significantly influenced by stereotype threat, men's performance remains unchangeable.

The current paper is structed as it follows. In Section 2, the theoretical literature of stereotype thereat, as well as, the empirical support of the theory are presented. Then, Section 3 provides a detailed elaboration of the methodology that has been followed, while in Section 4 the results of the analysis are presented. After that, the findings, the limitations of the current study and some suggestions for future research are discussed in Section 5. Finally, Section 6 covers the conclusion of the paper.

# 2. Literature review

## 2.1 Stereotype threat

The “stereotype threat” term was introduced in the academic world in 1995 by Steele and Aronson. During the past 27 years, hundreds of published studies have furtherly investigated this phenomenon, aiming to illuminate more aspects of it. According to the original definition of Steele and Aronson (1995), *“Stereotype threat is being at risk of confirming, as self-characteristic, a negative stereotype about one's group”.* It is a social-psychological phenomenon that emerges when an individual, who considers itself as part of a group, faces a widely famous negative stereotype about this group and (s)he is in “threat” of proving this stereotype.

Steele et al. (2002) identify three general characteristics of stereotype threat. According to the writers, stereotype threat is a situational phenomenon that occurs from signals that someone’s behaviour can be explained by one of his/her social identities (Steele et al, 2002). The existence of multiple social identities is a basic assumption in stereotype threat theory (Schmader and Inzlicht, 2012). In addition, stereotype threat can be experienced by everyone in different circumstances. All people are part of some social groups that attribute a social identity to them (Steele et al., 2002). For instance, I am a female, I am from Greece, and I am a millennial. Considering these characteristics, someone can recall numerous stereotypes about me. For all the social groups, negative stereotypes exist and when these stereotypes get triggered, everyone can feel the threat (Steele et al., 2002). Finally, Steele et al. (2002) emphasise the influence of the content of the related negative stereotype on the nature of the stereotype threat. Namely, one could say that the stereotype is getting activated when its content “matches” the relevant people, behaviour and situation. For instance, it is more likely that Black people can experience stereotype threat during a verbal test, where a negative stereotype about their intellectual ability applies (Steele and Aronson, 1995), rather than during a language test. At this point, it is important to mention that, according to Spencer et al. (2016), someone does not need to accept a negative stereotype about him/her, in order to experience stereotype threat. In fact, only being aware of this stereotype could be enough (Spencer et al., 2016).

Previous research (e.g. Aronson et al., 1999; Croizet and Claire, 1998; Keller, 2002; Steele and Aronson, 1995) has revealed that the fear of proving a negative stereotype, makes the person behave in a stereotype-consistent way. However, it is interesting to mention that the individual who feels the stereotype threat has a motivation to behave in a way opposed to this stereotype (i.e. stereotype avoidance). In one of the first experiments about stereotype threat, Steele and Aronson (1995) found that Black participants that were in the stereotype threat condition showed greater stereotype avoidance than Black participants in the non-threat condition. More specifically, in this research Black individuals faced a negative stereotype regarding the intellectual ability of Black people, with an evident activation of stereotype threat. This group of participants reported that they enjoyed fewer activities, music genres and sports that are stereotypically linked to African Americans (i.e. enjoying sports, rap music and basketball respectively), compared to the non-threatened participants (Steele and Aronson, 1995). A recent study by Pronin et al. (2004) about the stereotype threat and stereotype avoidance relationship provides consistent findings. Women under stereotype threat, with high exposure to math environment, reported fewer stereotypically feminine characteristics, compared to non-threatened women with the same math background.

## 2.2 Stereotype threat and performance

Published research on stereotype threat is characterised by great generalisability. As it is already mentioned, the phenomenon of stereotype threat can be associated with any negative stereotype regarding any social identity that a person possibly possesses in numerous different contexts. For instance, it has been demonstrated that stereotype threat affects loss-aversion and risk-aversion behaviour (Carr and Steele, 2010), social sensitivity (Koenig and Eagly, 2005) and leadership (Hoyt and Murphy, 2016). However, there is a large volume of published studies describing the negative consequences of stereotype threat to performance specifically. This finding has been replicated for multitudinous stereotypes (e.g., related to race, gender, age, socio-economic status, sexual orientation, even to health history) using various and numerous tasks (e.g., golf tasks, math tests, political knowledge tests, memory and attention tests, driving tasks, intellectual tasks, activities with children; e.g., Bosson et al., 2004; Croizet and Claire, 1998; Gonzales et al., 2002; Joanisse et al., 2013; McGlone et al., 2006; Rahhal et al., 2001; Schmader, 2002; Stein et al., 2002; Stone et al., 1999; Suhr and Gunstad, 2002).

### 2.2.1 The automatic process of stereotype threat

Although extensive research has been carried out on the effect of stereotype threat on performance, a clear view of this process is still lacking. Schmader et al. (2008) suggest an interesting approach to the automatic activation of stereotype threat. The stereotype threat process starts with the person being in contact with a situational cue that activates a relevant negative stereotype (Schmader and Beilock, 2011). The activation of the negative stereotype is unconscious, and the person is not aware of it (Steele, 1997; Steele et al., 2002). However, it is critical to mention that the situational cue -the source of the stereotype threat-, does not have to be something obvious or extreme (Spencer et al., 2016). Indeed, evidence suggests that the manner that a test is characterised (e.g. as a predictor of intellectual abilities), or a message about gender differences in previous similar tasks, or even the predominant number of men in a setting, compared to women, can trigger the stereotype threat process with further impact on the target's performance (Steele and Aronson, 1995; Spencer et al., 1999; Sekaquaptewa & Thompson, 2003).

Once the situational cue triggers stereotype threat, the activation of the *propositional relation* (Gawronski & Bodenhausen, 2006) of the following three concepts happens, *"the concept of one's ingroup, the concept of the ability domain in question and the self-concept"*, based on Schmader et al., 2008. For better comprehension of the aforementioned concepts, we can take as a reference the experiment of Steele and Aronson (1995) about the negative effect of stereotype threat on Blacks' performance when describing a verbal test as diagnostic. In this case, the positive condition consists of the three concepts: I am Black, Blacks are less intelligent compared to Whites, but I believe I am equally intelligent. However, during the stereotype threat phenomenon, imbalances between these three concepts occur, creating logical inconsistency among them and disrupting their relations (Schmader et al., 2008). The cognitive imbalance causes uncertainty and self-doubt, by creating a negative link between oneself and the domain (Schmader, 2010; Schmader and Beilock, 2011). Another negative consequence of stereotype threat and the automatic cognitive inconsistency that it activates, is that the individual shows increased vigilance to obtain more clues, which confirm or contradict the stereotype (Schmader and Beilock, 2011). Both outcomes can impair one's performance (Schmader and Beilock, 2011).

### 2.2.2 Mediators[1] of stereotype threat and performance

Apart from the automatic process of stereotype threat, Schmader and Beilock (2011) suggest some additional mediational mechanisms that negatively affect the performance, i.e. effort and working memory. Indeed, Smith (2004) argues that there are two competing theories regarding effort. The first hypothesis refers to poor performance due to stereotype threat, mediating by less effort spent to the task. While, the other one is about putting more effort as a reaction to the negative stereotype, which finally leads to underperformance. Furthermore, working memory has been indicated from several experimental studies as a potential cognitive mediator (e.g. Beilock et al., 2006; Beilock et al., 2007; Johns et al., 2008). The intuition behind this supposition is that stereotype threat causes stress-related thoughts and destructions, while the individual processing the stereotype-activator information. Therefore, these thoughts lessen working memory capacity, as they expend cognitive resources (Beilock et al., 2006; Schmader and Johns, 2003). Furthermore, several other studies have examined a diverse range of affective, cognitive and motivational mechanisms as stereotype threat mediators (for a comprehensive review, see Pennigton et al., 2016). For instance, there is evidence that stereotype threat can negatively affect personal expectations and consequently decreases performance (Cadinu et al., 2003). While Stone (2002) found that individuals use self-handicapping as a protective mechanism to defeat stereotype threat, yet with negative consequences to their performance. Overall, more than 15 different mediators have been researched (Pennigton et al., 2016), however anxiety has been one of the most popular among these mechanisms, with literature providing mixed results about the mediational effect of anxiety on the stereotype threat-performance relationship.

The impact of test anxiety on academic performance has been reported to several studies (e.g. Culler and Holahan, 1980; DordiNejad et al., 2011; Rana and Mahmood, 2010) with women to be more vulnerable to test anxiety compared to men (Chapell et al. 2005). According to Sarason (1984), test anxiety causes worries and preoccupations to the person during the evaluation process with negative consequences to his/her concentration and finally to performance. In addition, test anxiety may affect one's ability to process information with analogous results to performance (Hembree, 1988). Due to the strong relationship with the academic performance, anxiety has been investigated several times as a mediational

[1] According to the APA Dictionary of Psychology, the world "mediator" is defined as "an intermediary or intervening variable that accounts for an observed relation between two other variables".

mechanism between stereotype threat and test performance, using various and different methods and measures. Steele (1997) and Steele et al. (2002) argue that stereotype threat causes high anxiety on the "fear" of confirming the negative stereotype. However, the experimental findings are controversial.

Initially, this relationship was tested by Steele and Aronson (1995). In their experiment about stereotype threat and academic performance of African Americans, self-reported anxiety did not play a significant role on the effect of stereotype threat. Following up these findings, the results of Spencer et al. (1999) could not provide clear evidence that self-reported anxiety is a mediator between stereotype threat and women's math performance. However, according to Spencer et al. (1999), anxiety can be considered as a plausible mediator of stereotype threat effects since the effect of anxiety was partially significant. Some years later, Osborne (2001) found analogous results, with stated anxiety to explain a part of the underperformance of high school students. Further support on the "anxiety" hypothesis is provided by some recent studies, with anxiety linking stereotype threat and performance, either as an independent variable (Lu et al., 2015) or sequentially (Chung et al., 2010; Mrazek et al., 2011). Despite the theoretical background and the empirical support, anxiety was not a predictive factor of the negative effect of stereotype threat to performance in numerous studies (e.g. Aronson et al., 1999; Mayer and Hanges, 2003; Keller and Dauenheimer, 2003). Given that anxiety could be challenging to be detected and reported (Bosson et al., 2004) and even the position of the anxiety-related question(s) plays a role (Seipp, 1991), one can argue that the inconsistent results regarding anxiety as a mediational mechanism may derive from numerous sources.

## 2.3 Empirical Support for Stereotype Threat

### 2.3.1 Stereotype threat and women's math performance

In addition to the previous theoretical and empirical support, several experimental studies have researched the relationship between the phenomenon of stereotype threat and mathematics performance, principally focusing on women. This relationship has been tested in different settings and frameworks. Spencer et al. (1999) investigated the performance of male and female psychology students, with a good mathematics background, on a challenging math exam. The effect of stereotype threat on performance was tested with an experimental design consisting of two conditions. In the "relevance condition", participants were informed that the prior test results had shown a significant gender gap against women. While, in the alternative condition, participants got informed that in the past, there was never a score difference between women and men. The researchers found that when the participants were informed that the test showed gender differences, women significantly underperformed in comparison to equally qualified men. On the contrary, women performed as good as men did, when stereotype threat was not triggered (Spencer et al., 1999).

In an experimental study of stereotype threat and solo status, Sekaquaptewa and Thompson (2003) provided more evidence about stereotype threat and women's math underperformance. More precisely, women's performance was shown significantly lower under stereotype threat than in the control condition, on an oral examination (Sekaquaptewa & Thompson, 2003). It is interesting to mention that a woman's performance under stereotype threat was even lower when she was taking the test surrounded by male participants only, compared to the condition of being in a group composed of females exclusively (Sekaquaptewa & Thompson, 2003). Analogous results have also been found from Cadinu et al. (2005), Inzlicht & Ben-Zeev (2000) and Shih et al. (1999).

Additionally, Spencer et al. (1999) investigated the stereotype threat as a potential performance mitigator for women on both easy and difficult math exercises. After several studies, Spencer et al. (1999) found that stereotype threat influences women's math scores differently, concerning the difficulty of the test. In a randomised control experiment, they found evidence that women underperform, compared to equally qualified men, only on difficult tests, while the performance of the two genders is equal on easy tests. Some years later, O'Brien and Crandall (2003) investigated further the effect of stereotype threat in women's performance in both easy and difficult mathematical tests. In this research, the increased arousal during the maths test is considered as a significant component of stereotype threat. As arousal, they define *"the heightened activity, primarily in the sympathetic nervous system (SNS), that energises behaviour*" (O'Brien and Crandall, 2003). By following the method of Spencer et al. (1999), O'Brien and Crandall (2003) found that threatened women outperformed on an easy maths test and underperformed on a challenging maths test, compared to women under no stereotype threat conditions. The potential explanation of the relation between stereotype threat and task difficulty is that the presence of a gender-biased situation confirms a negative stereotype to women. Consequently, this can cause arousal that this arousal positively impacts performance on easy tasks and negatively on difficult ones. (O'Brien & Crandall, 2003).

Overall, we could say that the previous research findings into the stereotype threat phenomenon have been consistent and provide strong empirical support for the theory. Therefore, taking all the aforementioned into consideration, the first two hypotheses are the following:

**H1:** Providing information about gender differences against women, before a difficult math test leads STEM female students to underperform, compared to STEM male students.

**H2:** Providing information about gender differences against women, before a difficult math test leads STEM female students to underperform, compared to STEM female students who did not get this piece of information.

### 2.3.2 Stereotype threat and men's math performance

Although the relation between the arousal of stereotype threat and women's underperformance seems quite strong according to the existing literature, the effect of this gender-biased information on men's performance is not clear. In the aforementioned studies, the performance outcome of men differed significantly. In the series of experiments that they conducted, Spencer et al. (1999) found that the men's performance was higher in the "gender gap" condition. Reversely, men's score was slightly worse, when the math test was described as free from gender bias. However, it is important to mention that the differences in men's performance in the experiment of Spencer et al. (1999) were not significant. Additionally, the experiment conducted by Sekaquaptewa and Thompson (2003), showed that the performance of men remained overall the same, regardless the stereotype threat cue. We see analogous results in the experiment conducted by O'Brien and Crandall (2003), with men to perform equally good to a difficult math test, across the threat and non-threat conditions. Given these insights, I arrive to the following hypothesis to test:

**H3:** Providing information about gender differences against women before a difficult math test does not affect STEM male students' performance.

In addition to the hypotheses, I will investigate the self-reported anxiety as a potential mediational mechanism on the relationship of stereotype threat and academic performance, conducting an online experiment. In the following chapter, a detailed elaboration of the methodology is discussed.

# 3. Methodology

For this study, I tested if a message regarding gender differences before a math test would impact the participants' performance. In order to test my hypothesis and further investigate anxiety as a mediator, I conducted an online randomized controlled experiment, with participants being both male and female undergraduate students at Greek STEM universities.

## 3.1 Experimental Design

The online randomized controlled experiment took the form of a 2 X 2 (gender X gender gap message) design. The choice of a randomized controlled experiment allows control over the selection bias and draw causal inferences. The purpose of conducting the experiment online was the access to a wider sample of students from multiple universities, and of course the safety of the participants due to the Covid pandemic. The participants got introduced to the survey, using the Qualtrics software. The survey was in the Greek language for more convenience for the participants. The main part of the survey was consisted of a math test, while the other parts, before and after, were demographic and personal questions. For a comprehensive presentation of the survey, see Appendix I. The performance of each participant, in terms of correct answers in the test, was the tested (dependent) variable. As the main independent variables, I considered the interaction between the gender and the gender gap message (treatment).

At the beginning of the survey, there was a brief message regarding some general information regarding the study. As I was not able to reveal the real purpose of the study, the message informed the participants that the current survey is part of my master's thesis and that aims to measure the mathematics performance of students at Greek Universities. Then, the parts of the survey got shortly presented, as well as who is considered as the target group, the monetary incentive-prize and the reward mechanism. The message closed by highlighting the anonymity and confidentially of all the answers and thanking the participants for their time.

In total, the survey consisted of 18 questions, which are identical for both the control and the treatment group. Initially, all the respondents were asked about socio-demographic information, such as gender, age, nationality and place of being raised. Gender is a very important variable in this experiment, as I want to test the effect of the treatment on performance considering the gender of the participants. In order to secure that the participants will provide their gender, I forced the response to this specific question. In addition, age and nationality worked as typical controls, with the "age" question having an answer restriction in terms of content, meaning that the answer should be a digital number for more convenience in the analysis later on. Finally, the question regarding the place the participants are coming from, followed.

After the socio-demographic questions, the questions regarding education follow. The first one is about whether each participant is a currently enrolled undergraduate student or recent graduate of public Polytechnic Universities (in Greek: "Πολυτεχνεία") and public Higher Education Universities (in Greek: "Ανώτατα Εκπαιδευτικά Ιδρύματα") related to mathematics, physics and computer science. If any of the participants did not belong to this target group, (s)he was not able to proceed further with the survey. The next question was about the specific university that the participants are currently enrolled on or recent graduates from. This question was adapted to the response of the previous question. For instance, if a respondent indicated that (s)he is a student in a mathematics/applied mathematics university, the following question was about Greek universities with departments of this specific discipline. After that, the participants had to mention their GPA, which will be considered as a control variable in the analysis of their performance. Furthermore, the participants were asked to indicate at which level they agreed with the following statement *"I am confident about my level in math"*, as the level self-confidence has been shown to affect performance (e.g. Stankov et al., 2014; Tavani and Losh, 2003).

Following these questions, the participants got one of the control or treatment messages. The two messages were identical, apart from two additional sentences in the treatment message. In the control message, each participant got thanked for his/her answers so far and informed that (s)he is suitable to further proceed with the math test, which contained seven multiple-choice mathematics questions from the GRE Mathematics Test Practice Book (ETS, 2017). Furthermore, there was a reminder of the prize and the reward mechanism. In addition to this message, the treatment message contained the following text: "In similar previous tests, based on the latest GRE Worldwide Report, men had higher performance in quantitative/mathematics questions compared to women. This is the case for all the Undergraduate Major Fields including Engineering and Physical Sciences as well." The design of the additional message intended for the treatment group was mostly based on Spencer et al. (1999). In the paper of Spencer et al. (1999), the treatment group received a message indicating that previous research has sometimes shown gender differences in math performance. This method of stereotype threat activation has been used by other researchers as well (e.g. Laurin 2013, Cadinu et al., 2003). In addition, Pavlova et al. (2014) found that females are inclined to be stronger negatively affected by implicit negative messages, compared to explicit ones. Considering these insights, I designed the treatment message highlighting the gender gap that exists in mathematics performance, however, I kept it more specific, compared to the message of Spencer et al. (1999), by mentioning explicit past results from the GRE test, to make this sentence as disputable as possible. Indeed, according to the ETS organization (2022), which provides the GRE test, in the period July 2020-June 2021, women test-takers in the general GRE test had a performance n mean of 153.2 on the Quantitative Reasoning part, compared to male participants with an average score on the same part of the test of 158.7. Although in the fields of Engineering and Physical Sciences the means of both genders were higher, the gap of the 2-3 points on the average score between women and men still existed. In addition, the treatment message deliberately indicated that "men's performance is better compared to women's", rather than "women's performance is worse compared to men's", as according to Pavlova et al. (2014) highlighting past men's outperformance has a greater negative impact to women's performance, compared to highlighting past women's underperformance. After the control and treatment messages, the math test followed.

Participants had to solve seven mathematics questions, taken from the GRE Mathematics Test Practice Book (ETS, 2017). Exercises from the GRE test have been used in stereotype threat experiments several times (e.g. Aronson et al., 1999; Brodish and Devine, 2009; Cadinu et al., 2005; Jamieson and Harkins, 2012). I preferred to choose questions that correspond to the GRE Mathematics Test, compared to the general GRE test, as the math test should be difficult considering the strong mathematics background of the sample. However, from the 66 solved problems that the book provides, I picked a list with the 20 easiest questions, considering the percentages of test-takers answering each question correctly. Then, from the remaining 20 questions, I furtherly picked seven, considering including questions that the majority of the participants will be able to answer, and questions that do not need extensive calculations or memorization of special formulas (e.g. trigonometry formulas). The intuition behind this process is to keep the test difficult, however not too much that people would drop out from the experiment. Thus, the seven final questions appeared to the participants. In each question, the participants had five alternative answers to choose from and the option to indicate that they do not know the answer. In every question, the provided answers were identical to the alternatives that the GRE Mathematics Test Practice Book (ETS, 2017) provides respectively. Each question had only one correct answer. In addition, each participant could see every question individually on his/her screen, going back and forth to the questions and changing his/her answer unlimited times.

After the end of the math test, the final three questions followed. The first question after the test was about how difficult the participants perceived the test, considering their mathematics knowledge. This question aimed to check that, overall, the difficulty of the test was not extremely high or low for the participants. Then, there was a question about the feelings that the participants had during the test, focusing on anxiety-related statements, such as “tense”, “under pressure”, “nervous/jittery”, “uneasy”, “afraid of not doing well” and “uncomfortable”. This approach was initially followed by Osborne (2001) to measure self-reported anxiety after a similar experiment on stereotype threat, providing statistically significant results of anxiety as a mediator. After the anxiety-related statements, a question on who is believed to get more correct answers on the math test appeared. Finally, participants were asked to optionally give their email addresses to participate in the lottery.

## 3.2 Sample

The respondents were randomly equally assigned to the control and the treatment group, using the Qualtrics randomizer. In each group, there were both male and female respondents. The experiment had a between-subject design. Thus, each participant was a member of only one group. The optimal sample size, to minimize the Type I and II errors, was calculated by conducting a priori power analysis with the G*Power 3 statistical tool (Faul et al., 2007). According to a meta-analysis on stereotype threat existing literature, the average effect size of stereotype threat on women’s mathematics performance is estimated at 0.22 (Flore and Wicherts, 2015). Thus, with an effect size of d=0.22 and considering an α (Type I) error probability equal to 0.05 and a β (Type II) error probability of 0.2, a sample of 538 participants is essential to reach a power of 80%. This optimal sample size refers to the Wilcoxon-Mann-

Whitney means test of two independent equal groups. Unfortunately, considering the limited time this has not been reached, with the final sample to reach the 396 participants.

A random sample of 748 participants was recruited from February 5th until February 25th, 2022. Eligibility criteria required individuals to be enrolled students or recent graduates of Greek Universities of one of the following STEM disciplines: Mathematics and Applied Mathematics, Physics and Applied Physics, Computer Science, Civil Engineering, Mechanical Engineering, Electronic Engineering, Chemical Engineering, Environmental Engineering, Architecture and other Polytechnic sciences. The survey got distributed through social media platforms and principally through Facebook university students' groups. Prior entering to the survey, all the potential participants were informed about the goal and the content of the survey, the eligibility criterion, the monetary incentive and the reward mechanism. As already mentioned, the message at the beginning of the survey was repeating this information and additionally indicating the anonymity and confidentiality of the responses (see Appendix I).

From the 748 recorded responses, only 396 were valid. The 325 people that did not complete the survey got excluded from the analysis, with 133 participants being from the control group, 147 belonging to the treatment group and the remaining 45 were not categorized in any group as they dropped from the survey before the randomization took place. In addition, the 21 participants that did not meet the eligibility criterion of being a student/graduate of a Greek STEM-related university were dropped from the sample, as well as 6 respondents, who did not want to indicate their gender or mentioned a non-binary gender. As one of the main goals of this study is to compare the stereotype threat effect on the performance of men and women, it is essential for the analysis to keep only these two genders. Overall, the final sample consists of completed responses, however, there are individual missing values for some participants. Of the 396 respondents, 61.11% is male. The average age of the sample is 23.2 (*SD*=3.70) with a minimum of 18 and a maximum of 47. In addition, the majority of the participants are Greek (96.97%) and the most popular fields across the sample are the following, Mechanical Engineering (32.83%), Physics (25.25%) and Civil Engineering (24.49%). Further information on demographic data can be found in Table 3.1.

Table 3.1*: Descriptive statistics and randomization*

| Variable | Control group (n=209) | Treatment group (n=187) | Total sample (N=396) | | Chi-squared (p-value) |
|---|---|---|---|---|---|
| **Gender** | | | | | |
| Male | 127 | 115 | 242 | 61.11% | 0.0222 (0.881) |
| Female | 82 | 72 | 154 | 38.89% | |
| **Age Mean (SD)** | 23.34 (3.61) | 23.15 (3.80) | 23.25 (3.70) | | 22.7077 (0.250) |
| **Nationality** | | | | | |
| Greek | 203 | 181 | 384 | 96.97% | |
| Albanian | 1 | 1 | 2 | 0.51% | 2.5461 (0.467) |
| Cypriot | 3 | 5 | 8 | 2.02% | |
| Other | 2 | 0 | 2 | 0.51% | |
| **City** | | | | | |
| Athens | 20 | 18 | 38 | 9.90% | |
| Thessaloniki | 78 | 80 | 158 | 41.15% | |
| Patra | 3 | 0 | 3 | 0.78% | 7.4559 (0.383) |
| Herakleion | 6 | 5 | 11 | 2.86% | |
| Larisa | 6 | 4 | 10 | 2.60% | |
| Volos | 1 | 5 | 6 | 1.56% | |
| Other city | 83 | 65 | 148 | 38.54% | |
| Other country | 5 | 5 | 10 | 2.60% | |
| **Field** | | | | | |
| Mathematics/Applied Mathematics | 7 | 2 | 9 | 2.27% | |
| Physics/Applied Physics | 51 | 49 | 100 | 25.25% | |
| Computer Science | 10 | 8 | 18 | 4.55% | 12.6704* (0.081) |
| Civil Engineering | 48 | 49 | 97 | 24.49% | |
| Electronic Engineering | 28 | 10 | 38 | 9.60% | |
| Chemical Engineering | 2 | 0 | 2 | 0.51% | |
| Mechanical Engineering | 62 | 68 | 130 | 32.83% | |
| Architecture | 1 | 1 | 2 | 0.51% | |
| **University** | | | | | |
| Aristotle University of Thessaloniki | 166 | 154 | 320 | 81.01% | |
| University of Ioannina | 32 | 30 | 62 | 15.70% | |
| University of Patra | 3 | 2 | 5 | 1.27% | 4.4105 (0.492) |
| University of Crete | 1 | 0 | 1 | 0.25% | |
| University of Thessaly | 4 | 1 | 5 | 1.27% | |
| University of Thrace | 2 | 0 | 2 | 0.51% | |
| **Grade** | | | | | |
| [5, 6.49] | 44 | 35 | 79 | 20.20% | 0.4763 (0.788) |
| [6.5, 8.49] | 139 | 129 | 268 | 68.54% | |
| [8.5, 10] | 22 | 22 | 44 | 11.25% | |

| | | | | | |
|---|---|---|---|---|---|
| **Confidence** | | | | | |
| Strongly disagree | 3 | 4 | 7 | 1.78% | |
| Somewhat disagree | 25 | 18 | 43 | 10.91% | 4.6639 (0.324) |
| Neither agree nor disagree | 53 | 52 | 105 | 26.65% | |
| Somewhat agree | 103 | 79 | 182 | 46.19% | |
| Strongly agree | 24 | 33 | 57 | 14.47% | |

*Notes.* For the categorical variables, the absolute numbers indicate the frequency, while the percentages refer to the proportion of each category in the total sample. For the interval variable *Age*, the absolute numbers refer to the Mean, while the numbers in the parentheses represent the Standard Deviation for the control group, treatment group and total sample, respectively. For all the variables, in the "Chi-squared" column, the numbers in the parentheses represent the p-value of the Chi-squared test for each variable. ***p.<0.01, **p<0.05, *p<0.1

### 3.2.1 Randomization check

The randomization of the participants and their assignment into the control and the treatment group took place by the Qualtrics software. In order to test if the randomization was conducted correctly the Chi-squared test has been used. The Chi-squared tool is a non-parametric test of independence, meaning that it compares the distribution of a categorical variable between two independent samples (Kim, 2017). The null hypothesis of the Chi-squared test is the following, $H_0$: Independent distributions across the samples (no association). As Table 3.1 shows, there is no evidence of association for any of the tested variables between the control and the treatment group, at the 5% significance level. For instance, *Gender*, which is the main variable of interest, on the Chi-squared test, showed a p-value of 0.881, indicating no statistical significance for $\alpha=0.05$. Considering that the p-values of all the control variables are statistically insignificant, I can conclude that there is no evidence for incorrect randomization of the participants into the control and the treatment group, at the 5% significance level. Since the expected frequencies of some cells are <5, or even <1, it is possible that the results of the Chi-squared test are not reliable. Thus, I additionally conducted a Fisher's exact test to ensure the relevance of the results (Kim, 2017). Again, there was no evidence of non-accurate randomization (Appendix II).

## 3.3 Materials

Participation in the online experiment was feasible with the use of a PC or a smartphone with access to the Internet. Going through the whole survey took approximately 35 minutes, while the time that the participants spent on each mathematics question was on average 1.3 minutes. The software recorded their answers, and the time spent on each of them as well. In order to compensate people's participation, I added a monetary incentive in the experiment, as a show-up fee. The respondents who completed the whole survey were able to participate in a lottery with a €30 voucher prize from a famous online department store. As already mentioned, the participants who wanted to join in the lottery had to give their e-mail addresses, to contact the winner later on. As mentioning one's email address reduces respondents' anonymity, since usually in the email addresses people mention their name, the participation in the lottery was optional. Furthermore, I explicitly mentioned that the collection of the email addresses is only for the lottery purpose and that they will be confidentiality treated and will be deleted after the end of the lottery. All the respondents had equal probabilities to win the prize, considering their participation. However, in order to give them an extra incentive to focus on their assigned task (i.e. the mathematics test) and think harder, I added in the experiment a reward mechanism. The binary lottery incentive, as it is called, rewards the participant for every correct answer or good decision with an additional probability to win a fixed amount. Considering this, for every correct answer in the mathematics test, participants gained a greater probability to win the voucher. The winner of the lottery got randomly selected and contacted to the email address that (s)he has already provided during the survey.

### 3.3.1 Variables Description

A comprehensive description of the variables that have been used in the analysis follows. For a summary of the variables, see Table 3.2.

*Gender*

After excluding from the sample the individuals whose gender was not reported as male or female, *Gender* is a binary variable, which equals 1 when the participant is a female and 0 when the participant is a male.

*Age*

The variable *Age* is interval, taking values from 18 to 40. It is measured in years.

*Nationality*

The nominal categorical variable *Nationality* takes value 1 if the participant is Greek, 2 if the participant is Albanian, 3 if the participant is Cypriot and 4 if the participant is anything else.

*City*

The variable named *City* is nominal categorical and refers to the place that the participant comes from. The variable takes the value 1 when the participant is coming from Athens, the value 2 when (s)he is coming from Thessaloniki, the value 3 when (s)he is coming from Patra, the value 4 when (s)he is coming from Heraklion, the value 5 when (s)he is coming from Larisa, the value 6 when (s)he is coming from Volos, the value 7 when (s)he is coming from another place of Greece and finally the value 8 when (s)he is coming from another country.

*Field*

The variable *Field* is nominal categorical and represents the science that each participant mentioned that is related to. The variable equals 1 when a participant has mentioned "Mathematics/Applied Mathematics" as his/her science of study, 2 when the response is "Physics/Applied Physics", 3 when it is "Computer Science", 4 when it is "Civil Engineering", 5 when it is "Electronic Engineering", 6 when it is "Chemical Engineering", 7 when it is "Mechanical Engineering" and finally 8 when it is "Architecture".

*University*

The nominal categorical variable *University* takes the value 1 when a respondent is an enrolled student/recent graduate of the Aristotle University of Thessaloniki. The value 2 represents the University of Ioannina, the value 3 the University of Patra, the value 4 the University of Crete, the value 5 the University of Thessaly and finally the value 6 the University of Thrace.

*Grade*

The ordinal categorical variable *Grade* refers to the current (in case of a student) or the final (in case of a graduate) GPA of each respondent. It consists of three categories. Thus, it takes the value 1 when the participant has mentioned a GPA equal to [5, 6.49], the value 2 when a GPA is in the [6.5, 8.49] interval and finally the value 3 when a GPA is somewhere between 8.5 and 10. These categories are aligned with the Academic grading system in Greece for public universities and polytechnics.

*Confidence*

The variable *Confidence* is ordinal categorical and indicates to which degree a participant agreed with the following statement *"I am confident about my level in math"*. This statement was measured on a 5-point Likert scale from "Strongly disagree" to "Strongly agree".

*Score*

The variable *Score* is interval and indicates the correct answers that each participant gave in the questions of the math test. Therefore, the variable takes integral values from 0 (when a participant has no correct answers in the math test) to 7 (when a participant has completed the test with full success).

*Difficulty*

The ordinal variable *Difficulty* refers to how much difficult the respondent has perceived the math test and was measured on a 5-point Likert scale, from "Extremely easy" to "Extremely difficult".

*Anxiety Score*

The variable *Anxiety Score* is interval and takes integral values from 0 to 24. This variable has been composed using a method similar to Osborne's (2001) to indicate the stated anxiety of each participant. As already mentioned, the respondents had to report to which degree they had six anxiety-related feelings during the math test, using a 5-points Likert Scale. Although usually the points of this type of response scale, take quantitative values from 1 to 5, for analysis purposes I recoded the most negative alternative (i.e. "Definitely not") to 0 and the most positive answer (i.e. "Definitely yes") to 4. The Cronbach's Alpha of 0.87 was estimated, indicating that the reliability of self-reported anxiety measurement is satisfactory. Then, for each respondent, I created his/her personal "anxiety score" by summing up the recoded quantitative value for every anxiety-related feeling and standardized it. An increasing score indicates increasing self-reported anxiety right after the test. For instance, someone whose response was "Definitely not" to each of the six anxious-related feelings, his/her anxiety score was 0, while someone who reported three "Definitely not" responses and three "Probably not" had a score of 3.

*Check*

The variable *Check* is a nominal categorical variable that takes the value 1 when the answer is "Men", the value 2 when the answer is "Women" and the value 3 when the answer is "Both will have the same performance on average".

*Threat*

The variable *Threat* is binary and indicates the absence/presence of the treatment (i.e. the gender-gap message). Thus, it takes the value 0 when the respondent is part of the control group and 1 when the respondent belongs to the treatment group.

Table 3.2: *Summary of the variables used in the analysis.*

| Variable | Definition of variable | Measurement level/Values |
|---|---|---|
| Gender | What is your gender? | Binary: Female (1), Male (0) |
| Age | What is your age? | Interval: 18-40 |
| Nationality | What is your nationality? | Nominal: Greek (1), Albanian (2), Cypriot (3), Other (4) |
| City | From which of the following places are you coming? | Nominal: Athens (1), Thessaloniki (2), Patra (3), Herakleion (4), Larisa (5), Volos (6), Other city (7), Other country (8) |
| Field | Are you a Greek university student or recent graduate of one of the following (or similar) disciplines? | Nominal: Mathematics/Applied Mathematics (1), Physics/Applied Physics (2), Computer Science (3), Civil Engineering (4) Electronic Engineering (5), Chemical Engineering (6), Mechanical Engineering (7), Architecture (8) |
| University | Which is the Greek University that you are a student/recent graduate of? | Nominal: Aristotle Univerisity of Thessaloniki (1), University of Ioannina (2), University of Patra (3), University of Crete (4), University of Thessaly (5), University of Thrace (6) |
| Grade | What is your current total grade? | Ordinal: [5, 6.49] (1), [6.5, 8.49] (2), [8.5, 10] (3) |
| Confidence | To what degree, do you agree with the following statement? *"I am confident about my level in math."* | Ordinal (5-points Likert scale): Strongly disagree (1) - Strongly agree (5) |
| Score | The sum of the correct answers in the math test | Interval: 0-7 |
| Difficulty | Based on your mathematics knowledge you would characterize this test as: | Ordinal (5-points Likert scale): Extremely easy (1) - Extremely difficult (5) |
| Anxiety score | The sum of the 5-points Likert scale scores on the anxiety-related statements | Interval: 0-24 |
| Check | Which gender do you believe will get a higher average number of correct answers? | Nominal: Men (1), Women (2), Both (3) |
| Threat | Presence of the treatment | Binary: Yes (1), No (0) |

## 3.4 Analysis

In order to test my Hypotheses, I conducted a statistical analysis, using the statistical software Stata. To capture the effect of stereotype threat (i.e. treatment), with respect to the gender of the participants, a linear regression model has been used. The interaction term between *Gender* and *Threat* (i.e. *Gender*X*Threat*) indicates the effect of the combination of these two variables on one's score. In addition to the interaction term, I considered the unique variables and some further control variables as mentioned below.

(1) $Score = \alpha_0 + \alpha_1 Threat + \alpha_2 Gender + \alpha_3 Gender*Threat + \alpha_4 Nationality + \alpha_5 City + \alpha_6 Field + \alpha_7 Grade + \alpha_8 University + \alpha_9 Age + \alpha_{10} Confidence + \varepsilon_i$

To capture the effect of stereotype threat specifically for women, the following analysis has been conducted. Initially, controlling only for women participants (i.e. *Gender*=1), a Mann-Whitney U (or Wilcoxon-Mann-Whitney test or Wilcoxon rank-sum test) was performed. The Mann-Whitney U test is a non-parametric technique that aims to explore if two independent samples (in this case the treatment and the control group) come from the same population. The samples are compared to each other and not to a theoretical value. The null hypothesis of the Mann-Whitney U is the following, $H_0$: $\theta_x=\theta_y$, where $\theta_x$ and $\theta_y$ are the distributions of the independent samples x and y respectively. Furthermore, three assumptions need to hold in order to perform a Mann-Whitney test, which provides reliable results (Nachar, 2008).

Firstly, the dependent variable should be at least on an ordinal scale. This assumption holds as *Score* is a continuous variable. Then, the two samples should be randomly drawn, which is something that one's can assume considering that the participants were randomly selected and allocated in the control and the treatment group. Finally, each observation should be independent of the others, at every level of independence, namely subject, group and session level, and each observation should correspond to one person. Again, one can argue that this assumption holds as the experiment had a between-subject design and the respondents participated in the experiment individually using their personal computer or smartphone. In addition, it was restricted to participate more than once in the experiment. Considering that the aforementioned assumptions hold, I conducted a Mann-Whitney U test on *Score* for only the female participants of both control and treatment groups.

Although, the Mann-Whitney U test implies if any differences between the groups exist, it does not reveal the actual size of this difference. Thus, after the test, I conducted a linear regression model analysis to furtherly investigate the effect of the treatment (i.e. *Threat*) on women's math performance (i.e. *Score*). The linear regression is the following.

(2) $Score = \beta_0 + \beta_1 Threat + \beta_2 Nationality + \beta_3 City + \beta_4 Field + \beta_5 Grade + \beta_6 University + \beta_7 Age + \beta_8 Confidence + u_i$

Here, the interaction term *Gender*X*Threat*, as well as the *Gender* variable are not needed as I controlled only for women (*Gender*=1). I repeated the same process controlling only for men (*Gender*=0) for both the Mann-Whitney U test and a linear regression, using the following model.

(3) $Score = \gamma_0 + \gamma_1 Threat + \gamma_2 Nationality + \gamma_3 City + \gamma_4 Field + \gamma_5 Grade + \gamma_6 University + \gamma_7 Age + \gamma_8 Confidence + \kappa_i$

Finally, in order to furtherly explore the role of self-reported anxiety to the relationship between stereotype threat and math performance, the following linear regression model was considered.

(4) $Anxiety\ Score = \delta_0 + \delta_1 Threat + \delta_2 Gender + \delta_3 Gender*Threat + \delta_4 Nationality + \delta_5 City + \delta_6 Field + \delta_7 Grade + \delta_8 University + \delta_9 Age + \delta_{10} Confidence + \lambda_i$

# 4. Results

## 4.1 Hypotheses testing

The first hypothesis supports that providing information regarding gender differences against women (a cue that activates stereotype threat based on the past literature) negatively affects female participants' math performance compared to males. The simultaneous effect of gender and stereotype threat message to one's math performance can be captured, as already mentioned in the analysis section, with the interaction term *Gender*X*Threat*. Thus, the results of the linear regression model (1) are illustrated in Table 4.1.

Table 4.1: *Output of linear regression analysis of the treatment on the math score*

| Variable | Coefficient | Robust SE | t-value | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Gender | | | | | | |
| Female | -0.715** | 0.288 | -2.48 | 0.014 | -1.283 | -0.148 |
| Threat | | | | | | |
| Yes | -0.453* | 0.260 | -1.74 | 0.083 | -0.965 | 0.058 |
| Gender#Threat | | | | | | |
| Female#Yes | 0.974** | 0.385 | 2.53 | 0.012 | 0.216 | 1.732 |
| Nationality | | | | | | |
| Albanian | -0.867 | 0.783 | -1.11 | 0.269 | -2.408 | 0.673 |
| Cypriot | 0.944 | 1.040 | 0.91 | 0.365 | -1.103 | 2.991 |
| Other | -2.995*** | 0.478 | -6.26 | 0.000 | -3.937 | -2.054 |
| City | | | | | | |
| Thessaloniki | -0.363 | 0.345 | -1.05 | 0.293 | -1.044 | 0.316 |
| Patra | -0.533 | 0.714 | -0.75 | 0.456 | -1.939 | 0.872 |
| Herakleion | -0.454 | 0.491 | -0.92 | 0.356 | -1.422 | 0.513 |
| Larisa | -1.538** | 0.562 | -2.74 | 0.007 | -2.645 | -0.432 |
| Volos | -0.657 | 0.967 | -0.68 | 0.498 | -2.561 | 1.246 |
| Other city | -0.284 | 0.343 | -0.83 | 0.407 | -0.960 | 0.390 |
| Other country | -1.411** | 0.659 | -2.14 | 0.033 | -2.709 | -0.113 |
| Field | | | | | | |
| Physics/Applied Physics | 0.077 | 0.638 | 0.12 | 0.903 | -1.178 | 1.334 |
| Computer Science | -0.747 | 0.775 | -0.96 | 0.336 | -2.272 | 0.777 |
| Civil Engineering | -1.208* | 0.667 | -1.81 | 0.071 | -2.521 | 0.104 |
| Electronic Engineering | 1.205 | 0.733 | 1.64 | 0.101 | -0.237 | 2.648 |
| Chemical Engineering | 0.553 | 1.681 | 0.33 | 0.742 | -2.755 | 3.862 |
| Mechanical Engineering | 0.317 | 0.676 | 0.47 | 0.639 | -1.012 | 1.647 |
| Architecture | -1.697** | 0.726 | -2.34 | 0.020 | -3.126 | -0.268 |

| | | | | | | |
|---|---|---|---|---|---|---|
| University | | | | | | |
| University of Ioannina | -0.533 | 0.422 | -1.26 | 0.208 | -1.365 | 0.298 |
| University of Patra | -1.123* | 0.678 | -1.66 | 0.099 | -2.458 | 0.211 |
| University of Crete | 2.458*** | 0.552 | 4.45 | 0.000 | 1.372 | 3.545 |
| University of Thessaly | -1.479*** | 0.413 | -3.58 | 0.000 | -2.293 | -0.666 |
| University of Thrace | -1.239** | 0.418 | -2.96 | 0.003 | -2.061 | -0.416 |
| Grade | | | | | | |
| [6.5,8.49] | 0.509* | 0.304 | 1.67 | 0.096 | -0.090 | 1.108 |
| [8.5,10] | 1.504*** | 0.377 | 3.98 | 0.000 | 0.760 | 2.247 |
| Age | -0.005 | 0.021 | -0.25 | 0.805 | -0.047 | 0.036 |
| Confidence | | | | | | |
| Somewhat disagree | -0.062 | 0.388 | -0.16 | 0.873 | -0.826 | 0.701 |
| Neither agree nor disagree | 0.098 | 0.319 | 0.31 | 0.757 | -0.530 | 0.728 |
| Somewhat agree | 0.111 | 0.317 | 0.35 | 0.727 | -0.514 | 0.736 |
| Strongly agree | 0.351 | 0.378 | 0.93 | 0.355 | -0.394 | 1.096 |
| Constant | 3.403*** | 0.996 | 3.42 | 0.001 | 1.443 | 5.363 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

As Table 4.1 shows, on average, being a female, compared to being a male, decreases the math test performance by 0.715 points on average, ceteris paribus. This effect is statistically significant at the 5% significant level. Although the treatment has a non-significant effect on the average math score (p-value=0.083>0.05), at the 5% significance level, ceteris paribus, one can argue that this completely logical as a different effect of the treatment is expected based on the gender. Furthermore, by combining the treatment with gender with the interaction term between *Gender* and *Threat,* we have a statistically significant effect of 0.974 with a p-value of 0.012, at the same significance level. This effect can be interpreted as follows. On average, being in the treatment group (Threat=1) and a female (Female=1), increases the math score by 0.974 points, ceteris paribus, compared to being a threatened male (Threat=1, Female=0). Thus, although women generally have a lower average score compared to men, the presence of the treatment actually drive them to outperform. These results are contradictory to the first hypothesis (H1) and they do not provide any evidence to support this. In addition to these findings, all the control variables showed a partly significant effect on the average math performance, at the 5% significance level, apart from *Age* and *Confidence,* which are non-significant. Moreover, the constant of the regression model has a statistically significant effect of 3.4 points, ceteris paribus, at a significance level of 5%. This coefficient represents the average score a participant, who belongs to the base categories.

The second hypothesis refers to the comparison of the average math score of threatened and non-threatened women. To detect whether there is a significant difference between the female participants of the control and the treatment group, a Mann-Whitney U test was conducted on the 154 female participants of the sample. The table below (Table 4.2) illustrates the results of the test. Since the *p-value* (0.1752) is smaller than 0.05, I fail to reject the null hypothesis of the test, meaning that there is no statistically significant difference between the two groups, at a significance level of 5%, ceteris paribus. Based on these results, there is no evidence that "stereotype threat" treatment has a significant effect on women's math performance.

Table 4.2: *Mann-Whitney U tests results comparing Control and Treatment group*

| Tested variable | Tested groups | P-value | Z |
|---|---|---|---|
| Score of female participants(n=154) | Control(n=82) & Treatment group(n=72) | 0.1752 | -1.356 |
| Score of male participants (n=242) | Control(n=127) & Treatment group(n=115) | 0.1690 | 1.375 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

However, in order to furtherly investigate any potential relationship between the variables, I conducted a regression analysis on the math score of only the female participants. The results of the linear regression (Table 4.3) reveal that there is a significant effect of stereotype threat to women's math score. Namely, being a female student under the stereotype threat condition increases the math performance by 0.65 points, ceteris paribus, compared to being a female student under a non-threat condition, at a significance level of 5%. Hereby, this result suggests that there is no evidence on the second hypothesis (H2). Interestingly and contrastingly to the H2, the treatment has a positive effect on the female students' math performance. In addition, the control variables *City*, *Field* and *Grade* are partly statistically significant at the 5% significance level, while there is a total statistically significant effect of *Nationality* on the math score of women participants.

Table 4.3: *Output of linear regression analysis of the treatment on the math score, only for females*

| Variable | Coefficient | Robust SE | t-value | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Threat | | | | | | |
| Yes | 0.655** | 0.310 | 2.11 | 0.037 | 0.040 | 1.271 |
| Nationality | | | | | | |
| Cypriot | -1.155** | 0.557 | -2.07 | 0.040 | -2.259 | -0.051 |
| City | | | | | | |
| Thessaloniki | -0.706 | 0.474 | -1.49 | 0.139 | -1.646 | 0.233 |
| Patra | -1.078 | 1.014 | -1.06 | 0.290 | -3.088 | 0.930 |
| Herakleion | -1.231** | 0.585 | -2.10 | 0.038 | -2.390 | -0.072 |
| Larisa | -1.283* | 0.741 | -1.73 | 0.086 | -2.752 | 0.184 |
| Other city | -0.977** | 0.466 | -2.09 | 0.039 | -1.902 | -0.052 |
| Other country | 0 (omitted) | | | | | |
| Field | | | | | | |
| Physics/Applied Physics | 0.022 | 0.717 | 0.03 | 0.975 | -1.399 | 1.443 |
| Computer Science | 0.245 | 0.780 | 0.31 | 0.754 | -1.301 | 1.792 |
| Civil Engineering | -1.466* | 0.816 | -1.80 | 0.075 | -3.083 | 0.150 |
| Electronic Engineering | 1.518 | 1.079 | 1.41 | 0.162 | -0.619 | 3.656 |
| Chemical Engineering | 3.213** | 1.059 | 3.03 | 0.003 | 1.114 | 5.312 |
| Mechanical Engineering | 0.146 | 0.815 | 0.18 | 0.857 | -1.468 | 1.762 |
| University | | | | | | |
| University of Ioannina | -0.306 | 0.611 | -0.50 | 0.617 | -1.518 | 0.904 |
| University of Patra | -0.703 | 0.987 | -0.71 | 0.478 | -2.659 | 1.252 |
| University of Thrace | -1.391*** | 0.386 | -3.60 | 0.000 | -2.158 | -0.625 |
| Grade | | | | | | |
| [6.5, 8.49] | 0.884* | 0.504 | 1.75 | 0.082 | -0.114 | 1.883 |
| [8.5, 10] | 2.585** | 0.922 | 2.80 | 0.006 | 0.757 | 4.413 |
| Confidence | | | | | | |
| Somewhat disagree | -0.277 | 0.604 | -0.46 | 0.647 | -1.475 | 0.920 |
| Neither agree nor disagree | -0.121 | 0.416 | -0.29 | 0.771 | -0.946 | 0.703 |
| Somewhat agree | -0.335 | 0.464 | -0.72 | 0.471 | -1.255 | 0.584 |
| Strongly agree | -0.368 | 0.575 | -0.64 | 0.524 | -1.507 | 0.771 |
| Age | 0.026 | 0.043 | 0.60 | 0.547 | -0.060 | 0.112 |
| Constant | 2.439 | 1.494 | 1.63 | 0.105 | -0.521 | 5.400 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

For testing the third hypothesis (H3), an analogous process has been followed. A Mann-Whitney U test was performed on the male portion of the total sample (i.e. n=242) in order to test for any potential differences between the math scores of the control and the treatment group. According to the Hypothesis 3, the score of the male participants remains on average the same with the presence and absence of the treatment. The results of the Mann-Whitney U test, illustrated in Table 4.2, show that indeed there is no statistically significant difference in the distribution of the two groups, at the 5% significance level, with a *p-value* of 0.1690. These results provide some evidence on the third hypothesis, however, a linear regression analysis on the male student's scores followed to build upon.

According to the results of Table 4.4, the stereotype threat message (treatment) has a non-significant effect on the average math test score of the males, at a significance level of 5%, with a *p-value* of 0.07. This provides additional evidence on the H3, meaning that men's math performance remains unaffected by the presence of the treatment. Furthermore, none of the additional control variables is totally statistically significant, at the 5% significance level, while the variables *Nationality*, *City*, *Field* and *Grade* show significant effects only for some of their categories, at the 5% significance level. For instance, a man with a total grade between 8.5 and 10, compared to a man with a total grade in the [5, 6.49] interval, has an average math score increased by 1.3 points, ceteris paribus, at the 5% significance level.

Table 4.4: *Output of linear regression analysis of the treatment on the math score, only for males*

| Variable | Coefficient | Robust SE | t-value | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Threat | | | | | | |
| Yes | -0.487* | 0.270 | -1.80 | 0.073 | -1.022 | 0.046 |
| | | | | | | |
| Nationality | | | | | | |
| Albanian | -0.889 | 0.723 | -1.23 | 0.220 | -2.315 | 0.536 |
| Cypriot | 1.109 | 1.237 | 0.90 | 0.371 | -1.331 | 3.551 |
| Other | -2.629*** | 0.659 | -3.99 | 0.000 | -3.931 | -1.328 |
| | | | | | | |
| City | | | | | | |
| Thessaloniki | -0.085 | 0.497 | -0.17 | 0.864 | -1.066 | 0.895 |
| Herakleion | 0.251 | 1.080 | 0.23 | 0.816 | -1.880 | 2.382 |
| Larisa | -1.601** | 0.741 | -2.16 | 0.032 | -3.063 | -0.139 |
| Volos | -0.204 | 1.068 | -0.19 | 0.848 | -2.312 | 1.902 |
| Other city | 0.173 | 0.504 | 0.34 | 0.731 | -0.820 | 1.167 |
| Other country | -1.174 | 0.812 | -1.45 | 0.150 | -2.777 | 0.427 |
| | | | | | | |
| Field | | | | | | |
| Physics/Applied Physics | 1.976*** | 0.560 | 3.52 | 0.001 | 0.870 | 3.082 |
| Computer Science | 1.080 | 0.868 | 1.24 | 0.215 | -0.632 | 2.794 |
| Civil Engineering | 1.020 | 0.818 | 1.25 | 0.214 | -0.593 | 2.634 |
| Electronic Engineering | 3.008*** | 0.845 | 3.56 | 0.000 | 1.339 | 4.676 |
| Chemical Engineering | 0.555 | 0.859 | 0.65 | 0.519 | -1.139 | 2.250 |
| Mechanical Engineering | 2.290** | 0.806 | 2.84 | 0.005 | 0.700 | 3.881 |
| Architecture | 0.451 | 0.909 | 0.50 | 0.620 | -1.343 | 2.246 |
| | | | | | | |
| University | | | | | | |
| University of Ioannina | -0.461 | 0.651 | -0.71 | 0.480 | -1.745 | 0.823 |
| University of Patra | 1.029 | 0.961 | 1.07 | 0.285 | -0.865 | 2.925 |
| University of Crete | 2.118* | 1.097 | 1.93 | 0.055 | -0.046 | 4.284 |
| University of Thessaly | -1.108* | 0.572 | -1.94 | 0.054 | -2.237 | 0.020 |
| University of Thrace | -0.717 | 0.464 | -1.55 | 0.124 | -1.634 | 0.198 |
| | | | | | | |
| Grade | | | | | | |
| [6.5,8.49] | 0.382 | 0.410 | 0.93 | 0.352 | -0.427 | 1.192 |
| [8.5,10] | 1.377** | 0.452 | 3.05 | 0.003 | 0.485 | 2.268 |
| | | | | | | |
| Age | -0.016 | 0.027 | -0.61 | 0.543 | -0.070 | 0.037 |
| | | | | | | |
| Confidence | | | | | | |
| Somewhat disagree | 0.045 | 0.572 | 0.08 | 0.937 | -1.084 | 1.175 |
| Neither agree nor disagree | 0.114 | 0.557 | 0.21 | 0.837 | -0.984 | 1.214 |
| Somewhat agree | 0.244 | 0.528 | 0.46 | 0.645 | -0.798 | 1.287 |
| Strongly agree | 0.643 | 0.610 | 1.06 | 0.293 | -0.559 | 1.847 |
| | | | | | | |
| Constant | 1.361 | 1.017 | 1.34 | 0.183 | -0.645 | 3.367 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

## 4.2 Self-reported anxiety results

As already mentioned, the potential anxiety that one can experience during a test or an assessment process (i.e. test anxiety) can negatively impact performance. The theory supports that test anxiety can work as a mediational mechanism in the stereotype threat-performance relationship, however, there is a lack of strong empirical support. In order to investigate whether the self-reported anxiety works as a mediator in the stereotype threat-performance relationship, I conducted an additional regression analysis, however on the anxiety score this time. The anxiety score of each participant has been calculated by summing up his/her scores of the six anxiety-related statements (measured by 5-point Likert Scales). Then, I standardised the variable, by subtracting the mean anxiety score of the control group from the anxiety score of each participant and dividing this outcome by the standard deviation of the anxiety score of the control group. After this adjustment, I used the linear regression model (4) to detect any significant effect between of the treatment to the anxiety score. The respective results are illustrated in Table 4.5.

Based on the regression's results, there is no significant effect of the treatment on the anxiety score, at the 5% significance level, as the variable *Threat* has a *p-value* of 0.4. Thus, these results do not provide any evidence that self-reported test anxiety works as a mediational mechanism at the stereotype threat-performance relationship. Additionally, not the interaction term *Threat*X*Gender* nor the rest of the control variables are statistically significant, at a significance level of 5%, apart from *Gender*. More specifically, being a female, compared to being a male, increases the anxiety score by 0.4 points, ceteris paribus. This effect is statistically significant at the 5% significance level with a *p-value* of 0.004. A Mann-Whitney U test and an additional linear regression analysis was furtherly conducted on the female sample, however, there is no significant difference in the anxiety score of control and treatment female groups or a significant effect of the treatment on their anxiety score. For the illustration of the results, see Appendix III.

Table 4.5: *Output of linear regression analysis of the treatment on the anxiety score*

| Variable | Coefficient | Robust SE | t-value | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Threat | | | | | | |
| Yes | 0.098 | 0.137 | 0.72 | 0.473 | -0.172 | 0.369 |
| Gender | | | | | | |
| Female | 0.458** | 0.156 | 2.93 | 0.004 | 0.151 | 0.766 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Threat#Gender | | | | | | |
| Yes#Female | -0.083 | 0.229 | -0.36 | 0.717 | -0.535 | 0.368 |
| Nationality | | | | | | |
| Albanian | -0.977** | 0.347 | -2.81 | 0.005 | -1.661 | -0.293 |
| Cypriot | -0.015 | 0.524 | -0.03 | 0.977 | -1.047 | 1.017 |
| Other | 2.546*** | 0.221 | 11.49 | 0.000 | 2.110 | 2.981 |
| City | | | | | | |
| Thessaloniki | 0.064 | 0.159 | 0.41 | 0.686 | -0.248 | 0.377 |
| Patra | 0.072 | 0.453 | 0.16 | 0.872 | -0.819 | 0.965 |
| Herakleion | -0.051 | 0.408 | -0.13 | 0.899 | -0.854 | 0.750 |
| Larisa | 0.076 | 0.366 | 0.21 | 0.835 | -0.644 | 0.797 |
| Volos | -0.240 | 0.415 | -0.58 | 0.562 | -1.057 | 0.575 |
| Other city | 0.166 | 0.158 | 1.05 | 0.296 | -0.146 | 0.478 |
| Other country | -0.097 | 0.285 | -0.34 | 0.734 | -0.658 | 0.464 |
| Field | | | | | | |
| Physics/Applied Physics | -0.162 | 0.452 | -0.36 | 0.720 | -1.052 | 0.728 |
| Computer Science | 0.157 | 0.531 | 0.30 | 0.767 | -0.887 | 1.202 |
| Civil Engineering | 0.195 | 0.461 | 0.42 | 0.672 | -0.711 | 1.102 |
| Electronic Engineering | -0.213 | 0.483 | -0.44 | 0.659 | -1.165 | 0.737 |
| Chemical Engineering | 0.442 | 0.776 | 0.57 | 0.569 | -1.084 | 1.970 |
| Mechanical Engineering | -0.255 | 0.461 | -0.55 | 0.581 | -1.163 | 0.652 |
| Architecture | 0.666 | 1.490 | 0.45 | 0.655 | -2.267 | 3.599 |
| University | | | | | | |
| University of Ioannina | 0.211 | 0.218 | 0.97 | 0.333 | -0.217 | 0.640 |
| University of Patra | 0.355 | 0.612 | 0.58 | 0.562 | -0.848 | 1.559 |
| University of Crete | -0.010 | 0.425 | -0.02 | 0.980 | -0.847 | 0.826 |
| University of Thessaly | 0.801*** | 0.160 | 4.98 | 0.000 | 0.484 | 1.117 |
| University of Thrace | 0.820 | 0.593 | 1.38 | 0.168 | -0.347 | 1.988 |
| Grade | | | | | | |
| [6.5,8.49] | -0.092 | 0.168 | -0.55 | 0.583 | -0.424 | 0.239 |
| [8.5,10] | -0.042 | 0.216 | -0.19 | 0.846 | -0.468 | 0.384 |
| Age | -0.006 | 0.015 | -0.43 | 0.667 | -0.037 | 0.024 |
| Confidence | | | | | | |
| Somewhat disagree | 0.221 | 0.462 | 0.48 | 0.632 | -0.688 | 1.132 |
| Neither agree nor disagree | 0.094 | 0.441 | 0.21 | 0.831 | -0.773 | 0.962 |
| Somewhat agree | -0.247 | 0.438 | -0.56 | 0.573 | -1.109 | 0.615 |
| Strongly agree | -0.366 | 0.463 | -0.79 | 0.430 | -1.278 | 0.545 |
| Constant | 0.109 | 0.753 | 0.15 | 0.885 | -1.373 | 1.592 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

## 5. Discussion

Stereotype threat has been identified as an influential factor in performance. This research aimed to determine the effect of stereotype threat on math performance of female students from STEM universities in Greece. In addition, it furtherly investigated the role of self-reported test anxiety on this relationship. This happened by conducting an online experiment. Students and recent graduates of STEM-related faculties were invited to participate in an online test of seven multiple-choice math questions. Participants were randomly assigned into a control and treatment group. Before the test, the participants of the treatment group got a message revealing the existence of a gender gap against women, in performance, particularly in similar math tests. This message aimed to cause stereotype threat to the female participants.

Interestingly, the results show that on average women in the stereotype threat condition have a higher score compared to the non-threatened women. The positive effect of stereotype threat to women's performance is the most striking result that emerged from this study since it is inconsistent and contradictory with the previous research on stereotype threat. Based on the theory, which was initially introduced by Steele and Aronson (1995) and is characterised by great empirical support, the fear of not confirming a negative stereotype about one's social identity, makes the person behave in a stereotype-consistent way. Thus, considering this, the current results are contradictory with both the theory and the two first hypotheses based on the it. More specifically, the current findings did not prove any evidence to support that gender gap information against women lead female STEM students to underperform at math tests compared to male STEM students (H1) and female STEM students in the control condition (H2). As already mentioned, although stereotype threat has negative consequences on performance on difficult tasks, it has a positive effect on easy ones (O'Brien & Crandall, 2003; Spencer et al., 1999). In the current study, there is no evidence that the test was easy. Since, at the question regarding the difficulty of the test, the majority of the participants overall (75.96% of the total sample), as well as, 84.41% of the women respondents specifically, characterized the test as *"Neither easy nor difficult"* or *"Somewhat difficult",* based on their mathematics knowledge. Based on this, I do not believe that the outperformance of the threatened women comes from the perceived difficulty of the test.

A previous study by Crisp et al. (2009) revealed analogous to the current results. In their experiment, Crisp and his colleagues compared the effect of stereotype threat activation, by triggering the negative gender math stereotype, on the math performance of female psychology and engineering students. Similarly, the engineering female students under stereotype threat outperformed compared to the non-threatened. While the effect of stereotype threat on the female psychology students' performance was negative and consistent with the theory. According to Crisp et al. (2009), this positive effect may occur because women in math-related domains (such as STEM fields) are familiar with constant exposure to stereotype threatening environments, so they have developed internal mechanisms to face analogous situations. Another potential reason for the positive effect on women's performance could be that the gender gap information actually motivates women to try harder and perform better. Experimental evidence by Jamieson and Harkins (2007 and 2009) and McFall et al. (2009) supports this theory. Stereotype threat enhances the motivation and willingness of the targets to perform well, which *"potentiates prepotent responses"* and furtherly increases the given effort when the individual recognizes that (s)he initially has made a mistake, however (s)he has the knowledge and the time to correct it (Jamieson and Harkins, 2007). Considering that the math test was designed based on basic math knowledge (for the STEM domains) and that the time to solve it, was unlimited, this could explain the increased performance of the female participants in the treatment group.

In addition, the present study provides additional evidence that suggests that the performance of men stay unaffected after a message regarding gender differences against women. Both the Mann-Whitney U test and the linear regression analysis revealed no statistical differences in the average math score of male participants between the control and the treatment conditions. These results provide support for my third hypothesis and confirm previous findings of Sekaquaptewa and Thompson (2003) and O'Brien and Crandall (2003). The evidence that men's average math performance remained unaffected is additionally consistent with the theory of stereotype threat. According to Steele et al. (2002), stereotype threat gets activated when the individuals, who identify themselves as members of a group (e.g. males and females), are exposed to a negative stereotype regarding a behaviour or an attitude of their group. In the current study, the gender-gap message is orientated towards women's underperformance. Thus, it is logical that it did not influence men's performance.

Regarding the role of anxiety in the stereotype threat-performance relationship, the previous experimental findings are controversial. In this study, anxiety was measured using a self-reported method that aimed to capture the experienced anxiety during the test. Using a similar approach to Osborne's (2001), I created an individual anxiety score for each participant. After running a linear regression of the stereotype threat (i.e. treatment) on the anxiety score, no significant effect of the treatment was detected. In addition, neither the interaction term between the treatment and the gender was significant. The absence of significant effects indicates that there is no evidence of anxiety being a mediator of stereotype threat. These results are consistent with previous findings of stereotype threat literature (e.g. Bosson et al., 2004; Hess et al., 2003, Stone, 2002). However, as already mentioned, considering that anxiety can be challenging to be cognisable, the estimation of instant anxiety (during a test, for instance) through self-reported approaches may not be accurate. This topic is furtherly discussed in the next section, as one of the limitations of this study. Furthermore, an additional finding that occurred from the anxiety analysis as a mediator is the following. Although the interaction term *Threat*X*Gender* was not significant, the results showed a significant positive effect of gender on the anxiety score, meaning that overall being a woman increases the self-reported test anxiety. This finding is consistent with the previous test-anxiety literature, which supports that woman report higher conscious and self-reported anxiety during assessment processes and tests (Everson et al., 1991; Núñez-Peña et al., 2016; Rezazadeh and Tavakoli, 2009).

## 5.1 Limitations and future research

The findings in this paper are subject to at least three limitations. The first weakness of this study was the size of the sample, as the optimal sample, based on the a priori power analysis, was 538 participants. In addition, the sample consisted of only 154 female respondents, a number that is disproportional compared to the males of the sample (61.11% of the total sample) and far below the recommended sample size. Although efforts were undertaken by distributing the survey to students from multiple universities and fields, the sample was not representative of the Greek STEM university students, considering that many fields were underrepresented, such as the fields of Architecture and Chemical Engineering (with only two participants each). Thus, a larger sample with more homogeneity or better-distributed diversity would increase the statistical power of the analysis and enhance the generalizability of the results. Unfortunately, the number of participants who did not finish the whole survey was significant, with almost half of the recorded responses being non-completed. One potential reason would be the nature of the task, the design of the experiment or even the treatment itself. Although monetary incentives can reduce drop-out (Crump et. al, 2013), considering the budget restriction it was not possible to compensate each respondent individually or even better to provide a direct reward for the correct answers.

Another limitation, which is also linked to the high attrition from the study, was that the experiment was conducted online, while the majority of similar stereotype threat experiments took place in a lab. The multiple environments that the respondents fill in the survey and specifically the math test could be variant, considering that they had access to this by their laptop/PC and their mobile phone. Thus, it might be the case that some participants did not give accurate responses in the math test not because they did not have the knowledge or due to the treatment. It is likely that the circumstances, in which the participants took the math test, did not enable them to completely focus on the task. In addition, although I tried to include questions that did not require extensive calculations and I advised the respondents to have a pen, a piece of paper and a calculator for their convenience, maybe they did not have access to this equipment at the time of the test. In a lab setting, I would have better control over these conditions. However, collecting responses offline was not something possible mainly considering the Covid-19 pandemic situation.

Furthermore, a limitation that has already been highlighted by the existing stereotype threat and anxiety literature is the disputable accuracy of self-reported methods to measure anxiety. As already mentioned, measuring anxiety via these kinds of methods can be challenging (Bosson et al., 2004; Steele and Aronson, 1995). More specifically, Bosson et al. (2004), in their experimental study, found that people under stereotype threat showed higher non-verbal anxiety (considering their behaviour and movements), compared to the non-threatened participants. The non-verbal anxiety appeared to work as a mediator of the stereotype threat effect. However, the two groups had a non-significant difference in their self-reported anxiety. In another experiment, Osborne (2007) aimed to capture anxiety during the stereotype threat experience by measuring the physiological reactance of "threatened" participants. Indeed, those who were under stereotype threat manipulation showed evidence of higher anxiety, such as higher blood pressure and skin conductance (Osborne, 2007). Thus, anxiety can be characterised as a plausible mediational mechanism of stereotype threat, however, self-reported methods flounder to capture it. Nevertheless, considering the nature of the current study, it was not possible to measure anxiety using non-self-reported methods, due to ethical and limited time concerns.

Finally, the current study account for the effect of the stereotype threat to female students' math performance in STEM disciplines. Although the majority of the existing literature supports that stereotype threat disrupts the math performance of targeted women (e.g. Cadinu et al. 2005; Sekaquaptewa & Thompson, 2003; Spencer, 1999), the findings of this study reveal a positive relationship between stereotype threat and math performance of females. Unfortunately, there is limited literature that aims to explain this ironic phenomenon. Therefore, more research is required on stereotype threat positive effects on the performance of the targeted groups. Future studies should furtherly investigate the reasons that account for this reversed relationship and the factors that influence it.

## 6. Conclusion

This study aimed to investigate the effect of negative stereotypes on performance. More specifically, in this research, I focused on the phenomenon of stereotype threat, which can be described as a situational experience when a person, who identifies him/herself as a member of a negatively stereotyped group, is "threatened" by the idea of confirming the specific negative stereotype. Although stereotype threat can be related to numerous stereotypes and can target a broad number of social groups, I decided to research the effect of stereotype threat activation on the math performance of female students of Greek STEM higher education faculties. Based on the existing theoretical and empirical support, I stated three hypotheses, regarding the math performance of both male and female STEM students. Additionally, considering the previous contradictory findings regarding the involvement of test anxiety in this relationship, the current study furtherly investigated the mediational role of self-reported anxiety. This has been done by means of an online experiment.

Both male and female students and recent graduates from a broad range of STEM faculties from different Greek universities participated in the experiment, which had the form of a mathematics short test. The participants got randomly assigned into the control and the treatment group keeping the groups as homogenous as possible. The treatment group received a message regarding the existence of a gender gap in the scores of past participants, undermining the women's performance, prior to the test. Based on previous experimental studies, this kind of information activates the stereotype threat phenomenon to women regarding their math ability and decreases their math performance.

The empirical findings of the present study, however, do not provide further evidence on the stereotype threat theory. Ironically, the "threatened" female participants outperformed the women, who did not have any relevant information, meaning that stereotype threat, in fact, had an enhancive effect on the performance of the former. There are variant reasons that can explain this contrary effect. Returning to the research question posed at the beginning of this study, it is now possible to state that indeed stereotype threat influences women's math performance in Greek STEM universities, however towards an unexpected direction. In addition, as hypothesised, men's performance remained consistent, overall. This is consistent with the theory; namely, one can experience stereotype threat if the stereotype is relevant to at least one of his/her identities. Finally, self-reported anxiety plays a limited role in this relationship.

Nevertheless, one should consider the limitations of the current study before drawing any conclusions.

To conclude, the current research provides an additional insight into the influence of negative stereotypes on performance. Comprehending the source of this effect, and the potential personal and environmental factors that enhance, and more importantly, mitigate it, can contribute to limiting the negative consequences of stereotypes to academic performance.

## 7. Bibliography


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.

Beilock, S. L., Jellison, W. A., Rydell, R. J., McConnell, A. R., & Carr, T. H. (2006). On the causal mechanisms of stereotype threat: can skills that don't rely heavily on working memory still be threatened?. Personality and Social Psychology Bulletin, 32(8), 1059-1071.

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

Bell, A. E., Spencer, S. J., Iserman, E., & Logel, C. E. (2003). Stereotype threat and women's performance in engineering. Journal of Engineering Education, 92(4), 307-312.

Ben-Zeev, T., Fein, S., & Inzlicht, M. (2005). Arousal and stereotype threat. Journal of Experimental Social Psychology, 41(2), 174-181.

Bosson, J. K., Haymovitz, E. L., & Pinel, E. C. (2004). When saying and doing diverge: The effects of stereotype threat on self-reported versus non-verbal anxiety. Journal of experimental social psychology, 40(2), 247-255.

Brodish, A. B., & Devine, P. G. (2009). The role of performance–avoidance goals and worry in mediating the relationship between stereotype threat and performance. Journal of experimental social psychology, 45(1), 180-185.

Carr, P. B., & Steele, C. M. (2010). Stereotype threat affects financial decision making. Psychological Science, 21(10), 1411-1416.

Cadinu, M., Maass, A., Frigerio, S., Impagliazzo, L., & Latinotti, S. (2003). Stereotype threat: The effect of expectancy on performance. European journal of social psychology, 33(2), 267-285.

Cadinu, M., Maass, A., Rosabianca, A., & Kiesner, J. (2005). Why do women underperform under stereotype threat? Evidence for the role of negative thinking. Psychological science, 16(7), 572-578.

Chapell, M. S., Blanding, Z. B., Silverstein, M. E., Takahashi, M., Newman, B., Gubi, A., & McCann, N. (2005). Test anxiety and academic performance in undergraduate and graduate students. Journal of educational Psychology, 97(2), 268.

Chung, B. G., Ehrhart, M. G., Holcombe Ehrhart, K., Hattrup, K., & Solamon, J. (2010). Stereotype threat, state anxiety, and specific self-efficacy as predictors of promotion exam performance. Group & organization management, 35(1), 77-107.

Crisp, R. J., Bache, L. M., & Maitner, A. T. (2009). Dynamics of social comparison in counter-stereotypic domains: Stereotype boost, not stereotype threat, for women engineering majors. Social Influence, 4(3), 171-184.

Croizet, J. C., & Claire, T. (1998). Extending the concept of stereotype threat to social class: The intellectual underperformance of students from low socioeconomic backgrounds. Personality and Social Psychology Bulletin, 24(6), 588-594.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. PloS one, 8(3), e57410.

Culler, R. E., & Holahan, C. J. (1980). Test anxiety and academic performance: the effects of study-related behaviors. Journal of educational psychology, 72(1), 16.

DordiNejad, F. G., Hakimi, H., Ashouri, M., Dehghani, M., Zeinali, Z., Daghighi, M. S., & Bahrami, N. (2011). On the relationship between test anxiety and academic performance. Procedia-Social and Behavioral Sciences, 15, 3774-3778.

ETS. (2017). GRE Mathematics Test Practice Book. Educational Testing Service https://www.ets.org/s/gre/pdf/practice_book_math.pdf

ETS. (2022). A Snapshot of the Individuals Who Took the GRE General Test. July 2016-June 2021. Educational Testing Service. https://www.ets.org/s/gre/pdf/snapshot.pdf

European Institute for Gender Equality. (2018). Study and work in the EU: set apart by gender. Publications Office of the European Union.

European Institute for Gender Equality. (2021). Gender Equality Index | 2021 | Greece. EIGE. https://eige.europa.eu/gender-equality-index/2021/country/EL

Everson, H. T., Millsap, R. E., & Rodriguez, C. M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the Test Anxiety Inventory. Educational and Psychological Measurement, 51(1), 243-251.

Faul, F., Erdfelder, E., Lang, A.-G. & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods, 39, 175-191.

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

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: an integrative review of implicit and explicit attitude change. Psychological bulletin, 132(5), 692.

Gonzales, P. M., Blanton, H., & Williams, K. J. (2002). The effects of stereotype threat and double-minority status on the test performance of Latino women. Personality and social psychology bulletin, 28(5), 659-670.

Harrison, L. A., Stevens, C. M., Monty, A. N., & Coakley, C. A. (2006). The consequences of stereotype threat on the academic performance of White and non-White lower income college students. Social Psychology of Education, 9(3), 341-357.

Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. Review of educational research, 58(1), 47-77.

Hess, T. M., Auman, C., Colcombe, S. J., & Rahhal, T. A. (2003). The impact of stereotype threat on age differences in memory performance. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58(1), P3-P11.

Hilton, J. L., & Von Hippel, W. (1996). Stereotypes. Annual review of psychology, 47(1), 237-271.

Hoyt, C. L., & Murphy, S. E. (2016). Managing to clear the air: Stereotype threat, women, and leadership. The leadership quarterly, 27(3), 387-399.

Inzlicht, M., & Ben-Zeev, T. (2000). A threatening intellectual environment: Why females are susceptible to experiencing problem-solving deficits in the presence of males. Psychological Science, 11(5), 365-371.

Jamieson, J. P., & Harkins, S. G. (2007). Mere effort and stereotype threat performance effects. Journal of personality and social psychology, 93(4), 544.

Jamieson, J. P., & Harkins, S. G. (2009). The effect of stereotype threat on the solving of quantitative GRE problems: A mere effort interpretation. Personality and Social Psychology Bulletin, 35(10), 1301-1314.

Jamieson, J. P., & Harkins, S. G. (2012). Distinguishing between the effects of stereotype priming and stereotype threat on math performance. Group Processes & Intergroup Relations, 15(3), 291-304.

Joanisse, M., Gagnon, S., & Voloaca, M. (2013). The impact of stereotype threat on the simulated driving performance of older drivers. Accident Analysis & Prevention, 50, 530-538.

Johns, M., Inzlicht, M., & Schmader, T. (2008). Stereotype threat and executive resource depletion: examining the influence of emotion regulation. Journal of Experimental Psychology: General, 137(4), 691.

Keller, J. (2002). Blatant stereotype threat and women's math performance: Self handicapping as a strategic means to cope with obtrusive negative performance expectations. Sex Roles, 47(3), 193-198.

Keller, J., & Dauenheimer, D. (2003). Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women’s math performance. Personality and Social Psychology Bulletin, 29(3), 371-381.

Kim, H. Y. (2017). Statistical notes for clinical researchers: Chi-squared test and Fisher's exact test. Restorative dentistry & endodontics, 42(2), 152-155.

Koenig, A. M., & Eagly, A. H. (2005). Stereotype threat in men on a test of social sensitivity. Sex Roles, 52(7), 489-496.

Laurin, R. (2013). Stereotype threat and lift effects in motor task performance: The mediating role of somatic and cognitive anxiety. The Journal of social psychology, 153(6), 687-699.

Lesko, A. C., & Corpus, J. H. (2006). Discounting the difficult: How high math-identified women respond to stereotype threat. Sex Roles, 54(1), 113-125.

Lu, A., Feng, Y., Yu, Z., Tian, H., Hong, X., & Zheng, D. (2015). Anxiety and mind wandering as independent consequences of stereotype threat. Social Behavior and Personality: an international journal, 43(4), 537-558.

Mayer, D. M., & Hanges, P. J. (2003). Understanding the stereotype threat effect with" culture-free" tests: An examination of its mediators and measurement. Human Performance, 16(3), 207-230.

McFall, S. R., Jamieson, J. P., & Harkins, S. G. (2009). Testing the mere effort account of the evaluation-performance relationship. Journal of Personality and Social Psychology, 96(1), 135–154.

McGlone, M. S., Aronson, J., & Kobrynowicz, D. (2006). Stereotype threat and the gender gap in political knowledge. Psychology of Women Quarterly, 30(4), 392-398.

Mrazek, M. D., Chin, J. M., Schmader, T., Hartson, K. A., Smallwood, J., & Schooler, J. W. (2011). Threatened to distraction: Mind-wandering as a consequence of stereotype threat. Journal of Experimental Social Psychology, 47(6), 1243-1248.

Nachar, N. (2008). The Mann-Whitney U: A test for assessing whether two independent samples come from the same distribution. Tutorials in quantitative Methods for Psychology, 4(1), 13-20.

Núñez-Peña, M. I., Suárez-Pellicioni, M., & Bono, R. (2016). Gender differences in test anxiety and their impact on higher education students’ academic achievement. Procedia-Social and Behavioral Sciences, 228, 154-160.

O'Brien, L. T., & Crandall, C. S. (2003). Stereotype threat and arousal: Effects on women's math performance. Personality and Social Psychology Bulletin, 29(6), 782-789.

Osborne, J. W. (2001). Testing stereotype threat: Does anxiety explain race and sex differences in achievement?. Contemporary educational psychology, 26(3), 291-310.

Osborne, J. W. (2007). Linking stereotype threat and anxiety. Educational psychology, 27(1), 135-154.

Pavlova, M. A., Weber, S., Simoes, E., & Sokolov, A. N. (2014). Gender stereotype susceptibility. PloS one, 9(12), e114802.

Pennington, C. R., Heim, D., Levy, A. R., & Larkin, D. T. (2016). Twenty years of stereotype threat research: A review of psychological mediators. PloS one, 11(1), e0146487.

Pronin, E., Steele, C. M., & Ross, L. (2004). Identity bifurcation in response to stereotype threat: Women and mathematics. Journal of Experimental Social Psychology, 40(2), 152-168.

Rahhal, T. A., Hasher, L., & Colcombe, S. J. (2001). Instructional manipulations and age differences in memory: now you see them, now you don't. Psychology and Aging, 16(4), 697.

Rana, R., & Mahmood, N. (2010). The relationship between test anxiety and academic achievement. Bulletin of Education and research, 32(2), 63-74.

Rezazadeh, M., & Tavakoli, M. (2009). Investigating the Relationship among Test Anxiety, Gender, Academic Achievement and Years of Study: A Case of Iranian EFL University Students. English Language Teaching, 2(4), 68-74.

Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. Journal of personality and social psychology, 46(4), 929.

Schmader, T. (2002). Gender identification moderates stereotype threat effects on women's math performance. Journal of experimental social psychology, 38(2), 194-201.

Schmader, T. (2010). Stereotype threat deconstructed. Current Directions in Psychological Science, 19(1), 14-18.

Schmader, T., & Beilock, S. (2011). An Integration of Processes that Underlie Stereotype Threat. In Stereotype Threat: Theory, Process, and Application. : Oxford University Press.

Schmader, T. & Inzlicht, M. (2012). Stereotype threat: Theory, process, and application. Oxford University Press.

Schmader, T., & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. Journal of personality and social psychology, 85(3), 440.

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

Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. Anxiety research, 4(1), 27-41.

Sekaquaptewa, D., & Thompson, M. (2003). Solo status, stereotype threat, and performance expectancies: Their effects on women's performance. *Journal of experimental social psychology*, *39*(1), 68-74.

Shih, M., Pittinsky, T. L., & Ambady, N. (1999). Stereotype Susceptibility: Identity Salience and Shifts in Quantitative Performance. Psychological Science, 10(1), 80–83.

Smith, J. L. (2004). Understanding the process of stereotype threat: A review of mediational variables and new performance goal directions. Educational Psychology Review, 16(3), 177-206.

Spencer, S. J., Logel, C., & Davies, P. G. (2016). Stereotype threat. Annual review of psychology, 67, 415-437.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. Journal of experimental social psychology, 35(1), 4-28.

Stankov, L., Morony, S., & Lee, Y. P. (2014). Confidence: the best non-cognitive predictor of academic achievement?. Educational psychology, 34(1), 9-28.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. American psychologist, 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.

Steele, C. M., Spencer, S. J., & Aronson, J. (2002). Contending with group image: The psychology of stereotype and social identity threat. In Advances in experimental social psychology (Vol. 34, pp. 379-440). Academic Press.

Stein, R., Blanchard-Fields, F., & Hertzog, C. (2002). The effects of age-stereotype priming on the memory performance of older adults. *Experimental aging research*, *28*(2), 169-181.

Stone, J. (2002). Battling doubt by avoiding practice: The effects of stereotype threat on self handicapping in white athletes. Personality and Social Psychology Bulletin, 28(12), 1667-1678.

Stone, J., Lynch, C. I., Sjomeling, M., & Darley, J. M. (1999). Stereotype threat effects on Black and White athletic performance. Journal of personality and social psychology, 77(6), 1213.

Suhr, J. A., & Gunstad, J. (2002). "Diagnosis threat": The effect of negative expectations on cognitive performance in head injury. Journal of Clinical and Experimental Neuropsychology, 24(4), 448-457.

Tavani, C. M., & Losh, S. C. (2003). Motivation, self-confidence, and expectations as predictors of the academic performances among our high school students. *Child study journal*, *33*(3), 141-152.

The Nobel Prize. (n.d.). All Nobel Prizes in Chemistry. NobelPrize.Org. https://www.nobelprize.org/prizes/lists/all-nobel-prizes-in-chemistry/

The Nobel Prize. (n.d.). All Nobel Prizes in Physics. NobelPrize.Org. https://www.nobelprize.org/prizes/lists/all-nobel-prizes-in-physics/

UNESCO. (2017). Cracking the code: girls' and women's education in science, technology, engineering and mathematics (STEM) (Revised ed.). United Nations Educational, Scientific and Cultural Organization. https://unesdoc-unesco-org.eur.idm.oclc.org/ark:/48223/pf0000253479

# 8. Appendices

## Appendix I. Survey

*Message:*

The current questionnaire is part of my Master's Thesis in Behavioural Economics MSc program at Erasmus University, in Rotterdam. The current study aims to measure the mathematics performance of students from Greek Universities. The questionnaire consists of three parts:

1. Demographic questions
2. A math test of 7 multiple choice questions
3. Questions about the test

After the end of the questionnaire, you can enter your email address, in order to participate in a lottery with price a €30 voucher from Public stores. With every correct answer in the math test, you increase your probability to win.

**In order to complete the questionnaire, you should be a student or a recent graduate of a Greek polytechnic/mathematics/physics/computer science university.**

All your answers are anonymous and will be handled confidentially. You can quit the survey anytime. If you prefer not to answer a certain question, that is possible.

Thanks a lot for your time!

Kind regards,
Ourania

1. What is your gender?

- ○ Male
- ○ Female
- ○ Non-binary / third gender
- ○ Prefer not to say

2. What is your age?

(Use number digits to indicate your age)

___________________________________________

3. What is your nationality?

- ○ Greek
- ○ Albanian
- ○ Bulgarian
- ○ Cypriot
- ○ Italian
- ○ Other (please mention below)

___________________________________________

4. In which of the following places are you coming from?

- Athens
- Thessaloniki
- Patra
- Heraklion
- Somewhere else (please mention below)

______________________________________________

- Computer Science
- Civil Engineering
- Mechanical Engineering
- Electronic Engineering
- Chemical Engineering
- Environmental Engineering
- Architecture
- Other polytechnical discipline that is not in the list
- No. None of the above (or similar) disciplines

*(If "No. None of the above (or similar) disciplines" is selected, survey ends.)*

*(If the answer to question 5 is "Mathematics/ Applied Mathematics" the following question is displayed.)*

5a. Which is the Greek Mathematics/ Applied Mathematics University that you are a student/recent graduate of?

- University of Crete (Heraklion) - Department of Mathematics
- University of Crete (Heraklion) - Department of Applied Mathematics
- National and Kapodistrian University of Athens - Department of Mathematics
- Aristotle University of Thessaloniki - Department of Mathematics

- University of Ioannina- Department of Mathematics
- University of Western Macedonia (Kastoria) - Department of Mathematics
- University of Thessaly (Lamia) - Department of Mathematics
- University of Patras - Department of Mathematics
- University of the Aegean (Samos) - Department of Mathematics
- National Technical University of Athens - Department of Applied Mathematics and Physical Sciences
- None of the above (please mention below)

______________________________________________

---

*(If the answer to question 5 is "Physics/ Applied Physics" the following question is displayed.)*

5b. Which is the Greek Physics/ Applied Physics University that you are a student/recent graduate of?

- National and Kapodistrian University of Athens - Department of Physics
- Aristotle University of Thessaloniki - Department of Physics
- University of Patras - Department of Physics
- University of Ioannina- Department of Physics
- University of Crete (Heraklion) - Department of Physics
- University of Thessaly (Lamia) - Department of Physics
- National Technical University of Athens - Department of Applied Mathematics and Physical Sciences
- None of the above (please mention below)

______________________________________________

*(If the answer to question 5 is "Computer Science" the following question is displayed.)*

5c. Which is the Greek Computer Science University that you are a student/recent graduate of?

- University of Crete (Heraklion) - Department of Computer Science
- Aristotle University of Thessaloniki - Department of Computer Science
- Ionian University (Corfu) - Department of Computer Science
- University of Piraeus - Department of Computer Science
- University of Thessaly (Lamia) - Department of Computer Science in Biomedicine
- University of Patras - Department of Computer Engineering and Informatics
- None of the above (please mention below)

______________________________________________

*(If the answer to question 5 is "Civil Engineering" the following question is displayed.)*

5d. Which is the Greek Civil Engineering University that you are a student/recent graduate of?

- National Technical University of Athens - Department of Civil Engineering
- Aristotle University of Thessaloniki - Department of Civil Engineering
- Democritus University of Thrace - Department of Civil Engineering
- University of Thessaly (Volos) - Department of Civil Engineering
- None of the above (please mention below)

______________________________________________

*(If the answer to question 5 is "Mechanical Engineering" the following question is displayed.)*

5e. Which is the Greek Mechanical Engineering University that you are a student/recent graduate of?

- ○ National Technical University of Athens - Department of Mechanical Engineering
- ○ Aristotle University of Thessaloniki - Department of Mechanical Engineering
- ○ University of Patra - Department of Mechanical Engineering and Aeronautics
- ○ University of Thessaly - Department of Mechanical Engineering
- ○ None of the above (please mention below)

____________________________________________

*(If the answer to question 5 is "Electronic Engineering" the following question is displayed.)*

5f. Which is the Greek Electronic Engineering University that you are a student/recent graduate of?

- ○ National Technical University of Athens - Department of Electronic and Computer Engineering
- ○ Aristotle University of Thessaloniki - Department of Electronic and Computer Engineering
- ○ University of Patras - Department of Computer Engineering and Informatics
- ○ University of Patras - Department of Electronic and Computer Engineering
- ○ Democritus University of Thrace - Department of Electronic and Computer Engineering
- ○ University of Crete (Heraklion) - Department of Electronic and Computer Engineering
- ○ University of Thessaly - Department of Electronic and Computer Engineering
- ○ None of the above (please mention below)

____________________________________________

*(If the answer to question 5 is "Chemical Engineering" the following question is displayed.)*

5g. Which is the Greek Chemical Engineering University that you are a student/recent graduate of?

- ○ National Technical University of Athens - Department of Chemical Engineering
- ○ Aristotle University of Thessaloniki - Department of Chemical Engineering
- ○ University of Patras - Department of Chemical Engineering
- ○ None of the above (please mention below)

____________________________________________

*(If the answer to question 5 is "Environmental Engineering" the following question is displayed.)*

5h. Which is the Greek Environmental Engineering University that you are a student/recent graduate of?

- ○ University of Patras - Department of Environmental Engineering
- ○ Democritus University of Thrace - Department of Environmental Engineering
- ○ University of Crete (Heraklion) - Department of Environmental Engineering
- ○ None of the above (please mention below)

____________________________________________

*(If the answer to question 5 is "Architecture" the following question is displayed.)*

5i. Which is the Greek Architecture University that you are a student/recent graduate of?

- ○ National Technical University of Athens - Department of Architectural Engineering
- ○ University of Patras - Department of Architectural Engineering
- ○ Democritus University of Thrace - Department of Architectural Engineering
- ○ University of Crete (Heraklion) - Department of Architectural Engineering
- ○ University of Thessaly (Volos) - Department of Architectural Engineering
- ○ Aristotle University of Thessaloniki - Department of Architectural Engineering
- ○ None of the above (please mention below)

________________________________________________

*(If the answer to question 5 is "Other polytechnical discipline that is not in the list" the following question is displayed.)*

5j. Which is the Greek Polytechnical University that you are a student/recent graduate of?

- ○ National Technical University of Athens - Department of Rural, Surveying and Geoinformatics Engineering
- ○ National Technical University of Athens - Department of Mining & Metallurgical Engineering
- ○ Aristotle University of Thessaloniki - Department of Rural and Surveying Engineering
- ○ Aristotle University of Thessaloniki - Department of Spatial Planning and Development Engineering
- ○ University of Thessaly (Volos) - Department of Planning and Regional Development
- ○ University of Crete (Heraklion) - Department of Production and Management Engineering
- ○ University of Crete (Heraklion) - Department of Mineral Resources Engineering

○ Democritus University of Thrace - Department of Production and Management Engineering

○ None of the above (please mention below)

________________________________________________

6. What is your current total grade?

○ 5 - 6.49

○ 6.5 - 8.49

○ 8.5 - 10

7. To what degree, do you agree with the following statement?
"I am confident about my level in math."

○ Strongly disagree

○ Somewhat disagree

○ Neither agree nor disagree

○ Somewhat agree

○ Strongly agree

*Treatment message*

Thank you for your answers. You are suitable to proceed with the test. The test contains 7 multiple-choice Mathematics questions from the GRE* Mathematics Test Practice Book. You can go back and forth and change your answers. It is recommended to use paper, pen, and calculator for your own convenience.

In previous similar tests according to the latest GRE Worldwide Report, **men had higher performance in quantitative/mathematics questions compared to women**. This is the case for all the Undergraduate Major Fields including Engineering and Physical Sciences as well.

**Remember, with every correct answer you increase your probability to win the 30€ prize. Good luck!**

*The GRE (Graduate Record Examinations) is a standardized test that is an admissions requirement for many universities around the world for postgraduate studies.

*Control message*

Thank you for your answers. You are suitable to proceed with the test. The test contains 7 multiple-choice Mathematics questions from the GRE* Mathematics Test Practice Book. You can go back and forth and change your answers. It is recommended to use paper, pen, and calculator for your own convenience.

**Remember, with every correct answer you increase your probability to win the 30€ prize. Good luck!**

*The GRE (Graduate Record Examinations) is a standardized test that is an admissions requirement for many universities around the world for postgraduate studies.

*Math test*

i) $\lim_{x \to 0} \frac{\cos(3x)-1}{x^2} =$

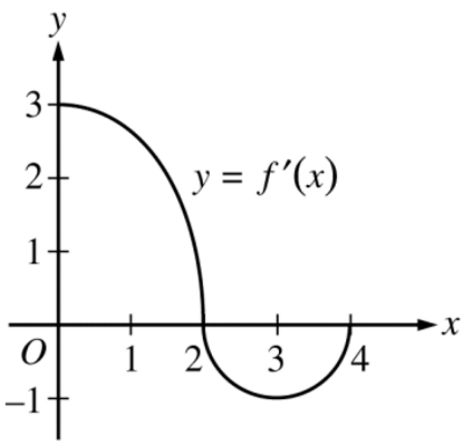
- A) 9/2
- B) 3/2
- C) -2/3
- D) -3/2
- E) -9/2
- F) I don't know

ii) Sofia and Tess will each randomly choose one of the 10 integers from 1 to 10. What is the probability neither integer chosen will be the square of the other?

- A) 0.64
- B) 0.72
- C) 0.81
- D) 0.90
- E) 0.95
- F) I don't know

iii) Which of the following shows the numbers $2^{1/2}$, $3^{1/3}$ and $6^{1/6}$ in increasing order?

- A) $2^{1/2} < 3^{1/3} < 6^{1/6}$
- B) $6^{1/6} < 3^{1/3} < 2^{1/2}$
- C) $6^{1/6} < 2^{1/2} < 3^{1/3}$
- D) $3^{1/3} < 2^{1/2} < 6^{1/6}$
- E) $3^{1/3} < 6^{1/6} < 2^{1/2}$
- F) I don't know



iv)

The figure above shows the graph of the **derivatives *f′*** of a function f, where *f* is continuous on the interval {0,4} and differentiable of the (0,4). Which of the following gives the correct ordering of the values *f(0)*, *f(2)* and *f(4)*?

○ A) f(0) < f(2) < f(4)

○ B) f(0) < f(4) = f(2)

○ C) f(0) < f(4) < f(2)

○ D) f(4) = f(2) < f(0)

○ E) f(4) < f(0) < f(2)

○ F) I don't know

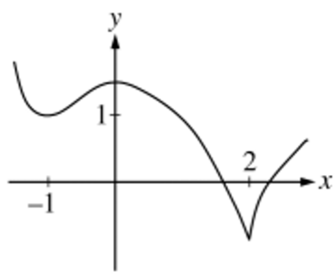
v) Let *g* be a continuous real-valued function defined on R with the following properties.

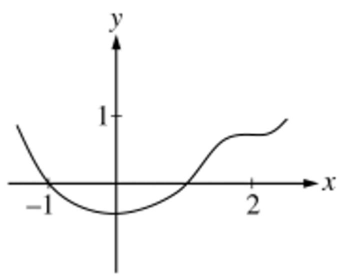
$g'(0) = 0$

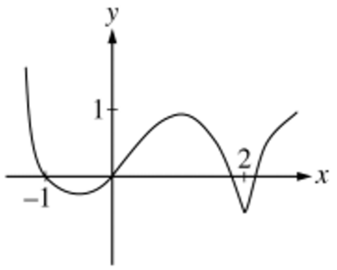
$g''(-1) > 0$

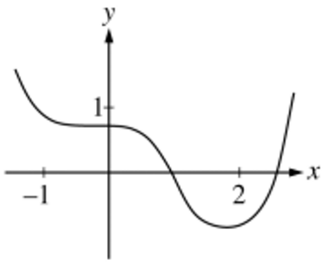
$g''(x) < 0$ if $0 < x < 2$

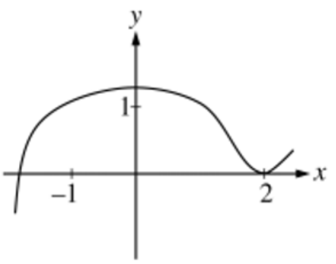
Which of the following could be part of the graph of *g*?

○ A) 

B) 


C) 


D) 


E) 


F) I don’t know

vi) $\sqrt{(x+3)^2+(y-2)^2} = \sqrt{(x-3)^2+y^2}$

In the *xy*-plane, the set of point whose coordinates satisfy the equation above is

- A) A line
- B) A circle
- C) An ellipse
- D) A parabola

- E) One branch of a hyperbola
- F) I don't know

vii) If *f* is a continuously differentiable real-valued function defined on the open interval (-1,4) such that $f(3) = 5$ and $f'(x) \geq -1$ for all x, what is the greatest possible value of $f(0)$?

- A) 3
- B) 4
- C) 5
- D) 8
- E) 11
- F) I don't know

*Message*

Thank you for completing the test. You are almost done!

Some questions about how you perceive the test follow.

8. Based on your mathematics knowledge you would characterize this test as:

- Extremely easy
- Somewhat easy
- Neither easy nor difficult
- Somewhat difficult
- Extremely difficult

9. How did you feel while you were taking the test?

| | Definitely not | Probably not | Might or might not | Probably yes | Definitely yes |
|---|---|---|---|---|---|
| Tense | ○ | ○ | ○ | ○ | ○ |
| Under pressure | ○ | ○ | ○ | ○ | ○ |
| Nervous/ jittery | ○ | ○ | ○ | ○ | ○ |
| Uneasy | ○ | ○ | ○ | ○ | ○ |
| Afraid of not doing well | ○ | ○ | ○ | ○ | ○ |
| Uncomfortable | ○ | ○ | ○ | ○ | ○ |

10.Which gender do you believe will get a higher average number of correct answers?

- ○ Men
- ○ Women
- ○ Both will have the same performance on average

*Message*

By mentioning your email address below, you participate in a lottery with a €30 voucher from Public stores. With every correct answer in the math test you increased your probability to win the price. In case you win, the voucher will be sent to the email you will provide below.

All the email addresses will be treated confidentially and ONLY for lottery purposes. After the lottery, they will be deleted.


**(This step is optional. In any case, please press the arrow to the right, to save your responses and complete the questionnaire.)**

---

*Message*

We thank you for your time spent taking this survey.

Your response has been recorded.

## Appendix II. Supplementary randomization check – Fisher's exact test

Table II: *Output of Fisher's exact test on the categorical variables used in the analysis*

| Variable | Fisher's exact test p-value |
|---|---|
| Gender | 0.918 |
| Nationality | 0.586 |
| City | 0.429 |
| Field | 0.060* |
| University | 0.594 |
| Grade | 0.782 |
| Confidence | 0.319 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

## Appendix III. Supplementary anxiety analysis

Table III.a: *Mann-Whitney U tests results comparing Control and Treatment group*

| Tested variable | Tested groups | P-value | Z |
|---|---|---|---|
| Anxiety score of female participants(n=154) | Control(n=82) & Treatment group(n=72) | 0.5409 | 0.611 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1

Table III.b: *Output of linear regression analysis of the treatment on the anxiety score, only for females*

| Variable | Coefficient | Robust SE | t-value | p-value | 95% Confidence Interval | |
|---|---|---|---|---|---|---|
| Threat | | | | | | |
| Yes | 0.075 | 0.202 | 0.37 | 0.711 | -0.326 | 0.477 |
| Nationality | | | | | | |
| Cypriot | -0.027 | 0.316 | -0.09 | 0.930 | -0.654 | 0.599 |
| City | | | | | | |
| Thessaloniki | 0.242 | 0.281 | 0.86 | 0.391 | -0.315 | 0.801 |
| Patra | -0.272 | 0.623 | -0.44 | 0.662 | -1.507 | 0.962 |
| Herakleion | -0.038 | 0.405 | -0.09 | 0.925 | -0.841 | 0.765 |

| | | | | | | |
|---|---|---|---|---|---|---|
| Larisa | -0.488 | 0.689 | -0.71 | 0.480 | -1.854 | 0.877 |
| Other city | 0.384 | 0.283 | 1.36 | 0.178 | -0.177 | 0.946 |
| Other country | 0 | (omitted) | | | | |
| | | | | | | |
| Field | | | | | | |
| Physics | -0.027 | 0.492 | -0.06 | 0.956 | -1.003 | 0.949 |
| Computer Sc | -0.069 | 0.692 | -0.10 | 0.920 | -1.442 | 1.302 |
| Civil Eng | -0.013 | 0.460 | -0.03 | 0.976 | -0.926 | 0.898 |
| Electronic Eng | -0.083 | 0.578 | -0.14 | 0.886 | -1.229 | 1.062 |
| Chemical Eng | -0.933 | 0.603 | -1.55 | 0.125 | -2.129 | 0.262 |
| Mech Eng | -0.630 | 0.474 | -1.33 | 0.186 | -1.570 | 0.309 |
| | | | | | | |
| University | | | | | | |
| University of Ioannina | -0.320 | 0.354 | -0.90 | 0.368 | -1.021 | 0.381 |
| University of Patra | 0.859 | 0.500 | 1.72 | 0.089 | -0.132 | 1.851 |
| University of Thrace | 1.581*** | 0.235 | 6.71 | 0.000 | 1.114 | 2.048 |
| | | | | | | |
| Grade | | | | | | |
| [6.5,8.49] | -0.353 | 0.301 | -1.17 | 0.244 | -0.951 | 0.244 |
| [8.5,10] | -0.935** | 0.469 | -1.99 | 0.049 | -1.865 | -0.005 |
| | | | | | | |
| Age | -0.020 | 0.038 | -0.54 | 0.590 | -0.095 | 0.054 |
| | | | | | | |
| Confidence | | | | | | |
| Somewhat disagree | 0.399 | 0.624 | 0.64 | 0.524 | -0.838 | 1.636 |
| Neither agree nor disagree | 0.461 | 0.576 | 0.80 | 0.426 | -0.681 | 1.603 |
| Somewhat agree | 0.194 | 0.584 | 0.33 | 0.740 | -0.963 | 1.352 |
| Strongly agree | 0.084 | 0.671 | 0.13 | 0.900 | -1.246 | 1.415 |
| | | | | | | |
| Constant | 0.798 | 1.242 | 0.64 | 0.522 | -1.662 | 3.259 |

*Note.* ***p.<0.01, **p<0.05, *p<0.1