Aalto University
School of Business

Gender differences in competition

Master's Thesis

Xinyuan Zhou

Aalto University School of Business

Master's programme in Economics

Spring 2019

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076 AALTO
www.aalto.fi
Abstract of master's thesis

Author Xinyuan Zhou

Title of thesis Gender differences in competition

Degree Master

Degree programme Economics

Thesis advisor(s) Tuomas Pekkarinen

Year of approval 2019 **Number of pages** 87 **Language** English

Abstract

Recently, a growing body of research explores the effects of competition preferences on wages. This thesis attempts to review important studies which look into how competitiveness affect education and labor market outcomes. I start my thesis with traditional explanations for gender pay gap. In Section 2 I provide some facts to show gender differences in real outcomes in general. Section 3 discusses Gneezy et al. (2003) and Niederle and Vesterlund (2007), which pioneer this line of researches. Then I analyze possible affecting factors by using these two methods. Even though there is only one study which compares countries directly, to my knowledge, overall it's much easier to detect gender differences for competition preferences in developed countries than in developing countries. This section is also designed to shed light on further cross-country comparison.

Section 4 lists papers which try to use this emerging theory to explain gender differences in outcomes of education and labor market. The phenomenon that women are better educated but still shy away from STEM fields could be explained by competition preferences to some extent. Moreover, I put forward policy implications for education and labor market respectively. However, the two policy recommendations need to be treated with caution as empirical evidences, especially those about quota, seems not to support conclusion drawn from experimental results.

Section 5 assesses the relative importance and unsolved issues of competition preferences, to shed light on future research. It's inspiring that this psychological factor seems to play a role outside the lab, however compare with other factors such as occupation and industry, and considering key questions such as how preferences evolve remain to be determined further studies still have a long way to go.

Section 6 concludes this master's thesis by summarizing current achievements and pointing out possible improvement for later research.

Keywords gender, competition, education, labor market

Contents

List of Figures

1. Introduction

The gender pay gap is persistent and universal despite women's advancement and advancement. According to Altonji and Blank (1999), traditional explanations for the wage gap are as following:

The human capital investment model argues that the return to skills based on expected labor force participation if these skills enhance market productivity more than non-market productivity and hence predicts that people who are going to spend less time in the labor market would reduce investment, such as related education and training, in valuable marketplace skills. Therefore, the return would be higher for full-time workers and for people who are less likely to do non-market work.

The comparative advantage theory contends that men and women are advantaged in different industries, which could lead to gender differences in the allocation of time. However, as the focus of labor market transit from physical strength to cognitive skills, now comparative advantages are less effective in explaining gender differences in labor market outcomes.

Discrimination is defined as a situation in which equal productive people are treated unequally (receive different wages or face different demands) that has to do with an observable characteristic. The main types of discrimination are taste-based discrimination and statistical discrimination. The former refers to distaste against members of minority groups, such as employer, employee and consumer discrimination. Ideally employer and employee discrimination would lead wages to converge in the long run whereas in case of consumer discrimination, wage differentials depend on customer contacts and distribution of just and prejudiced customers. The latter one is driven by imperfect information and subdivided into two strands. The first examines the effects of different prior beliefs about productivity in which members of minority groups earn less than equally productive counterpart from majority groups. The second investigates the impacts of differences in the precision of information about individual productivity. In this case, if the measure, say the test, predicts productivity more precisely for males, high-scoring males would be paid more than their female counterpart whereas the pattern is reversed among workers with low test scores.

The group preferences highlight the trade-off between work and leisure and tastes for different kinds of jobs. For instance, the compensating differentials theory maintains that jobs which are less preferable due to danger or dirty would offer a wage premium. The key

question is, as I would discuss below, how and why preferences might evolve throughout the life span, which relevant evidences are scarce.

2. Gender differences in outcomes: facts

2.1 Education

2.1.1 Educational attainment

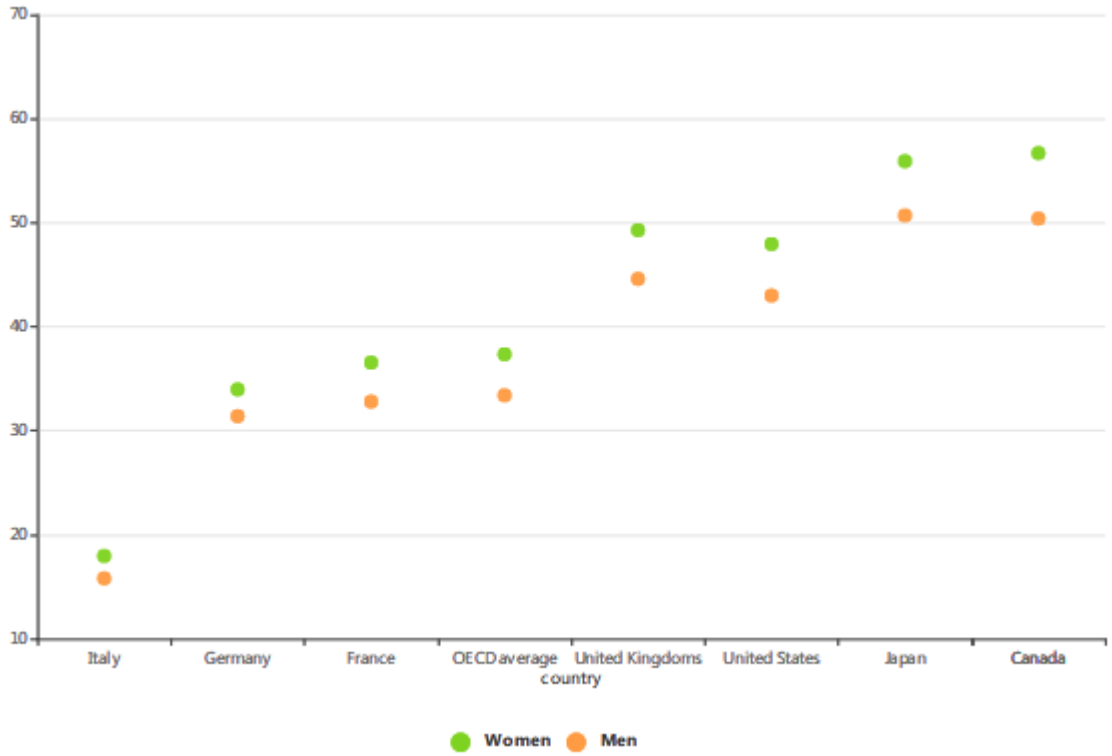

Figure 1: population who aged 25-64 and attend tertiary education[1]

In all G7 countries, women attend tertiary education more often than male counterpart. On average, women from OECD countries are more likely to receive tertiary education, too. The gender gap is even larger among the young generation, who is aged between 25 and 34 years with the exception of Germany.

[1] The data of Japan is from another category.

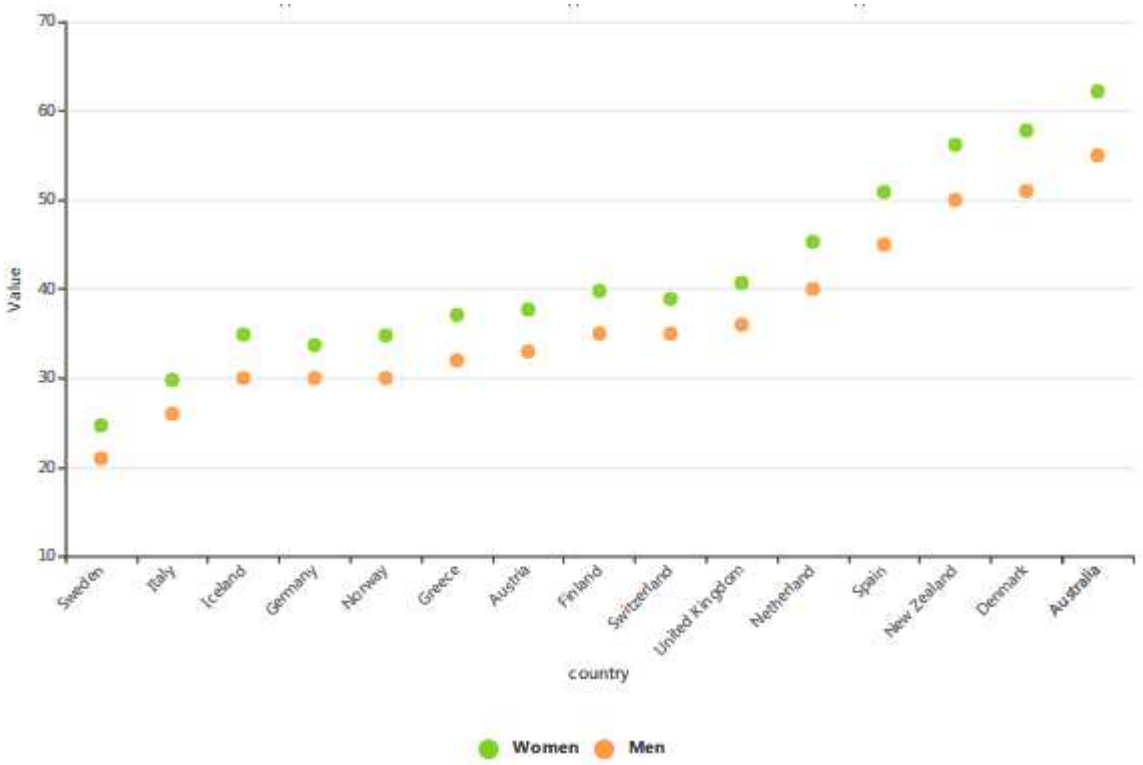

Figure 2: graduation rate in tertiary education for people younger than 30-year-old

As we can see, in all countries which have relevant data, young women have a higher overall graduation rate in tertiary education[2].

2.1.2 Academic performance

PISA is the Programme for International Student Assessment launched by OECD and held every three years. 15-year-old students would be tested in reading, mathematics and science[3]. OECD chooses 15-year-old on purpose as in majority of countries, students can decide whether to continue their study at this stage. Therefore, PISA scores of students from different countries are relatively comparable.

[2] Tertiary education in this graph stands for ISCED 2011 level 5-8, hence includes short-cycle, Bachelor's, Master's and Doctoral level.

[3] I don't include science scores for two reasons. First, compare with reading and math performance, science scores are mixed. Second, as I would discuss throughout the whole thesis, especially in section 3.2.7, the focus would be math (male-favoring task) and verbal task (female-favoring task).

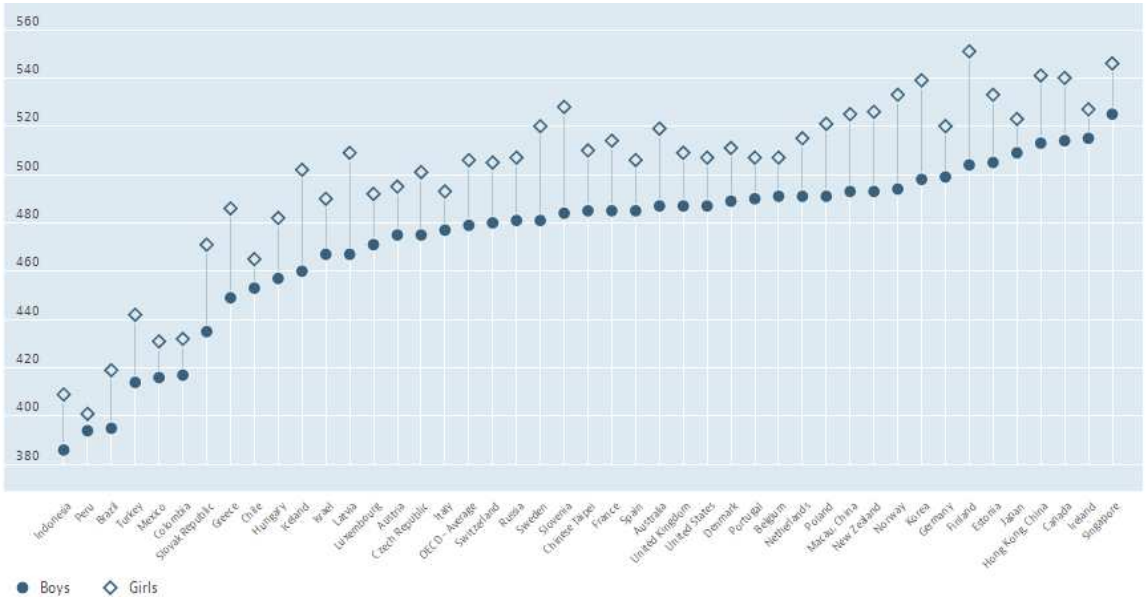

Figure 3: PISA reading performance in 2015 by sex, by country

In all countries, girls outperform boys and the female advantage is significant at 5 % level of significance.

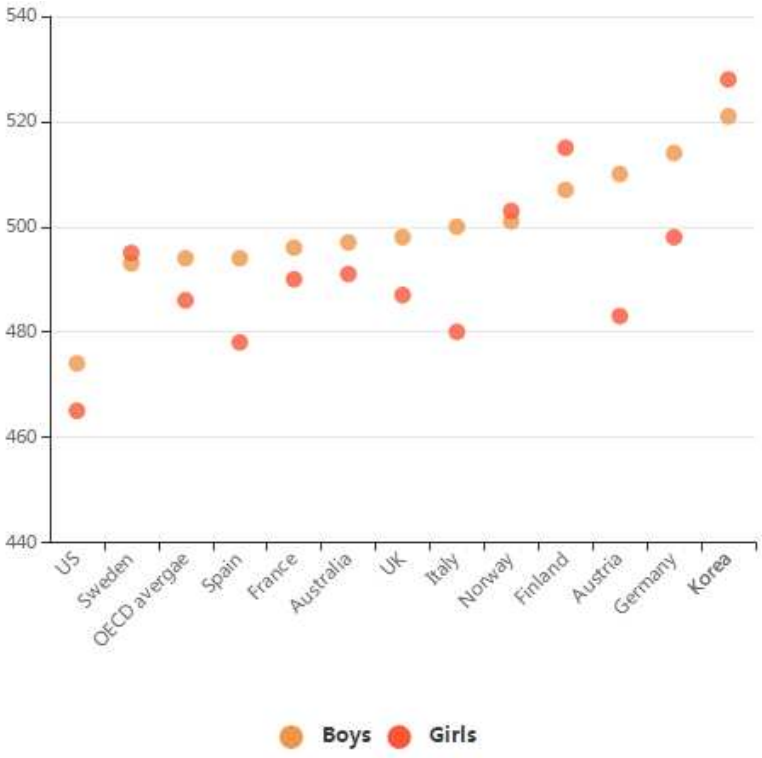

Figure 4: PISA mathematical performance in 2015 by sex, by country

In majority of countries, boys have higher scores in mathematical part of PISA with the exception of Norway, Sweden, Korea and Finland. However, only in Finland girls outperform boys at 5 % level of significance. On average, boys have significant better mathematical performance.

2.2 Employment

2.2.1 Gender wage gap

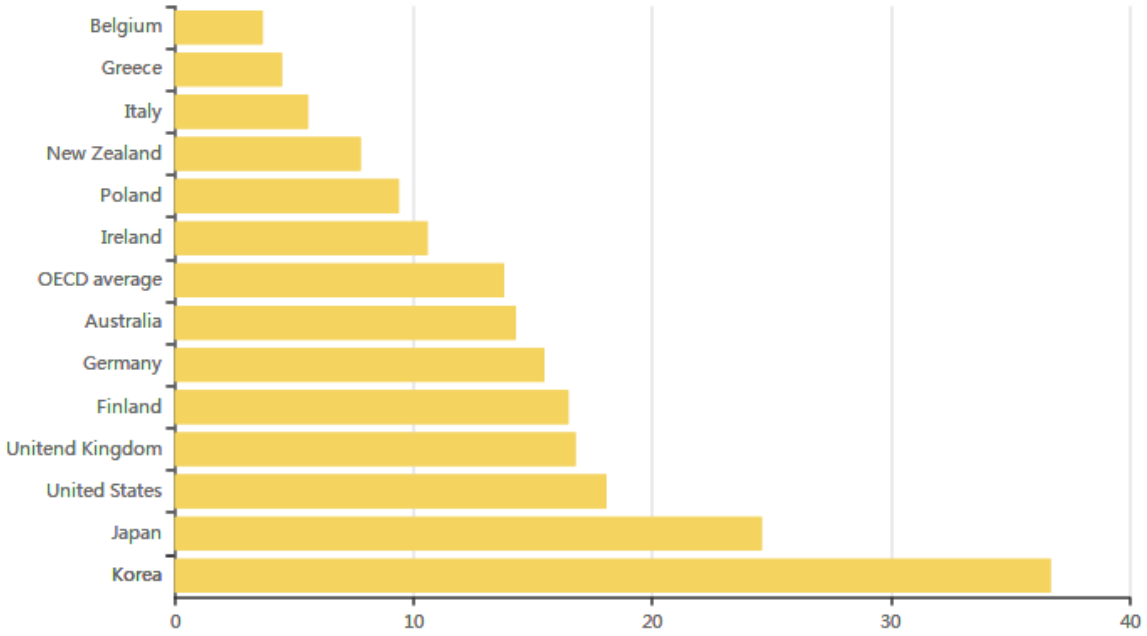

Figure 5: the difference between median earnings of men and women relative to median earnings of men in 2016 (percentage)

The gender pay gap is a universal phenomenon, ranging from 3.7 % in Belgium to the 36.7 % in Korea. On average, the gender pay gap of OECD countries is 13.8 %.

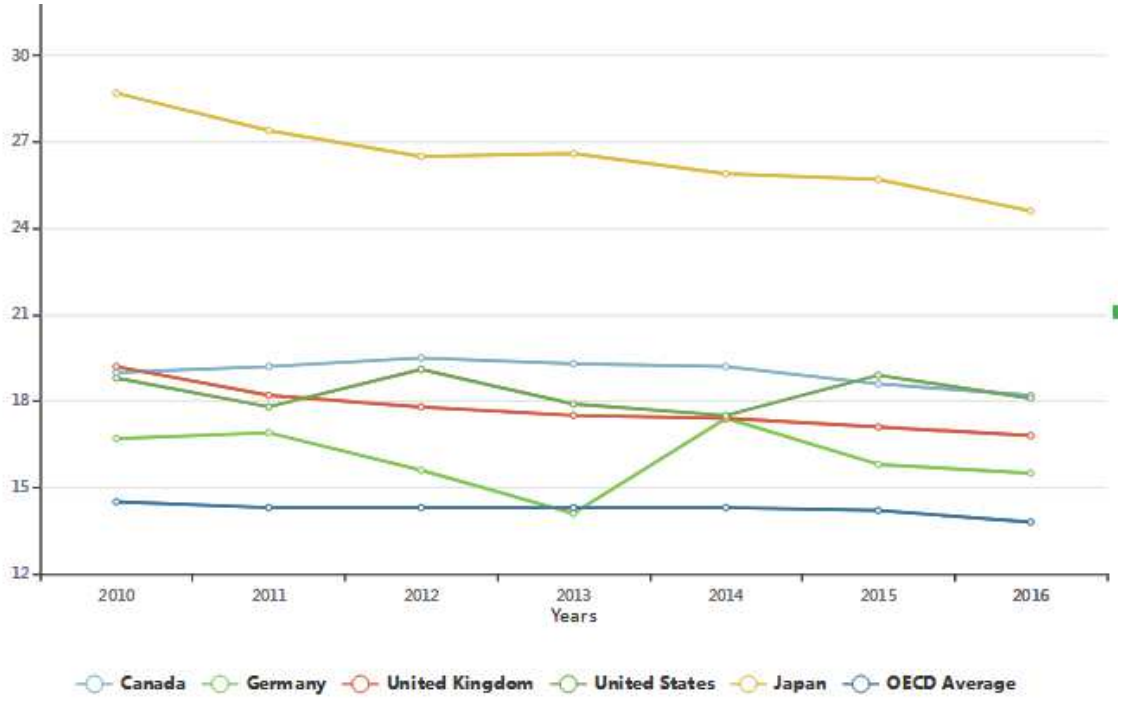

Figure 6: the difference between median earnings of men and women relative to median earnings of men in 2010-2016 (percentage)[4]

The gender pay gap is longstanding and stagnant. On average, the gender pay gap of OECD

[4] Italy and France, two member countries of G7, are excluded because data is not available.

countries is about 14 percentage points.

2.2.2 Female share of seats on boards of the largest publicly listed companies

Compare with Italy, France and Germany[5], the female share in Nordic countries evolve steadily except Iceland, which experience a sharp increase from 2011 to 2013. The historic high appears in Iceland in 2013, yet still lower than 50 %.

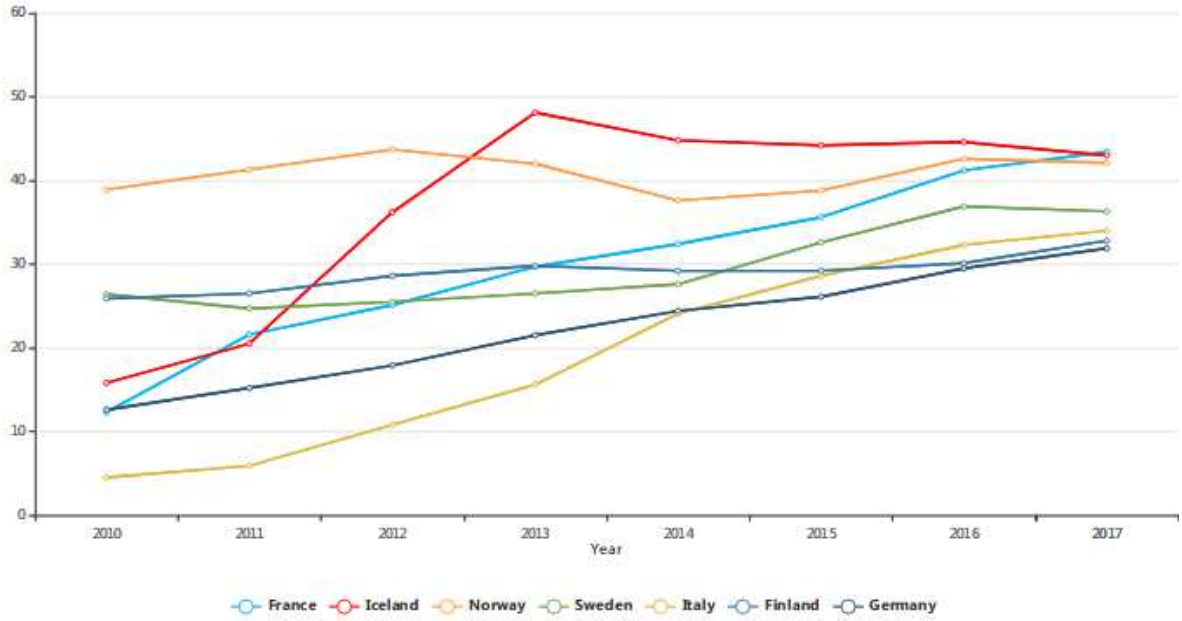

Figure 7: Female share of seats on boards of the largest publicly listed companies

3. Gender and competition

3.1 Experimental Designs

The experimental approach is now gaining popularity. Charness and Kuhn (2011) argue that laboratory experiments have several advantages. First, this setting allows labor economists to have more control over variables. It's difficult to isolate the effects of environmental factors from empirical data. Second, it enhances the ability to put a specific theoretical model in practice and enables labor economists to compare the prediction of theoretical model with the results which could be in conflict with each other even experimenters attempt to create an ideal context. Third, laboratory experiments have lower costs and are relatively easy to elicit individual beliefs. Finally, labor economists are able to investigate strategies used by agents in solving dynamic problems.

[5] At first I intend to include G7 and Nordic countries. The United States, Canada and Japan are not included due to the lack of data in 2011-2015. The United Kingdom is also excluded for simplicity as it shares a similar trend with Germany.

3.1.1 Performance in a competitive environment

Gneezy et al. (2003) is the first paper to examine gender differences in performance in a competitive environment. Participants are divided into six-people groups and required to perform the maze-solving task, which has five difficulty levels. Only in the last treatment could participants set the difficulty level. In the first three treatments, the groups are gender balanced. After students write down their scores, the experimenter would check the records to ensure accuracy. Although gender is not pointed out during the experiment, participants could see each other in the laboratory and hence could determine the gender composition of their group. Participants don't know the performance of others across treatment.

The first treatment, piece rate serves as a benchmark for gender differences in performance. In this treatment, participants could get 2 shekels for each correctly solved maze. In the second treatment, participants would perform in competitive tournament and be paid according to their performances: only the best performer could be rewarded 12 shekels for each correctly solved maze. The tournament has an uncertain payoff and payoff depends on others' performance. Thus the authors introduce the third treatment named as random pay, in which the only winner is selected randomly, to disentangle the effects of uncertainty and interdependence from the effects of competitive environments. In treatment 1 and 3, men perform better but the gender differences in performance are not significant. The gender differences in mean performances of two treatments are both 1.5 mazes. Hence the authors conclude that risk aversion doesn't affect performance. On average, men solve significantly more mazes than women in mixed tournament (4.2 mazes), which is caused by dramatic increases in men's performance and no significant changes in women's performance.

The authors list four explanations for the large gender gap in competitive environments: 1) Women don't compete against men, if, for example, women think men are more adept at solving mazes. Stereotype threat, put forward by psychology literatures, provides additional support for this explanation; 2) Women just don't compete. This might because women have higher costs of efforts, lower sensitivity to compensation schemes, and stronger aversion towards competition; 3) Men compete too much. It could due to additional utility men gain from winning or overconfidence.

In order to find out the reason, the authors conduct the following two treatments in which participants are divided into six-people single-sex groups. In the fourth treatment, single-sex tournament, the performance of men is not significantly different from that in mixed tournament and thus not affected by the gender of their competitors. By contrast,

women significantly improve their performance in a single-sex competitive environment compare with in non-competitive environment. The gender gap in mean performance is somehow higher in single sex tournaments (1.7 mazes) while not significant. The increase in women's performance could due to the compensation scheme, or to the absence of male competitors. Thus the authors use the fifth experiment, single-sex piece rate to disentangle these two effects. On average, women solve 10 mazes in the single-sex piece rate, 9.73 in the mixed piece rate, and 12.6 in single-sex tournaments. The results show that women's performance increase in single-sex tournament is due to the change in compensation scheme rather than the absence of male competitors.

Then the authors examine how men and women respond when moving from noncompetitive to competitive environments. For example, a man who solved 15 mazes in the noncompetitive treatments is expected to solve at least 17 mazes in the mixed tournament. If women receive the exactly same boost, then she would solve at least 17 mazes in the mixed tournament if she solved 15 mazes in the noncompetitive treatments. By comparing the expected performance of women with actual performance of women, the author concludes that women improve their performance significantly less than men. The results provide supports to the explanation that men and women face different competitors as men perform better across treatments. Thus, in the following analysis, the authors find that for all performance levels, a woman who solved x mazes always has a higher chance of winning the mixed tournament than a man who solved x-1 mazes in the noncompetitive treatments. Therefore, the authors assume both gender receive the same boosts and compare the performance change of woman who solved x mazes with men who solved x-1 mazes in the noncompetitive treatments. The expected mixed tournament performance of women which yield in this way, is still higher than actual performance of women in mixed tournaments. Hence, the poor performance of women in mixed tournaments is not solely because a man and a woman of same performance level behave differently in mixed tournaments as man faces an easier competition than the women.

The authors consider two explanations for the differences in behaviors. First, both genders know little about their performance but realize the potential gender gap in ability. Second, women are relatively less overconfident about their ability, compare with men. Therefore, in the final treatment, participants are allowed to set a difficulty level of mazes after solving a maze of difficulty level 2. The payoff is a function of difficulty level. The mean chosen difficulty level is 3.4 for men and 2.6 for women, the difference is significant. Thus, men are overconfident and prefer challenging tasks than women.

Finally, the authors compare mean performance across treatments. The mean performance is 12.95 in mixed tournament, and 13.47 in single-sex tournament. The difference is insignificant. Participants perform significantly better in competitive environments than in noncompetitive environments. Therefore tournaments have a strong positive effect on performance.

In order to find out gender differences in competition at young age, Gneezy et al. (2004) choose the running task instead. Participants are all fourth graders and 9-10 years old. The variable of interest is the speed in a 40 meter race. The experiment was conducted during a physical education class and the speed data is recorded by a teacher to make it like a normal practice in the class and prevent students from realizing they are taking part in an experiment. Each participants need to run twice. At first, each child runs alone. In the second round, students in the treatment group are paired according to their speed in the first round, regardless of their gender. After the match, two students run side by side. Children of the control group still run alone. Unlike Gneezy et al. (2003), in this paper, subjects know not only their own performance but also their competitors' performance which is close to theirs. They find in the first round, girls outperform boys (7.672 seconds vs. 7.693 seconds) insignificantly while in the second round, the female advantage reversed. The average running time of the control group decreases by 0.037 seconds and no significant gender gap in the improvement of performances. In contrast, the treatment group experiences a 0.081 second reduction in running time. However, after they divide improvement by gender, they find on average boys improved (-0.163 seconds) while girls actually slow down (+0.015 seconds). Boys benefit from the competition: the running time of the one who runs faster decreases by 0.182 seconds, and 0.088 seconds declination for another one. When girls are matched with other girls, they experience performance decrease. The running time increases by 0.066 seconds for the girl who performs better in the first round and 0.0333 seconds for the girl who performs worse in the first round. Gneezy et al. (2004) attribute the opposite findings to the peer pressure from the rest of children.

There're several differences between these two papers. In Gneezy et al. (2004), students are not aware of the field study and younger than college students, which are subjects of Gneezy et al. (2003). Furthermore, children have information about own and others' performance while college students know nothing. Last but not least, college students receive monetary reward based on their show-up and performance while children only receive intrinsic motivation.

To summarize, both Gneezy et al. (2003) and Gneezy et al. (2004) support the statement that competition settings have positive effects on the overall performance, and the performance increase is larger for males.

3.1.2 Tournament entry decision

Niederle and Vesterlund (2007) use tournament entry, instead. Participants are divided into different groups, consist of 2 male and 2 female. The task is to add up sets of five two-digit numbers for five minutes without calculator. In the first two tasks, participants perform this addition task in the context of noncompetitive piece rate and competitive tournament, the only difference is the compensation scheme. In the noncompetitive piece rate, participants can get 50 cents for each correct answer while in the competitive tournament, only the best performer could be rewarded for 2 dollars per correct answer. Participants are then asked to select which of these two compensation schemes they would like to apply to their next performance. In task 3, participants who choose piece rate earn 50 cents per correctly solved problem. If they choose tournament and only if their performance in task-3 is better than the performance of their group members in task-2, they would be rewarded 2 dollars for each correct answer. This has several advantages: first, the performance of a player who enters the tournament is evaluated against the performance of other participants who also performed under tournament. Second, beliefs regarding the choices of others won't affect the entry decision. Third, the entry decision is an individual decision problem and won't impose negative externality on others since it compares with the prior performance of competitors. Therefore gender differences in social preferences won't cause gender gaps in tournament entry.

They put forward four possible explanations for gender differences in tournament entry among equally able participants: 1) Men prefer competition; 2) Men are more confident; 3) Men are more risk prone; 4) Men are less feedback aversion. Since individuals could deduce their relative performance according to the monetary reward, how they think about feedbacks provided by the payoffs could affect the entry decision. Niederle and Vesterlund (2007) consider the first explanation as preferences for performing in a competitive environment and the remaining three explanations as general factors. The question of interest is to what extent competition preferences and general factors, separately, could explain the gender differences in tournament entry. Therefore, in order to isolate the effect of general factors from that of competition preferences, they develop task 4: participants choose one compensation scheme that they would like to apply to their past piece-rate

performance in task 1 and don't have to perform. The decision does not affect the earnings of any other participant, nor does it depend on the entry decisions of others. The decision of task 4 only affected by general factors since participants don't need to do the task. Finally, participants are asked to guess their ranks in task 1 and 2.

The mean performance is 10.15 problems for women and 10.68 for men in the piece rate. In the tournament, it's 11.8 for women and 12.1 for men. In both tasks, the gender differences in performances and in performance improvement don't reach significance. The authors attribute this performance increase to learning or tournament incentives. The chance of winning the tournament is 26 % for a male participant and 24 % for a female participant after controlling for gender only. The probability of winning is similar for each gender if they are equally able. Hence the authors don't expect gender gaps in propensity of tournament entry.

According to the compensation schemes of task 3, a risk neutral agent would choose to enter the tournament if the chance of winning is more than 25 %. If in task 3, participants behave exactly like what they perform in task 2, then this corresponds to participants who solve more than 13 problems correctly. Hence, 30 % of the women and 30 % of the men should enter. If include people who are indifferent between these two options, that is, people who solve 13 problems correctly, then 40 % of women and 45 % of men should enter.

However, the actual entry ratio is strike and significantly different from the expected ratio. 35 % of women and 73 % of men opt for the tournament. Among women who choose piece rate and tournament, there's no performance difference in the Task-1, Task-2 or in improvement when move from Task-1 to Task-2. The performance in Task-2 is marginally higher for men who select the tournament ($p = 0.14$). When only gender, performance and performance increase are controlled, the results show gender has a significant and substantial effect. Being a woman reduce the probability of entering the tournament by 38 %, ceteris paribus. The performance in Task-3 parallels performance in Task-2 and can't explain the substantial gender gap. Participants who enter the tournament don't have better past performance or improve their performance.

Then the authors look into four quintiles according to their intergroup ranks in Task-2 and Task-3 and observe for each group, the performance have limited effect. For every performance level, men are more likely to enter the tournament. Surprisingly, the entry probability of the top 25 % of female participants is even lower than that of the bottom 25 % of male participants. Risk attitudes alone are unable to explain the entry gap. As

mentioned above, subjects who solve more than 13 problems should enter the tournament as the payoff is higher. Of those participants, about two-thirds of women and only one fourth of men shy away from the competition. Of participants who solve less than 12 problems, about 61 % of men and 29 % of women embrace the competition.

There is no doubt that the over-entry of low-performing participants and under-entry of high-performing generate costs. After ignoring performance costs and assuming that there's no correlation between performance and the chosen incentive scheme, the authors argue that the cost of under-entry is higher for women (99.4 dollars vs. 34.5 dollars), while the cost of over-entry is higher for men (32.9 vs. 56.5). Compare with over-entry, under-entry has higher costs which are not surprising as under-entry profits are forgone by high-performing participants. Women bear much higher costs. Calculations based on performance in Task-2 show that in total, the costs are 132.3 for women and 91.0 for men. Replace it with performance in Task-3 doesn't alter the result (113.5 vs. 93.3).

Both gender are overconfident as they expect themselves to have better ranks in Task-2, men are more overconfident however. Three-fourths of men and 43 % of women think they are the best performer of their teams. Women are significantly less confident than men ($p = 0.016$). Including the guessed rank into regression close 27 % of the gender gap.

In Task-4, subjects are required to choose one of the incentive schemes to apply to their performance in Task-1. In the Task-1, the chance of being the best performer is 29 % for a man and 21 % for a woman. Individuals who solve more than 11 problems should select the tournament, which refer to 30 % of women and 40 % of men. If individuals who are indifferent between these two inventive schemes are included, then 40 % of women and 45 % of men should choose the tournament. However, actually 55 % of men and only 25 % of women decide to apply the tournament to their performance in Task-1. Again, there's no performance difference between women who choose the tournament and women who choose the piece rate. Men who choose the tournament have significantly better performance than the rest who choose piece rate (12.05 vs. 9, $p = 0.004$). There is a positive association between men's performance and the chosen incentive scheme, but not for women. The largest gender gap appears among participants who solve the most problem correctly. For participants who should choose the tournament, 75 % of women and only 12.5 % of men made the wrong decision. For participants who have higher expected payoffs from the piece rate, the likelihoods of making wrong decision for both genders are similar. As before, both women and men are overconfident about their ranks in Task-1 and men are significantly more confident. Confidence has a positive effect on the probability of

selecting the tournament. The gender gap in tournament entry in Task-4 is insignificant when performance and guessed rank in Task-1 and female dummy are included. Therefore, the authors argue that risk attitudes and feedback aversion only play negligible roles.

Controlling for the dummy which indicate the participant choose the piece rate in Task-4 or not, the authors find that subjects who select the tournament in Task-4 are significantly more likely to enter the tournament in Task-3, but the gender gap remains. Being a woman reduce the odd of choosing the tournament by 16 %.

The authors allege that the effects of gender on competition should be treated with caution. First, factors such as risk attitudes and feedback aversion may have greater impact when people need to compete with others. Thus Task-4 may only capture part of effects and the results overestimate the importance of competition preference. Second, result would be upward biased if men are more optimistic about their future performance. The guessed ranks not only demonstrate how confident the participant is, but also show how the participant perceives past success/failure. If women attribute their past success to luck rather than individual ability while men do the opposite, the study would exaggerate the impact of willingness to compete.

To summarize, 57 % of the original gender gap can be explained by general factors (confidence, risk attitudes and feedback aversion). Competition preferences, which means men and women have different preferences when perform in a competitive environment, contribute to the residual 43 % of the gap. The low probability of choosing to compete not only reduces the number of women who enter tournaments, but also the number of female winners in the competition.

3.1.3 Link two approaches

Niederle (2011) argue that both approach describe similar phenomena, especially if we consider high performance in a competitive environment as the decision to work hard. Hence, the two approaches would lead to the same conclusion that men are more competitive than women. However, Niederle and Vesterlund (2007) select a task in which performance is independent of the chosen compensation scheme. Further studies should build connection between these two approaches and assess for women whose performance in a competitive context does not fully reflect their piece-rate performance, would they avoid competition.

Saccardo et al. (2018) introduce a new procedure in order to measure the size of competitiveness. The authors argue that the tournament entry could be seen as extensive

margin, while to what extent the earnings depend on tournament could be viewed as intensive margin and then examine them separately. In the extensive margin treatment, participants who choose piece rate would earn 1 dollar for each successful toss. For participants who choose the tournament, he would be paired and only if he outperforms another opponent could he earn 3 dollars for each successful toss. In the intensive margin treatment, participants need to pin down the composition of their payoffs: the participant would receive 100 points and assign t points to the tournament. If the participant wins the tournament, then the payoff would be (100-t+3t) per successful toss; if the participant loses, the monetary reward for each successful toss is 100-t. The greater share the participant invests in the tournament, the more competitive the participant is. Hence this original design enables the authors to determine 101 levels of competitiveness rather than just the binary measure. In both treatments, decisions need to be made before performing the task and then perform ball-tossing tasks ten times.

In the extensive margin treatment, in line with Gneezy et al. (2003), men outperform women in the competitive context. 32.4 % of female participants enter the tournament while the number is 78.2 % for men, a substantial and significant gender gap. Without any controls, women are 36.4 % less likely to participate in the tournament. Adding the variable which indicates the sex ratio widens the gap to 39.9 %. In line with Niederle and Vesterlund (2007), each gender is overconfident and men are more confident than women about their expected successful toss (5.79 vs. 4.51) and the expected probability of winning (63.13 % vs. 43.29 %). Adding the expected chance of winning to the model decreases the gender gap by 9 % while the expected successful toss and risk attitudes only have negligible effects.

In the intensive margin, on average, 50.11 points are allocated to the tournament. Women devote 35.27 points whereas men devote 66.43 points, the allocation decision significantly differs by gender. The cumulative distribution function depicts the gender differences in allocate decision more evident: the distribution for male participants is visibly shifted to the right. The most competitive 25 % of female participants invest over 50 points in the tournament, while only the least competitive 25 % of male participants invest fewer than 50 points. The median points are 70 for male participants however the most competitive woman only invests 80 points. Without any controls, female participants allocate 31.15 points fewer than men. Adding the sex ratio to the model doesn't alter the result, which suggests gender composition affects both gender similarly. Still, both men and women are overconfident and men are significantly more confident about their expected performance

(5.75 vs. 4.28) and the expected probability of winning (65.15 % vs. 43.70 %). Adding the expected chance of winning to the model decreases the gender gap in allocation decision by 14.66 points, which closes 45 % of the gap. However, the gender gap remains significant. The expected performance and risk attitudes only have negligible effects.

Then the authors compare extensive margin with intensive margin. For the extensive margin treatment, the female-to-male ratio means divide the ratio of female participants who choose the tournament out of all women by the ratio of males who choose the tournament out of all men. For the intensive margin treatment, the female-to-male ratio corresponds to the ratio of female participants to male participants who allocate more than certain points. The authors observe that among participants whose allocated points are fewer than median points, the female-to-male ratio is almost same in two treatments, which is 0.41 to 1. The upper tail of the distribution however differs. The women-to-men ratio reduces to 0.09 among the most competitive 25% of participants and become 0 at the top 10% of participants because they are all men. There are correlations between confidence and risk attitudes and the likelihoods of the subject among the most competitive 25% and confidence alone could explain whether individuals are in the least competitive 25% of the sample. However, risk attitudes and confidence are unable to account for the gender gap of the whole distribution.

From the threshold of tournament entry perspective, 32.4% of women participate in the tournament in the extensive margin treatment while about 32% of women assign over 45 points to the tournament, hence the authors consider 45 points as the cutoff points for women to enter the tournament. Similarly, the cutoff point for men is 40 points as in the extensive margin treatment, approximately 78% of men enter the tournament while 77.5% of men assign over 40 points to the tournament. Therefore, both gender face similar cutoff point when make the entry decision and gender differences in competition preference are driven by the different distribution of competitiveness level.

To conclude, the in-or-out tournament entry dummy fails to capture the competitiveness level among highly competitive people. The linear measure (intensive margin), instead, shows that the ratio of women among the most competitive participants is even smaller than what is estimated by the extensive margin treatment. If candidates for well-paid jobs are mainly drawn from the pool of highly competitive individuals, then the underrepresentation at the top of the distribution could explain the dramatic gender gap in the real world to some degree. This gap could emerge at the start of career path. Therefore, the novel measure not only offers a more precise method for measuring the intensity of

competitiveness level and for estimating the magnitude of the gender differences among different percentiles, but also allows for deeper investigation into gender pay gap in real world. The methodological advantage is that it requires fewer subjects for a determined power level to reject a null hypothesis.

3.2 Affecting Factors

By analyzing possible affecting factors, we could have a better understanding of competition preferences, how it works, would it be nature or nurture, if there are gender differences when competition preferences come into effect and hence put forward policies aimed at closing the gender gap. Also, would institutions, both in the real world and in the lab itself, has an impact on competitiveness?

3.2.1 Biological Factors

Buser (2012) explores the role played by hormone in competition preferences. The 28-day menstrual cycle is categorized into five phases and therefore subjects are assigned to one of five groups according to the phase they are experiencing. The author states that that the hormone level and the use of hormonal contraceptives have strongly and significantly negative effects on women's entry decisions but not on confidence, risk attitudes and performance in a competitive environment.

In contrast, Wozniak et al. (2014) find women whose hormone levels are high enter the tournament more often than those who have low hormone levels. Except differences in experimental designs and participants[6], Buser (2012) argues that in Wozniak et al. (2014), entry decisions may depend on factors which are related to hormone levels but have nothing to do with competition preferences.

3.2.2 Culture

Gneezy et al. (2009) compare the Maasai tribe in Tanzania, which is patriarchal and the Khasi tribe in India, which is a matrilineal society. Gneezy et al. (2009) classify the Khasi as a female-dominated society and list several features: 1) Inheritance and clan membership always follow the female lineage through the youngest daughter; 2) Women usually live in separate households which next to their families, thus they never join their husbands' households; 3) Women make the majority of household decisions while men assume

[6] Subjects in Buser (2012) are all women while subjects perform in a mixed-sex context in Wozinak et al. (2014).

responsibilities that similar to those of women in male-dominated societies; 4) Families could keep the investment of raising girls in their households thus have the freedom to raise the daughter at will. Contrary to Khasi, the status of Maasai men depends on their ages and the number of cattle they have. The most common type of marriage is polygamy.

In the experimental task, subjects are required to throw a tennis ball into a bucket which is 3 meters away ten times. The authors argue the unfamiliarity toward tasks guarantees that traditional skills won't lead to male advantages, thus results could provide insights into competitive inclinations. Participants are divided into two groups and know they would compete with a person from another group simultaneously. However, participants don't know the identity of their competitors. Before performing the task, participants need to pin down the compensation schemes, whether to be paid by piece rate or by tournament.

Results show there is no gender gap in success rates in each society. About 50 % of subjects in the Khasi decide to compete, while the number is 39 % in the Maasai. As expected, male in the Maasai are more likely to opt for competition as 50 % of men and 26 % of women choose to competition. In the Khasi, 54 % of female participants select to compete, which is even slightly higher than the ratio of Maasai men and 39 % of Khasi men made the same decision. The authors rule out the possibility of biological evolution and attribute the gender gaps to the culture.

Even though participants don't know the gender of their competitors, they can deduce by observing the gender composition of their own group. Hence, the authors use four ways to control the effect of gender composition. First, the author includes a dummy variable (male = 1) that represent the gender of the one who stands in front of the person. The second variable is the arithmetic average of the gender of two participants who stands in front of and behind the subjects. The third variable is about the 4 nearest subjects and the fourth is about the 8 nearest subjects. The fifth is about all others in the group. They argue that if subjects make deduction according to the gender composition of their own group, then, overall, women are inclined to compete against other women, especially in the Khasi, the matrilineal societies.

Anderson et al. (2013) follow the experimental design developed by Gneezy et al.(2009) closely and replicate the results among 13-15 years old kids. They recruit subjects who live in Meghalaya, India but belong to different tribes to eliminate possible biological factors such as genetic distance. They don't observe gender differences in competition in the matrilineal society while girls in patriarchal society show a significantly strong aversion toward competition (67 % vs. 19 %).

Based on papers list above, we can see that culture does shape people's competitiveness to some extent. Moreover, it seems to be straightforward to draw the conclusion that more gender equal society would have smaller gender differences in competition. However, some papers, in which results are counterintuitive, deserve more attention.

Cárdenas et al. (2012) compare Swedish children with Colombian children of same age, and surprisingly, they find that Colombian boys and girls are equally competitive across tasks and mixed results in Sweden, despite of its much higher gender equality index. There are in total four tasks: running, skipping ropes, math and verbal tasks. Competitiveness when perform running and skipping ropes tasks are measured by performance change, while in math and verbal tasks is additionally measured by tournament entry. Swedish girls experience greater improvements in rope-skipping and math tasks. Speaking of choices, 44 % of Swedish boys and only 19 % of Swedish girls select to compete in math, difference is significant at 1 % level, whereas the entry rates of verbal task are 39 % for Swedish boys and 27 % for Swedish girls ($p = 0.045$). The gender of opponent only plays a role when Colombian girls and Swedish boys perform running tasks, both of them benefit from female opponents. They explore the attitudes towards these tasks, the scores range from 0, completely feminine to 10, completely masculine. Generally, children in both countries consider math and running tasks are significantly more masculine whereas rope-skipping and verbal tasks are more girlish. They also ask student how important it is to compete against boys and girls. In both countries, boys attach much more value to competition as than girls. In Colombia, both girls and boys believe that it is more important to compete against a boy than against a girl. Girls in Sweden rate competing against a boy as being more important compared to competing against a girl, whereas boys are indifferent to whom they are going to compete with.

Dariel et al.(2017) focus on the a minority group, Emirati in the United Arab Emirates (UAE henceforth), a traditionally male-dominated society but currently experiencing a transition as it has enforced a number of policies to empower women recently. Therefore, their sample is composed of college students who were kids when some of supportive policies were introduced and the average age is 22.5-year-old. Participants know their individual performance and whether they would compete in a single-sex or a mixed group. Overall there are gender differences in performance across tasks: men solve 36.1 % more problems in the piece rate task, 38.1 % more in the tournament task, 22.0 % more in the choice task, all significant at p value < 0.001. The gender composition impacts performance in neither task 1 nor task 2. Only in task 3, when participants are able to

choose incentive schemes, men in mixed group significantly outperform men in single-sex group (17.63 vs 15.00). The raw female tournament entry rate is 4 % higher than that of male (54.3 % vs. 50.0 %) while insignificant. The gap widens and reaches significance once control for tournament performance, performance increase, risk aversion and overconfidence. Women behave similarly no matter what kind of groups they are assigned to. Contrarily, men in single-sex group are 12.6 % less likely to enter tournament than men in mixed-group.

Zhang (2019) compare ethnic groups live in one Chinese county: Patriarchal Yi and Matrilineal Mosuo. Even there's no significant gender difference in the likelihood of winning the tournament for the Yi and the Mosuo, the author find substantial gender differences in raw tournament entry rates. The Yi men are 23 % more likely and the Mosuo men are 28 % more likely to enter the tournament.

3.2.3 Institutions

Both Dariel et al.(2017) and Zhang (2019) adopt the approach put forward by Niederle and Vesterlund (2007) but fail to find competition distaste for female who live in patriarchal society. Dariel et al. (2017) state that results may be partly driven by institutional changes happened in recent years. Zhang (2019) provide supports for the argument. If institutions help explain gender differences in competition, then this offers a prospective for different results in developing and advanced countries, since developing countries are still at an early stage of development and more politically volatile.

From 1949, China launched several institutional reforms, which is characterized by covering broadly, from economy to education, from employment to social norms. A series of reforms set the minimum age for marriage, grant women the right to divorce, extricate women from unpaid household production and enable them to enter labor markets, etc. On the one hand, Han Chinese, the ethnic majority are exposed to reforms. On the other hand, due to political reasons, ethnic minority groups who live in mainland China and work in non-public sectors were exempted from the reforms starts in 1949, even though they experienced institutional changes several decades later while Taiwanese are exempted from the communist reforms completely and still exposed to traditional Confucian ideology. Hence, China serves as a good example to provide insights about the effects of institutions.

Zhang (2019) look into three different ethnicities: patriarchal Han and Yi and matrilineal Mosuo. Mosuo child automatically became a member of the mother's family. Unlike the Khasi bribe, the head of household could be male but each family member has control over

his or her labor income and is able to join the important household decisions. Mosuo society is a monogamous society, which differs from traditional Han and Yi society. The author observe no gender differences in performance for Yi and Mosuo, and Han Chinese men outperform only marginally significant at 13% level. Also, when move from piece-rate to tournament, there are no gender or ethnic differences in performance changes. To be more specific, all three ethnicities don't improve their performance under the competitive context. Without any controls, men from all 3 groups are more inclined to enter the tournament. The raw gender gap in tournament entry rates are 15% for Han, 23% for Yi and 28% for the Mosuo. The gap for the Yi and the Mosuo persists after controlling for overconfidence and risk attitudes, while the gap for the Han lose its significance after condition on overconfidence.

The author contends that differences in competition aren't caused by socio-economic status, educational level or one child policy. Instead, the difference in the gender gap in competitiveness between the Han Chinese and the Yi is mainly due to national marriage law, and particularly the enacted greater marriage autonomy and delayed marriage. Mosuo women compete as equally as Han Chinese woman, which indicates that culture doesn't generate an additional effect compare with compulsory regulation, whereas Mosuo men enter the tournament significantly more than the Han Chinese men, possibly because Mosuo has a unique way of marriages and Mosuo male enjoy higher returns to competitive behavior.

Booth et al. (2019) not only compare Taiwanese with mainland Chinese, but also compare people of different generations. In 1978, China started to transit from central planning to market economy and Marxist ideology fade out due to the immeasurable damage caused in the Cultural Revolution (1966-1976) and the influx of individualistic and money-oriented ideology. The Cultural Revolution put a lot of emphasis on gender equality, shown in the high frequency of relevant words that appeared in 'People's Daily', the official CCP[7] newspaper. The treatment group consists of three groups: (1) individuals born in 1958 thus receive primary education during the Cultural Revolution complete; (2) individuals born in 1966 thus are partially exposed to the Cultural Revolution and the market economy; (3) individuals born in 1977, thus grow up in an economic reform era. As expected, Chinese women show stronger willingness to compete than Taiwanese women. Chinese women who are born in 1958 are more competitive compare with male of same age and the two younger groups. They don't find gender or generation gaps in competiveness among

[7] Chinese Communist Party.

Taiwanese.

There're concerns that Taiwanese and minority groups who work in non-public sectors are more or less affected by the reform, which could be true especially for minority groups since there are no visible segregation. The problem is less severe when compare with Taiwanese and Chinese due to geographic location[8] and political factors[9].

3.2.4 Socioeconomic status

Almas et al. (2015) conduct the Niederle and Vesterlund (2007) experiment among 9th grade Norwegian students. In general, boys are much more willing to select to competition (51.6 % vs. 32.2 %). Interestingly, there are no gender gap in confidence, risk preference, social preference and family backgrounds. A low SES family is defined as the family that is in the bottom 20 % of both the income and education distributions in the sample. Only 23.1 % of children from Low SES family enter the tournament while the number for children from medium and high SES family is 46.9 %. Controlling for performance closes 24.2 % of the entry gap.

Family background has a strong and highly significant negative impact on the competition preferences of boys from low SES families but not for girls. The coefficient on boys is 26.6 % and 2.4 % on girls. The effects on boys remain even take confidence, probability of winning, social preferences, personality traits and risk attitudes into account. The negative effects are due to the strong association between father's SES and sons' competition preferences. The possible explanation maybe that medium and high SES fathers spend more time with their sons during their crucial developmental age and, meantime, serve as role models.

Findings of Almas et al. (2015) could explain the counterintuitive results of Cárdenas et al. (2012), that the gender differences in Colombian children are smaller than Swedish children, to some degrees. In the Scandinavian countries, fathers are allowed and encouraged to spend more time and engage more in parenting. In the more gender equal societies, the father plays a more important role in the offspring personality traits. The gender gap in competitiveness only appear among children from medium and high SES families, which shed lights on papers which conduct studies in developing countries and

[8] As Taiwan is an inland, it's much harder for interaction and communication.

[9] On 19 May 1949, the governors of Taiwan announce the Taiwan marital law would be imposed from 20 May 1949. The law strictly forbids and severely punishes the spread of communism-related information and any kinds of cross-traits exchange, particularly direct links of mail, air and shipping services, and trade. The law is abolished in 15 July 1987 but the earliest communication starts from 1994. Therefore Taiwanese samples in Booth et al. (2019) are expected to be unaffected by the reform almost completely.

usually observe no gender differences.

Bartling et al. (2012) find a similar pattern among German preschoolers. Participants are asked to throw toy frogs into a toy pond and opt to be paid by piece rate or tournament. However, in this analysis, family background comes into effect via health issues, which strongly and negatively impacts the willingness to compete of boys from low SES family whereas doesn't kick in for boys from high SES families.

By contrast, Khadjavi et al. (2018) find a reverse association among 3-6 years old German children. Preschoolers are required to perform running task but have the freedom to choose who he/she is going to compete with. If the participant chooses to compete with others, then he would win a large reward plus a small reward or nothing. If the participant select self-competition, he would get a large reward if he succeed or a small reward if he loses. Results show that children from low SES family are more willing to compete, which is due to parents' ambition. Parents' ambition is measured by the extent they agree that "success is based on hard work" and the value they attached to the future job (both professional and sports) success of their kids. Ambitious parents, however, earn less and have lower education levels. Children of ambitious parents choose to compete even though the odd of winning is thin and actually better off if they change their decisions. In this analysis, neither age nor gender influences the competition decisions. Indeed, children whose parents are less ambitious obtain more rewards.

At first glance, the results of Almas et al. (2015) and Khadjavi et al. (2018) contradict with each other. However, the findings of Almas et al. (2015) could be seen as a consequence of findings of Khadjavi et al.(2018): low SES children overinvest in competition which result in a waste of effort in the early stage. The past failure may be detrimental in the long run and hence lead children of low SES families avoid competition after they grow up.

3.2.5 Age

Flory et al. (2018) explore how competitiveness level changes with age. They simulate the four-round experimental design of Niederle and Vesterlund (2007) but the task is modified as to arrange shapes in a row according to the size. Participants have no idea about other competitors in their groups. They first conduct the experiment in Malawi, a sub-Saharan African country. They recruit more than 700 subjects which span a broad age range and are gender-balanced. They include a dummy intentionally which indicate whether the subject is younger or older than 50-year-old, as in general women at this age have experienced menopause symptoms and hormone may have effects on competitiveness (Wozniak et al.,

2014; Buser 2012). The raw gender difference in tournament entry of the whole Malawi sample is significant but relatively small. After controlling for age and gender together, they find the gender gap is driven by the youth, who have a similar gap like subjects of other papers, and disappears in the elderly. On the one hand, older men are not significantly different from younger men, whereas older women are significantly more competitive than younger men and have similar competition preferences as older and younger men. Being female could reduce the likelihoods of tournament entry by 10.5 percentage points while being age 50 or older is strong enough to close the gender gap in competitiveness(the coefficient: +0.141). On the other hand, for female subjects, risk preferences or feedback aversion doesn't increase with age. Later on, they conduct the same experiment in the United States intended to test external validity and find similar results except competition preference is not significant in the full sample.

It's equally important to figure out when the gender gap emerges which could help researchers get a better grasp of the source and provide a guide for policymakers when devise policies.

Sutter et al. (2010) examine this problem among 1035 3-18 years old Austrian. 717 of them who are over 8 are required to perform the math tasks and the younger children would perform running tasks. They find that across the whole age spectrum, boys are more competitive, measured by either performance increase or tournament entry, than girls even condition on risk aversion and confidence, hence summarize that gender gap emerges at an early stage of life, as young as 3-year-old.

By contrast, Khachatryan et al. (2015) investigate the gender gap in Armenia, a form Soviet Union country. Participants are 7-15 years old and asked to perform running, rope-skipping, math and verbal task. Armenia is still dominated by matrilineal kinship system despite of the encouragement and support for women when Armenia belongs to Soviet Union. Results show that boys and girls compete equally in rope-skipping, math and verbal tasks in the terms of performance changes while girls improve more than boys did in the running task. There are no gender differences in tournament entry when students are offered two payment options in math and verbal tasks. They choose grade 7 specifically as children of this age hit their puberty. They don't find any associations between the competitiveness level and age.

Andersen et al. (2013) examine the gender gap among 7-15 years old children in matrilineal and patriarchal villages in northeast India. They consider age 13 as the starting age of puberty. In both contexts, among children who aged 7-12, boys are more competitive

than girls (patriarchal: 2 %, matrilineal: 6 %), but neither gender nor society difference is significant. However, gender differences emerge around puberty. On the one hand, older boys who live in patriarchal villages become more competitive (increase from 46 % to 67 %) while older girls become less competitive (decrease from 44 % to 19 %), a gender gap which is similar to the commonly found gap in adults and hence may reflect the mixed effects of culture and biological changes. On the other hand, boys and girls from matrilineal society compete equally across age despite the reversed gender gap (9 %) among children who are over 13. Older girls in matrilineal society exhibit a significantly higher propensity to compete than their female counterpart in patriarchal society (50 % vs. 19 %). Results are still robust when replace the age 13 with age 12 or 14. Unlike Khachatryan et al. (2015), Andersen et al. (2013) suggest that policies which aim to close the gender gap should target children around the age of puberty to generate the most effective effects.

However, studies which compare the competitiveness level of people at different ages are subject to several problems. First, life experiences help shape competition preference. Hence the effects of age may be overestimated. Second, the potential selection effects which caused by participants.

3.2.6 Personality

As competition preference is one kind of personality traits, there are some papers talk about the relationship between personality and competitiveness.

Bartling et al. (2009) find a significant negative association between egalitarian preferences and competition preferences. Egalitarian preferences are motivated by behindness aversion (aversion to negative payoff inequality) and aheadness aversion (aversion to positive payoff inequality).

Social preferences are elicited by four simple binary choices which is a combination of the participant's income and the opponent's income. For example, in the pro-sociality game, participants need to choose between (10, 10) and (10, 6), the former number represents monetary incentives for the participant while the latter is for the opponent. The options in the costly pro-sociality game are (10, 10) and (16, 4). Therefore, in this two game, participants need to select between egalitarian and egotistic decisions. However, compare with pro-sociality game, the costly pro-sociality impose financial burden on participants if he/she prefer equal pay. Subjects who select the egalitarian options in these two games would be categorized as aheadness averse. Payoffs in the envy game are (10, 10) and (10,

18) and in the costly envy game are (10, 10) and (11, 19), the second option, now, favors the opponent. The decision-maker could choose to decrease the opponent's income and only incur the cost in the costly envy game. Subjects who select the egalitarian options in these two games would be categorized as behindness averse. For people who choose egalitarian options over the four games, they would be classified as egalitarian.

In this paper, competition preferences are elicited by using the approach of Niederle and Vesterlund (2007). Results show the probability of tournament entry is 55 % for egalitarian subjects and 72 % for non-egalitarian subjects, significant with p-value of 0.06. The chance of entering is 55 % for aheadness averse participants, in contrast, the number for subjects who select the unequal payoffs at least once in the pro-sociality and costly pro-sociality games is 87 % ($p < 0.01$). The odd of tournament entry is 59 % for behindness averse participants while 67 % for subjects who choose the unequal payoffs at least once in the envy and costly envy game. The gap is not significant ($p = 0.46$).

Using probit models, the author find that being equalitarian reduces the probability of tournament entry by 30 percentage points, ceteris paribus. This effect is mainly driven by aheadness averse subjects while there's no association between behindness averse and self-selection into competition. Not surprisingly, being risk-averse and unconfident also decreases the chance of self-selection into the competition, but the magnitude is much smaller than that of aheadness averse. Finally, they take the Big Five into consideration and argue that subjects scoring high in Agreeableness are less likely to enter the competition.

Similarly, Balafoutas et al. (2012) examine the association between distributional preferences and competitive preferences and select the Niederle and Vesterlund's approach to measure willingness to compete. The focus on this paper is the elicitation of distributional preferences. Each participant faces with ten binary choices which need them to determine how to allocate money between self and another person. The ten questions are classified as two types: disadvantageous inequality block and advantage inequality block. There's always an egalitarian option in each questions, that is, both the participant and the opponent could get two euro. The non-egalitarian options of disadvantageous (advantageous) inequality block offer the opponent constant monetary reward, 2.6 euro (1.4 euro) whereas payoffs of the participant are 1.6, 1.8, 2, 2.2, 2.4 euro, respectively. Hence in each block, a profit-oriented agent switches at most once and only in a direction from the egalitarian to the egalitarian allocation. In the disadvantageous inequality block, the decision maker would be considered benevolent if he/she starts to choose the unequal payoff option in the third question or earlier. Otherwise the decision maker would be

classified as malevolent subjects. In the advantageous inequality block, the decision-maker would be considered benevolent if he/she switches to the asymmetric allocation in the fourth question or later. Otherwise the decision maker would be classified as malevolent subjects. Participants who are benevolent in both blocks are referred to as efficiency loving, participants who are malevolent in both blocks are called spiteful, participants who are benevolent when advantaged, but malevolent in the disadvantageous block are classified inequality averse, participants who is benevolent when lag behind but malevolent in the advantageous block are inequality loving.

As expected, women are 26.1 % less likely to enter the tournament. Men are more likely to perceive themselves as winners of the tournament than women (49 % vs. 19 %, $p < 0.001$) despite of the similar probability of winning (47.7 % vs. 46.3 %). This is caused by women whose performance is above median are strongly underconfident. However, there's no gender difference in the distribution of social preferences. Of 132 participants, 49 male and 45 female participants are efficiency loving, 9 males and 12 females are inequality averse whereas 7 males and 10 females are spiteful. No one is consistent with the definition of inequality loving.

When perform under piece rate, on average, spiteful and efficiency loving participants have similar performance (5.78 vs. 5.76) while inequality averse participants perform slightly worse, only solve 4.95 questions correctly. There are no differences in performance. Therefore, the types of distributional preferences only have little effects when subjects perform in a non-competitive context.

When move from piece rate to tournament, spiteful people solve two questions more correctly, while the performance improvement is about 1 for inequality averse subjects and only about 0.6 for efficiency-minded people. Spiteful people improve significantly more than efficiency lovers ($p < 0.05$) and marginally significant than inequality averse subjects ($p < 0.1$). 71 % of spiteful subjects, 47 % of efficiency-oriented subjects and only 29 % of inequality averse win the tournament.

When participants are able to selection between two incentive schemes, 28.6 % of inequality averse and 29.4 % of spiteful subjects decide to enter competition. Efficiency-oriented people enter much more often, the entry rate is 51 %.

It's intriguing that spiteful subjects shy away from competition despite of the much higher success rate. The authors argue it doesn't result from the low confidence level of spiteful subjects. Hence, the behaviors of spiteful subjects seem to be deliberate considering their preferences and correct beliefs about individual performance rather than a mistake caused

by biased expectations.

Those two papers are closely related with each other, even though name types of other-regarding preferences differently. In Balafoutas et al. (2012), benevolent in the advantageous inequality block could be seen as an aheadness averse action while malevolent in the disadvantageous inequality block could be seen as a behindness averse action. Therefore, compare with Bartling et al. (2009), Balafoutas et al. (2012) have different findings that aversion to be left behind rather than be high-performing participants plays a more role in entry decision. The difference results could due to several reasons: First, the sample pools of Bartling et al. (2009) are German preschoolers while in Balafoutas et al. (2012) participants are all college students. Second, Bartling et al. (2009) don't require participants to perform under piece rate or tournament. Participants make choices after they experience a practice round. The way participants make decisions itself could lead to differences in self-selection into competition. Finally these two papers use different ways to elicit social preferences. Thus, this line of research just gets started.

3.2.7 Stereotype threat

As researches listed above shows, in some tasks it's harder to detect gender differences. Also, males and females perceive a given task differently (Cárdenas et al. (2012)). Women's performance could be task-dependent. Niederle and Vesterlund (2007) maintain that men and women perform similarly in arithmetic or algebra despite the better performance of men in abstract math problems. Stereotype threat, according to the definition of Steele and Aronson (1995), refer to the concern generated from a situation where an agent worried about his/her self-characteristics would confirm a negative stereotype about his/her social group. Even minor manipulations could activate stereotype threat and negatively affect performance. Therefore, some papers alter tasks to investigate the effects of stereotypes about a given tasks on competition preference.

Iriberri et al. (2017) requires all participants to perform each tasks twice, first under piece-rate and then under the tournament. In the tournament, participants would be matched with an equally able competitor, measured by performance in piece rate. Participants only know the matching protocol after they finish performing under piece rate. The reasons for the pair-wise tournament are that the competition between people of similar performance enables to determine the pure impact of competition on all participants. Also this eliminates the effect of gender differences in beliefs about ability. There are 8 treatments. Subjects in the *Control* groups received no information before they participated

in the tournament. The remaining participants are told different information according to the groups they are assigned to. The details are shown in Figure 8.

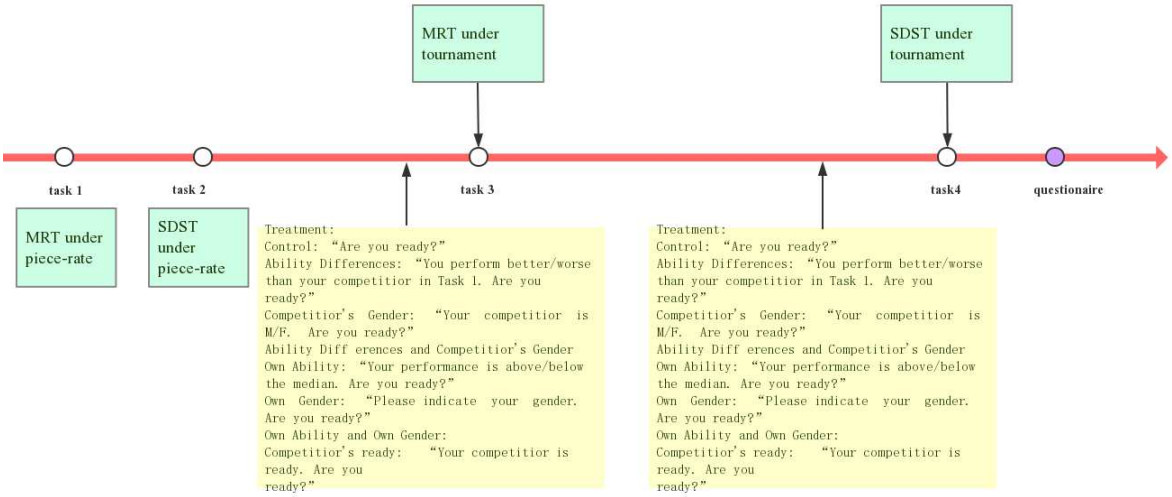

Figure 8: Timeline of the experiment

The experimental tasks are the focus of this kind of research, therefore are chosen according to the meta-analysis of Psychology literature on gender differences in performance in the absence of incentives. Men are believed to outperform women in spatial tasks while women are perceived to have advantages in verbal and memory tasks. Therefore, they choose a mental rotation task (MRT) as the 'male task' and a symbol digit substitution task (SDST) as the 'female task'. In the MRT, participants need to answer whether the figures are identical after seeing three-dimensional figures. In the SDST, subjects have to translate the three-letter sequences into numbers according to information which show the letters and the corresponding number. The survey which reveals how participants perceive these two tasks confirms the expectation of the authors. Compare with men, Women are more likely to view SDST as a female-favoring task whereas men and women are equal likely to view MRT as a male-favoring task. In this paper, whether a task favors (wo)men or not depends on the perceptions rather than actual performance. Overall, male participants perform significantly better in MRT while there is no gender difference in performance in SDST. In MRT, the average improvement is 5.23 when move from piece rate to tournament. The improvement is significant in three treatments: Competitor's Gender (coefficient: -6.48[10]), Ability Differences and Competitor's Gender (coefficient: -4.557) and Competitor Ready (coefficient: -4.655). Compare with the control group, revealing competitor's gender boosts men's performance improvements by 75 % in the

[10] The coefficients list here indicate the interactions between female and specific treatment.

tournament (8.4 vs. 4.77), while women's performance increase is reduced by about 50 % (3.13 vs. 5.76). In SDST there are no gender differences in performance improvement across treatment.

The findings suggest that when compete with someone who is in favorable position, women are more likely to prone to underperform when pre-existing negative thoughts are primed by revealing the identity or existence of competitors. The information provided before starting competition is crucial. In this case, policymakers should be cautious and try to correct the wrong previous perception about women's relatively lower productivity at jobs which are normally considered as favor men.

There are some other papers find similar patterns. Wieland and Sarin (2012) find gender differences in tournament entry as a whole are not significant (46 % for men and 45 % for women). However, they do find gender differences in some tasks. In this paper, subjects are required to answer quizzes cover four subjects and decide whether to be paid by piece rate or the tournament. Subjects are categorized into male-favoring (Math), female-favoring (Fashion), gender-neutral (Verbal and Crafts). As expected, women have significantly higher propensity of selecting the tournament when it comes to Fashion (26 % vs. 46 %) and men and women have the same entry rate in Craft (47 %). Contrary to expectation, there are no gender differences in tournament entry in Math (48 % vs. 45 %) and men enter the tournament more often then women in Verbal (62 % vs. 42 %). Familiarity about the certain subject and self-perception about competition are two key determinants of entry decision. Greater familiarity results in better performance. The authors also measure how important competition is for participants to gain a sense of self-worth.

Günther et al. (2010) argue that men also shy away from competition in area which they think they would lose, even though to a lower degree. They argue the task in Gneezy et al. (2003) favors male, so they also organize a neutral task(write words start with a given letter) and female task(memorize and answer). The results of the male task (maze) are in line with Gneezy et al. (2003). When solve mazes under piece rate, compare with the average absolute performance, women solve 10.4 % fewer while men solve 1.7 % more, the gender difference in performance is not significant. When solve mazes under the tournament, women solve 9.5 % fewer while men solve 20.1 % more. The performance increase is insignificant for women and only marginally significant for men ($p = 0.077$). Hence, men improve their performance under competitive environment when perform a male tasks. When perform the neutral task, consistent with expectation, men and women react similarly to the competitive pressure, measured by absolute performance and performance

improvement. The results of the female task are a little tricky. The female tasks ask participants to memorize short descriptions and then answer questions based on the given material. However, there's a distraction task between this two stages, which require participants to distinguish the different picture from the remaining several times. Men and women perform almost identically in the memory task and the gender differences are actually driven by the distraction task. Therefore the further studies need to proceed carefully as women indeed perform worse than men in the memory task, which is supposed to favor women. Contrary to expectation, both gender perform worse under competitive pressure, which may explained by the efforts put into the distraction task. The authors conclude that women may suffer double-stereotype threat: first, it's inappropriate for a woman to being competitive. Second, the male settings may impose a negative stigma on women as it makes gender salient.

Balafoutas et al. (2018) examine the effect of priming policy on gender gap. Participants are randomly assigned to one of three treatments: Neutral, High-power or Low-power priming treatment. All participants are required to perform a calculation task in 4-people groups which designed to measure competition preference. Additionally, before the calculation task, the High-power priming (HIGH) treatment asks subjects to describe a personal experience about control other people while the Low-power priming (LOW) treatment asks subject to describe a personal experience about being controlled by others. After finish writing, participants need to rate how powerful they feel they are on a scale from 1 (absolutely powerless) to 9 (absolutely powerful). On average, subjects in HIGH (6.22) consider themselves much more powerful than subjects in LOW (3.41). There's no gender difference in self-perception in LOW (3.43 vs. 3.38) and HIGH (6.16 vs. 6.29).

There're three stages: piece rate, tournament, choice. At the end of each stage, participants are told about their individual performance. Only men in Neutral perform better than women at the choice stage ($p = 0.077$).

In the Neutral treatment, men enter the tournament about three times as often as women (40.3 % vs. 13.9 %). The gender gap is smaller in LOW is 8.3 percentage point lower but remains significant (38.1 % vs. 20 %). However, the gap in the HIGH decreases dramatically to 8.2 % (27.9 % vs. 19.7 %). High-power prime treatment has a substantial and significant impact on the gender gap by influencing risk attitudes and helping people make rational decisions. Male participants in High are significantly more risk averse than males in Neutral. From payoff-maximization perspective, the average entry rate should be 25 %. The entry rates of both genders are insignificant different from 25 % while only

women in Low has a similar entry rate. Using performance at the second stage as a proxy, the expected entry rates are 28.3 % for men and 21.7 % for women. Women in all treatments enter at an optimal rate while it's not the case for men in Neutral and LOW. Therefore, the priming comes into effect by adjusting men's actions. This is noteworthy because unlike previous studies, this analysis intentionally manipulates the payoffs of tournament so that tournament yields a lower payoff than piece rate. Except priming, performance and confidence also play a role in entry decision.

Both and Günther et al. (2010) cast doubt on men's performance in verbal tasks, in which men are believed to be disadvantaged and attribute it to self-perception. Günther et al. (2010) maintain the results could result from the relevance of winning for self-esteem. Aronson et al. (1999) argue that stereotype threat also impede performance for people who view the task as relevant for self-esteem. In this paper, male participants may consider verbal task irrelevant therefore don't suffer stereotype threat. In contrast, Wieland and Sarin (2012) argue that the self-worth control variable fully mediates the relationship between gender and competition preference. Not surprisingly, men attach more value to competition than women and the effect of gender on competition become insignificant after adding the self-worth variable to the model.

Build on data of Gneezy et al. (2003) and Cotton et al. (2013)[11], Cotton et al. (2015)[12] develop a simple theoretical model. They find that the explanation that men enjoy competition match with the findings of Gneezy et al. (2003) and Cotton et al. (2013) perfectly. Compare with women, men obtain more enjoyment from competition, which reduces costs of the efforts. The other explanation is that men are more able than women. For example, men are better at handling competitive pressure, which could because men are less susceptible to stereotype threat.

4. Real outcomes

4.1 Education outcomes

Recently, the gender gaps in educational attainment have reversed. In United States, women

[11] Cotton et al. (2013), which require subjects to perform math task, is reviewed in detail in section 5.5.1.

[12] Even though the model help connect theory with empirical evidences, the model only apply for certain types of competition with fixed performance rating. In Cotton et al. (2013), it's one-on-one competition with males in mixed groups perform better than females in single-sex groups and worse than males in sing-sex groups. Females in mixed groups have the worst performance. In the six-people competition of Gneezy et al. (2003), participants ranked by tournament performance are males in gender-balanced groups, males in all-male groups, female in all-female groups, female in gender-balanced groups.

have earned more bachelor's degrees than men since 1982, more master's degrees since 1987, more PhD degrees since 2006. Similar phenomena appear in many other developed nations and some developing countries. Women not only outnumber men on college, but also tend to have better academic performances during high school. The human capital theory predicts that individual should make educational decisions which can maximize their lifetime incomes. However, pecuniary factors such as expected returns to labor market are not the only determinant of educational choices because compare with equally able male counterpart, female students are concentrated in fields of study like Humanities and Language, while still lag behind in fields such as Science, Technology, Engineering, Mathematics (STEM). Therefore, in this section, I will start with how existing literatures use gender differences in competition to explain gender differences in major choices and then performance in admission exam and response to competition. In section 4.1.4, I will give a policy recommendation and evaluate its effect. Finally I would conclude and link education outcomes to labor market outcomes.

4.1.1 Choice of academic track

Using the approach developed by Niederle and Vesterlund (2007), Buser et al. (2014) link the willingness to compete to choice of academic track in order to fill the gap between experimental results and real-life outcomes. After the experiment, participants are asked to fill in questionnaires which are designed to collect information about name, gender, self assessment of math ability and expected track choice. The experiment is conducted in four pre-university-level secondary schools a few months before the track choice to avoid reverse causality. In Netherlands, after graduating from primary school (grade 1 to 6), students need to make their first track decision: six-year pre-university, five-year general level and four-year vocational level. This track decision is mainly determined by scores of a nationwide test administered at the end of primary school. One fifth of Dutch students would go to six-year pre-university and be taught in the same class from grade 7 to grade 9. By the end of grade 9, students are offered four track options: Nature & Technology (NT); Nature & Health (NH); Economics & Society (ES); Culture & Society (CS). Students can choose tracks at will despite of the poor academic performance. Even though each track has its corresponding subjects, mathematics is the only one taught at a different level in each track. The order of math difficulty is NT>NH>ES>CS. Students are able to choose the combination of two tracks, therefore Buser et al. (2014) use the track stated in the questionnaires. There is a strong association between the study track at pre-university and

the choice of major in tertiary education.

Consistent with Niederle and Vesterlund (2007), Buser et al. (2014) find that boys are also substantially more competitive than girls and competitiveness is strongly positively correlated with choosing more math-intensive academic tracks even conditional on academic ability. They conclude that the gender difference in competitiveness could explain about 20 % of the gender difference in track choice by the end of grade 9.

Buser et al. (2017) find a similar pattern in Switzerland. In Swiss educational system, the compulsory schooling is composed of six-year primary school and three-year lower-secondary school. At the end of grade 9, most students either opt for vocational education within the Swiss apprenticeship system or for Baccalaureate school, an academic high school which prepares students for university. At the start of their high school education, Baccalaureate school students are offered 7 specialization options and need to choose one of them. In this paper, specializations' math intensity levels are determined by specializations' average math scores of a project administered by a canton-level Ministry of Education by the end of Baccalaureate school. The researchers go deeper in Physics and Math as it's the only specialization which requires students to take more Math courses. The gender gap in competitiveness is 20 % after controlling for piece-rate performances and tournament performances. They find that Swiss students who choose to enter tournament are significantly more likely to choose math-intensive track, controlling for grades and performance in the experiment. Boys are substantially more likely to enter tournament and gender gap in competitiveness can explain a significant portion (9 %-17 %) of the gender difference in specialization choices even condition on socioeconomic backgrounds. Condition on academic performances and school, the propensity of opting for Physics and Math is 15.6 percentage points lower for girls, a large gap considering only 14 % of students would choose this major. While being competitive reduce the gap to 13 percentage points, which reduces about 17 % of the gap.

Further, Buser et al. (2017) subdivide Swiss students into top, medium-high, medium low and bottom ability groups based on the level of their lower-secondary school. Students from bottom ability groups can only choose vocational track, while there's no limitation for students from other three groups. The gender gap in willingness to compete is essentially zero among the lowest-ability students, but increases steadily with ability and reaches a peak for the highest-ability students. Willingness to compete predicts career choices along the whole ability distribution. Top-ability students who compete are more likely to choose a math or science-related academic specialization and girls who compete are more likely to

choose academic over vocational education in general. At the middle, competitive boys are more likely to choose a business-oriented apprenticeship, while competitive girls are more likely to choose a math-intensive apprenticeship or an academic education. At the bottom, students who compete are more likely to succeed in securing an apprenticeship position.

In contrast, Reuben et al. (2017) fail to find the effect of competitiveness on college major choices, even though competitiveness has explanatory power on future earnings, which is a key determinant of major choices. All participants are students of New York University but from different grades and majors. In the post-experiment survey, participants are asked about their expected earnings after graduation and expected earnings if they were to complete different majors. They interpret the results as competitiveness has a muted impact on major choices because association between competitiveness and expected earnings exist across majors, not only in the major which students chose. Thus, competitiveness may be able to explain gender differences within major.

Both Buser et al. (2017) and Reuben et al. (2017) use addition task as Niederle and Vesterlund (2007) did and tournament entry to measure competitiveness, though slightly difference in experimental designs. There are several reasons for these different results.

First, samples are not comparable (European high school students versus American college students). Buser et al. (2014) contend that their samples are close to the national averages in terms of track choices by gender, sample sex ratio, average scores and pass rates of national test at the end of grade 6. However, in Reuben et al. (2017), samples are made up of high-ability students with high socioeconomic background, shown in their SAT scores and parental characteristics. Evidence from some countries suggests that track placement has at least as much to do with socio-economic status as with actual student achievement. Kids from better off family might have higher risk tolerance. Saks and Shore (2005) find that students from wealthier backgrounds choose majors such as business more often than disadvantaged students. Also, individuals from poorer households may lack important information when make their major decisions. Hoxby and Avery (2013) find that bright students from disadvantaged backgrounds often fail to apply to selective colleges that have lower out-of-pocket costs than less selective schools. This occurs despite the fact that information about various schools, programs and costs is available freely online. Family background affects not only major choices but also competitiveness. Almås et al (2015) find that students from high socio-economic family background are more willing to compete. Therefore, factors such as family background that could obscure the real situation need more attention.

Second, the experiment of Buser et al. (2017) is done several months before the major choice decision, while the majority of participants of Reuben et al. (2017) have already made the decision before they participate in the experiment, which could lead to different study experiences and further affect their compensation scheme choice. After focusing on freshman and sophomore who have more similar coursework experiences and get similar results, Reuben et al. (2017) rule out the possibility of reverse causality. Neither did they find results biased in the direction of the relationship between competitiveness and major choice. However, Kamas and Preston (2012) find no gender difference in willingness to compete among students with STEM backgrounds and gender gaps among students in humanities and social sciences disappear after accounting for differences in confidence. However, even controlling for risk aversion and confidence, gender differences persist among business students. Male business students are more likely to enter tournament than men from other majors while female business students are slightly less likely to choose to compete than other women. They conclude different major choices might be potential explanation for the much larger gender gap (38 %) found by Niederle and Vesterlund (2007) compare with results of other studies. Therefore, the relationship between competitiveness and major choices remains to be determined.

Last but not least, settings matter. Betts (2011) contends that in North America, within-school tracks enable students to move between tracks or ability groups relatively easily over time; students could be in a high track in one subject and a low track in another subject; mistakes could be fixed relatively easily. By contrast, in many European countries, students are streamed into different schools, with either vocational or academic emphases. Brunello and Checchi (2007) compare tracking across countries, report that the age at first tracking in the United States is 18, despite Betts (2011) argue it's not true unless ignore the within-school ability grouping and streaming that occurs as early as middle school. Niederle (2016) maintain that American educational system are more flexible about math-intensity taken at school while in continental Europe, these decisions are very binding. If American students have difficulties in a math class, they could switch to take an easier one next semester in high school. This difference in the way the decisions are made may in itself affect gender differences in choices of math intensive courses. This argument is supported by Fanny et al. (2018), who investigate whether track choice depends on the high school to which students are assigned. In France, students have to choose a major field of study at the end of their first year of high school, namely grade 10. Each field of study corresponds to a very specific curriculum, specific high school examinations, and specific opportunities after

high school. Therefore this decision is a key determinant of the undergraduate programs because science-related subjects only account for a small share of Humanities fields, which make it almost impossible for Humanities students to enter an engineering school. Additionally, students have little leeway to modify their major field of study during the two last years of high school. In France, selective schools are usually oversubscribed, therefore tend to have higher minimum admission scores, using regression discontinuity design, they compare students just above and below the cutoff. This analysis reveals that there is no gender gap in the willingness to attend more selective schools and enrollment at a more selective high school has little effect on students' performance on high school exit exams and on boys' track choice. However girls turn away from scientific fields and settle for less competitive and prestigious ones, that is, humanities. Effects are even larger for girls at the top of the ability distribution. They reject the possibility that enrollment in Humanities is due to high-quality courses provided by selective high school. Also, the fact that enrollment at selective schools has much stronger effects on girls' choices does not seem to be a mere consequence of the fact that girls are relatively stronger in humanities and boys relatively stronger in science. Fanny et al. (2018) argue that the introduction of the centralized process shift peer competition from lower ability students to higher ability ones and induce opposite changes in the level of competition faced by the two groups of students. Thus gender gap becomes larger and significant among high-ability

To sum up, Betts (2011) contend that obviously different countries have its own definitions of tracking and ability grouping, while existing measures of tracking do a rather poor job of identifying what these differences are. Therefore further work should put more emphasis on better measurement and endogenous problems, especially for cross-country comparison, better information on what tracking really is and careful thought about why countries vary in tracking policy would help convince readers that something close to causal is being measured.

4.1.2 Entry exam

Entry exam is usually highly competitive, especially those for selective programmes or schools. It's straightforward to consider scores of entry exam as performance under competitive context, the approach introduced by Gneezy et al.(2003). Ors et al. (2013) select applicants for HEC in 2005-2007, the top business school in France ranked by Financial Times and compare scores of the same cohort of people in three exams: relatively low competitive national baccalauréat exam (serves as high school exit exam, results don't

depend on others' performance and success rates are over 82 % for the whole sample) and first year exams of program with approximately 99% of master students validate their first year study at HEC. Instead, the entry exam of HEC is highly competitive only 11% of candidates would get accepted. They find that men perform better than women in the competitive entry exam while women performed significantly better at baccalauréat exam and, once admitted and after finishing first-year studies at HEC, women perform better at nonmathematics-oriented course. Among the top 10 % of applicants, namely selected students, men outperform women and there's no significant gender difference in performance among the bottom 10 % of applicants. Their findings confirm that females tend to perform worse in more competitive environment. They rule out the possibility that women's underperformance in entry exam is driven by gender differences in innate abilities and mathematics and view risk aversion as possible explanation. This paper complements experimental results as monetary rewards in the lab are quite small, for instance, in Niederle and Vesterlund (2007), winners of tournament only get 2 dollars for each correct answer, in case of a tie, winners only get 1 dollar. The average annual salary of HEC MSc alumni is approximately 13,000 dollars higher than the average annual salary of the alumni of the second best business school, which accounts for 24.5 % of the latter group of people. Thus, this paper shows gender differences in competition at large stakes.

Follow Ors et al. (2013), Jurajda and Munich (2011) study the performance of Czech secondary-school graduates who apply for universities. At the end of secondary school, both vocational program students and academic program students could take the national secondary school exit test, a prerequisite for university. Shortly after the national test, over half of students would apply for university and each university program selects its students in a separate competition, almost always using a program-specific written admission exam. They compare the admission chances at a given university of equally male and female applicants, measured by scores of noncompetitive national test. In order to avoid problems driven by gender differences in program-specific unobservables such as program-specific skills and by discrimination in grading university admission test, they focus on gender-balanced program which the share of women is between 25 % and 75 % of applicants and control for subject-of-study preferences. They find that men significantly perform better than women when admission rate is below 19 % and no significant gender difference in performance when admission rate is higher. This gender differences holds for both the most and the least able groups of applicants. In the robustness check, they report that selection on unobservable subject-specific skills is unlikely to be an explanation.

In comparison to the paper of Ors et al. (2013), which claim that gender differences in performance of three exams aren't caused by differences in test-taking strategy, Pekkarinen (2014) argue that gender differences in performance of entrance examination for nine Finnish universities can be partially explained by differences in answering decisions. In line with Ors et al. (2008), they find that women have better performance in high school and thus, higher starting points. Gender differences in performance of entry exam emerge upon controlling for starting points. Use the Rasch Model to derive the predicted probabilities of answering items correctly for each applicant, they show that women deviate more from the optimal number of items than men and that they do so because they tend to answer too few items.

Instead of entrance examination for university in western high income countries, Zhang (2012) turn her attention to the choice of entry exam for high school in China. Zhang (2012) argues that participants who choose tournament are more likely to take the exam, controlling for prior test scores and that neither exam fee nor parents attitudes toward entry exam could explanations the gap in exam choices.

4.1.3 External changes

Men and women may react differently to changes in external learning environment. Unlike admission exam nor track and major choice, which enable students to make preparations and put efforts during a long period of time, those changes usually occur suddenly.

Positive change could be scholarship. Price (2008) investigates the effect of Graduate Education Initiative (GEI), a competitive fellowship aimed at increasing graduation rates and decreasing time to degree. This fellowship was announced in March 1991 and launched later that year. The awards are very generous as it increased the stipend of recipients by 64%. Unlike one-shot tasks in the lab, the awards require long-term effort, which is closer to the reality. By comparing outcomes of pre-GEI students with students who entered between 1989 and 1990 so there could be no selection about competition preference, the author concludes that men's performance, measured by time to candidacy, increased after the implementation of the program, with the largest gains for men in departments with the highest proportions of female students. Women did not increase performance on average and only start to compete when they are in groups that are at least 58% female. Both male and female students benefit from the proportion of female in groups.

Negative change could be a reduction in admission rates or an increase in applicants, both intensify competition. Morin (2015) evaluate the response of both gender to an educational

reform which shorten the 5-year secondary school to 4-year. This reform is announced in 1997 and the first cohort of the new educational system starts in September 1999, hence create a 'double cohort' in 2003. The increase in the number of applicants is significantly greater than the increase in enrollment, therefore reduce the enrollments-to-applicants ratio from 71% to 55%. Because the educational reform changes the teaching content of mathematics, the author excludes mathematics courses from the analysis. Comparing performance of pre-reform students with that of post-reform student, the author finds that male average grades and the proportion of male students graduating "on time" increased relative to females. The author rule out the possibility that girls who are supposed to graduate in 2003 applied for university in 2002 in order to avoid competition. Further the analysis does not suggest that the results presented are due to gender differences in the dropping out decision process.

Booth et al. (2018) investigate the speedboat racing in Japan in which participants are randomly assigned to mixed groups or single-sex groups. People of both genders receive same training from the Yamato Boat School, the only training school. Also, participants compete under the same conditions. They find that female participants are more likely to be assigned to single-sex groups whereas men are more likely to be assigned to mixed-sex groups. Hence in mixed-sex groups, the sex ratio is skewed towards men. When being assigned to mixed groups, men behave more aggressive while women act less aggressive, measured by lane changing. Women have better performance in a single-sex environment even they have chosen a competitive career.

4.1.4 Policy implication: Single-sex school

Based on results above, one of the possible policy implications would be single-sex school. Akerlof and Kranton (2002) argue that students care about the extent to which their behavior deviates from that of their social group. The presence of men might make women be more alert to their gender, and if 'competitiveness' is considered as a masculine characteristic, women in mixed environment may prone to stereotype threat. In this context which female students pay too much attention to their negative identity, investments in education depend not only on individual benefits, such as test scores and grades, but also on social benefits, such as whether a particular level of effort is consistent with the behavior of one's social group. Walton et al. (2015) administer the experiments to foster a sense of belonging among female students from engineering, a typical male-dominated field of study. They report that GPA in Engineering courses of female students in treatment group increase

by 4 %.

The optimal level of competition in class may be different for females and males and the relative homogeneity of single-sex school make it easier for teachers to adjust the competition level and tailor their pedagogical approaches for the given set of students. However, the relationship between single-sex school and competitiveness shown in existing literatures is mixed and less conclusive.

Booth et al. (2012) show that compare with girls from coeducational school, girls from single-sex schools are more likely to select tournament even when randomly assigned to mixed experimental groups. Girls who are assigned to single-sex experimental groups choose to enter the tournament more than girls in mixed groups even they stay in sing-sex group for only 20 minutes. It thus suggests that observed gender differences might reflect socialization rather than inherent gender attributes.

Lee et al. (2014) find that in Seoul, South Korea, single-sex schooling does not reduce the gender gap in competitiveness conditional on student and parental characteristics. They maintain endogeneity won't be a problem because within a school district, a student is randomly assigned to a single-sex or coeducational school; the assignment rule does not take students' preferences into consideration. All school districts in Seoul have both single-sex and coeducational schools, hence students can't choose schools they want by moving to other districts. Another advantage is that all middle schools are subject to the same educational policies, such as curriculum, number of school days, and teacher hiring which allow them to identify the causal effect of single-sex schooling on competitiveness. However, Park et al.(2012) argue that single-sex schools in Seoul are predominantly private and enjoy a great autonomy in teacher selection and teacher tenure policies. Differences in teachers' characteristics and the interaction between teachers and students need to be taken into account.

Studies about single-sex school are almost always plagued with (self-) selection problems. Thus Booth et al.(2012) use instrumental variables (post codes and distance between home and school) to predict the probability of attending a single-sex school as sensitivity analyses and the coefficients to single-sex schooling are still of significance. Therefore, they conclude that schooling environment has an effect on individuals' propensity to choose a tournament. Niederle and Vesterlund (2010) point out that it is not possible to determine whether this response results from the superior performance by students at the single-sex schools. This could be the case. In Booth et al. (2012), samples are from Suffolk and Essex, two British counties. The United Kingdom used to channeled students into different types of

schools based on a test students take when they aged 11(11+ exam), but began to move away from this system in the 1960s. In this educational reform, central government also requires local authorities to establish coeducational comprehensive schools. The old educational system survived in some areas but was abolished in most others. Thus there is no single-sex public schools in Suffolk, all single-sex school in this paper are from Essex. In Essex, students still need to take 11+ exam if they want to study at sing-sex school. In order to make samples comparable, Booth et al. (2012) only recruit higher-ability students from academic track of coeducational school.

Hence, some papers use random assignment to try to solve this problem. Eisenkopf et al. (2015) look into students from the same high school. Students are randomly assigned to mixed class or all-girl class and all face the same curriculum. The students' academic performance is observed up to four years. They show that single-sex class improves the performance of female students in mathematics, especially for girls who had high scores in entry exam. However, there's no significant difference in Germen test scores among two kinds of girls. They explain the improvement as single-sex schooling strengthens female students' self-confidence and make girls to perceive their mathematics skills more accurately and positively. Indeed, Students are assigned to classes randomly but might choose the school intentionally because this school focus on pedagogical subjects therefore attract students aspire to study in Education major. Therefore this school is more attractive for female students which make 80 % of students are female.

4.1.5 Conclusion

The majority of papers listed above refer to the gap in math, one of the hottest issues in gender differences. Even if women now have greater educational attainment in many industrial countries, men still tend to outperform women in Maths. Schrøter Joensen and Skyt Nielsen (2010) report that maths skills have a causal effect on labor market outcomes and there is evidence that the individual returns to maths skills are higher than the returns to other skills (Buonanno & Pozzoli, 2009; Koedel & Tyhurst, 2012).

Niederle and Vesterlund(2010) argue that if women can't handle competitive pressures as well as men did, then test scores may be unable to measure truly math skills, and may exaggerate the male math advantage. Noncognitive factors may not only affect an individual's investments in cognitive skills, but also the individual's test score performance. Follow the experiments designed by Niederle and Vesterlund (2007) completely, Masayuki (2018) confirms the argument by investigating the effect of competitiveness on math

achievement among Japanese public middle school students. Consistent with previous studies, girls are less competitive and men have higher math scores. Being competitive improve the math achievement, controlling for prior achievements and demographics characteristics. The coefficient on the entry dummy is 0.105 and significant at 10 % level. The average growth in math scores from 8^{th} to 9^{th} grade is 0.5 hence it accounts for over 20 % of the growth. Being competitive could reverse the gender gap as the coefficient on female dummy is -0.083.Further analysis concludes that competitiveness has an effect independent of risk attitudes and other factors and could explain about 9.2 % to 13.5 % of the gender gap in math achievement. Iriberri et al. (2018) conduct 3-stage competition to select winners and show that despite the higher math scores at school, girls perform worse in competitive environment. The increase in competitive pressure widens the gender gap. The bad performance is not caused by selection, discrimination or differences in reaction to increasing difficulty.

Fryer et al. (2010) find earlier results that countries with higher gender equality index tend to have a smaller gender gap in math are sensitive to the inclusion of Middle East countries, where there's a high-degree of sex-segregated education and no gender gap in Math.

In conclusion, further research needs to focus on selection problems of single-sex schools and try to find which group of people is most likely to be affected. If high-ability women would benefit most from single-sex schooling without hurting others, then it could enhance efficiency and reduce gender gap and inequality.

4.2 Labor market outcomes

Female students tend to have better academic performance, but this advantage at school doesn't translate into advantages at labor market, measured by earnings. In this section, I would go through several crucial stages of employment to see how competitiveness shapes women's employment decisions.

4.2.1 Job sorting and gender segregation

Gender differences in preferences for competition offer explanations for job sorting — the first step of entering labor market. Also, especially in some European countries, the major choice at high school is decisive for career path.

Kleinjans (2009) finds that women's reluctance to competition can explain part of the gender segregation in occupational fields. Based on the assumption that occupational

expectations are strong predictors of later career attainment of professionals, the author uses occupational expectations as a proxy for occupational choice. In the survey, 2312 19 years old Danish students are asked about their occupational expectations when they are about 30 years old. The author doesn't use the performance or entry of choice to measure competition preference. Instead, students are asked to what extent they agree with the following statement: ''Outside the world of sports, people should compete as little as possible". The 5-point scale options range from 1 (completely disagree) to 5 (completely agree). According to Danish educational system, the author divides educational levels required for the expected occupation into 3 types: vocational training or less, short- or medium-cycle education, and college or more. The author summarizes that a greater distaste for competition is related to a lower fraction of women expecting to work in an occupation requiring a college degree. After controlling for ability and socioeconomic background, competitiveness reduce gender segregation in the fields of Law, Business & Management, Health, and Education. If an average woman increases her competitiveness to an average man level, the predicted probabilities to work in education field decrease by 4.6 %, in health fields decrease by 4.2 %, in law, business and management fields increases by 9.2 %.

There are two main concerns. First, women who have strong preferences for kids may prefer flexible working hours and family friendly, normally less challenging jobs. In the robustness check, the author includes job characteristics described by working hours and job safety as proxies and finds similar results. The author attributes to the high female labor force participation, generous maternal leave and universal childcare. Also, gender difference in labor force participation rate in Denmark is quite small. However, if women with lower fertility expectations have stronger willingness to compete, the effect of competitiveness would be biased upwards. The author thinks it won't be a problem because about 90 % of Danish women aged 49 have kids.

The second concern is omitted variable bias. In this analysis, main control variables are ability (measured by test scores) and family background. It could be possible that self-confidence and risk aversion could affect both competition preference and occupational expectation. However, including self-confidence variable did not alter the outcome. Unfortunately, there is no measure of risk aversion in the data even distaste for competition is likely to reflect at least some of these attitudes indirectly, according to previous studies.

However, it should be noted that in this paper, both Skilled Trades, the occupational field with highest level of gender segregation and Science and Engineering, the traditional male-dominated field, are not affected by competitiveness. Therefore, gender segregation in

these fields is likely to stem from other sources. Also the author drops about one fourth of total 3084 observations due to the missing important information, mainly about occupational expectations. This leads to a slightly positive selection in terms of family background and ability. Therefore, results might underestimate the effect of competitiveness on job selection if children from low socioeconomic backgrounds are less willing to compete.

Reuben et al. (2015) investigates whether competition preference explains subsequent gender differences in earnings and industry choice in a sample of high-ability MBA graduates from University of Chicago, one of the top business schools in United States. Top business positions is one of the areas with dramatic gender gap, as documented by Equilar (2015), who reports that women represented only 6.5% of the best-paid CEOs in 2014 and were paid 9.9 % less than their male counterparts. Therefore, focusing on this group of people, who would be candidates for these positions in the future, could help us better understand gender differences in this field.

During the first month of their MBA study, students are invited to join in a survey which collects information about demographic characteristics and personality traits. After the survey, students would participate in experiments that were designed to measure individual characteristics: risk preferences, other-regarding preferences, and competition preferences. In the first year of the program, career services would hold job fairs to help MBA students obtain an internship in a firm. In the second year, MBA students can decide whether to stay at the interned firm if they were offered a job or look for other jobs.

Industry categories are based on NAICS industry codes: finance (code 52), professional services, which they refer to as "consulting" (code 54), and "other" (the remaining two-digit codes). The information included data on financial earnings, which includes salaries as well as yearly and one-time bonuses.

This paper has multiple advantages: First, earning data is provided by university career services and students' corresponding employers, rather than self-reported, therefore reduce measurement errors caused by overconfident people overstate their earnings. Second, students are tracked over 7 years after they enter labor market, which allows them to determine the long-term effect of competitiveness on earnings and taking factors such as marriage, pregnancy and childrearing into consideration. Also the authors have employment information of MBA students before they start studying. Hence, it's able to indentify the impact of competition preferences in different stages of the job market. Last but not least, competition preference is measured at the beginning of the MBA program therefore it's not

affected by MBA study experiences. Even though men have a longer work experience before getting admitted, Reuben et al. (2015) don't find any correlations between competition preference and participants' characteristics at the admissions stage, so former work experience is unlikely to cause gender differences in competitiveness.

They report that competition preference comes into effect when participants decide in which firm to intern. Participants who choose to enter tournament are 8.4 % more likely to intern in consulting and 8.5 % less likely to intern in other industries, compare with individuals who choose to piece rate. Thereafter, the gap grows as firms decide formal employees and participants decide whether to continue working at the interned firm. At graduation, a competitive participant is 7.1 % more likely to work in consulting, 4.3 % more likely to work in finance, and 11.5 % less likely to work in another industry. Hence, competition taste is also a good predictor of the industry participants choose to work in at graduation and several years later. Competitive individuals are more likely to start and keep working in consulting and finance (to a less extent, though) rather than in other industries. However, even after controlling for competition preferences, risk attitudes, and individual characteristics, gender gap still remains and gender still play an important role in the industry participants work in.

Men and women self select into different industries and this difference persists. Female MBAs are 8 % more likely to work in other industries at graduation and 12 % more likely to work in other industries seven years later. At graduation, male MBA students are about 50 % more likely to work in finance industries; this gender difference persists over time even though both men and woman shift to other industries. Women who start their careers in finance are significantly more likely to move to other industries than men. Second, individuals who started their career in consulting are significantly more likely to transition to finance and less likely to transition to other industries if they chose the tournament compared to individuals who chose piece-rate. The gender gap in selection into industries emerges at the very beginning of the MBAs' job market. This gender difference is more clear if pay attention to the application stage. Compare with men, women send fewer applications for internships at finance industries and more applications for consulting and other industries. Hence, women receive fewer invitations from finance firms and more invitations from consulting and other firms in consulting and other industries.

With respect to wage, individuals who chose to enter the tournament earn 9 % more than their less competitive counterparts do with the largest difference for the top earners. Moreover, gender differences in competitiveness could explain around 10 % of the overall

gender gap which is half as much a rich set of variables such as demographic characteristics, academic performance, etc did. Also, competitive individuals are more likely to work in high-paying industries nine years later, which suggests the persistent effect on earnings. Reuben et al. (2015) also try to figure out why competitive individuals earn more money. If competitive individuals are simply better at the tasks demanded by better-paid jobs, then this positive selection widens the pay gap. Even they don't detect the correlation between competition preferences and GPA and GMAT scores, they believe it is possible for taste for competition to reflect differences in ability that are not captured by conventional measures. This is confirmed by the fact that competition preference has relatively weak effects when interactions between participants and firms are limited. At first, it's difficult for firms to select competitive individuals and for participants to identify their comparative advantages caused by competitiveness based on existing information. After two parties have meaningful contact with each other, employers are able to observe the higher productivity of competitive individuals and competitive participants feel comfortable and they are qualified for this job. The positive impact of competitiveness increase as matching process proceeds. Another potential explanation is that competitive individuals are more willing to negotiate their earnings with their employers, which is not supported by analysis. A third explanation is tastes. If people gain disutility from working under competitive environment, the wage gap would be premium compensation for workers. Therefore future research needs to identify the specific characteristics of competitive workplaces that competitive individuals find less distasteful.

In line with Reuben et al. (2015), in Reuben et al. (2017), students are asked about their expected wage of full-time works, to eliminate bias introduced by gender differences in labor supply. They find that even though both gender expect to earn much higher than average US college students of same gender and same major, female students clearly expect to earn less than male students and this difference increases with age: from 31 % less at age 30 to 39 % less at age 45, on average. They report that gender differences in overconfidence and competitiveness combined explain 18 % of the gender gap in wage expectation. The two measures explain in total as much of the gender gap in earnings expectations as a rich set of control variables. Reuben et al. (2017) rule out the possibility that gender gap in expected wage is driven by misperception of labor market. Students' beliefs about the gender pay gap among general population are quite close to the true gender gap. Also competitive and overconfident students do not expect to work longer than their counterpart. However, competitive and overconfident students do show less aversion to earnings

uncertainty. Earnings expectations are a significant determinant of major choice as differences in major choice can explain about one third of gender gap in wage expectation. Kamas and Preston (2018) look into the specific effects of these two measures, and find neither competition preference nor confidence alone could cause an increase in wage. Like Reuben et al. (2017), this paper only includes full-time workers. Men not only have a higher annual wage but also work longer weekly, compare with women. The gender pay gap, which already appears at the beginning of careers, increases from 8.8 % to 13.6 % if bonus is included. Women who are more willing to compete and more confident earn substantial more than other women while there's no significant effect on men's wage. Half of the positive effect on female wage is explained by college major and labor market controls such as employment status. However even after conditioning on these characteristics, a higher taste for competition for the most confident women results in more than a 7 % increase in compensation, the magnitude is able to reverse the negative effect of gender on compensation differential (−6.2%). While other factors, such as family responsibilities, may lead to gender differentials in compensation later in their careers, for the years immediately after graduating from college, confidence and preferences for competition do lead to higher earnings for women. In line with Kamas and Preston (2012), there is a correlation between decisions to compete and major choice in Kamas and Preston (2018). Kamas and Preston (2018) maintain that women who have preference for competition may select into well paid sectors of employment.

4.2.2 Performance pay

Section 2.1 describes that men and women tend to have different job preferences, however it's hard to compare competitive pressure across industries. Thus Flory et al.(2015) restrict jobs to administrative assistant to examine whether the compensation scheme, by itself, can cause differential job entry. They posted employment advertisements online and use applicants' resumes to gather information about personal characteristics. This experiment is held in 16 major American cities therefore factors like local labor market conditions can be taken into account. Competitiveness of the jobs is manipulated by changing the degree of dependence on relative performance. There're in total 5 types compensation schemes: fixed-wage individual work (T1), fixed-wage group work (T2), lightly depend on personal relative performance (T3), heavily depend on personal relative performance (T4), and heavily depend on team relative performance (T5). Job-seekers know compensation schemes only after they express interest in the position.

Results show that women disproportionately shy away from performance-based compensation. Conditional on expressing initial interest in the job, in general, men are significantly more likely to apply. This gender gap is most pronounced in the T4, the treatment with highest level of competition. Thus the authors argue that women are daunted by competitive workplaces when apply for jobs. It might be straightforward to conclude that men are motivated by competition while women choose to avoid competition, which is documented by Niederle and Vesterlund (2007). However the analysis shows that both men and women shy sway from the competitive compensation schemes, but women show a significantly stronger aversion to competitive workplaces, compared with men.

There are some factors that influence gender gap except the degree of performance pay. First, using local median wage of administrative assistant and gender pay gap as proxies for outside options, they find that when women have less alternative options, they show less aversion to competitive work settings. Second, even though both compensation schemes are heavily reliant on relative performance, when women work in team to against other groups, they are more likely to apply for this position.

In Samek's paper, college students receive e-mail invitations about an open position job, temporary research assistant. The author set the compensation scheme as (1) flat rate, (2) performance-based bonus or (3) non performance-based bonus. In order to isolate the effect of social preferences, the job is described in 2 ways, non-profit and profitable, holding all else constant. The charity manipulation was conducted for both the flat and performance-based compensation schemes. The author maintains that both genders are discouraged by the performance-based compensation schemes from applying, but the effect is larger and statistically significant for women. The effect is not caused by aversion to uncertainty in payoff. Social preferences increase application rates but don't reduce the gender gap in competitiveness. The willingness to compete is mediated by career choice, Business and Engineering students show preferences for competition.

Lavy (2013) examine how teachers' performance, measured by scores of their classes, is affected by a competitive environment by providing cash bonuses. The author finds no evidence of gender differences in performance under competition in any gender composition of class. Women, however, were more pessimistic about the effectiveness of teachers' performance pay and less confident about their chance of winning.

However, the evidence is far from conclusive if use national level data. Manning and Saidi (2010) estimate the degree that a worker's compensation is based on performance as a proxy for a job's competitiveness. Using data from the 1998 and 2004 British Workplace

Employment Relations Survey, they find that British women are only slightly less likely than men to be paid according to relative performance, and performance pay has similar effect on both gender's wages.

There're several potential criticisms. This paper is based on the assumption that performance pay occurs exogenously. Bryson et al. (2014) use data from the 2011 British Workplace Employment Relations Survey and contend that performance pay contracts are more likely to be used in higher paying sectors, which more able people are likely to work in. The premium rises along with the wage distribution: it is seven times higher at the 90th percentile than it is at the 10th percentile in the wage distribution. Hence, in my opinion, if performance pay is a good indicator of competitiveness, further research should link the self-selection of women into low-paying jobs to the degree of competitiveness in compensation schemes, which might find a larger impact of competition preference on wages.

Also, Manning and Saidi (2010) ask workers about how much effort they think they need to put in the job and find that men's effort response to performance pay is largest in female-dominated jobs while women seem to respond less to such an incentive in male-dominated jobs. Thus further studies need to look into the performance pay in male-dominated and female-dominated jobs specifically

McGee et al. (2015) using data from the National Longitudinal Surveys of Youth (NLSY) 1979 and1997 cohorts to estimate the relevance of competitive compensation to gender pay gap in United States. In the NLSY1979, performance pay included piece-rates, commissions, bonuses, stock options and other performance pay; in the NLSY97, it included commissions, bonuses and incentive pay. Unlike Manning and Saidi (2010), this paper distinguishes the competitiveness of different types of performance pay. Also McGee et al. (2015) don't include firm fixed effects which are impossible to detect in NLSY. Consistent with the laboratory studies, they find that women are less likely to work in jobs whose compensation schemes depend heavily on performance pay, despite the magnitude are smaller than that of gender differences in competition. However, gender differences in compensation schemes can only explain 6.2 % of the gender wage gap in NLSY 1997 and 2.9 % of in the NLSY1979, controlling for personal characteristics. Results show that the most competitive performance pay measure is bonus while the least is piece rate.

4.2.3 Promotion

Women might be disadvantaged in two aspects: first, if women have worse performances in

competitive environment, then they are less likely to be promoted as promotion usually based on productivity. Second, even if women are equally able as men, if they have much lower chances of applying, then fewer women would get promoted.

De Paola et al. (2017) analyze the Italian promotion system and find significant but small gender gap in the likelihood to apply for academic promotion and conclude that the application rate gap is mainly driven by gender differences in risk aversion and confidence and the fear of discrimination.

In Italy, individuals who want to get associate or full professor positions must take a nation-wide and field-level competition (NSQ). The applicant would be evaluated by a 5-people committee randomly selected and formed from the full professor pool. The committee use general criteria like CV and publications and are allowed to adopt additional criteria. However, no teaching lecture or oral presentation is required. Hence there's no direct contact between the committee and the applicant, hence gender difference in performances in the competitive contexts won't be a problem. Failed candidates are not able to apply in the following two years. Candidates who reach the minimum threshold and obtain the qualification will be considered for promotion. 99 % of new positions are filled through internal promotions due to the much lower cost, about 70 % less. Moreover, higher positions don't require for extended work hours. As a result, the Italian promotion system allows researchers to eliminate the effect of factors such as unobserved ex-ante abilities, unwillingness to work longer and move to another university and city.

Except productivity(measured by the quantity and quality of publications) and experience(the number of years the applicant has been hired), the authors also include controls like possible connections between the applicant and professors who are eligible evaluators and effective connection between applicant and committee members, take the value of one if they belong to the same university. Controlling for those variables, they report that being a women reduce the propensity of entering the competition by four to five percentage points, which make up 8 %-10 % of male application rate.

However, it's unable to measure competition preferences of applicants according to existing data. Therefore the authors divide applicants into 4 groups according to their productivity and they find the propensity of applying increases with productivity. Among the bottom 25 % of applicants, women are 6.3 % less like to apply, the gender gap is substantial considering the chance of applying for this group of people is only 28%. However, this gap decreases with productivity. A female applicant with average productivity is 3.4 % less likely to apply. The gender gap completely disappears among high ability group. The

authors also divide fields according to the measurability of productivity and report that women are 3 % less likely to apply in fields which productivity is easily measured and 5.4 % less likely in other fields. The authors also investigate the impact of women's success in the field, measured by the share of female professor at the time of application and the share of female assistant professor. The results show that in a field in which women's success is one SD below the mean, women are 9.5 % less likely to apply, ceteris paribus. This gender gap reduces to 3.3 % when women's success is 1 SD above the mean. They don't find any evidence for gender discrimination in NSQ, but can't rule out the possibility of downward biased results as applicants have been self selected.

Bosquet et al.(2018) assess the gender gap in French promotion system, which is quite similar to the Italian system. There are two types of academic positions in France, university researchers and researchers of public research organization (in this paper, the authors only focus on researchers from CNRS, the largest research organization, for simplicity), lead to the dual track promotion system. University researchers need to spend a lot of time preparing for 4-stage evaluation, consists of the research seminar and three oral exams which would take half of year while CNRS researchers only need to submit CV and a research project. Also, researchers who get promoted are usually unable to stay in the original university. Women are less often to apply for promotion. Then the authors decompose gender differences in the probability of being promoted into two factors: differences in the probability of applying and success. They report for university researchers, almost 70 % of the raw gender gap in promotion are due to the application gap, and it widens even larger (76 %) when characteristics are accounted for. For CNRS researchers, the gender gap in promotion that can be explained by gap in apply is about 55 %. Therefore, the application stage is crucial. Publications, which favor men, have a moderate effect. Bosquet et al. (2018) contend that gender gap in application rates could be the following reasons: 1) the higher cost for female if female are less geographically mobile; 2) female researchers value the promotion less; 3) fear of possible discrimination; 4) distaste for competition. Men are very likely to apply when the success rate is low and the top 5 % of women, even are two times more likely to get promoted than the bottom 25 % of women, the number of applicants only increase by 10 %. However, the existing data is insufficient to determine whether men apply too much and women apply not enough.

By contrast, research output is the key determinant of the probability of success regardless of the type of researchers. The main gender difference is in the first stage in which women are less likely to enter next round. While in the final stage, being a women increase the

likelihood of being promoted. The authors interpret it as the negative unconscious discrimination at the beginning and the positive conscious discrimination in the final stage.

However, Booth et al. (2003) find that full-time women are promoted more often than full-time men with unadjusted data from British Household Panel Survey. Controlling for individual heterogeneity, they find no gender difference in the probability of being promoted. However, men would receive larger wage rewards. They propose a sticky floor model to explain the phenomenon. Sticky floor could be result from the worse market alternatives faced by women. If it's the case, then more women are promoted and promoted and retained women experience small wage increases upon promotion. It could also due to the unfair treatment which discriminates against women by firms. Under this circumstance, men end up with a higher wage while promoted women are more likely to quit than promoted men. Both models show that women benefit less from promotion but be promoted at the same rate as men.

There're also some evidences for gender differences in task choices, such as women allocate their time to tasks in a different way. Lazear and Rosen (1990) allege that if women have better non-market outside options, then women would reduce the investment of human capital which could contribute to promotion. Engaging in non-promotable tasks would take time away from promotable tasks, widening the gap in productivity and put oneself in a relatively disadvantaged position. For example, for research-oriented jobs like assistant professors, working on research papers is more relevant than holding an orientation party, with regard to the likelihood of being promoted. Therefore, Babcock et al. (2017) examine the choice of performing an effortless task that everyone prefers be completed by someone else. The experiments are conducted in mixed sex groups and then single-sex groups. There're 10 rounds and in each round, participants are assigned to groups of three and then decide whether to invest during the following two minutes. If no one invests, each group member could get 1 dollar. If someone invests, the investor would get 1.25 dollar while the payoff for other two people is 2 dollars. Results show that in the first experiment, women invest 3.4 times and men invest 2.3 times, the 48 % difference is significant. Controlling for risk and other regarding preferences doesn't eliminate the gender differences, suggests a gender difference in belief that women are more likely to invest. The second experiment, in which participants have group members of their own sex, the investment rate is similar. Then the authors evaluate from demand side, whether women are more likely to be asked to invest than man. In the third experiment, subjects are divided into groups of four, and then decide who to be investor after seeing the photos of other three people. The requestor would

be randomly selected from these four people after each member submit their decision. The result would be published on the screen while the chosen investor still has discretion over the acceptance of the request. Results show the mean and median numbers of requests for men are 8.7 and 9, respectively, while for women they are 11.1 and 12. There's no gender difference in investment rates for people who are not requested, both are 14%. While for people who are chosen, even though the request significantly increases the likelihood of investment, men still have lower probability of investment (51 % versus 76 %). In the last experiment, the authors rule out the possibility that participants are more generous toward men and confirm the effect of belief. Babcock et al. (2017) summarize that if gender differences in task allocation result in gender differences in productivity, then the unbalanced allocation would persist as it would be efficient to assign less promotable tasks to less productive female workers.

There are two papers analyze from the employers side that needs to be noted and support the argument of Babcock et al.(2017). Peterle et al. (2017) use calculation task and design a 4-step experiment: piece rate, tournament, promotion game, tournament/piece rate (depends on position). In the promotion game, participants would be either an employee or an employer. Experimental firms are composed of one employer and three workers who compete for one position. The employee who gets the position would be paid according to a winner-take-all tournament, compete with two employees from other firms. The other two are paid according to a piece rate. Employees could send a costly signal to indicate their willingness to compete but are not guaranteed to get the place. In the Baseline treatment of promotion game, employers make the decision based on the signal and demographic characteristics. In the No Selection treatment, the decision is made by the computer, randomly select a person who expressed the willingness to compete. If no one asks for the promotion, the computer would randomly choose one. In the No Signal, employers make the decision only based on demographic characteristic while workers can't send signals. Thus, only the gender differences in attitudes toward competition could affect the promotion decision of No Selection treatment.

The raw gender gaps in Baseline, No Selection, No Signal are: 41.66 %, 25 %, and 9.1 %, respectively. Hence the highest and only significant gap occurs when employees could send signal and employers have the right to choose. The gender gap is almost not existent when workers are unable express their willingness to compete and employers select only according to individual characteristics. Women are significantly less likely to send signals than men. The gender gaps in Baseline and No Selection are quite similar: 27.77 % and

26.22 %, respectively. Compare these two treatments, the authors conclude that gender matters only when employers can select as female think they are subject to potential discrimination. In both treatments, the sending decision could be explained by risk preferences. Compare the Baseline treatment with No Signal treatment, the authors argue that employees who send signals are significantly more often to be promoted and employers discriminate against women if they are able to send signals but don't ask for the promotion.

Heinz et al. (2016) find that employers determine the wage as in a dictator game. Gender discrimination is non-existent if employees select tournament or let the computer do the decision. However, women who choose to be paid by piece rate earn significant less both in employee share (18 % versus 36 %) and income (2.03 vs. 3.09 euros) even no gender differences in performances, regardless of compensation schemes. As women are less likely to apply for tournament (40 % versus 64 %), this discrimination behavior of employers widens the pay gap.

Dasgupta et al.(2015) evaluate the self-selection into the six-month training program provided by NGO which aims at improving sewing skills of Indian women aged between 18 and 39 by providing stitching and tailoring courses. It's compulsory to spend 10 hours every week. They argue that socioeconomic background and demographic characteristics can't fully explain the differences in self-selection. Women who are more risk prone and competitive apply for the program more often than those who are not. Risk preferences have a pronounced effect possibly due to the imperfect and underdeveloped financial market in developing countries. This paper should be noted as findings could be applied to other developing countries. First, the sample is representative of the population structure in developing countries. Second, people live in developing countries, women in particular, usually have low levels of human capital and hence low levels of productivity. Thus, this kind of training program which could would be more relevant and beneficial for women from developing countries than their male counterpart and women from economically advanced countries. Last but not least, garment factories now relocated in developing countries and offer a lot of job opportunities for women who used to be exclude from local labor market. Figuring out the mechanism could better promote gender equality, boost economic growths and reduce global inequality to get sustainable development under way.

4.2.4 Policy implication

Niederle et al.(2013) introduce a gender quota which requires at least one of the winners must be women in experiment to evaluate its effect. Participants are divided into 6-people

gender balanced groups. There're two types of tournament, standard tournament (the top 2 performers would be winners) and affirmative action tournament (the top 2 performers would be winners, but at least one of them must be women), AA tournament for short. There're 6 treatments: piece rate; standard tournament; choose piece rate or standard tournament for personal performance in treatment 3; choose piece rate or AA tournament for personal performance in treatment 4; choose piece rate or standard tournament for personal performance in treatment 1; choose piece rate or AA tournament for personal performance in treatment 1. The first four treatments are real-effort tasks, the last two treatments don't require subjects to really perform task. Thus, gender differences in entry tournament caused by competition preferences are eliminated in the last 2 treatments. After controlling for the chance of winning and confidence, the difference in the final two choices, is a measure of the effect of mentioning affirmative action.

The quota policy makes competition more gender-specific: a woman wins if she is either the best female performers or among the two best performers in the group, but a man needs to be not only be the best male performers but also among the two best performing participants in the group. In consistent with previous studies (Niederle and Vesterlund, 2007), without quota, high-performing women don't enter the tournament enough and this could be attributed to gender differences in confidence and competition preference.

In the standard tournament, the largest gender gap in tournament entry appears among those participants who would earn more if they choose standard tournament rather than piece rate: among those who have higher expected earnings in the tournament than in the piece, all men and only one-third of the women enter the tournament. As expected, affirmative action substantially increases tournament entry by women and deter men from entering tournament. This time, women who can benefit from tournament, all choose to enter AA tournament. The introduced quota boosts women's confidence significantly.

Next, Niederle et al. (2013) assess the cost of affirmative action. In the standard tournament, winners on average solve 19.2 questions correctly, while in the AA tournament, the number is 18.7. This difference is not significant. Niederle et al.(2013) evaluate both ex ante cost and ex post cost. The ex ante assess expected effect of quota if participants don't change their behaviors. The ex post evaluates realized costs of affirmative action, in which quota policy is announced and individuals have adjust their behaviors accordingly.

Niederle et al. (2013) also contend that quota doesn't cause efficiency loss as the pool of high performing participants is rarely affected. Even though some high performing men are dropped out, some high performing women are encouraged to enter the tournament by the

quota. Therefore, the ex posts performance costs decrease substantially. Quota is also criticized due to possible reverse discrimination by lowering the standards. However, in this analysis, the requirement is almost same for winning the tournament with and without quota, thus there's limited space for reverse discrimination. By contrast, ex ante costs are expensive, which imply the importance of making quota public.

Balafoutas et al. (2016) compare the effects of gender quota and color quota. In color quota, people are divided into Team Green and Team Pink, while members of Team Pink benefit from the quota policy. The quota isn't compulsory, participants could demonstrate their attitudes by voting. Voting is either costless or require participants to pay 1 euro. The quota would only be implemented if over half of subjects agree to this policy. On the one hand, advantaged (disadvantaged) subjects support (oppose) the policy more often when voting doesn't require participants to pay anything and almost all (92.9 %) women support the gender quota. In contrast, only 57.8 % of participants who belong to Team Pink agree to the color quota. On the other hand, when voting is costly, both the ratio of female supporters (46.7 %) in the gender quota and of pink supporters (28.9 %) decrease. Overall, imposing costs reduces the proportion of disadvantaged people who vote against AA by over 60 percentage points.

The mean performance in the gender quota slightly decreases from 9.22 to 8.49 but the difference is insignificant ($p = 0.46$). Moreover, both genders don't change their behaviors. While in the color quota, of which the criterion is arbitrary, gain less support and cause efficiency losses. In the absence of color quota, on average, participants need to solve 11.75 problems correctly to win. With the implantation of costly voting and color quota, the threshold lowers to 10.73 and the difference is marginally significant ($p = 0.07$). People in pink groups perform significantly better from 7.33 to 8.96 ($p < 0.01$) when AA is enforced, while the disadvantaged perform significantly worse from 9.81 to 8.43 ($p = 0.03$). This paper implies that the effects and efficiency of quota policies largely depend on whether these policies are viewed favorably within the affected groups.

However, results of papers which analyze the impacts of existing quota policies are less optimistic than researchers' prediction based on laboratory experiments. Besley et al. (2017) evaluate the effects of gender quotas adopted by Social Democratic in 1993 on the competence of party members. Unlike other papers which use educational attainment or income, this analysis introduces a novel approach: an individual's earnings relative to people of similar age and similar labor-market characteristics. They argue that this new measure is able to predicts political success for politicians and have strong and positive

associations with enlistment tests correlates and with policy success. Results show that the increase in competence of candidates is driven by the new way to select male politicians which force mediocre male leaders to resign, while there are no significant changes in female competence.

Ahern and Dittmar explore the gender quota which adopted in Norway in 2003 and requires that two-fifths of Norwegian firms' directors be female. The number was only 9% at that moment. They maintain that this quota policy lead to a dramatic decrease in the stock price upon announcing and a significant drop in Tobin's Q over the following years. Companies increase the share of female directors by reducing the number of male directors as there are no large changes in the average size of corporate boards. The new female directors are less experienced as only 31.1 % of them have CEO experience, compare with retained male directors (69.41%), existing male directors (65.2%) and retained female directors (43%). The new female directors are also younger (45.8 years old), compare with the exiting male directors (52.9 years), the retained male directors (54.1 years), and the retained female directors (48.2 years). By comparing the male directors who left and who retained, the authors conclude that firms value older and more experienced directors. The analysis shows that a 20% increase in female share results in a near 12 percentage points reduction in the ratio of directors who have CEO experience.

5. Relative importance

Gender differences in competitiveness have been replicated in many papers which use the same or slightly changed experimental designs. However, the relative importance of this emerging explanation is not remarkable as it was in the lab. Also, there're several unclear questions remain to be figured out. In this section, I will first go through several papers which assess and compare the effect of different factors in explaining gender pay gap. Then I will list existing problems and difficulties which further research should take into account and enrich this theory. So far we have been talking about the advantages of being competitive, choose a more prestigious major, have a higher possibility to be promoted, etc. However, competitive people may be more likely to engaged in tasks which beyond them, too. Therefore, in the final part, I will introduce some negative sides of competition.

5.1 Relative importance

Blau and Kahn (2017) use published data to explore the gender pay gap in US and argue

that even though psychological factors such as gender differences in competition preference provide a new perspective, the quantitative evidences show that these factors only contribute to a small to moderate portion of gender pay gap. The impacts are smaller than that of occupation and industry, which could be affected by psychological factors, though. They take educational attainment and labor market experience as 2 measures of human capital. In 1981, on average, men had nearly 7 more years of full-time work experience than women do. By 2011, the gap is only 1.4 years. At the same time, women study longer than man and are more likely to hold an advanced degree. They state that by 2010, the traditional reason, that is, human capital now only explained little of the gender pay gap: female-to-male log wage ratio is from 79.3 % with no adjustment to 82.1 % after controlling human capital. The full specification, in which variables such as race, occupation, industry and collective bargaining are included, the ratio is 91.6 % and considerably higher than 82.1 %. This result suggests gender differences in occupation and industry continued to be important in explaining gender pay gap. Using the Oaxaca–Blinder decomposition, they show that in human capital specification, male advantage in experience contribute about 25 % of the gap in 1980 while this number reduces to 16 % in 2010 as the reversal of schoolings. These figures are quite similar in the full specification which suggests that the impact of human capital operates primarily within industries and occupations. The portion of gaps that could be explained by the differences in industry and occupation increases substantially: nearly 20 % in 1980 and over a half in 2010, making occupation and industries the two most relevant factors for the gender pay gap. In the full specification, the unexplained factors remain to account for a remarkable share of gender gap: 49 % in 1980 and 38 % in 2010.

In order to better understand the exact effects of changes in the unexplained factors and other coefficients on changes in the gender wage gap, Blau and Kahn (2017) adopt an approach developed by Juhn, Murphy, and Pierce (1991). In the human capital specification, women's improvements in education and experience together help narrow about 38 % - 40 % of the gender pay gap. Therefore, human capital did play a relevant role in explaining the wage gap. In the full specification, changes in occupation, industry and unionism in total close about 26 % of the gender gap. Women shift from administrative and services jobs to managerial and professional jobs. The role of unions in accounting for gender differences in wages has virtually disappeared, as have gender differences in unionization. However, in contrast with the convergence in occupation and union status, gender differences in industry remain. In both specifications, the decline in the unexplained factors account for 58% of the

closing. It's an open question what factors contribute to the reduction. Based on the results of Blau and Kahn (2006), Blau and Kahn (2017) attribute to decrease in discrimination, gender differences in unobservable characteristics and changes in labor demand.

Moreover, the gender pay gap declined much more slowly at the top-income level than at the middle or bottom and by 2010 was noticeably higher at the top. Women's career interruptions and fewer working hours remain significant in high-skilled occupations, possibly due to compensating differentials. Gender differences in occupations and industries, as well as differences in gender roles and the gender division of labor remain relevant, and research based on experimental evidence strongly suggests that discrimination cannot be discounted.

Goldin (2014) finds that the gender pay gap is much lower than it used to be and the decline could be attributed to an decrease in gender gap in human capital as education at all levels increased for women relative to men. The gradually converged human capital investments of both genders make the remaining unexplained portion rise relatively. Therefore, the author develops a framework which shows that the position with the highest output per unit time is also the one with the highest penalty with regard to reduced hours. An employee who wants work shorter hours would turn to jobs that have a lower penalty but also a lower output per unit time. If worker keeps pursuing the shorter work times, eventually they would take the reservation job, of which pay with respect to hours is completely linear.

In this paper, job flexibility includes the number of hours, precise times, predictability and ability to schedule one's own hours as not only the length of work times matter, but also employers want their employees to work at the particular moment and work continuously. If it's easy for an employee to switch to other similar jobs without costs, then there is no premium in earnings with respect to the number or the timing of hours, which is linear. If not, then pay with respect to work time tends to be with nonlinearity.

This compensating differentials model explains earning differences caused by the costs of flexibility. Hence it supports that cross occupation earning differences is due to productivity differences in the workplace, rather than differences in human capital. Goldin (2014) use business (which I will discuss in detail below) and law-related occupations as the one with stiff punishment for reduced work times and pharmacy as jobs which have linear elasticity of earnings with regard to work times.

By focusing on MBAs graduates from the University of Chicago from 1990 to 2006, Bertrand, Goldin, and Katz (2010) find that there's no significant gender pay gap at the beginning of their career, but the gender earning differences soon emerge and then develop:

there's a 30 log point difference in annual earnings at five years after MBA study, grows to 60 log points (namely women earn 45 % less than men do) at 10 to 16 years graduation. Three factors which can explain 84% of the gap are: 1) (24 % explained by) pre-MBA training, such as finance courses; 2) (30 % explained by) career interruptions and job experience, career interruption refers to six-month or longer period during which the subject is not working; 3) (30 % explained by) weekly working hours. Thus, Goldin (2014) concludes that about two third of the penalty from job interruptions is due to take any time out.

At 10 to 16 years after the completion of MBA, 23 % of female graduates who stay in the labor market work part-time and, interestingly, over half of those part-time female workers are self-employed. Around 17 % of female graduates are not currently employed. In contrast, less than 1 % of males are not working in any following year. About 92 % to 94 % of males work full-time in any given post-graduation year but the ratio of females from 89 % at the outset of career but reduces to 62 % 10 years later. For women who have child, this ratio is even lower as 52 %. The portion of males who had at least one career interruption is 4 % one year after graduation and 10 % at 10 years after graduation. The corresponding figures for women are 9 % and 41 %. On average, the cumulative not-work-year of women is 0.28 years by 6^{th} year, and 0.57 years by 9^{th}; for men, the numbers are 0.07 and 0.10. Ten years or more after graduation, the figures are 1.05 for women and just 0.12 for men.

Child is the main reason for changes in female labor supply. Women with children work 24 percent fewer hours per week than men or than women who are childless. The magnitudes of children effect depend on spousal incomes. MBA moms who marry high-earning husbands are 18.5 % less likely to enter labor market than those who marry low-income husbands. If MBA moms who marry high-income husbands work, their working hour is only 81 % of work time of those who have low-income spouses.

The impact of the first birth on female labor supply increase over time. For working women, their working hours and participation decrease by 17 % and 13 % respectively one year after the first birth, and increase to 24 % and 18 % three or four years later. Some women with children have difficulties in work-life balance if work is quite inflexible.

Goldin (2014) criticizes that both field and laboratory experiments which show that women are more competition averse usually forget to take jobs that value competition the most. The author argues that these types of jobs have such following features: 1) Pay with respect to working hours is highly convex. 2) Employees who work longer are more likely to receive

the reward; 3) Working hours alone, could be awarded. Therefore, staying in these positions and work continuously might have a much larger effect than the willingness to compete. In line with Blau and Kahn (2017), the author maintains that the majority of the current gender pay gap results from within occupation earnings differences rather than from between occupation differences. The within occupation gender pay gap is largely explained by the emphasis put on the work times and job continuity of employees. The gender gap is low in occupations which earnings are linear with respect to work time; the gender gap is larger when there is nonlinearity.

To summarize, both Bertand (2011) and Niederle (2016) argue that studies which attempt to establish the empirical relevance of competitiveness for economic outcomes usually fail to detect significant effects found in the lab. This also applies to some papers which use lab experiment but don't find the similarly substantial gender gap (38 %) in Niederle and Vesterlund (2007). The economical meanings for actual outcomes remain to be determined.

5.2 Unsolved Issues

Bertand (2011) calls for the robustness checks which take repetition and learning into account. Several papers have shown that to some extent socialization shapes the competition preferences. The boost in female labor force participation rate shakes up the traditional female roles: homemaker and caregiver. However, in most papers, experiments are one-shot. Even in a few of them, subjects are tracked over time, those analyses don't record the change in competition preference and the measure of competitiveness solely bases on results of an experiment several years ago. It's straightforward that women might reduce work times after having a baby, but would they become more unwilling to compete because they need to take care of their babies or become more competitive to fight for more resources to feed their babies? Hence, Future work should look into the question of how this trait evolves over time. When analyze time series data, specific changes in environments, such as quota policy, labor demand changes and technological advances are unneglectable.

The second issue is to what degree gender composition could affect participants' preferences. Stereotype threat has strongest impacts when women perform a task which is considered to favor men in mixed environment. At the first glance, Gneezy et al. (2003) seems to document that women only react to competition in a single sex environment. However, after in-depth analysis, it's the compensation scheme rather than gender competition that change women's behaviors. Hogarth et al. (2012) measure the

competitiveness by the leave decision instead. For women, the chances of making wrong leave decisions decrease when more women present. When the proportion of women in group is about 65 %, the probability of incorrect withdrawal is approximately the same for women and men. If the ratio keeps increasing, men would make more false exit decision. Price (2008) shows that women only start to compete when the ratio exceeds 58%.

Further research need to investigate competition preferences in high-stake experiment. There are several ways to do it. First, researchers could conduct experiments in developing countries where the pricing level is much lower. However, people from developing countries usually have low educational level, which makes researchers have to adopt simple tasks or enlarge sample size if they insist on complex tasks such as calculation. The cost advantages is about to disappear if more participants are recruited. Second, researchers could use opportunity cost of attending the specific university. For example, Ors et al. (2013) show the average salary of HEC alumni. The corresponding weakness is that researchers can't detect risk preferences of participants. Third, look into TV shows which involve real and substantial payoffs and participants of different backgrounds. However, TV channels usually seek public attention which could recruit participants selectively and exaggerate the conflict. Also, future studies should aim to confirm that competition preferences have independent effects, rather than reflections of already identified attributes such as risk attitudes and social preferences.

Niederle (2016) points out two methodological issues when studying gender differences in competitiveness: First, unlike traditional economic experiments, this type of experiments usually require participants to put effort in performing tasks such as calculation or solve mazes, while it's complicated to measure participants' efforts. For example, if women bear higher costs in improving performance, then one can expect that women might choose to put less effort as payoffs are quite small(in Niederle and Vesterlund (2007), winners only get 2 dollars for each correctly solved problem in tournament). Furthermore, it's unknown the how efforts affect performance. If efforts are not sensitive to performance, which means a low elasticity, then changes in compensation schemes may only lead to changes in efforts. Charness, Gneezy and Henderson (2018) assert that a skill or ability task would lead to greater variance in performance, hence larger sample sizes are needed. Skills may be positively correlated with utility of performing task, therefore less cost of effort. That's why different tasks which, according to stereotype, favor different group of people are used. A lot of related papers attribute the performance improvement from piece rate to tournament to learning, which is also a problem. Participants may improve at translating effort into

productivity over time.

Second, not only the gender of the participant but also the gender of the other subjects matter. There're several solutions. Some researchers try not to mention any thing about gender before, during and after experiments which might let participants unaware that they are taking part in a sex related experiment. The other way is to allow participants to see people in their groups, and then they can determine the gender of their competitors.

Also Niederle (2016) state that results of field experiments, mirror results in the lab though, cannot directly assess the role this newly developed explanation play in explaining for gender differences in educational or labor market outcomes.

First, the representativeness of sample is unknown. In this field experiments, participants are drawn from nonstandard subject pool. Field evidence on this certain group of people, such as gender differences among Tanzanian entrepreneurs, may have limited applicability to American entrepreneurs or general population in Tanzania. Also, gender difference in a trait doesn't mean it's the cause of gender pay gap among this group of people, let alone among general population. External validity just shows this trait still exists outside of the lab, while is insufficient to conclude this trait is important for educational and labor market outcomes.

Enlarging sample size might result in different findings. For instance, both 9-12 years old Swedish kids in Cárdenas et al. (2012) and 7-10 years old Swedish kids in Dreber et al. (2011) are asked to perform running and rope skipping task. Both of these two papers find no gender differences in running competition while in Cárdenas et al.(2012), girls are more competitive when skip rope. Cárdenas et al. (2012) attribute the differences in results to their lager sample size. Niederle (2016) consider Cohen's d (Cohen, 1988) as a useful tool for evaluating the economical relevance in differences between male and female means.

On the other hand, Fréchette (2016) assert that unlike students, the nonstandard subjects allow researchers to better test the robustness of findings and understand the nature of this certain group. Experiments which use representative sample have four main disadvantages. First, the high cost. The cost of recruiting students is lower than recruiting average men. Thus the same amount of money might generate different incentives for different people, reduce the validity of results. Large sample size would make the experiment less affordable. Second, the low availability. It's much easier to recruit students while researchers are lack of access to representative people. Those two factors combined result in the third problem: low replicability. Last but not least, the limits to control the experimental environment. Subjects may understand the instructions in a different way, introducing noise in the

observed data. In order to make sure that every participant understands and complete regardless of literacy, the optional tasks are limited. Experiments conducted in developing countries are usually physical and easy tasks, such as throw balls into bucket (Gneezy et al., 2009) or arrange shapes of small wooden blocks (Flory et al., 2018).

Second, it's relatively difficult to find field evidence, based on existing data. Thus, further studies need to combine good measures of competitiveness with real world outcomes in databases. For example, in some papers, competitiveness degrees are measured by the degree of consent to one competition-related question. Hence, create data sets which include both laboratory and field measures would be helpful for determining the explanatory power of this psychological trait.

5.3 Is competitiveness a desirable trait?

5.3.1 Efficiency Loss and Individual Wellbeing

Competitive people may gain more disutility from losing the competition. Also social comparison could put competitive people under huge pressure and hence hurt mental health and personal wellbeing. Grasseni et al. (2018) study the relationship of competition preferences, positionality and individual wellbeing. People who are competitive may concern about their relative position in the society and compare themselves to others more frequently. In this paper, competition preferences are measured by to what degree participants agree with the following two statements:1) Competition is good because it motivate people to work harder and ignite new ideas; 2) A person be rich only by sabotaging others. The first statement considers competition as positive sum game while the second considers competition as zero-sum game. Participants are therefore grouped into three types: positive, negative, remaining. Wellbeing information is collected by the survey which includes questions about life satisfaction and about satisfaction with interpersonal relationships. Participants are asked about evaluate their positional concerns with income, leisure, education, etc.

They find that distaste for competition increases the probability of being positional, especially on income while a positive attitude toward competition increases life satisfaction. There are no gender differences in competition preferences. However, distaste for competition is detrimental for women. The adverse effect of negative opinions of competition is stronger for women and the positive effect of competition preferences is stronger for man. They don't find correlations between the degrees of positionality and

happiness.

Niederle and Vesterlund (2007) shows that, from the society's point of view, there's efficiency loss due to the under-entry of high ability women and over-entry of low ability men. People who may irrational decision also suffer losses which could be avoided if they switch to another payment methods.

Eckel et al. (2015) study the relationship between gender composition and price bubbles. Basing on the assumption that men are more risk prone and competitive, they hypothesize that all-male market would generate larger bubbles. Results show that in financial asset markets run by female traders, bubbles don't occur. In some of them, there're even negative bubbles. They adopt an original approach, that is, Type A personality theory[13], to assess the competitiveness of participants. They do find that competitiveness seems to contribute to the bubbles even they don't detect any significant gender differences in competitiveness.

5.3.2 Cheating and Lying

Schwieren et al. (2010) study the relationship of competition and cheating. In the experiments, subjects are required to solve maze under noncompetitive and competitive context and self report the number of solved mazes to the experimenter. Subjects could take advantage of several tools to improve their results:1) indicate a larger number to the experimenter; 2) let the computer solve mazes by press the Auto-Solve button; 3) use instructive Path-Verify function; 4) modify the difficulty level. Consider subjects as a whole, they find no significant changes in the frequency of cheating, from 37.5 % to 42.4 %. However, women experience a dramatic increase in the frequency of cheating: from 29. 4 % to 60 % while men behave similar in the two contexts. The most popular way to cheat is to overstate their performance: 31 % in the non-competitive and 39 % in the competitive treatment. The authors argue that the increase in frequency and the way of the cheating behavior is entirely due to poor performance rather than sex. In both treatments, women solve mazes significantly fewer than men (22.19 versus 29). Bad performing participants might feel ashamed and therefore save face by cheating.

Similarly, Faravelli er al.(2015) require subjects to perform calculation tasks in three environments:1) self report the performance and be paid according to piece rate; 2) compete with another one and be paid according to self-stated performance; 3) choose one compensation scheme. Consistent with Schwieren et al. (2010), they report that competitive environment elicits more dishonesty behaviors, both the number of liars and the differences

[13] For details, see Friedman (1996).

between stated and actual performance increase. Moreover, people who cheat choose tournaments more often than honest people. However, there's no significant output difference between piece rate and tournament due to the high output of honest people.
Therefore, if competitiveness increases the likelihoods of cheating which mainly done by less productive people, then the gap in productivity would be broadened, making it harder to cheat and finally lead to reputational damage which participants definitely want to avoid.

5.3.3 Cooperation and Sabotage

If agents view competition as zero-sum game, then agents would be less willing to cooperate and more likely to do sabotage to make their stand out.
Dato et al. (2014) conduct a experiment which allow participants to decide the number they want to be deducted from their competitors' achieved points before each period starts. Participants get one point for each correct answer and compete in 2-people group. Each treatment consists of 8 periods. After each period, agents acquire information about their points after deduction, their choice of sabotage and the result of competition. The output of two people determined the payment of the third party, which is known by participants. There are four treatments: 1) Baseline; 2) Belief: like Baseline treatment, but participants are also asked about their beliefs about performance and the amount of sabotage they suffer; 3) Cheating: like Baseline treatment, but now the number is added to the participants' own points rather than deducted from competitors' points. 4) Gender: like Baseline treatment, but the gender of the opponent is revealed before treatment. They find no gender differences in performance except in the Gender treatment, in which men significantly outperform women in mixed groups. Males not only select significantly higher number but also believe that their opponent would choose larger sabotage. In the Baseline treatment, the average number of chosen sabotage is 26.66 points for men and 12.99 points for women, this is a substantial gap considering the achieved points are 114.87 for men and 111.43 for women. They don't find any significant gender differences in risk preferences, probably due to the homogeneity of the sample group. They also rule out the possibility of gender differences in human value and social preference.
The belief male participants hold that they would suffer severe loss needs to be treated with caution as it only explains the gender differences in sabotage decision to some extent, the authors maintain an additional explanation: joy of winning. Some agents would gain additional, and under most circumstances, non-monetary utility from winning a tournament. Men have higher likelihoods of winning but profits are not significantly different from that

of women because sabotage is costly.

Cárdenas et al. (2015) study the relationship of competitiveness and cooperativeness among young children and report the willingness to compete in a math task is correlated negatively with cooperativeness.

5.4 Robustness check: public observability

Experiments, both in the lab and field study, are typically conducted in anonymous environments. However it's not the case in real world, where competitions are often public and hence result in concerns about social image. If women care more about their social images than men, this could widen the gender gap. For example, women may shy away from competition because they view competitiveness as an aggressive and inappropriate characteristic for females. Therefore Buser et al. (2017) design an experiment to assess the effects of observability on the willingness to compete. Overall the experimental design closely follows Niederle and Vesterlund (2007) but insert an annoucement part between decision-making and task-performing. Subjects don't know their relative performance during the experiment. After participants pin down which compensation scheme they want to apply for their performance, participants in the *Public Choice* group need to announce their first names and their choices in front of the public. Participants in the *Control* group only need to announce their first names while Participants in the *Public Outcome* group only need to announce their decisions. Participants in the *Private* are not required to do anything.

Then participants would participate in a mock interview to decide whether candidates are qualified for an internship at a bank based on the CV and how they feel about these candidates. They manipulate the gender, competitiveness level and other demographic characteristics of candidates to elicit their attitudes toward competition as an outsider.

Overall men are much more likely to enter the tournament (57 % vs. 22 %) and have better performance under piece rate and the tournament. In the *Control* treatment, the chance of tournament entry is 55 % for men and 23 % for women. In the *Private* treatment, both genders are more likely to enter (61 % for men and 28 % for women) but the gender gap is similar to that in the *Control* treatment, which suggests introducing oneself has no effects on gender differences in competition. In the *Public Choice* treatment, the gender gap is 39 % and in the *Public Outcome* treatment, the gender gap is 26 %. As hypothesized, women are slightly less likely to select competition when names and choices are public and

men are slightly less likely to opt for competition when choices become public. However, the public observability only has limited effect, in terms of size and statistical significance. In the mock interview, participants are less willing to work with or hire highly competitive individuals, regardless of their gender. Hence there's no evident backlash against women.

5.5 Robustness check: repetition of competition

5.5.1 Performance

In existing studies, competition is usually one-shot, while in real world competition is continuous and repeated.

Cotton et al. (2013) conduct five-round competition. Student participants are matched with another classmate and answer math quizzes in 5 minute, only the person who solve more questions correctly within the time limit would be rewarded. If both solve the same amount of questions correctly before running out of time, the person who submits earlier would win. Plus, in the treatment group the winner could get a bonus as he submits before the time limit. In each round, participants are paired with a new opponent. There is a short break between each competition and the whole experiment takes about one hour. Participants know neither individual performance nor relative performance. There are two types of control groups: the one answer math quiz but without time pressure and the other do the same as the treatment group but answer 10-question reading quiz. At the beginning of the experiment, participants are informed of the number of rounds, the number and sources of questions, the identity of their competitor, procedure for determining their competitors and winners in each round. The questions of quizzes are randomly drawn from state assessment, as the scores could serve as a baseline of gender differences.

The quiz is 5 questions for 4 classrooms, 10 questions for 5 classrooms, 15 questions for 3 classrooms. Two weeks later, the 4 classrooms whose students answer the 5-question quiz participate in the competition in the second period and the number of questions remains unchanged.

The variable of interest is the student's score in each round, which could be affected by innate ability and gender of the student and gender of his competitor. The innate ability in certain round could be influenced by performance of the student in other rounds, while performance of the student in other round is, in reverse, influenced by innate ability and competition preference. Therefore they use the prior year's state assessment score as an instrument for innate ability. As questions are also drawn from the state assessment, the

prior scores tell about student's performance when in the non-competitive environment.

In the first round, male students outperform equally able female students significantly with the male coefficient equals to 0.34. Males perform somehow better when competing in a single-sex context. Similarly, females perform slightly better when they are paired with another female. The gender of the competitor only has limited effect on both genders' performance. The results are in line with previous studies which only conduct experiment once.

However, the male advantage only emerges at the first round and disappearing in the following four rounds. Males even perform worse than females of similar ability. The coefficient on male in the second round is −0.01 while ability has a stronger effect. Female performance decrease when move to the second round, while increase in the last three rounds. Females perform significantly better in the last three rounds than in the first two rounds. Men, however, decrease dramatically when moving to the second round (p-value < 0.001), and the performance is slowly improved in the following rounds. Further analysis show that the male advantage is caused by the overperformance of low-ability males and underperformance of high-ability females. For all quiz lengths, the male advantage only appears in the first round, with the largest appear in classrooms whose students are asked to solve 15 questions. Compare with females, males solve more questions while has a similar accuracy rate. In the first round of 15-question quiz, males solve 1.6 more questions (p = 0.036) but no gender difference in correct rate. In the following rounds, males answer 0.14 additional questions and the accuracy rate is only 1 percentage point higher than females in the first round. They argue that the closed and reversed gender gap is not because students view the task unimportant and therefore treat it as non-competition.

There's no significant gender difference in performance in the 2-week later competition. The coefficient on male is either zero or positive, both are significant. Therefore they conclude that male advantage disappeared for at least two weeks after the initial competition, and if it appears again, the magnitude is much smaller.

There is a control group to explore performance and gender gap without time pressure, that is, students won't get bonus from submitting before time runs out. Results show that there's no gender difference in performance in all five round, hence the authors argue that time pressure is important for the emergence of the initial male advantage and that females perform better in a competitive environment without time pressure. Moreover, the accuracy rate is higher (on average, 65 %) as the highest correct rate is 61% when students perform under time pressure. In this setting, students concentrate on skill rather than put undue

emphasis on speed.

For students who answer the reading quiz, the female advantage appears in the first round and persists across five rounds. These results show that any gender differences in the reading treatment do not change across periods of competition. However, there's no gender difference in response to competition as the gender gap in performance is almost same as the gender gap in reading part of prior state assessment. Combined the results found in two control groups, the authors maintain that both the design of the competition (with or without time pressure) and the task (math or reading) are crucial for the gender differences in competition.

5.5.2 Response to failure

Buser et al. (2018) build a connection between lab and Dutch Math Olympiad to examine the effect of past failure on re-entry to the competition. The lab experiment bases on the approach developed by Niederle and Vesterlund (2007) while students are divided into three groups which receive different treatment: main, feedback and risk.

The *Main* treatment is composed of six rounds. In each round, participants are matched with another new person without knowing the identity of their opponent and decide which compensation scheme (piece rate or tournament) they would like to apply to their performance. At the end of each round, only participants who opt for competition know their individual performance and the result of competition.

In the *Feedback* treatment, the only differences are that participants who select the piece rate now would receive same feedback as participants who opt for competition as they would be paired with a hypothetical opponent and that there are only four rounds.

In the *Risk* treatment, the winner of competition is randomly determined. Participants who choose the competition would 2 points with probability x and nothing with probability 1-x (x=0.3, 0.4, 0.5, 0.6, 0.7 and is fixed during the treatment) for each correct answer. Before getting started, participants are told about their probability x. The question of interest in the lab is: how would participants react to the outcomes of competition in the first round?

In the *Main* treatment, both males and females who win the competition in the first round insist on their choice in the following five rounds and there's no gender difference in response to the initial victory (4.6 out of 5 times for men and 4.2 out of 5 times for women). However, males and females react to losing in a very different way. On average males enter 2.6 times while women only enter 1.1 times ($p = 0.04$). The gender difference in the response to the failure is still significant after conditioning on to controlling for absolute

and relative performance. Compare with male winners, males losers are 24 % less likely to re-enter in the subsequent competition while females losers are about 59 % less likely. The initial failure has almost no effect on high-performing men[14] while strongly discouraged high-performing women from re-entering, the probability is 56 percentage points lower than high-performing female winners.

The reactions to competition are similar for both genders in the *Feedback* treatment. On average, men enter 2.9 times and women enter 2.8 times in the following three rounds if they win the competition. Women who experience failure are also much less likely to re-enter than men (1.5 vs. 0.6, $p = 0.02$). The propensity of participating in the subsequent competition for male losers reduced by 28 % while for female the data reduced by 56 %. For participants who opt for the piece rate but lose, there's no gender difference in response. Men enter the competition 0.2 times and women enter 0.3 times. However, males who perform better than their opponent in the piece rate enter the competition more frequently than women (2.4 vs. 1.5; $p = 0.02$). Piece rate male winners are 56 % more likely to enter the tournament, while the number for women is only 28 %. The gap even widens among high-performing participants.

In the *Risk* treatment, men are 21 % more likely to engage in the risky competition, this gender gap in tournament entry decreases and are no longer significant over the following rounds. For participants who opt for the competition in the first round, the gender gaps in tournament entry over the subsequent competition disappear. For participants who choose the safe piece rate in round 1, female are even more likely to take risk from round 2 onwards. Findings of Risk treatment suggest that risk attitudes don't play a role in tournament entry.

In both *Main* and *Feedback* treatments, the largest gender differences in the response to failure appear in round 2, and gradually decrease in the subsequent rounds but persist. The differences are not result from gender differences in confidence, belief updating, performance improvement, and risk attitudes. The initial failure has a strong adverse affect on competition preference for female loses in the round 1 competition. Consistent with Niederle and Vesterlund (2007), high-performing women bear higher under-entry cost than men. Among top 25 % of participants, women lose 57 cents per round while men lose 35 cents.

The Dutch Math Olympiad is an annually math competition and the ultimate goal is to select a national team for the International Mathematics Olympiad. Students in tenth or

[14] In this paper, high-perform participants refer to the group of people whose chance of winning exceeds 50%.

eleventh grade could register for the campaign. There are two rounds of competition and only a fixed amount of top performers could proceed to the second period of competition. However, only tenth grade students could register again in the next year no matter how their performance is. This institutional setting allows the author to do regression discontinuity design.

Among 11,591 candidates, 63.1 % of them are boys which in line with Buser et al. (2014) and Buser et al. (2017) that girls are less willingness to compete in mathematical areas. 14.4 % of male candidates and 9.0 % of female candidates advance to the second round. The probability of advancing is almost identical (91 % for men vs. 89.4 % for women) after controlling for performance. However, consistent with the lab results, the probability of re-enter the competition for girls who fail to reach the threshold girls reduced by 20 % to 40 % while boys seem to be not affected. As the highly competitive national and forthcoming international Olympiad mainly attract talented students, the findings may help in explaining why female students are still lag behind in STEM fields. Also, the effects may deter women from seizing the opportunity for promotion at the workplaces.

Similarly, Hoyer et al. (2016) also focus on Dutch university students and confirm that past failure would deter women from self-selection into competition. 51 % of male students and 33 % of female students select the tournament. The gender gap emerges among students who are not in the top 25 %, in particular there's a gender gap (59 % vs. 25 %) among the bottom 50 % of students. By contrast, 48 % of the top 25 % of female students enter the tournament, as often as male students do. As subjects don't know their absolute or relative performance throughout the experiment, low-ability female students make their entry decision based on the prior academic performance at high school. Specifically, they use prior Math scores to assess their probability of winning. However, men and talented female students are not affected by their high school experiences.

Buser et al. (2016) explore whether people would be more conservative and choose safer options, or keep seeking challenges after experiencing failure. In the experiment, after completing the first round two-people winner-take-all tournament, each participant is informed of absolute performance, and two-thirds of them would know the results of competition. The differences of available information create three groups: winner, loser, or no information (which serves as control group). In the second round, participants need to decide the incentive compensation. In addition, they are required to set a goal for their performance. The upper bound is 20 questions. They would be paid only if they live up to their expectation.

On average, subjects solve 10.4 questions correctly in the first round; the chosen target is 9.3 and participants solve 10.5 questions correctly in the second round. The success rate is 81 %.

There are no gender differences in performance or selected goals. Men and women of same ability pick on the same goals.

After controlling for performance in the first round, they find that losers select a significantly challenging goal than winners, with a coefficient value of 0.5. Winners and control group choose similar targets. Contrary to expectation, losers are not discouraged by failures. However, losers still perform significantly worse than winners, solve 0.8 fewer questions. The worse performance and higher target lead to a 17-percentage-point success rate gap between winners and losers.

However, male and female losers react to their failures differently. The higher target is driven by male losers. On average, compare with male winners, they expect themselves to solve about 0.95 more question ($p < 0.05$). The worse performance is resulted from female losers. On average, female losers solve 1.7 fewer questions than female winners ($p <0 .05$). Average women choose a significantly lower target than male losers (differences: 0.74, $p = 0.051$) whereas slightly higher than male winners differences (0.29, $p = 0.286$). Average males significantly outperform female losers (difference: 1.1, p = 00008) whereas female winners perform better than average males (difference: 0.54, p = 00293). This indicates that both the effect of on challenge seeking for men and on performance for women are caused by failure rather than success.

Berlin et al. (2016) provide support and explanations for Buser et al. (2016). They find women and men don't react in the same way. When participants receive negative feedbacks that their performance is below median, women update their beliefs more pessimistically than men of equal ability. The first four rounds are organized exactly same as Niederle and Vesterlund (2007), while before making decisions at the third round, students need to guess their ranks in the second round. Before the fifth round, students are told whether their performance is below or above median. After providing feedbacks, the authors again ask students to guess ranks in the second round and let subjects to choose compensation schemes for the performance in fifth round. In the sixth round, participants are divided into two groups and required to apply their performance in the first round in which they perform in a non-competitive piece rate environment. Participants in Ability Group compare their first-round performance with opponent of same ability, while there are no changes for control group consists of participants in Repetition.

There are no gender differences in performance in first two rounds, and also no differences in the beliefs about ranks among below-median participants before they receive feedback. However, below-median men become more confident than female counterparts after obtaining feedback. By contrast, above-median males are always more confident but the feedback closes the gender gap to some extent. Women become less confident after receiving negative feedback and more confident after receiving positive feedback. After being told they are below median, compared to an agent who his beliefs about performance in the second round in a Bayesian way, women are more likely to think they are in the bottom 25 %.

As the authors use diff-in-diff to estimate the effects of feedback on sixth-round decision, the variable of interest is thus the interaction term AbilityGroup * Decision. The coefficients are positive and significant for both below-median women (0.30, $p = 0.04$) and men (0.33, $p = 0.02$). The negative effect is limited on above-median males and females. The coefficient on the interaction term Belief* Decision is significantly negative for women, indicates that women react more strongly to feedback when they did not expect it. Further analysis suggests that the effects of feedback on sixth-round decision work through beliefs. Women respond to feedback on their individual performance whereas men respond mainly to the level of their opponent. Below-median men believe they would perform better over time after entering a tournament whereas it's not the case for females.

6. Conclusion

Despite considerable progress over the past several decades, the gender gap in STEM fields and wage remains persistent and substantial. Gender differences in competition offer a new perspective for gender differences in real outcomes apart from traditional factors such as gender differences in labor supply and discrimination. The experimental designs introduced by Gneezy et al. (2003) and Niederle and Vesterlund (2007) serve as building blocks for future researches to detect, evaluate and analyze this issue.

There is a broad consensus that men exhibit less competition aversion than women in economically advanced countries. Current literatures suggest that a number of factors help shape competition preferences at different stages in the human life cycle. Existing analyses also show that this psychological factor has potential to explain gender differences in education and labor market outcomes.

However, the relative importance remains unclear. Several papers argue that though this

explanation is intriguing, could only account for a small proportion of gender pay gap. Besides, there are three main problems. First, only a few papers try to connect willingness to compete with real life outcomes despite of despite abundant empirical evidence in the lab. Hence, further studies should prove this explanation still make sense outside the lab rather than just determine an already identified pattern. Second, robustness check is needed for lab experiments. How agent would behave at large stakes and in several rounds of competition is still an open question. Also, under most circumstance, the sample consists of high school or college students. To what extent the results derived from these two groups could apply to general population remain to be determined. The main driving forces of gender gap in willingness to compete are crucial for policymakers, yet still under debate. Moreover, before encourage women to enter the competition, it's important to figure out would competitiveness be a desirable trait? Last but not least, to my knowledge, until now there are no papers track changes in competition preference over time, which is unrealistic. The scarcity in field data causes troubles for building causal relationships.

Overall, the regarding literature is growing and rich. This new explanation provides the possibility to interpret gender gap from a psychological point of view. However, this line of research is just a beginning, there are still a lot of problems further researches have to cope with.

Reference

Ahern Kenneth R., Dittmar Amy A. 2012. The changing of the boards: the impact on frim valuation of mandated female board representation. The Quarterly Journal of Economics. Vol 127(1): 137-197

Akerlof George A, Kranton Rachel E, 2002. Identity and schooling: some lessons for the economics of education. Journal of Economic Literature. Vol 40 (4):1167–1201.

Almås, Ingvild, Cappelen Alexander W., Salvanes Kjell G.,Sørensen Erik Ø., Tungodden Bertil.2015. Willingness to Compete: Family Matters. Management Magazine. Vol 62(8), p.2149-2162

Altonji, Joseph, Blank, Rebecca, 1999. Race and gender in the labor market. Handbook of Labor Economics. Vol 3143-3259

Andersen, S., Ertac, S., Gneezy, U., List, J. A., & Maximiano, S. 2013. Gender, competitiveness, and socialization at a young age: Evidence from a matrilineal and a patriarchal society. Review of Economics and Statistics, 95(4), 1438–1443.

Antonovics, Kate, Arcidiacono, Peter and Walsh, Randall, 2009. The effects of gender interactions in the lab and in the field. The Review of Economics and Statistics. Vol 91 (1), 152–163

Babcock, Linda, Recalde, Maria P., Vesterlund, Vise and Weingart, Laurie.2017. Gender Differences in Accepting and Receiving Requests for Tasks with Low Promotability.American Economic Review. Vol 107:714-747

Balafoutas, Loukas, Davis, Brent J. and Sutter Matthias.2016. Affirmative action or just discrimination? A study on the endogenous emergence of quotas. Journal of Economic Behavior & Organization. Vol 127:87-98

Balafoutas Loukas, Fornwagner Helena,Sutter Matthias. 2018. Closing the gender gap in competitiveness through priming.Nature Communications. Vol 9(1):1-6.

Balafoutas Loukas, Kerschbamer Rudolf, Sutter Matthias. 2012. Distributional preferences and competitive behavior. Journal of Economic Behavior & Organization. Vol 83:125-133

Bartling, Björn, Ernst Fehr, Michel Andre Marechal,and Daniel Schunk. 2009. Egalitarianism and Competitiveness. American Economic Review, Papers & Proceedings 99(2): 93–98.

Bartling, Björn, Ernst Fehr, and Daniel Schunk. 2012. Health Effects on Children's Willingness to Compete. Experimental Economics. Vol 15(1): 58–70.

Berlin Noémi, Dargnies Marie-Pierre.2016. Gender differences in reactions to feedback

and willingness to compete. Journal of Economic Behavior & Organization. Vol 130:320-336

Bertrand, Marianne, Goldin, Claudia and Katz, Lawrence F.2010. Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors. American Economic Journal: Applied Economics Vol 2: 228–255

Besley, Timothy ; Folke, Olle ; Persson, Torsten ; Rickne, Johanna. 2017. Gender Quotas and the Crisis of the Mediocre Man: Theory and Evidence from Sweden. American Economic Review. Vol 107(8): 2204–2242

Betts, R. Julian. 2011. The Economics of Tracking in Education. Handbook of Economics of Education. Vol(3):341-380

Booth, Alison, Fan Elliott, Meng Xin and Zhang Dandan.2019. Gender differences in willingness to compete: the role of culture and institutions. The Economic Journal. Vol 129(618):734-764

Booth, Alison L., Francesconi, Marco and Frank Jeff.2003. A sticky floors model of promotion, pay, and gender. European Economic Review. Vol 47:295-322

Booth, Alison and Nolen, Patrick. 2012. Choosing to compete: How different are girls and boys?. Journal of Economic Behavior and Organization.Vol 81:542-555

Booth, Alison, Yamamura Eiji. 2018. Performance in mixed-sex and single-sex competitions:what we can learn from speedboat races in Japan. The Review of Economics and Statistics. Vol 100 (4): 581-593

Bosquet, Clement, Combes, Pierre-Philippe and Garcia-Penalosa Cecilia.2018. Gender and Promotions: Evidence from Academic Economists in France. IDEAS working paper series from RePEc.

Brunello, G., & Checchi, D., 2007. Does school tracking affect equality of opportunity? New international evidence. Economic Policy.Vol 22 (52):781–861.

Bryson, Alex, Forth, John and Stokes,Lucy. 2014. The Performance Pay Premium: How Big Is It and Does It Affect Wage Dispersion. IDEAS working paper series.

Buonanno, P., & Pozzoli, D. (2009). Early labour market returns to college subjects. Labour, Review of Labour Economics and Industrial Organisation. Vol.23(4), 559–588.

Buser Thomas. 2012. The impact of the menstrual cycle and hormonal contraceptives on competitiveness. Journal of Economic Behavior & Organization. Vol 83: 1-10

Buser Thomas. 2016. The Impact of Losing in a Competition on the Willingness to Seek Further Challenges. Management Science. Vol 62 (12): 3439-3449

Buser, Thomas, Niederle, Muriel and Oosterbeek Hessel.2014.Gender, competitiveness and

career choice. Quarterly Journal of Economics. 129(3):1409-1447.

Buser, Thomas, Peter, Noemi Peter and Wolter, Stefan.2017.Gender, competitiveness and study choices in high school: evidence from Switzerland.American Economic Review: Papers & Proceedings. 107(5): 125-130

Buser, Thomas, Peter, Noemi Peter and Wolter, Stefan.2017. Gender, willingness to compete and career choices along the whole ability distribution.Working paper.

Buser Thomas, Ranehill Eva, Van Veldhuizen Roel. 2017. Gender Differences in Willingness to Compete: The Role of Public Observability. Discussion Paper.

Buser Thomas, Yuan Huaiping. Do women give up competing more easily? Evidence from the lab and the Dutch Math Olympiad. American Economic Journal: Applied Economics. Forthcoming.

Cárdenas,Juan Camilo, Dreber Anna, Von Essen Emma, Ranehill Eva.2015. Cooperativeness and competitiveness in children. Journal of Behavioral and Experimental Economics. Vol 59:32-41

Cárdenas, Juan-Camilo, Dreber, Anna, Von Essen, Emma and Ranehill Eva. 2012. Gender differences in competitiveness and risk taking: Comparing children in Colombia and Sweden. Journal of Economic Behavior & Organization.Vol 83:11-23

Charness,Gary, Gneezy, Uri and Henderson, Austin. 2018. Experimental methods: Measuring effort in economics experiments. Journal of Economic Behavior and Organization.Vol 149:74-87

Charness Gary, Kuhn Peter. 2011. Lab Labor: What Can Labor Economists Learn from the Lab?. Handbook of Labor Economics. Vol 4A: 229-330

Cotton, S. Christopher, Li Cheng, McIntyre, Frank, Price, Joseph. 2015. Which explanations for gender differences in competition are consistent with a simple theoretical model?. Journal of Behavioral and Experimental Economics. Vol 59: 56-67

Cotton, S. Christopher, McIntyre, Frank, Price, Joseph. 2013.Gender differences in repeated competition: Evidence from school Math contests. Journal of Economic Behavior & Organization. Vol 86: 52-66

Dariel, Aurelie, Kephart, Curtis, Nikiforakis, Nikos ; Zenker, Christina. 2017. Emirati women do not shy away from competition: evidence from a patriarchal society in transition. Journal of the Economic Science Association.Vol.3(2):121-136

Dasgupta, Utteeyo, Gangadharan, Lata, Maitra, Pushkar, Mani, Subha and Subramanian, Samyukta.2015. Choosing to be trained: Do behavioral traits matter?. Journal of Economic Behavior & Organization.Vol 110:145-159

Dato, Simon and Nieken Petra. 2014.Gender differences in competition and sabotage. Journal of Economic Behavior & Organization.Vol 100:64-80

De Paola, Maria, Ponzo, Michela and Scoppa Vincenzo.2017. Gender differences in the propensity to apply for promotion: evidence from the Italian Scientific Qualification. Oxford Economic Papers.Vol 69(4):986–1009

Dreber, Anna, Von Essen, Emma and Ranehill, Eva.2011.Outrunning the gender gap—boys and girls compete equally.Experimental Economics.Vol 14:567–582

Eckel, Catherine C. and Füllbrunn Sascha C. Thar SHE Blows? Gender, Competition, and Bubbles in Experimental Asset Markets. American Economic Review.105 (2): 906–920

Eisenkopf, Gerald, Hessami,Zohal, Fischbacher,Urs and Ursprung Heinrich W.2015. Academic performance and single-sex schooling: Evidence from a natural experiment in Switzerland. Journal of Economic Behavior& Organization.115(2015):123-143

Equilar. 2015. 200 Highest-Paid CEO Rankings.

Faravelli, Marco, Friesen, Lana and Gangadharan, Lata.2015. Selection, tournaments, and dishonesty. Journal of Economic Behavior & Organization. Vol 110:160-175

Flory, Jeffrey A., Leibbrandt, Andreas and List, John. 2015. Do Competitive Workplaces Deter Female Workers? A Large-Scale Natural Field Experiment on Job Entry Decisions. The Review of Economic Studies. Vol 82: 122–155

Flory, Jeffrey A., Gneezy, Uri, Leonard,Kenneth L. and List, John. 2018. Gender, age, and competition: A disappearing gap?. Journal of Economic Behavior and Organization.Vol 150:256-276

Fréchette, Guillaume R.,2016. Experimental Economics across Subject Populations. The handbook of experimental economics. Vol 2:435-480

Friedman, Meyer. 1996. Type A Behavior: Its Diagnosis and Treatment. New York: Plenum Press.

Fryer, Roland G. and Levitt, Steven D. 2010. An Empirical Analysis of the Gender Gap in Mathematics. American Economic Journal: Applied Economics: 210–240

Gneezy Uri, Leonard Kenneth, List, John A.2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. Econometrica. Vol 77(5):1637-1664

Gneezy, Uri, Niederle, Muriel and Rustichini, Aldo. Performance In Competitive Environments: Gender Differences.Quarterly Journal of Economic.Vol 118: 1049–1074

Gneezy, Uri, Leonard, Kenneth L. and List, John. 2009. Gender differences in competition: evidence from a matrilineal and a patriarchal society. Econometrica. Vol. 77, No. 5： 1637–1664

Goldin Claudia.2014. A Grand Gender Convergence: Its Last Chapter. American Economic Review. Vol 104(4): 1–30

Grasseni, Mara and Origo Federica.2018. Competing for Happiness: Attitudes to Competition, Positional Concerns and Wellbeing. Journal of Happiness Study. Vol 19:1981–2008

Günther, Christina ; Ekinci, Neslihan ; Schwieren, Christiane ; Strobel, Martin.2010. Women can't jump? An experiment on competitive attitudes and stereotype threat. Journal of Economic Behavior & Organization.Vol.75(3), p.395

Heinz, Matthias, Normann, Hans-Theo and Rau, Holger A. 2016. How competitiveness may cause a gender wage gap: Experimental evidence. European Economic Review.Vol 90:336-349

Hogarth, Robin M., Karelaia, Natalia and Trujillo, Carlos Andrés. 2012. When should I quit? Gender differences in exiting competitions. Journal of Economic Behavior & Organization. Vol 83:136-150

Hoxby Caroline.M, Avery Christopher. 2013. The missing"one-offs": The hidden supply of high-achieving, low income students. Brookings Papers on Economic Activity.Vol.2013(1):1-65

Hoyer Britta ; van Huizen, T.M. ; Keijzer, L.M. ; Rezaei Khavas, T. ; Rosenkranz, S. ; Westbrock, B. 2016. Do talented women shy away from competition?. Discussion Paper Series.

Iriberri Nagore, Rey-Biel Pedro.2017. Stereotypes are only a threat when beliefs are reinforced: On the sensitivity of gender differences in performance under competition to information provision. Journal of Economic Behavior & Organization. Vol 135:99-111

Iriberri, Nagore and Rey-Biel Pedro.2018. Competitive pressure widens the gender gap in performance:evidence from a two stage competition in mathematics. The Economic Journal.

Jurajda Stepan and Munich Daniel. 2011. Gender gap in admission performance under competitive pressure. American economic review. Vol.101(3):514-518

Kamas, Linda and Preston, Anne.2012. The importance of being confident; gender, career choice, and willingness to compete. Journal of Economic Behavior and Organization. Vol 83:82-97

Kamas, Linda and Preston, Anne.2018. Competing with confidence: The ticket to labor market success for college-educated women. Journal of Economic Behavior and Organization.Vol 155:231-252

Khachatryan Karen, Dreber Anna, Von Essen Emma , Ranehill Eva.2015. Gender and preferences at a young age: Evidence from Armenia. Journal of Economic Behavior & Organization. Vol 118: 318-332

Khadjavi, Menusch and Nicklisch, Andreas.2018. Parents' ambitions and children's competitiveness. Journal of Economic Psychology. Vol 67:87-102

Kleinjans,Kristin J. 2009. Do gender differences in preferences for competition matter for occupational

expectations?. Journal of Economic Psychology.Vol 30:701-710

Koedel, Cory and Tyhurst, Eric (2012). Math skills and labour market outcomes: Evidence from a resume-based field experiment. Economics of Education Review.Vol 31:131–140.

Landaud Fanny, Ly Son-Thierry, Maurin Éric.2018. Competitive Schools and the Gender Gap in the Choice of Field of Study. Journal of Human resources.

Lavy Victor.2013. Gender Differences in Market Competitiveness in a Real Workplace: Evidence from Performance - based Pay Tournaments among Teachers. Economic Journal.Vol 123:540-573

Lazear, E.P., Rosen, S., 1990. Male–female wage differentials in job ladders. Journal of Labor Economics 8 (1, Part 2), S106–S123.

Lee, Soohyung, Niederle, Muriel and Kang, Namwook.2014.Do single-sex schools make girl more competitive?. Economics Letters, Vol 124(3):474–477.

McGee, Andrew, McGee, Peter and Pan, Jessica. 2015. Performance pay, competitiveness, and the gender wage gap: Evidence from the United States. Economics Letters.Vol 128:35-38

Morin, Louis-Philippe.2015. Do Men and Women Respond Differently to Competition? Evidence from a Major Education Reform. Journal of labor economics.Vol.33(2): 443-491

Niederle, Muriel.2016.Gender.The Handbook of Experimental Economics.Vol 2:481-562

Niederle, Muriel, Segal, Carmit and Vesterlund, Lise.2013. How Costly Is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness.Management Science. Vol 59:1-16

Niederle, Muriel and Vesterlund, Lise.2007. Do Women Shy Away from Competition? Do Men Compete Too Much?. Quarterly Journal of Economics, 122:1067–1101.

Niederle, Muriel and Vesterlund,Lise. 2010. Explaining the Gender Gap in Math Test Scores: The Role of Competition. Journal of Economic Perspectives.Vol.24(2):129–144

Ors E, Palomino F, Peyrache E. 2013. Performance gender-gap: Does competition matter?.Journal of labor economics.Vol.31(3):443-499

Park, H., Behrman, J.R., Choi, J., 2012. Causal effects of single-sex schools on college entrance exams and college attendance: random assignment in Seoul high schools. Demography.Vol 50 (2), 447–469.

Pekkarinen, T., Gender differences in behaviour under competitive pressure: Evidence on omission patterns in university entrance examinations. J. Econ. Behav. Organ. (2014)

Peterle, Emmanuel and Rau, Holger.2017. Gender Differences in Competitive Positions: Experimental Evidence on Job Promotion. IDEAS working paper series from RePEc.

Price, Joseph. 2008. Gender differences in the response to competition. Industrial and Labor Relation Review.Vol 61:320–333

Reuben, Ernesto, Sapienza, Paola, and Zingales, Luigi Zingales. 2015. Taste for competition and the gender gap among young business professionals. NBER working paper series

Reuben, Ernesto, Wiswall, Matthew and Zafar Basit. 2017.Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender. The Economic Journal. Vol. 127(604):2153-2186

Saccardo Silvia, Pietrasz Aniela, Gneezy Uri.2018. On the Size of the Gender Difference in Competitiveness. Management Science 64(4):1541-1554.

Saks, R.E., Shore, S.H., 2005. Risk and career choice. Advances in Econmic Analysis& Policy,Vol5(1): 1–43.

Samek Anya. Gender Differences in Job Entry Decisions: A University-Wide Field Experiment. Management Science. Forthcoming.

Schrøter Joensen, J., & Skyt Nielsen, H. (2010). Is there a causal effect of high school math on labor market outcomes. The Journal of Human Resources, Vol.44(1), 171–198.

Schwieren, Christiane and Weichselbaumer Doris.2010. Does competition enhance performance or cheating? A laboratory experiment. Journal of Economic Psychology.Vol 31:241–253

Sutter Matthias, Rützler Daniela. 2010. Gender Differences in Competition Emerge Early in Life. IZA Discussion Paper 5015

Walton Gregory M.,Logel Christine,Peach, Jennifer M.,Spencer Steven J.,Zanna Mark P.,2015.Two brief interventions to mitigate a “chilly climate” transform women’s experience, relationships, and achievement in engineering. Journal of Educational Psychology.Vol 107 (2):468–485

Wieland Alice, Sarin Rakesh.2012. Domain specificity of sex differences in competition. Journal of Economic Behavior & Organization. Vol 83:151-157

Wozinak David, Harbaugh William T. , Mayr Ulrich. 2014. Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback. Journal of Labor Economics. Vol 32 (1):161-198

Yagasaki Masayuki. 2018. Competitiveness, Risk Attitudes, and the Gender Gap in Math Achievement. IDEAS Working Paper Series from RePEc

Zhang, Y. Jane, ‘‘Can Experimental Economics Explain Competitive Behavior Outside the Lab?,’’ Unpublished manuscript (2012a).

Zhang Y.Jane. 2019. Culture, institutions and the gender gap in competitive inclination: evidence from the communist experiment in China. The Economic Journal. Vol 129 (617):509–5