Cancelling out early age gender differences in competition – an analysis of policy interventions[#]

Matthias Sutter, Daniela Glätzle-Rützler, Loukas Balafoutas, Simon Czermak [*]

This version: 19 May, 2015

Abstract

We study the willingness to compete of 588 children and teenagers aged ten to seventeen. We replicate the gender difference in tournament entry choices usually found in the literature for adults. We then show that policy interventions like quotas and preferential treatment help to close down the gender gap without leading to losses in efficiency, during or after a tournament. Given that differences in competitive behavior are prevalent from an early age, the application of interventions to promote females in competitions may be desirable already at early stages to promote equal chances for women on labor markets later on.

JEL Codes: C91, D03, D04

Keywords: tournaments; gender gap; affirmative action; experiment; children

Forthcoming in: Experimental Economics

[#] We thank Gary Charness, Uri Gneezy, Bill Harbaugh, Steffen Huck, Kai Konrad, Charles Noussair, Pedro Rey-Biel, Rupert Sausgruber, two anonymous referees, and seminar participants at the Economic Science Association Meeting in Kuala Lumpur and the Max Planck Institute for Tax Law and Public Finance in Munich for helpful comments and suggestions. Financial support through the Austrian Federal Ministry of Science program "Sparkling Science" (grant SPA/02-99 – Project: Gender and competition) and through the Austrian Science Fund (Project P22772-G11) is gratefully acknowledged. We thank Thomas Plankensteiner from the Tyrolean State Board of Education and headmasters Gerlinde Christandl, Max Gnigler and Hermann Lergetporer for making this study possible.

[*] Sutter: University of Cologne and University of Innsbruck. Email: matthias.sutter@uni-koeln.de. Glätzle-Rützler: University of Innsbruck. Email: daniela.ruetzler@uibk.ac.at. Balafoutas: University of Innsbruck. Email: loukas.balafoutas@uibk.ac.at; Czermak: Management Center Innsbruck. Email: simon.czermak@mci.edu. Balafoutas is corresponding author.

1. Introduction

Motivated by the well-known fact that labor market outcomes – even in highly developed countries – are often characterized by gender imbalances in terms of earnings or career advancement, a recent body of literature has investigated gender differences in competitive behavior as a potential explanation for these imbalances (see, e.g., Gneezy et al., 2003; Niederle and Vesterlund, 2007; Wozniak, 2012; Datta Gupta et al., 2013; Andersen et al., 2013). These studies show that men are generally much more likely than women to self-select into competitive environments, *ceteris paribus*.[1] The gender gap in competitive behavior appears to form already at a very young age and to persist throughout childhood and adolescence, especially in highly developed countries (Booth and Nolen, 2012; Cardenas et al., 2012; Almås et al., 2013; Andersen et al., 2013; Sutter and Glätzle-Rützler, 2015).[2]

In this paper, we examine the effects of policy interventions aimed at closing down this gap already in childhood and adolescence. In particular, we consider the following interventions in the spirit of affirmative action: *Minimum Quotas*, which determine a minimum representation of women among the winners of a tournament, and *Preferential Treatment*, which, as the name suggests, amounts to an explicit advantage given to female participants of a tournament over their male competitors.[3] We study the effects of these two kinds of policy interventions on the competitive behavior of girls and boys in three different age brackets between 10 and 17 years of age. The motivation behind our research is the following: Given that differences in competitive behavior are prevalent from a very early age, the application of interventions to promote females in competitions may be desirable already at early stages during a child's education and upbringing, insofar as these can bridge the gap between girls and boys in their attitudes towards competition.

---

[1] It must be noted, however, that a number of studies suggest that this gender difference may not be systematic and that it depends on various factors such as the sample considered. For instance, Charness and Villeval (2009) do not find evidence of a consistent gender gap in a non-student population including seniors, while Gneezy et al. (2009) show that competitive attitudes can vary greatly among different cultures.

[2] Gneezy and Rustichini (2004) and Dreber et al. (2011) study competitive behavior not in terms of self-selection but of performance in response to competitive pressure. While the former study finds a significant gender gap with boys increasing their performance more than girls in Israel, the latter fails to replicate this finding in a Swedish sample. As far as the willingness to compete is concerned, there also exist some mixed results. A number of studies show that the willingness to compete is influenced by the educational attainment of the parents (Almås et al. (2013) as well as the task (Cardenas et al., 2012, Dreber et al., 2014). Culture also plays a major role, with competitive behavior differing substantially among countries (Andersen et al., 2013, Cardenas et al., 2012). In developing countries, neither Zhang (2013) nor Khachatryan (2012) have found a systematic gender gap in the willingness to compete. Overall, in highly developed countries a gender gap in the willingness to compete typically exists among children, while in developing countries this is not the case or the gap emerges only later in life (Anderson et al., 2013).

[3] See Fang and Moro (2010) for a historical overview and formal treatment of affirmative action policies.

The existing literature has studied gender differences in the competitive behavior of children and adolescents (as in Sutter and Glätzle-Rützler, 2015), and it has made significant steps towards identifying the effectiveness and the efficiency properties of policy interventions to the benefit of women (as in Balafoutas and Sutter, 2012; Villeval, 2012; Niederle et al., 2013). It has, however, only evaluated such policies in laboratory experiments that use samples of adult participants.[4] In this present study, we aim to combine the approaches of the above studies in order to draw some lessons on the effects of policy interventions on competitiveness in a sample of children and adolescents. Compared to previously mentioned studies on adult populations, we are able to consider three different age groups and identify how the gender gap in the willingness to compete develops across age and which policy interventions work best at a particular age.

Our reasons for concentrating on two particular policies – minimum quotas and preferential treatment – are twofold: first, the literature has shown that these two policies are very effective in closing down the gender gap in competitiveness in the case of adults. Second, they are both often encountered in practice. For instance, quotas are sometimes encountered in civil service or in national parliaments, guaranteeing that a specified fraction of job positions is occupied by women. Preferential treatment schemes are implemented both in the public and the private sector. A tie-breaking rule favoring women in the case of two job applicants with equal qualifications can be interpreted as a form of preferential treatment. Children and teenagers do not typically encounter such interventions in their daily lives. Nevertheless, there exist many opportunities, especially within the education system, to implement such policies already for children and teenagers. For example, quotas could be introduced in the election of student councils to guarantee an equal representation of girls and boys. Furthermore, class ranks could be separately computed for girls and boys and awards given to both the best performing boy as well as the best performing girl. In fact, one very recent example of introducing quotas for 18 year olds in Austria is the introduction of an equal number of men and women to be admitted into the Medical University of Vienna as students of medicine. All over Austria, about 10,200 high school graduates compete for 1,530 slots to study medicine, and the Medical University of Vienna is the first university to apply a strict quota rule (see – in German – http://oesterreich.orf.at/stories/2540175/).

[4] A notable exception is Calsamiglia et al. (2013) who study the effects of preferential treatment among 10-13 year-old school children in Spain. They show that preferential treatment enhances incentives to compete and perform in a tournament environment. However, preferential treatment is not conditioned on gender or some other innate characteristic in their study, but on the amount of experience with Sudoku puzzles that children are exposed to in different schools.

We "take the lab to the classroom", running experiments with a sample of 588 children from secondary schools in Austria. The children are pupils in three different grades: the 5th grade corresponds to the age group of 10-11 years of age, the 8th grade to the age group of 13-14 years, and the 11th grade to the oldest age group in our sample, 16-17 years of age. After exposing our subjects to a piece-rate and a tournament compensation scheme for their performance in an experimental task, we let them choose in a third stage the compensation scheme that they prefer, thus eliciting their preferences with respect to competition. We differentiate between a control treatment of no policy intervention, a treatment with minimum quotas, and one with preferential treatment. Then, in order to estimate the impact of policy on productivity *after the tournament*, we let the subjects perform in a team task after they have been exposed once more to a competitive environment. In such a framework, total team production gives us a measure of productivity after the tournament has been concluded and its winners are known. The rationale behind this team task is to evaluate the effects of policy interventions on efficiency in the longer run, i.e., on the eagerness of team members to contribute to total production after the tournament under the three different institutional regimes that we consider (no intervention, minimum quotas, preferential treatment).[5]

The paper reports two main findings. First, while there is a strong and significant gender gap in the willingness to compete in the control treatment (with boys competing much more often than girls), this gap is small and generally insignificant in the treatments with a policy intervention. This result holds for all three age groups and indicates that the two policies under consideration are effective, that is, they achieve their goal of a more equal gender balance in competitive behavior. Digging a bit deeper into the effects of policy, we find that these mostly apply to girls in the two younger age groups, and boys in the older group. Besides these effects, competition entry choices are also strongly influenced by confidence (beliefs about one's relative performance) and, not surprisingly, by ability.

We then evaluate the efficiency of interventions along three dimensions. The first dimension relates to the ability of the two winners of a tournament, which depends on the profile of those participants who choose to compete as well as on the tournament's rules. The second dimension relates to the performance and earnings of participants under different regimes, and the third dimension considers productivity after the tournament. Then, the second important finding of the paper is that policy interventions do not entail efficiency costs. The higher entry rates of high-ability girls into the tournament are more than enough to

[5] Mollerstrom (2012), for instance, finds that a quota system based on arbitrary group assignment reduces cooperation as measured by contributions into a public good.

compensate for an inevitable cost due to the selection rules, so that at the end of the day preferential treatment as well as minimum quotas result in a slightly “better” (i.e., more productive) pool of tournament winners compared to the control treatment. Moreover, total team output after the tournament is, if anything, slightly higher in the two treatments with a policy intervention. Hence, we have gathered enough evidence to argue that changing the rules of the game in order to help girls is not detrimental to efficiency, and it is an effective instrument when the goal of policy is to close down the gender gap in competition entry rates.

The findings of this paper are in many ways similar to Balafoutas and Sutter (2012), who find that certain types of affirmative action – a minimum quota system and two forms of preferential treatment – have a strong impact on women’s willingness to enter competitions. Besides closing down the gender gap in tournament entry rates, these policies are then evaluated in terms of efficiency and they are found to be at least as efficient as the control treatment of no intervention – a result which is also found in Niederle et al. (2013) for the case of quotas. Hence, we are able to replicate the key messages in Balafoutas and Sutter (2012) for the case of adolescents. In light of the findings by Harbaugh et al. (2001), who show that children as young as age 7 already choose rationally and respond to incentives systematically, we believe it is not surprising that our results match those found in adults. However, it is interesting to note that our children and adolescents react to the given incentives (policy interventions) although they most likely have not encountered such interventions in real life before.

Some important differences to Balafoutas and Sutter (2012) remain, however: First, we use a different measure of post-tournament efficiency, namely a team production exercise. This exercise captures the extent to which group members can successfully work together after a tournament, and has the advantage that it is quite straightforward and easier for children to understand than the minimum effort game used in Balafoutas and Sutter (2012). Second, we study the effects of affirmative action for three age groups and detect some differential effects by age. Finally, we believe that it is important to know how interventions apply to the younger population under consideration. As a matter of fact, in most industrialized countries, teenagers around the age of 14 or 15 years often have to make important decisions, such as whether to continue in school – and in which educational direction – or enter the labor market and start working. Such decisions have long-lasting consequences, in particular for young person’s subsequent professional career and thus their life-time earnings and socio-economic status in society (Card, 1999). For instance, Buser et al.

(2014) and Zhang (2013) show that there is a link between subjects' willingness to enter a competition in experiments and their choice of education track.

The rest of the paper is organized as follows. Section 2 describes the experimental design. Section 3 presents the experimental results, and section 4 concludes the paper.

2. Experimental design

Our design builds on the seminal paper by Niederle et al. (2013) and on Balafoutas and Sutter (2012). In all three treatments described below, the task consisted of adding three two-digit numbers in a limited time period of two minutes. At the beginning of the experiment pupils were randomly assigned to groups of six, with three boys and three girls in each group. All groups went through five different stages and each stage was only introduced and explained after the previous one had ended. Subjects were not allowed to use calculators, but could use scratch paper or do the additions off the top of their head. After each calculation subjects were informed whether it was correct or not, and then the next problem was shown. The five stages in the experiment were the following:

Stage 1 – Piece rate. Each subject received €0.40 for each correct calculation; this payment was independent of the other group members' performance.

Stage 2 – Tournament. In this stage group members had to compete against each other. The two members who solved the most calculations correctly were paid €1.20 for each calculation. The other four group members received nothing. Ties were broken randomly in this stage as well as in stages 3 and 4.

Stage 3 – Choice. Every group member could choose whether he (or she) wanted to solve the calculations under a piece rate scheme (as in Stage 1) or a tournament scheme. If the tournament was chosen, then a subject's performance was compared to the other group members' performance in Stage 2[6] and the rules for determining the winners differed across **treatments** as follows:

1. **Control treatment (CTR)**. The winners were the two group members with the largest numbers of correct calculations, regardless of gender.

[6] Using the other group members' past performance has several advantages. First, tournament entry decisions do not depend on a subject's expectation about the other members' entry decisions, but only on the subject's beliefs about own ability. Second, Stage 2 performances are competitive performances, and thus a subject competes against others when they were also exposed to a competitive payment scheme. Third, entering competition does not impose externalities on others. In principle, this means that Stage 3 is an individual decision making problem. Note that this scheme also implies that it is possible that all group members entering the competition in Stage 3 may win or all lose since they are competing against the others' performance in Stage 2.

2. **Preferential Treatment (PTR)**. Each girl's performance was automatically increased by one unit (i.e., one correct calculation). Using the augmented scores for girls, the rules of the control treatment applied, meaning that there were no further restrictions on the gender composition of the two winners.
3. **Minimum Quotas (QUO)**. In this treatment, the rules required that there be at least one girl among the two winners of the tournament. This implied that the best performing girl was a winner for sure. The second winner could either be a boy – if he performed better than all other boys and better then the second-best girl – or a girl – if the second-best girl performed better than all boys.

All treatments were run in a between-subjects design. It is important to stress that subjects did not receive any feedback on the outcome of the compulsory competition in Stage 2 or the optional competition in Stage 3 until the end of the experiment. This was done in order to avoid that subjects condition their choices on previous outcomes of a competition. At the end of Stage 3 we elicited the beliefs of all subjects regarding their relative performance in Stages 1 and 2: for each stage subjects had to indicate their expected rank within the group of six members. Correct guesses were rewarded with €0.50 each, and feedback was given also only at the end of the experiment.

Stage 4 – Compulsory tournament with policy intervention from Stage 3. In this stage the same treatments as in Stage 3 applied. However, subjects were forced to compete, and hence could not opt out as in Stage 3. As before, the two winners received €1.20 for each correct calculation. At the end of Stage 4 we informed them whether they had won or lost before moving on to the non-competitive task of Stage 5. In order to make winning and losing in Stage 4 more salient, we gave each winner in Stage 4 an additional €3 as initial endowment in Stage 5, and each loser only €1[7]. Our purpose was to introduce a clear distinction between winners and losers before starting with the post-competition stage (team task).

Stage 5 – Team task. This task was identical in all treatments. Subjects had to add up two-digit numbers again. However, the payment scheme was such that each correct calculation was worth €0.40 for the group in total and then split equally among all group members. Hence, each member's effort within a team can be substituted by the effort of every other team member. We use the total group performance in Stage 5 in order to evaluate the impact of interventions on a group's joint performance after a competition has been

[7] These endowments for winners and losers apply to students in the 11th grade. The endowment was 60% thereof for 8th graders (thus €1.8 for winners and €0.6 for losers) and 40% thereof for 5th graders (thus €1.2 for winners and €0.4 for losers). These proportions were chosen since the average weekly pocket money of 8th graders (5th graders) is roughly 60% (40%) of 11th graders' pocket money.

concluded. While the monetary incentives are identical in Stage 5 across all treatments, the different experiences in the different treatments in Stage 4 might affect Stage 5 behavior in different ways. The choice of this team task was motivated by the fact that in reality individuals often have to work together in teams following tournament-based promotions. To implement such a situation in our experiment, we gave the two winners of the tournament in Stage 4 a higher endowment (salary) than the four losers in Stage 5, and had all group members work together in that stage in order to test for spillovers from the tournament to the team production task in this simple environment under different tournament rules (affirmative action versus no intervention).

We ran experiments in four different Tyrolean secondary schools. In total 588 students from the 5th, 8th and 11th grade participated in this study. The experiments were conducted during school hours. Participation in the experiments was voluntary, however all students of the selected classes chose to participate. The distribution of participants by age, gender and treatment is shown in Table 1. The three treatments per grade and school were run at the same time. Therefore, before the start of the experiments we randomly assigned students to one of the three treatments. Each treatment was run in a separate classroom, so that subjects of different classes (but of the same grade) participated in the same treatment and this was public information for all students. The explanation of the experiment was done orally but followed a fixed script (see Appendix). During the illustration the experimenters frequently paused and allowed students to ask questions, which were privately answered by the experimenters or one of the research assistants. Before the start of each stage a short summary of the most important facts of the next stage was distributed to each student. The duration of a typical session was roughly 100 minutes (which is equivalent to two school hours in Austria).

Table 1 about here

3. Results

3.1 Performance in the experimental task (Stages 1-5)

Consistent with previous studies, we find that the number-adding task is gender neutral in terms of performance. Figure 1 shows the performance of boys and girls by stage and treatment, separately for each grade. The figures reveal that boys generally give slightly more correct answers than girls do in grades 5 and 8, while girls outperform boys in the oldest age group (grade 11). In any case, these differences are very small and not statistically significant for any age group and in any of the five stages of the experiment. The fact that we do not detect any significant differences in performance means that eventual differences in competitive behavior must be attributable to factors other than ability. Not surprisingly, we observe that boys as well as girls improve their performance over time, i.e., from one stage to the next. This is to some extent driven by the stronger incentives in the stages with a tournament compared to Stage 1, but it is also indicative of learning effects. Indeed, there is a relatively strong increase in aggregate performance between stages 2 and 4, which both entail compulsory competition (6.65 vs. 7.38 correct answers; $p < 0.01$, Wilcoxon signed-rank test). Moreover, performance naturally increases with age, especially between grades 5 and 8. For instance, the mean number of correct answers in Stage 1 increases from 4.69 to 6.53 from the 5^{th} to the 8^{th} grade (+39%), and then to 7.40 in the 11^{th} grade (+13.3%). Finally, the weaker individual incentives to perform in Stage 5 due to the much lower per capita return on each correct answer do not seem to affect performance: performance in the team task of Stage 5 is on average slightly higher than in Stage 4.

Figure 1 about here

3.2. Competition entry choices and policy effectiveness (Stage 3)

In the absence of policy interventions (treatment **CTR**), we are able to replicate the findings of Sutter and Glätzle-Rützler for Austria (2015), in the sense that there is a significant gender gap in tournament entry choices in children and teenagers. In Stage 3, the proportion of boys who choose the tournament payment scheme for their performance in the control treatment lies between 35% (in grades 5 and 8) and 63% (in grade 11), while the share of competition-oriented girls never exceeds one quarter. These percentages are shown in the first column of Table 2. It is noteworthy that the gap widens with age and it becomes more strongly significant.

Comparing boys' and girls' competition entry choices across columns (treatments) in Table 2 gives an overview of the effectiveness of our two interventions in closing the gender gap in competitiveness. The magnitude of the effects is perhaps striking: Preferential treatment leads to a reverse gender gap, with girls competing much more than boys in grades 5 and 8, and slightly more in grade 11. Quotas also lead to slightly higher (albeit not significant) entry rates for girls compared to boys, with the largest difference in the eighth grade. Overall, we observe that both interventions, and particularly preferential treatment, have a strong impact on the choices of girls in our sample. Similarly, both treatments discourage boys from competing in Stage 3 (with the exception of **QUO** in the 5th grade). The overarching pattern is that the two policies under consideration are particularly effective in the 8th and 11th grade in terms of achieving their goal of eliminating the gender gap in competition entry choices and the resulting underrepresentation of girls in the tournament. Indeed, in those two grades, both interventions are associated with tournament entry rates that do not differ significantly by gender. For the younger pupils in the 5th grade, the quota restores the balance in entry choices, but preferential treatment leads to a large reverse gender gap. This can probably be explained – at least partly – by the fact that the one additional correct answer given to girls is a larger proportion of performance for that age group, meaning that the one additional correct answer has a larger impact on the likelihood to win in the youngest age group.

Table 2 about here

3.2.1. Gender differences in confidence

The fact that – barring policy interventions – boys are more likely than girls to choose the tournament in Stage 3 is to a large extent attributable to differences in self-confidence and self-assessment. Although we have seen that girls are not worse than boys in terms of performance, the differences in the self-assessed ranks for Stages 1 and 2 are very large, with boys reporting lower (i.e., better) ranks than girls ($p < 0.01$, Mann-Whitney tests for both elicited ranks). An interesting question is the origin of these differences in confidence: are boys (girls) over (under) confident, or is it both? The first two rows in Table 3 report the means of the variable *Δrank*, defined as the difference between the actual and the self-reported rank. We see that boys are overconfident while girls are slightly underconfident, particularly in their self-assessment for Stage 1. Broken down by age group, overconfidence for boys is mainly driven by the oldest group in Stage 2. Underconfidence in Stage 1 is, in turn, only significant for the girls of the oldest age group. Accordingly, the gender difference in confidence is increasing with age, which is consistent with the widening gap in tournament entry choices reported in Table 2.

Table 3 about here

In terms of competition entry choices in Stage 3, the self-assessed ranks for the Stage 2 tournament are of particular importance, since they are a very good indicator of whether a subject also expected to win the tournament in Stage 3 (given that there was no feedback between stages). We construct a dummy variable, which is 1 if a subject has reported a perceived rank of 1 or 2 for his (her) performance in Stage 2, and call it *guesswin2*. The proportion of boys who expected to have won in Stage 2 is 41.8%, compared to 29.6% for girls ($p < 0.01$, chi-square test). Disaggregating by age group we find the same qualitative result for all three groups, although the difference is not significant for the 5th grade. This is in sharp contrast to the proportions of boys and girls who actually had a rank of 1 or 2 in the tournament of Stage 2 and which are practically equal,[8] as shown in the fourth row of Table 3. A further useful, and remarkable, insight from the last three rows of Table 3 is that the gender gap in tournament entry is significant even for those subjects in the control treatment who rank first or second in the second stage of the experiment. This means that even the best girls (those performing in the best quartile) have a much larger aversion against competition than the best-performing boys. At the same time, there is a reverse gender gap for the two treatments with an intervention (significant in the case of **QUO**)

3.2.2. Regression analysis of competition entry choices

We proceed to show a number of probit regressions, with the purpose of econometrically identifying the effects of policy on competitive behavior by age group. Columns (1) to (3) in Table 4 fit the following regression specification to each of the three grades:

$$\Pr\,(choice_i = 1) = \Phi\left(\alpha + \beta\, female_i + \sum_j \gamma_j\, policy_j + \sum_j \delta_j\, female_i \times policy_j + C'\zeta\right).$$

The dependent variable *choice* (1 if a subject chooses competition in Stage 3, 0 otherwise) is regressed on a gender dummy, the two policy (treatment) dummies, the interaction terms between gender and policies, and the following two control variables denoted by ***C'***: *correct2*, which is a subject's competitive performance in Stage 2, and *guesswin2*, which captures the perceived probability of winning the tournament. $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution.

Table 4 about here

[8] We refer to pupils who had a rank of 1 or 2 in the tournament, regardless of whether they actually won, in order to ensure comparability with the variable *guesswin2* and to keep the results unaffected by the random tie-breaking rule. Note that since ties were possible, it is possible to have more than 33.3% of subjects ranked first and second.

The first thing to note is that, abstracting from the effects of policy and controlling for performance and the perceived probability of winning the tournament, the gender gap in competition entry choices – captured by the variable *female* – increases in strength and gains statistical significance as we move from the younger to the older pupils in our sample, with girls being almost 40% less likely than boys to choose the tournament in the eleventh grade, *ceteris paribus*. This is consistent with the frequencies reported in Table 2. Moreover, both control variables have the expected positive coefficient in all three regressions and they are significant: subjects with a high competitive performance (*correct2*) are more likely to enter the tournament in Stage 3, and so are those who believe they have won in the tournament of Stage 2 (*guesswin2*).

The introduction of policy interventions does not generally have a significant negative impact on boys' willingness to compete in Stage 3, with the exceptions of **PTR** in the 5th and **QUO** in the 11th grade. The latter effect, although only weakly significant, is of substantial magnitude indicating that boys in the older age group are almost 40% less likely to compete when a quota system is in place. In the case of girls, both interventions are particularly effective in terms of promoting competitive behavior in the 5th grade. In the 8th grade we observe a significant positive effect only in treatment **QUO**, while in the oldest age group there are no statistically significant treatment effects for females – even though the effects of both interventions have the expected positive coefficients. These effects are measured by means of the joint coefficient *PTR+female*PTR* (or *QUO+female*QUO*), whose magnitudes and significance levels are also shown in Table 4. Moreover, the joint coefficient *female+female*PTR* (*female+female*QUO*) measures the gender gap in tournament entry choices in treatment **PTR** (treatment **QUO**). As the significance levels of these coefficients reveal, and as we have already said earlier in this section, both policies generally lead to a closing down of the gender gap, with the notable exception of preferential treatment at the youngest age group, where a reverse gap emerges.[9,10] Note that Table 4 presents the magnitude and significance levels of the four joint coefficients mentioned above, but omits the values of the two interaction terms (*female*PTR, female*QUO*). The reason is that Stata does not correctly calculate marginal effects and p-values of interaction terms in nonlinear

[9] For robustness, we have replicated the Table 4 regressions using a linear probability model. All of our results remain qualitatively the same, with the minor exception of the female dummy in column 2 (8th grade), which turns marginally insignificant.

[10] In order to control for learning effects, we have run additional regressions including a variable which measures the increase in performance from stage 1 to stage 2. The results remain practically the same.

models (see Ai and Norton, 2003). We have corrected for this and we report directly the tests on the joint coefficients, which are relevant for our analysis.

Based on these results, one can discern the following pattern regarding the effects of policy interventions on competitive behavior. First and foremost, both policies are effective, which means that they achieve their explicit target of closing down the gender gap in tournament participation – sometimes even exceedingly so. In the two younger age groups (5th and 8th grade), the policy effects work their way mostly through the choices of girls, encouraging them to compete more often than they otherwise would. However, the picture changes in the 11th grade, where policy – and in particular quotas – affects boys more strongly than girls, reducing the proportion of tournament entrants among them. Even though both channels result in the same outcome of a more equal gender balance in competition, the qualitative difference is important because it may imply a very different profile of participants in the voluntary tournament of Stage 3. This, in turn, has implications for the efficiency of the tournament.

In general, it is noteworthy that the effects of the two policies on the willingness to enter the tournament largely differ by age. Although this result is interesting, it is not straightforward to explain with our dataset. One possibility could be that it is driven by heterogeneous effects of beliefs on entry decisions among different age groups; however, this explanation is ruled out if one considers full versions of the Table 4 regressions which include interactions between beliefs, policies, and age.[11] One explanation for the pattern we see in treatment PTR could be the following: In that treatment, the one bonus point given to girls is a larger share of total performance among younger pupils, and conversely a smaller share for the older girls, which means that the effect of the additional bonus point should generally be decreasing with age – as is indeed the case.

3.3. Policy efficiency

We consider three alternative notions of efficiency, namely candidate selection, total earnings and performance, and post-tournament performance.

3.3.1. Policy efficiency I: Candidate selection

From the standpoint of aggregate social welfare and abstracting from individual preferences regarding the actual process of competition, one would want the best performers within a

[11] Details are available from the authors upon request.

group to enter (and win) the competition. Whether winning the competition amounts to being employed for a job or given a promotion, it is profitable for the employer and efficient from an economic point of view if the winners are the candidates with the highest ability. The impact of our policy interventions on this aspect of efficiency cannot be known *a priori*. We have seen that entry rates for girls increase strongly as a result of policy, while the decrease in boys' entry rates is not as strong. Since girls are on average not less skilled than boys, the more equal gender balance in entry rates in treatments **PTR** and **QUO** should then also increase the proportion of high performers entering the competition. On the other hand, there is a countervailing effect of the policy mechanism *per se* on the selection of the winners: *given* the identities of the candidates, both policies favor girls and may result in high-performing boys being passed by.

We estimate the overall impact of policy interventions on candidate selection by comparing the mean ability of all winners in the tournament of Stage 3 in each of the three treatments. In Stage 3, both aforementioned effects – the entry effect and the pure selection effect – are at play, and therefore it is interesting to see which one prevails. The figures are shown in Table 5. We measure a candidate's ability by means of his (her) performance in Stage 1, which is unaffected by exposure to competition.

Table 5 about here

It becomes clear that efficiency does not suffer in the treatments with a policy intervention compared to the control. The Stage 1 performance of the tournament winners does not differ significantly across treatments in the 5th and 8th grade (p > 0.35, pairwise Mann-Whitney tests with **CTR**; p > 0.50, Kruskal-Wallis tests).[12] On the contrary, we observe moderate to strong efficiency *gains* for the older age group of the 11th grade, with the winners in treatment **PTR** (**QUO**) having more than one (two) more correct answers in Stage 1 compared to the control of no intervention. This difference is statistically significant for treatment **QUO** (p < 0.05, Mann-Whitney test), showing that minimum quotas not only do not lead to efficiency losses due to distortions in the selection process, but as a matter of fact increase efficiency in the oldest age group. Hence, it must be the case that quotas have

[12] As suggested by a referee we have also computed Table 5 with the Stage 2-performance of the winners in Stage 3. The results remain robust. We do not detect any significant differences between the three treatments in single age groups (p > 0.29, Kruskal-Wallis tests).

encouraged many high-performing girls to compete in Stage 3, when they would not otherwise have done so. We now proceed to investigate this mechanism a bit further.

Figure 2 shows the proportion of pupils in our sample who chose to enter the tournament in Stage 3, initially pooled and then broken down by gender, and classified by level of ability (low, intermediate, high) based on the number of correct answers in Stage 1.[13] Panel (a) shows that preferential treatment increased the share of tournament entrants among low performers and quotas increased the share of tournament entrants among high performers. Panels (b) and (c) show the proportion of entrants by ability level for each gender separately. The effects of policy interventions on the ability profile of tournament entrants are clearly more pronounced in the case of girls. Particularly for the subgroup of high-performing girls, the results produce a nice ranking of the three treatments in terms of encouraging the most productive girls to engage in competition. In the control, just 36% of those girls who belong to the highest performance quartile choose the tournament; this percentage rate increases to 52% in treatment **PTR** and to 72% in **QUO**. The difference to **CTR** is, however, statistically significant only in the case of **QUO** ($p < 0.02$, chi-square test). The mirror image of the increased tournament participation of high-ability girls is the slight decrease in the participation rates of high-ability boys, as seen in Panel (c).

Figure 2 about here

3.3.2. Policy efficiency II: Earnings and performance

Besides the success of a tournament in selecting the most qualified candidates as winners, different tournament formats may lead to variations in the total earnings of the participants. Earnings are a measure of efficiency in the sense that they represent the total surplus earned by all members of a group (those who entered competition and those who did not). For instance, one must account for the possibility that interventions lead to too much competition, so that too many subjects enter the tournament in Stage 3, earn nothing, and reduce efficiency (girls under preferential treatment may provide for such a scenario). Moreover, tournament efficiency may be measured by means of total output, which in our design is given by the performance of all participants. Compared to section 3.3.1, this measure takes into account

[13] The classification follows Balafoutas and Sutter (2012): The low- and high ability pupils correspond roughly to the first and fourth quartile in our sample respectively, while intermediate ability pupils correspond roughly to the two middle quartiles.

not only the identity and performance of the two winners in Stage 3, but also the performance of the remaining four members in a group.

Table 6 shows the mean earnings and performance of all participants in Stage 3 by treatment, broken down by grade. Our findings show that the policy interventions under consideration do not entail efficiency costs. Per capita earnings are practically identical across treatments in all age groups ($p > 0.28$, Kruskal-Wallis tests). Performance also does not significantly differ between the three treatments for 5th- and 11th-graders ($p > 0.45$, Kruskal-Wallis tests). The performance of 8th graders is in fact significantly higher in the PTR-treatment than the Control ($p = 0.04$, Mann-Whitney test), while there is no significant difference between the Control and the QUO-treatment ($p = 0.88$).

Table 6 about here

3.3.3. Policy efficiency III: Post-tournament performance

We now turn our attention to the last notion of efficiency under consideration, namely the total performance of a team after the tournament has taken place and the winners have been publicly announced. In the team task each member's output can be perfectly substituted by another member's output, so that all payoffs are determined entirely by the sum of efforts from all group members. Individual incentives for solving the number-adding problems are now much lower (one sixth) compared to Stage 1 (piece-rate) and even lower compared to the stages with a tournament, but – like in a social dilemma situation – each problem solved benefits all six group members.

Standard theory predicts that a subject's contribution into team production depends on his (her) own ability as well as on the cost of effort, so that subjects should either not try to solve any problems (if the cost of effort is high), or solve as many problems as possible (if the cost of effort is low), or attempt only a certain number of problems (with convex costs of effort). In any case, one should observe no differences across treatments, since the monetary incentives in Stage 5 are not affected by the policy implementation in Stage 4. However, one can come up with behavioral stories that could explain potential treatment effects in the team task. For instance, the winners of Stage 4 could invest a lower amount of effort in Stage 5 as a result of income effects; but the losers of Stage 4 could also contribute less to the team's production if they are disappointed by the result of the tournament in Stage 4, especially if they had expected to win and have bitter feelings as a result of the policy that was applied. On the other hand, it is possible that the Stage 4 winners work harder in Stage 5 in order to justify their higher endowment; and the losers may work harder in the team task in order to compensate to some extent for the foregone income in the previous stage.

Figure 3 shows the average team performance in the team task of Stage 5, by treatment and for each age group separately. The unit of observation this time is the team of six members, so that the total number of observations is one sixth of the full sample (N = 98 teams of size six). The main result that emerges from the aggregate figures is that efficiency does not decrease when a policy is in place compared to the control. The total team performance is actually *higher* with **PTR** as well as with **QUO** compared to **CTR** (47.1, 45.3 and 44.2 correct answers, respectively), which is an indication of efficiency gains, although the differences are not statistically significant ($p > 0.25$, pairwise Mann-Whitney tests with **CTR**; $p > 0.50$, Kruskal-Wallis test). The picture is very similar when we consider the disaggregated figures for each of the three grades. Hence, we arrive at the conclusion that the two policy interventions under investigation are not detrimental to post-tournament efficiency, defined here as the productivity of teams after the compulsory tournament of Stage 4.

Figure 3 about here

We conclude this section by briefly examining the behavior of the winners and losers of the Stage 4 tournament separately. For this reason, we look at the individual Stage 5 performances using the full sample. It turns out that the slight aggregate increase in team productivity in the treatments with a policy intervention is driven by both groups of subjects to a more or less equal extent – in treatment **PTR**, for instance, winners contribute to the team 7% more compared to **CTR**, while the same percentage difference is 6.3% for losers.

4. Conclusion

In a sample of 588 children and teenagers aged between ten and seventeen years we have been able to replicate the standard finding of a strong and significant gender gap in the willingness to compete in a math task, with boys choosing to enter a competitive environment much more often than girls. We have then examined two policy interventions. Under minimum quotas, it is guaranteed that within each experimental group of three girls and three boys, one of the two selected winners is female. Under preferential treatment, girls are given an extra bonus point when entering the competition.

In our treatments with a policy intervention the gender gap in tournament entry choices is small and generally insignificant. This result holds for all three age groups (5th graders, 8th graders and 11th graders) and indicates that the two policies under consideration are effective, in other words that they achieve their goal of a more equal gender balance in competitive behavior. The fact that these interventions are effective indicates that the gender gap in competitiveness is something that can be changed with appropriate instruments. For instance, one way to think about minimum quotas is that, by guaranteeing the best-performing girl to be one of the winners, they partly transform the tournament into a same-sex competition. This closes down or even reverses the gender gap in the willingness to compete, in line with the findings on competitive performance from Gneezy et al. (2003). A second important finding of the paper is that policy interventions do not lead to efficiency costs. We have considered three different notions of efficiency. Two of those refer to efficiency *during* the tournament, in the sense that they measure the ability of the two winners and the average per capita performance and earnings under different regimes. The third notion refers to efficiency *after* the tournament, measured by means of total output in a post-competition team task. Our results indicate that quotas as well as preferential treatment do not reduce efficiency – in fact, there even exists some evidence in support of efficiency gains.

Hence, we have gathered enough evidence to argue that changing the rules of the game in order to help girls is an effective policy instrument, which is not detrimental to efficiency. Given that differences in competitive behavior can be prevalent from a very early age (Sutter and Glätzle-Rützler, 2015), the application of policy interventions to induce females to enter tournaments may be desirable already at early stages during a child's education and upbringing, insofar as they may bridge the gap between girls' and boys' attitudes towards competition. Implementing educational programs in kindergarten or school could in fact be much easier than reaching grown- ups with policy interventions. A very important decision which adolescents face is the choice of their education. Buser et al. (2014) and Zhang (2013) show that subjects who have a higher willingness to enter a competition in experiments are more likely to choose a more demanding – and in expectation more profitable – education track than subjects who are less willing to compete. A natural policy implication of our results would be to intervene in this important decision in order to encourage girls to choose more competitive fields of education. One possible example would be to implement quotas when selecting candidates for male dominated studies such as engineering.

Finally, we would like to conclude with a word of caution: The above statements should not necessarily be interpreted as a normative appeal for the application of interventions favoring girls and encouraging them to compete at a young age. Such an appeal would have to evaluate the impact of interventions on some social objective function, which should include the utility of all parties involved. What our results have shown is that, *provided* that the goal of policy is indeed to close down the gender gap in competition entry rates, then quotas as well as preferential treatment in the school can be very good instruments in order to achieve this goal. Moreover, one must keep in mind that our study cannot shed light on the potential long-run implications of affirmative action policies for children, as it only shows how boys and girls change their behavior in response to the implementation of different rules during the experiment. A valuable extension would therefore involve investigating whether exposure to competing because of the interventions changes competitive preferences or whether the gender difference in tournament entry rates returns once the affirmative action is eliminated.

References

Ai, C., Norton, E. (2003). Interaction Terms in Logit and Probit Models. *Economics letters* 80: 123-129.

Almås, I., Cappelen, A. W., Salvanes, K. G., Sørensen, E. Ø., Tungodden, B. (2013). Willingness to Compete: Family Matters. Working Paper, NHH Norwegian School of Economics.

Andersen, S., Ertac, S., Gneezy, U., List, J., 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: 1438-1443.

Balafoutas, L., Sutter, M. (2012). Affirmative Action Policies Promote Women and Do Not Harm Efficiency in the Laboratory. *Science* 335: 579-582.

Booth, A. L., Nolen, P. J. (2012). Choosing to Compete: How Different are Girls and Boys? *Journal of Economic Behavior and Organization* 81: 542-555.

Buser, T., Niederle, M., Oosterbeek, H. (2014). Gender, Competitiveness and Career Choices. *Quarterly Journal of Economics* 129: 1409-1447..

Calsamiglia, C., Franke, J., Rey-Biel, P. (2013). The Incentive Effects of Affirmative Action in a Real-Effort Tournament. *Journal of Public Economics* 98: 15-31.

Card, D. E. (1999). The Causal Effect of Education on Earnings. In: Ashenfelter, O., Card, D. E. (eds.), *Handbook of Labor Economics*, vol. 3, Part A, chapter 30: 1801-1863.

Cardenas, J., Dreber, A., von Essen, E., Ranehill, E. (2012). Gender Differences in Competitiveness and Risk Taking: Comparing Children in Colombia and Sweden. *Journal of Economic Behavior and Organization* 83: 11-23.

Charness, G., Villeval, M.C. (2009). Cooperation and Competition in Intergenerational Experiments in the Field and the Laboratory. *American Economic Review* 99: 956-78.

Datta Gupta, N., Poulsen, A., Villeval, M.C. (2013). Gender Matching and Competitiveness: Experimental Evidence. *Economic Inquiry* 51: 816-835.

Dreber, A., von Essen, E., Ranehill, E. (2011). Outrunning the gender gap – boys and girls compete equally. *Experimental Economics* 14: 567-582.

Dreber, A., von Essen, E., Ranehill, E. (2014). Gender and Competition in Adolescence: Task Matters. *Experimental Economics* 17: 154-172.

Fang., H., Moro, A. (2010). Theories of Statistical Discrimination and Affirmative Action. In: Benhabib, J., Bisin, A., Jackson, M. (eds.), *Handbook of Social Economics,* vol. 1, Part A, chapter 5: 133-200. North-Holland.

Fischbacher, U. (2007). z-Tree: Zurich Toolbox for Readymade Economic Experiments. *Experimental Economics* 10: 171-178.

Gneezy, U., Leonard, K., List, J. (2009). Gender Differences in Competition: Evidence from a Matrilineal and a Patriarchal Society. *Econometrica* 77: 1637–1664.

Gneezy, U., Niederle, M., Rustichini, A. (2003). Performance in Competitive Environments: Gender Differences. *Quarterly Journal of Economics* 118: 1049-1074.

Gneezy, U., Rustichini, A. (2004). Gender and Competition at a Young Age. *American Economic Review, Papers and Proceedings* 94: 377-381.

Harbaugh, W. T., Krause, K., Berry, T. R. (2001). GARP for Kids: On the Development of Rational Choice Behavior. *American Economic Review* 91: 1539-1545.

Khachatryan, K. (2012). Gender Differences in Preferences at a Young Age? Experimental Evidence from Armenia. Working Paper, Stockholm School of Economics.

Mollerstrom, J. (2012). Quotas and Cooperation. Working Paper, Harvard University.

Niederle, M., Segal C., Vesterlund, L. (2013). How Costly is Diversity? Affirmative Action in Light of Gender Differences in Competitiveness. *Management Science* 59:1-16.

Niederle, M., Vesterlund, L. (2007). Do Women Shy Away from Competition? Do Men Compete Too Much? *Quarterly Journal of Economics* 122: 1067-1101.

Sutter, M., Glätzle-Rützler, D. (2015). Gender differences in the willingness to compete emerge early in life and persist. *Management Science*, forthcoming.

Villeval, M.-C. (2012). Ready, Steady, Compete. *Science* 335: 544-545.

Wozniak, D. (2012). Gender Differences in a Market with Relative Performance Feedback: Professional Tennis Players. *Journal of Economic Behavior and Organization* 83: 158-171.

Zhang, J. (2013). Can Experimental Economics Explain Competitive Behavior Outside the Lab? Working Paper, Hong Kong University of Science and Technology.

Tables and Figures

Table 1. Distribution of participants by age, gender and treatment

	CTR	**PTR**	**QUO**
5th grade (n=216)			
Boys	45	33	30
Girls	45	33	30
8th grade (n=228)			
Boys	54	30	30
Girls	54	30	30
11th grade (n=144)			
Boys	24	24	24
Girls	24	24	24
Sum (n=588)	246	174	168

Table 2. Share of boys and girls entering competition in Stage 3, in %.

	CTR	PTR	QUO
Panel (a): 5th grade			
Boys	35.6 *	18.2	36.7
Girls	20.0	60.6 ***	40.0
Panel (b): 8th grade			
Boys	35.2 **	23.3	26.7
Girls	16.7	40.0	43.3
Panel (c): 11th grade			
Boys	62.5 ***	41.7	33.3
Girls	25.0	45.8	37.5

*, **, *** indicate significant gender differences at the 10%, 5%, 1% level respectively (chi-square tests)

Table 3. Confidence

		Boys	Girls
Δrank, Stage 1		0.14	-0.18 **
Δrank, Stage 2		0.19 **	-0.01
guesswin2		41.8 ***	29.6
% with a rank of 1 or 2 in Stage 2		40.8	39.8
% choosing to compete, among subjects with a rank of 1 or 2 in Stage 2	**CTR**	53.9*	36.4
	PTR	43.3	59.0
	QUO	44.7*	64.7

*, **, *** denotes significance at the 10%, 5%, 1% level respectively.
Δrank: Wilcoxon signed-rank tests on the restriction *Δrank*=0.
guesswin2 and last four rows: chi-square tests on gender differences

Table 4. Regression results for the tournament entry choices in Stage 3 (corrected for interaction terms). Dependent variable: *choice*

	(1) 5th grade	(2) 8th grade	(3) 11th grade
Female	-0.136 (0.091)	-0.138 * (0.081)	-0.394 *** (0.151)
PTR	-0.177 * (0.095)	-0.117 (0.093)	-0.248 (0.152)
QUO	0.042 (0.113)	-0.062 (0.110)	-0.356 ** (0.156)
correct2	0.051 *** (0.018)	0.051 *** (0.015)	0.034 * (0.018)
guesswin2	0.198 *** (0.073)	0.198 ** (0.080)	0.410 *** (0.095)
Tests for specific treatment effects:			
*PTR+female*PTR* [A]	0.430 *** (0.111)	0.180 (0.115)	0.193 (0.146)
*QUO+female*QUO* [B]	0.253 ** (0.112)	0.309 *** (0.100)	0.168 (0.143)
*female+female*PTR* [C]	0.471 *** (0.113)	0.159 (0.124)	0.047 (0.147)
*female+female*QUO* [D]	0.075 (0.130)	0.233 * (0.127)	0.129 (0.146)
N	216	228	144

Probit regressions. Robust standard errors in parentheses. *, **, *** denotes significance at the 10%, 5%, 1% level respectively. The table presents marginal effects of the coefficients.

[A] Testing for a treatment effect for girls in treatment PTR while keeping the variables *correct2* and *guesswin2* at their mean values.

[B] Testing for a treatment effect for girls in treatment QUO while keeping the variables *correct2* and *guesswin2* at their mean values.

[C] Testing for a gender gap in treatment PTR while keeping the variables *correct2* and *guesswin2* at their mean values.

[D] Testing for a gender gap in treatment QUO while keeping the variables *correct2* and *guesswin2* at their mean values.

Table 5. Efficiency in candidate selection – Average Stage 1 performance of the winners in Stage 3, by treatment and grade (standard errors in parentheses)

	CTR	PTR	QUO
5th grade	5.93 (0.52)	5.8 (0.47)	5.71 (0.61)
8th grade	8.9 (0.52)	9.0 (1.0)	8.13 (0.36)
11th grade	8.27 (0.45)	9.3 (0.92)	10.38 (0.94)

Table 6. Earnings and performance – Average earnings and performance of all participants in Stage 3, by treatment and grade (standard errors in parentheses)

	Average per capita earnings (€)			Average performance (correct answers)		
	CTR	**PTR**	**QUO**	**CTR**	**PTR**	**QUO**
5th grade	3.23 (0.31)	3.47 (0.38)	3.74 (0.44)	5.37 (0.20)	5.79 (0.24)	5.45 (0.27)
8th grade	4.77 (0.39)	4.88 (0.53)	4.19 (0.49)	7.26 (0.27)	8.18 (0.36)	7.05 (0.29)
11th grade	5.45 (0.68)	4.93 (0.67)	5.19 (0.72)	8.58 (0.39)	7.98 (0.41)	8.77 (0.46)

Figure 1. Average performance (# correct answers) by stage, treatment and gender

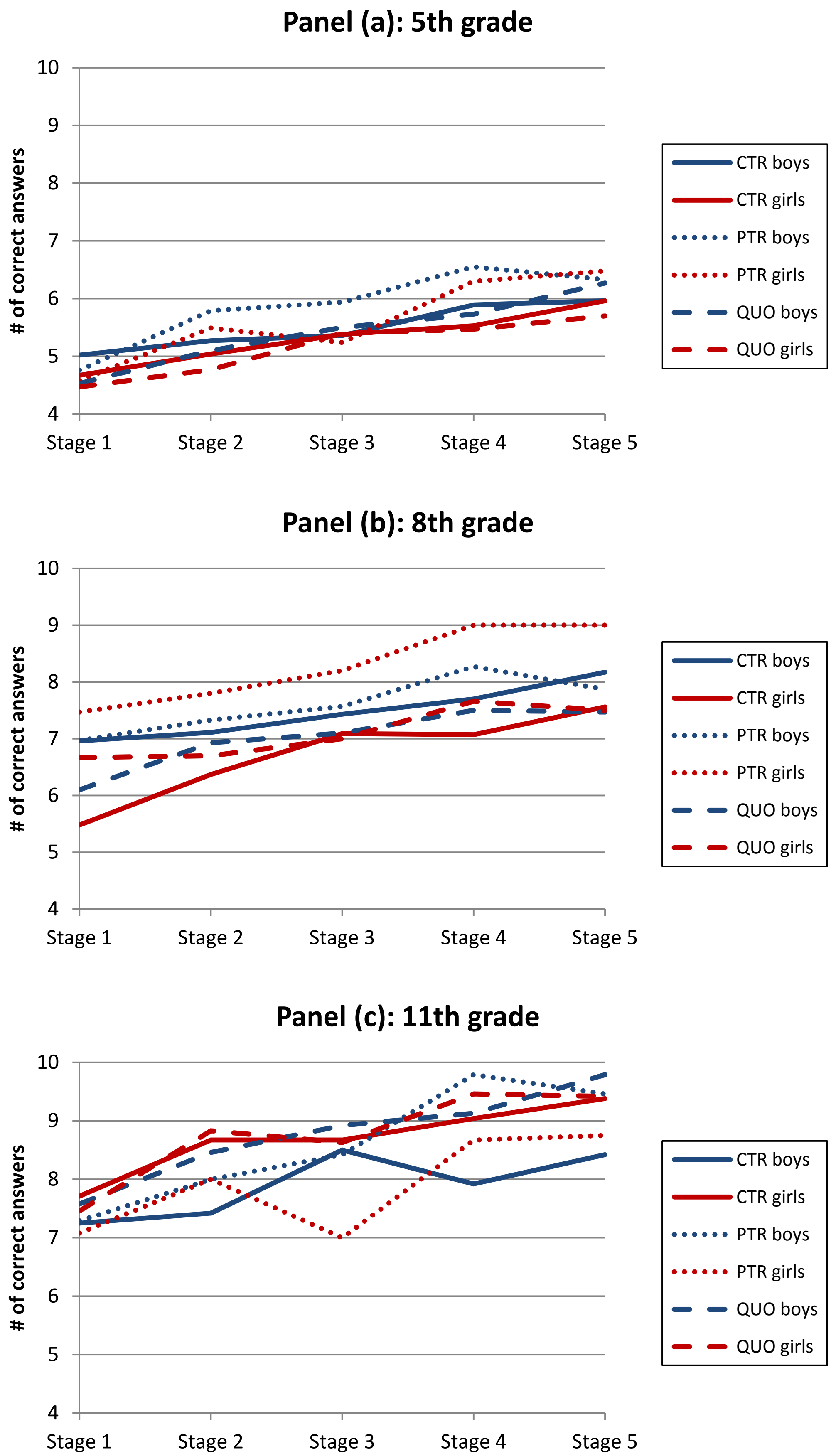


Figure 2. Competition entry choices, by ability level

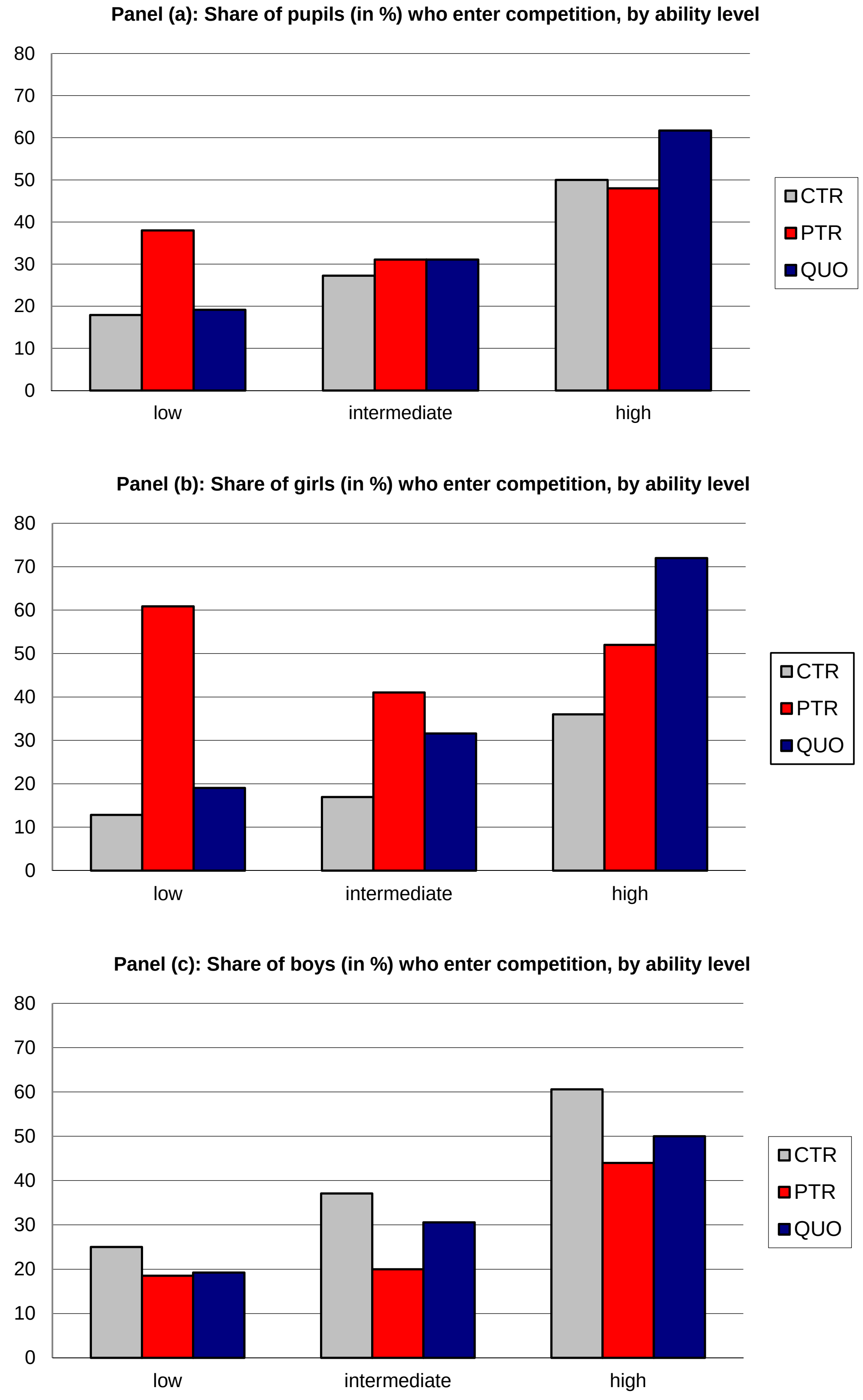
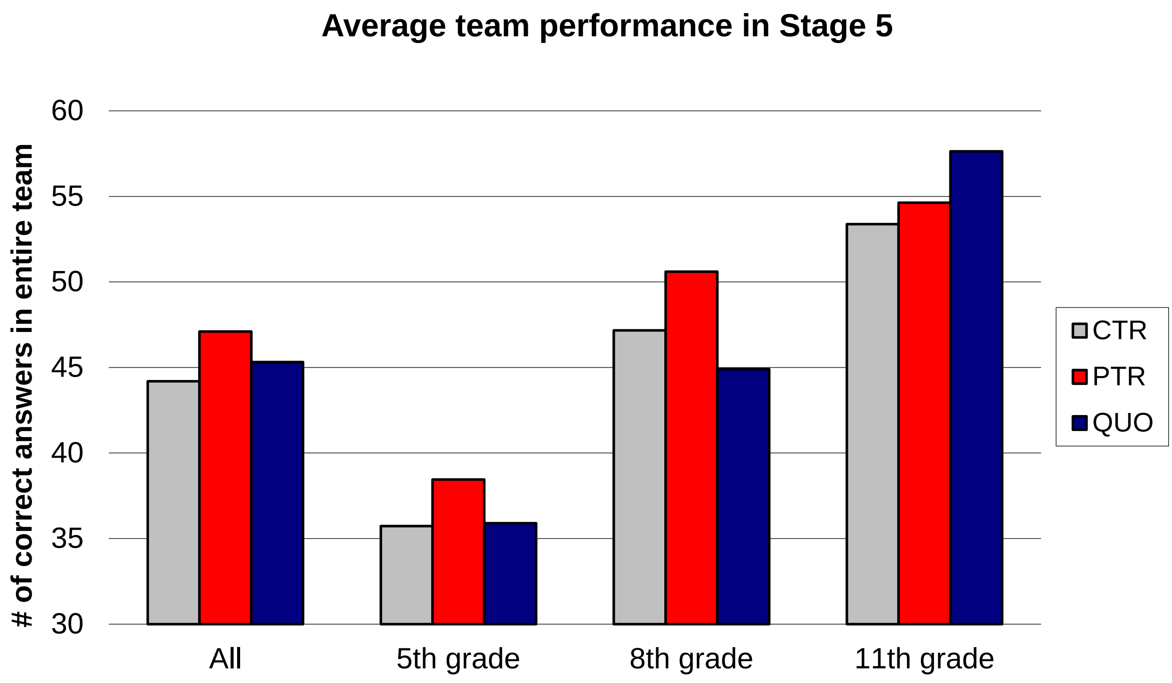


Figure 3. Post-tournament efficiency



Appendix: Experimental instructions [translated from German] – *Not intended for publication*

Instructions were memorized by the experimenter and orally presented in class. From time to time the instructor paused and let the subjects raise their hands for questions. All questions were answered privately. Originally the instructions were in German. An English translation of the instruction is presented below.

Welcome to our game. We thank you for your participation. This game will have a duration of 2 school hours with no breaks. In this game you can earn money. During the experiment, you and the other participants will be asked to make certain decisions. Your own decisions as well as the decisions of the other participants will determine your payment from the experiment, according to the rules that will be described in what follows. That's why it is important that you understand the rules of the game. I will frequently stop the explanation in order that you can ask your questions. If you do have a question, please raise your hand. One of us will then come to you and answer your question privately. From now on, please don't speak to your neighbor and listen carefully. The experiment will be conducted on the computer. You make your decisions on the screen. All decisions and answers will remain confidential and anonymous. You will receive your money at the end of the game.

Everybody ok so far? *Leave time for questions.*

The game consists of 5 stages. One of the first four stages (1-4) will be randomly selected for your payment. On top of that also the fifth stage will be paid out. The random selection of one of the first four stages will be done like this: We have brought along 4 cards, numbered from 1 to 4. At the end of the game one participant will draw one of the four cards. The card drawn will determine the stage which is relevant for payment. Your total earnings from the game will be the sum of your payments for the randomly selected stage plus your earnings from the fifth stage.

Everybody ok so far? *Leave time for questions.*

You will receive instructions for each of the five stages, one after the other.

Stage 1: Piece rate

Now I am going to explain stage 1. Your task in stage 1 is to solve correctly as many addition exercises as possible. To be more precise, you will have 2 minutes' time in order to solve as many additions of three randomly selected two-digit numbers as possible. The numbers are randomly picked by the computer software, whereby the numbers will be drawn from the interval of 10 to 99. You are not allowed to use calculators but you can use the provided scribbling paper for your calculations. You have to enter your answer on the decision screen and then click with the mouse on the "Confirm" button. When you enter an answer, you immediately find out on the screen whether it was correct or not. At the end of the 2 minutes you will be given a screen where you can see how many problems you solved correctly. If stage 1 is the stage selected for payment (among stages 1-4), then you will receive **€0.40 for each correct answer**. Your payment is not reduced when you enter a wrong answer. Directly before the start of this stage you will be given two minutes in order to familiarize yourselves with the screen: During this time you can solve addition exercises, which do not count for the payment. Afterwards, stage 1 will begin after a short break.

Everybody ok so far? *Leave time for questions.*

Stage 2: Tournament

Now I am going to explain stage 2. This stage is the same as stage 1, except that your payments will be computed differently. The computer will randomly match groups consisting of **6 participants, 3 of whom are men and 3 are women.** Groups are randomly formed at the beginning of this stage and

each participant stays in the same group until the end of the experiment. You will not find out the identity of the other participants in your group during or after the game, so that all decisions remain anonymous. If stage 2 is the stage selected for payment (among stages 1-4) then only the two winners in each group receive a payment. The two winners are the two group members who have entered the most correct answers. The two winners receive **€1.20 per correct answer** each, while the other four members **do not receive any payment**. In case of a tie, the ranking among the members with equal performances is determined randomly. You will not be informed about the outcome of the tournament until the end of the game.

Everybody ok so far? *Leave time for questions.*

Stage 3: Choice [Control treatment]

Now I am going to explain stage 3. As in stages 1 and 2, you will have 2 minutes' time in order to solve correctly as many addition exercises as possible. However, you must now choose yourself your preferred payment method for your performance in stage 3. You can either choose the Piece-rate payment as in stage 1 or the Tournament payment as in stage 2. If stage 3 is the stage selected for payment (among stages 1-4), then your payment is determined as follows:

- If you choose the **Piece-rate payment**, then you will receive **€0.40 per correct answer**
- If you choose the **Tournament payment**, then you will receive **€1.20 per correct answer**, but only if you are one of the two winners of your group. Otherwise you receive nothing. In order to find the two winners we will compare your performance in stage 3 to the performance of the other five group members **in stage 2**. This guarantees that there will be still 6 people in each group. In case of a tie, the ranking among the members with equal performances is again determined randomly.

The group composition (with 3 men and 3 women) is as in stage 2. If you choose the Tournament payment, you will not be informed about the outcome of the tournament until the end of the game. On the next screen you will be asked whether you want to choose the Piece-rate payment or the Tournament payment for your performance in stage 3. Afterwards you will have 2 minutes in order to calculate the sums of the two-digit numbers.

Everybody ok so far? *Leave time for questions.*

Stage 3: Choice [Preferential treatment]

Now I am going to explain stage 3. As in stages 1 and 2, you will have 2 minutes' time in order to solve correctly as many addition exercises as possible. However, you must now choose yourself your preferred payment method for your performance in stage 3. You can either choose the Piece-rate payment as in stage 1 or the Tournament payment as in stage 2. If stage 3 is the stage selected for payment (among stages 1-4), then your payment is determined as follows:

- If you choose the **Piece-rate payment**, then you will receive **€0.40 per correct answer**
- If you choose the **Tournament payment**, then you will receive **€1.20 per correct answer**, but only if you are one of the two winners of your group. Otherwise you receive nothing. In order to find the two winners we implement a new rule.

In the Tournament, the number of every woman's correctly solved exercises is automatically increased by 1. Thus, in each group, all 3 women receive one additional point each, while the 3 men receive no additional points. As in stage 2, the two winners of the Tournament are then the two group members with the best performances (taking the additional point for women into account). We will compare your performance in stage 3 to the performance of the other five group members **in stage 2** (also here we take on additional point for women into account). Comparing your performance to the performance of the other five group members in stage 2, guarantees, that there will be still 6 people in each group. The additional points are taken into account for the determination of the two winners, but not for the payment. In case of a tie, the ranking among the members with equal performances is again determined randomly. The group composition (with 3 men and 3 women) is as in stage 2. If you choose the Tournament payment, you will not be informed about the outcome of the tournament until the end of the game. On the next screen you will be asked whether you want to choose the Piece-rate payment or the Tournament payment for your performance in stage 3. Afterwards you will have 2 minutes in order to calculate the sums of the two-digit numbers.

Everybody ok so far? *Leave time for questions.*

Stage 3: Choice [Minimum quotas treatment]

Now I am going to explain stage 3. As in stages 1 and 2, you will have 2 minutes' time in order to solve correctly as many addition exercises as possible. However, you must now choose yourself your preferred payment method for your performance in stage 3. You can either choose the Piece-rate payment as in stage 1 or the Tournament payment as in stage 2. If stage 3 is the stage selected for payment (among stages 1-4), then your payment is determined as follows:

· If you choose the **Piece-rate payment**, then you will receive **€0.40 per correct answer**

· If you choose the **Tournament payment**, then you will receive **€1.20 per correct answer**, but only if you are one of the two winners of your group. Otherwise you receive nothing. In order to find the two winners we implement a new rule.

In the Tournament, the two winners are determined as follows. In each group, one of the two winners is in any case the woman with the best performance (of all three women). The other winner is the group member with the best performance among the remaining members (i.e., excluding the best-performing woman). In order to find the two winners we will compare your performance in stage 3 to the performance of the other five group members **in stage 2**. This guarantees that there will be still 6 people in each group.

Let's go through some examples: If 2 girls are the best performers within a group then these two girls are the winners. If one girl and one boy are the best performers within a group then these two people are the winners. If 2 boys are the best performer within a group then only the better performing boy is the winner. The second winner is the best performing girl in the group. Therefore there will be always at least one girl among the winners and at most one boy.

In case of a tie, the ranking among the members with equal performances is again determined randomly. The group composition (with 3 men and 3 women) is as in stage 2. If you choose the Tournament payment, you will not be informed about the outcome of the tournament until the end of the game. On the next screen you will be asked whether you want to choose the Piece-rate payment or the Tournament payment for your performance in stage 3. Afterwards you will have 2 minutes in order to calculate the sums of the two-digit numbers.

Everybody ok so far? *Leave time for questions.*

Expectations:

Before we are turning to stage 4, you are asked to evaluate your performance in stage 1 and stage 2. For each correct guess you will additionally earn €0.50. Here I show you how the decision screen will look like *(point to the overhead projector)*. This slide refers to stage 1. First, you are asked for your belief concerning your ranking in the whole group. E.g. if you think that you have solved the most correct answers in stage one, than you click on the button saying "First place". The group composition is the same as in stage 2 and 3. Next you are asked to state your beliefs concerning your ranking within your gender. This means that a girl gives her ranking within the three girls of her group. It's not possible, e.g. to belief to be first ranked in the whole group and second ranked within your gender.

Everybody ok so far? *Leave time for questions.*

Stage 4: Tournament [Control treatment]

Now I am going to explain stage 4. This stage is the same as stage 2. I quickly repeat how it works: If stage 4 is the stage selected for payment (among stages 1-4), then the two winners receive **€1.20 per correct answer** each, while the other four members of the group receive **no payment**. In case of a tie, the ranking among the members with equal performances is again determined randomly.

Furthermore, the tournament in this stage also serves the purpose of determining the initial endowment of every group member in stage 5. This is done as follows: The two winners of stage 4 receive then in stage 5 an initial endowment of **€3.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]; the other four group members (i.e., the non-winners) receive an initial endowment of **€1.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]. At the end of this stage you will be informed about the outcome of the tournament (whether you have won or lost the tournament) and thereby about your initial endowment in stage 5.

Everybody ok so far? *Leave time for questions.*

Stage 4: Tournament [Preferential treatment]

Now I am going to explain stage 4. This stage is similar to stage 3. The group composition is also the same. There is only one difference between stage 3 and stage 4 which is that there is no decision between a Piece-rate payment and a Tournament payment. Instead all participants are paid according to a Tournament payment if stage 4 is the stage selected for payment (among stages 1-4). As in stage 3 all girls receive an additional point also in stage 4. The two winners receive **€1.20 per correct answer** each, while the other four members of the group receive **no payment**. In case of a tie, the ranking among the members with equal performances is again determined randomly.
Furthermore, the tournament in this stage also serves the purpose of determining the initial endowment of every group member in stage 5. This is done as follows: The two winners of stage 4 receive then in stage 5 an initial endowment of **€3.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]; the other four group members (i.e., the non-winners) receive an initial endowment of **€1.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]. At the end of this stage you will be informed about the outcome of the tournament (whether you have won or lost the tournament) and thereby about your initial endowment in stage 5.

Everybody ok so far? *Leave time for questions.*

Stage 4: Tournament [Minimum quotas treatment]

Now I am going to explain stage 4. This stage is similar to stage 3. The group composition is also the same. There is only one difference between stage 3 and stage 4 which is that there is no decision between a Piece-rate payment and a Tournament payment. Instead all participants are paid according to a Tournament payment if stage 4 is the stage selected for payment (among stages 1-4). As in stage 3 at least one of the two winners has to be a girl. The two winners receive **€1.20 per correct answer** each, while the other four members of the group receive **no payment**. In case of a tie, the ranking among the members with equal performances is again determined randomly.
Furthermore, the tournament in this stage also serves the purpose of determining the initial endowment of every group member in stage 5. This is done as follows: The two winners of stage 4 receive then in stage 5 an initial endowment of **€3.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]; the other four group members (i.e., the non-winners) receive an initial endowment of **€1.00** each [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]. At the end of this stage you will be informed about the outcome of the tournament (whether you have won or lost the tournament) and thereby about your initial endowment in stage 5.

Everybody ok so far? *Leave time for questions.*

Stage 5

In stage 5, the two winners of the tournament in stage 4 have an endowment of € 3 [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]; the non-winners have an endowment of € 1 [60 % thereof for participants in grade 8; 40 % thereof for participants in grade 5]. As in stages 1-4, you will have 2 minutes' time in order to solve correctly as many addition exercises as possible. The group composition (with 3 men and 3 women) is the same as before. Your payment is determined as follows: The group receives € 0.40 per correct problem solved within the group. Thus, everybody receives 40Cent/6=**6.67 euro cent** for each correct answer that a member of your group has entered in the 2 minutes. You can think of this like this: Your group forms a team. At the end we will count how many correct problems have been solved within the team. For each correct answer everybody will receive 6.67 euro cent. This means that all members of a group receive the same payment in this stage.
Your payment in stage 5 is the sum of your endowment and your payoff from the performance of your group. At the end of the experiment you will be informed about the total performance of your group in this stage.

Everybody ok so far? *Leave time for questions.*