Essays on Gender, Competition and Status

Essays on Gender, Competition and Status

Eva Ranehill

Dissertation for the Degree of Doctor of Philosophy, Ph.D. in Economics
Stockholm School of Economics 2011

Keywords: Gender; Competition; Risk; Status; Third Party Punishment

Essays on Gender, Competition and Status

ISBN 978-91-7258-852-3

Printed in Sweden by:
Intellecta Infolog, Göteborg 2011

Distributed by:
The SSE Research Secretariat
Stockholm School of Economics
Box 6501, SE-113 83 Stockholm, Sweden
www.hhs.se

To my family

Preface

This report is a result of a research project carried out at the Department of Economics at the Stockholm School of Economics (SSE).

This volume is submitted as a doctor's thesis at SSE. The author has been entirely free to conduct and present her research in her own ways as an expression of her own ideas.

SSE is grateful for the financial support provided by the Jan Wallander and Tom Hedelius Foundation as well as the Carl Silfvén Foundation.

Filip Wijkström
Associate Professor
SSE Director of Research

Acknowledgements

This thesis would neither have been started nor finished without the help and support of many people. First of all I want to thank Magnus Johannesson, my supervisor, without whom I would not have started, certainly not pursued, and hence not finished my PhD. Thanks for believing in me, for great discussions and advice, and for a lot of your time! All PhD students should have a supervisor like you.

Tore Ellingsen has also provided invaluable feedback on ideas and drafts. Tore has an astonishing capacity to provide relevant feedback on the most varying topics.

Many fellow PhD students made life on the 8^{th}, and later 7^{th}, floor easier. Thanks to all of you for stimulating conversations, even (or especially perhaps) when we disagreed, and thanks for contributing to the social and academic environment at Stockholm School of Economics! To mention a few, Robert Östling provided invaluable help when it was the most needed (on various topics, most critically some reinforcement in my ongoing battle with stata commands). Erik Lindqvist kindly took the time to discuss methodological issues with me. Linus Siming always shared some new gossip and made us all laugh. Johan Almenberg kindly read many of the papers included in this volume and gave many insightful comments as well as English corrections. Mark Bernard, Damian Migueles Chazarreta, Amanda Jakobsson, Kristin Magnusson Bernard, Ignat Stephanok, and Claudia Wolff provided great company. During my last coursework, I had the opportunity to study with Abel Schumann, Palle Elger and Ronny Freier. Thanks for making the last course work a lot more fun! Karin Hederos Eriksson and Anna Sandberg were also a great asset to the PhD program. It was great to get to know you all! For many inspiring moments, I also want to thank Anne Boschini, Peter Langenius, Astri Muren, and Mats Persson. I always found our discussions interesting, and I am thankful for your encouragement and support.

Some of my colleagues have also become close friends. Tobias Lundquist, without you I had probably not started my PhD. Thanks for being there encouraging me in the beginning, and thanks for being such a great person and friend! Emma von Essen and Anna Dreber Almenberg, without you I never would have continued. You both made a huge difference, both personally and professionally. Thanks for being there during ups and downs. My roommate Erik Mohlin has also been very important to me. Besides providing an impeccable judgment on most things, Erik has been a valuable friend, and a patient guide in the world of economic theory.

I also want to thank my collaborators with whom I have worked with on the papers in this thesis as well as on upcoming ones: Konstanze Albrecht, Anne Boschini, Juan-Camilo Cárdenas, Anna Dreber Almenberg, Tore Ellingsen, Emma von Essen, Armin Falk, Magnus Johannesson, Karen Khachatryan, Astri Muren, Niklas Zethraeus, and Robert Östling. Thanks for great collaborations and inspiring conversations.

The administrative staff at the Department of Economics also deserves to be mentioned. Anna Angerman, Carin Blanksvärd, Ritva Kiviharju, Anneli Sandblad, and Lilian Öberg provide efficient and friendly help. Without them the department would not function.

I spent almost two years of my doctoral studies at the Department of Economics at the University of Zürich. I wish to thank Ernst Fehr for hosting me and for giving valuable comments on my research. The Department of Economics at the University of Zürich provides an outstanding academic environment, and I have benefited a lot from discussions with many of the people working there.

Financial support from the Jan Wallander and Tom Hedelius Foundation as well as the Carl Silfvén Foundation is also gratefully acknowledged. It made much of my research, and my stay abroad, possible.

Finally I want to thank my family for always being there for me no matter what. My husband Ludovic may not always have understood why I worked long hours for no money, but never stopped supporting me in doing what I want. My son Viggo also helped by providing a good work life balance. Since he was born nine months ago he has provided the most decisive distraction from work. And finally, maybe now my parents will stop asking whether I am "in school" today or not…

Zürich, May 2011
Eva Ranehill

Table of Content

1 Introduction

This thesis consists of 6 papers, all of them using experimental methods. The experimental approach appeals to me, since a good design may allow a clear cut identification of the effects under study. In spite of all its advantages, however, the typical experimental study performed in the laboratory also has drawbacks due to its stylized setting and often restricted subject pool. I believe that enlarging the value of laboratory experiments by understanding more about the link from the lab to "real life" is important. I attempt to do this in my research by combining lab and field studies in order to allow comparison of the results, by using non-standard samples, or by staging experimental studies outside of the lab in a more natural setting. The papers included in this thesis range from natural field experiments, where the participants did not know that they were participating in a study (Dreber, von Essen and Ranehill, 2009), across mixtures of field and lab studies (Cárdenas, Dreber, von Essen and Ranehill, 2010), framed field experiments (Johannesson, Östling and Ranehill 2010), and to pure lab experiments (von Essen and Ranehill, 2011; and Dreber, von Essen and Ranehill, 2011).

The first three studies in this volume investigate gender differences in competitiveness across cultures, age groups, and tasks. Competitiveness in economics is measured either as a performance increase as a response to a competitive setting compared to an individual setting or as self-selection into a competitive setting. Typically, an individual setting is staged in such a way that participants in a study perform a task, and are paid independently of the other participants' performance, for example according to a piece rate payment scheme. In a competitive setting, on the other hand, the payment structure mimics a tournament, where only the best in a randomly selected group is paid a premium. Gender differences in preferences for competition are interesting because they have been proposed as one mechanism through which gender differences in the labor market arise. Men are typically more competitive and risk seeking than women, and are therefore hypothesized to perform better under competitive pay contracts and to self-select into highly competitive professions to a larger extent than women. Occupations with varying pay tend to pay more on average, due to compensating wage differentials. The original economic studies that brought this hypothesis forward and illustrated the gender gap in competitiveness did so mainly in samples of university students solving math exercises or mazes under time pressure and under different pay schemes (Gneezy, Niederle and Rustichini, 2003, Datta Gupta, Poulsen and Villeval, 2005, Niederle and Vesterlund 2007) .

The first paper in this volume, "Outrunning the gender gap" came about as a critique and extension to this rather stereotypic setting by providing variation in the sample, the setting, and the tasks. Previous studies have shown that if anywhere, gender differences exist in spatial ability and possibly some areas of mathematics (Voyer, Voyer and Bryden, 1995, Else-Quest, Shibley Hyde and Linn. 2010). In addition, stereotypes about the general capacity of different groups have been shown to influence the performance of the disadvantaged group negatively (Spencer, Steele, & Quinn, 1999). Hence, male associated areas such as mathematics and spatial ability may not be suitable for investigating gender differences in competitive performance and choice. One previous study that looked at competitiveness in running in a sample of Israeli children aged 10-11 found that only boys reacted to a competitive setting by running faster than they did in an individual setting (Gneezy and Rustichini, 2004). We let children aged 8-10 years in Sweden, one of the world's most gender neutral countries according to gender equality indices, compete in running, skipping rope, and dancing during PE classes. We thus introduced two tasks in an area associated with female skills to investigate whether this caused girls to compete relatively more. Contrary to previous literature, we did not find any gender differences in performance increase under a competitive setting compared to an individual in any task. Boys and girls competed equally.

Encouraged by this result, we moved on with studies number two and three, and investigated competitive behavior and risk taking across cultures and ages. We examine the impact of culture and gender norms on girls' and boys' competitive behavior in chapter two, "Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden", where we look at competitiveness in a large sample of children aged 9-12 in Sweden and Colombia. Our hypotheses were that gender differences in behaviors would be more pronounced in the Colombian sample than in the Swedish one. As in the first paper in this volume, this study investigates competitive behavior among children during their physical education class in skipping rope and running. In addition, it also includes competition in a math and a language task during a regular school class as well as a task that measures risk behavior. The choice to compete implies a riskier choice than the choice not to do so, and controlling for risk is necessary since men and boys are often found to be less risk averse than women and girls (Croson and Gneezy, 2009). If not controlled for, potential gender differences in competitive behavior may simply be due to underlying differences in risk preferences. In many ways, the results in this study did not correspond to our hypotheses. Colombian boys and girls competed equally in both measurements of competitiveness mentioned above, namely performance and choice, and we found mixed results in Sweden. We found an indication that girls increase their performance more under tournament incentives, but that boys are more likely to select competition when the choice is available. This illustrates that gender differences in competitiveness do not generalize easily across cultures, tasks, and samples.

The second implication of our first study was that competition may not only vary with culture, but also with age. We did not find any gender differences among young boys and girls, but would we get the typically gendered behavior found in previous studies among Swedish adolescents? If gender differences in preferences can explain part of the gender gap in labor market outcomes, assessing gender differences before individuals enter the labor market is relevant. Many important decisions that have implications for labor market outcomes, such as education choices, are taken during adolescence. For this reason we set out to study gender differences in economic preferences among individuals aged 16-18 years in Sweden. Apart from studying competitiveness in math and language, we also measured altruism and risk taking, thus focusing on the three areas where the most robust gender differences in preferences are found in the economic literature. They are all potentially important for labor market behavior. Our results largely support previous literature. Adolescent girls are more altruistic and risk averse than adolescent boys. In terms of competitiveness, we find no gender difference in performance change as a response to competitive settings. We do find, however, that boys more often select a competitive payment scheme in the math task, but not the language task. Hence gender differences in competitive behavior are task dependent in this case.

Both we and other researchers presented new studies after we began these research projects. Now knowing much more about the shape of gender differences in competitiveness across ages, tasks, and cultures, I look forward to continuing this strand of research to add to our knowledge regarding the determinants of these differences, as well as their implications for behavior in real life settings and individual economic outcome.

The three last papers in this volume are on different themes. Still on the theme of competition, paper four deals with one of today's major health challenges: the trend in increasing BMI and decreasing physical activity observed in Western countries. I study how the introduction of the competitive element in a "step contest" can be used to promote physical activity in a large framed field experiment. The use of pedometers has previously been found to increase physical activity by itself (Bravata et al. 2007). Participants in a step contest not only use a pedometer to record their steps, but register the number of steps they walk daily and compete both individually and in teams for a symbolic prize. We find that in addition to the base line effect of pedometer use found in Bravata et al. (2007), step competitions increase the number of steps walked per day by about 1000 steps. Compared to the average effect of pedometer use found in previous studies, this represent an additional increase in the number of steps walked per day of 40%. This is a sizeable additional effect on physical activity. One long term goal of my research is to apply knowledge from behavioral economics to real world problems. The increasing trend in obesity is one area where I think behavioral economics may provide useful insights and it is an area where I would like to continue working.

The next study was the first project I began as a PhD student. When I began my PhD, I found myself in what I considered a rather peculiar male hierarchy. Having always been interested in the effect of status and hierarchy on peoples' perception, I decided to do a study on this theme. Status has been largely ignored in economics, apart from research on consumer patterns (Veblen, 1899, Frank, 2000). In contrast, status is treated in sociology and psychology as a fundamental framework through which we understand our social environment and assign entitlements and appropriateness. As such, social status is also likely to influence our decisions in economic interactions. A few previous economic studies indicated that high status individuals made more money in economic experiments (Ball and Eckel, 1996 and 1998, Ball et al., 2001, and Glaser et al., 2000). Interestingly, the driving factor in these studies did not seem to be high status individuals behaving differently, but rather that low status individuals differentiated between high and low status individuals by proposing better deals, or by acting more trustworthily towards high status opponents than low status opponents. Participation in these studies was not anonymous, however, and the experimental design thus did not exclude strategic concerns in relation to future interaction. These studies also all pertained to the domain of giving. I was interested in whether high social status also "buys you a larger action span", that is if high status individuals are less likely to be punished for the same actions as low status individuals. Social hierarchy is inherently linked to punishment. You may perhaps avoid being generous to your superior, but you would not punish him or her. We therefore introduced social status in a dictator game with third party punishment. Social status was inconspicuously implemented by adding the names of the dictator on the third party's decision sheet. The participants in the dictator game never knew the name of the third party. Half of the third parties then faced a noble dictator, and half faced a common dictator. We find that only men react to social status, and mainly in male to male interactions. Male third parties punish common males almost twice as much as noble males. This result suggests that social status has important implications for men's decisions to use economic punishment, and that this holds true in situations where reputation or strategic concerns have no importance. Though the results from this study should be interpreted with caution due to the small sample size, I think our finding is an interesting indication for future studies.

The last chapter in this volume touches on a slightly different subject than the other chapters. Research in sociology and development psychology suggests that pubertal timing is correlated with educational outcomes. In particular, some studies show that girls that mature early have, on average, lower grades (Simmons and Blyth 1987, Semon Dubas et al. 1991, Cavanaugh et al. 2007), lower academic goals (Graber et al. 1997), and a higher probability of dropping out of school early (Cavanaugh et al. 2007). We use a survey of Swedish adolescents aged 15-16 years to investigate some of the potential mechanisms through which this effect may be mediated. In particular, we measure attitudes to risk, time preferences, and priorities of school versus friends. Some of the most salient characteristics of adolescence include an increase in the importance attached to peer relations and an increase in behaviors with inherently risky and impulsive elements, such as drinking, smoking, and having unprotected sex (Arnett 1999, Steinberg 2010, Boyer 2006). These behaviors have previously also been linked to low academic achievement among adolescents. We find that girls that mature early have lower grades as well as lower educational aspirations. However, we do not find any evidence that risk attitudes, time preferences or changes in priorities regarding school versus friends mediate the relation between puberty and educational outcome. There is no correlation between the potential mediating factors and pubertal development in our data. Future studies should attempt to further investigate the mechanisms behind the correlation between pubertal timing and educational outcomes, preferably in a large longitudinal sample.

1.1 References

Arnett J., 1999, "Adolescent storm and stress, reconsidered", *American Psychologist,* 54:317–26.

Ball S., Eckel C., 1996, "Buying status: experimental evidence on status in negotiation", *Psychology and Marketing*, 13, 381-405.

Ball S. Eckel C., 1998, "Stars upon thars", Working Paper. The University of Texas at Dallas.

Ball S., Eckel C., Grossman P. Zame W., 2001, "Status in markets", *The Quarterly Journal of Economics*, 116:1, 161-188.

Boyer T., 2006, "The development of risk-taking: a multi-perspective view" *Developmental Review*, 26: 291–345.

Bravata D.M., Smith-Spangler C., Sundaram V., Gienger A. L., Lin N., Lewis R., Stave C. D., Olkin I., Sirard J. R., 2007, "Using pedometers to increase physical activity and improve health", *Journal of the American Medical Association* 298, 2296–2304.

Cárdenas J.C., Dreber A., von Essen E., Ranehill E., "Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden", Forthcoming in *Journal of Economic Behavior and Organization*.

Cavanagh S., Riegle-Crumb C., Crosnoe R., 2007, "Puberty and the education of girls" *Social Psychology Quarterly*, 70:2, 186-198.

Croson R., Gneezy U., 2009, "Gender differences in preferences." *Journal of Economic Literature*, 47(2): 1-27.

Datta Gupta, N., Poulsen, A., Villeval, M.C., 2005, "Male and female competitive behavior – experimental evidence". IZA Discussion Paper, No. 1833.

Dreber, A., von Essen, E., Ranehill, E., "Outrunning the gender gap: boys and girls compete equally", Forthcoming *Experimental Economics*.

Dreber, A., von Essen, E., Ranehill, E., 2011, "In bloom: gender differences in preferences among adolescents", Working Paper Series in Economics and Finance 734, Stockholm School of Economics.

Else-Quest, N.M., Shibley Hyde, J., Linn, M.C., 2010, "Cross-national patterns of gender differences in mathematics: a meta-analysis." *Psychological Bulletin*, 136 (1): 103-127.

Von Essen, E., Ranehill E., 2011, "Status and submission: social status biases economic sanctions", Working Paper Series in Economics and Finance 732, Stockholm School of Economics.

Frank, R., 2000, *Luxury Fever: Money and Happiness in an Era of Excess,* Princeton: Princeton University Press 2000.

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. and Soutter, C. L., 2000. "Measuring trust". *Quarterly Journal of Economics*, 115, 811-846.

Gneezy, U., Niederle, M., Rustichini, A., 2003, "Performance in competitive environments: gender differences." *Quarterly Journal of Economics*, 118(3): 1049-1074.

Gneezy, U., Rustichini, A., 2004, "Gender and competition at a young age." *American Economic Review*. 94(2): 377-381.

Graber, J. A., Lewinsohn, P. M., Seeley J. R., Brooks-Gunn, J., 1997, "Is psychopathology associated with the timing of pubertal development? *Journal of the American Academy of Child and Adolescent Psychiatry* 36 (12)

Johannesson, M., Östling, R., Ranehill, E., 2010, "The effect of competition on physical activity: a randomized trial", *The B.E. Journal of Economic Analysis & Policy*: Vol. 10 : Iss. 1 (Topics), Article 91.

Niederle, M. Vesterlund, L., 2007, "Do women shy away from competition? Do men compete too much?." *Quarterly Journal of Economics*, 122, 1057–1101.

Semon Dubas, J., Graber, J. A., Petersen, A.C., 1991, "The Effects of Pubertal Development on Achievement during Adolescence", *American Journal of Education*, 99(4): 444-460.

Simmons R. G., and D. A. Blyth (1987). *Moving into adolescence.* Hawthorne, N.Y.: Aldine De Gruyter.

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). "Stereotype threat and women's math performance." *Journal of Experimental Social Psychology*, 35, 4–28.

Steinberg, Laurence. 2010. "Commentary: a behavioral scientist looks at the science of sdolescent brain development." *Brain Cognition*, 72(1): 160-164.

Veblen, T., 1899, *Theory of the Leisure Class: An Economic Study in the Evolution of Institutions*. New York: Macmillan.

Voyer, D., Voyer, S., Bryden, P.M., 1995, "Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables." *Psychological Bulletin*, 117(2):250-270.

2 Outrunning the Gender Gap – Boys and Girls Compete Equally

By: Anna Dreber, Emma von Essen, and Eva Ranehill

We are grateful for comments from Johan Almenberg, Anne Boschini, Hannah Riley Bowles, Armin Falk, Uri Gneezy, Moshe Hoffman, Magnus Johannesson, Astri Muren, Paul Nystedt, Mats Persson, David G. Rand and seminar participants at the Gender in Negotiation and Decision Making Seminar at Harvard Kennedy School, Harvard Business School, Harvard University, Ratio Institute, Santa Fe Institute, Stockholm School of Economics, Stockholm University, the 3rd Nordic Conference on Behavioral and Experimental Economics, University of Bonn and University of Zürich. Financial support from the Jan Wallander and Tom Hedelius Foundation (A.D., E.R.) and the Carl Silfvén Foundation (E.R) is gratefully acknowledged, as well as financial support from the Swedish Council for Working Life and Social Research (FAS).

ABSTRACT: Recent studies find that women are less competitive than men. This gender difference in competitiveness has been suggested as one possible explanation for why men occupy the majority of top positions in many sectors. In this study we explore competitiveness in children, with the premise that both culture and gendered stereotypes regarding the task at hand may influence competitive behavior. A related field experiment on Israeli children shows that only boys react to competition by running faster when competing in a race. We here test if there is a gender gap in running among 7-10 year old Swedish children. We also introduce two female sports, skipping rope and dancing, to see if competitiveness is task dependent. We find no gender difference in reaction to competition in any task; boys and girls compete equally. Studies in different environments with different types of tasks are thus important in order to make generalizable claims about gender differences in competitiveness.

2.1 Introduction

Men occupy the majority of top positions in most societies, both in the private and in the public sector. The proposed reasons for this remain highly controversial within academia as well as politics (Ceci and Williams 2006). Today, women in many countries are at least as likely as men to pursue higher education, and female labor force participation has risen to levels similar to that of men. Meanwhile, a number of recent studies show that women compete less than men. Competitiveness is typically measured as either a preference for competition, such as self-selecting into a tournament instead of a piece-rate payment scheme, or by the performance response as a reaction to a competitive setting compared to a non-competitive setting. Many studies find that only males perform better under competition (Gneezy et al. 2003, Gneezy and Rustichini 2004a). It has also been shown that women tend to prefer the non-competitive setting even when there is no gender gap in performance in the competitive setting and that men compete more than what is optimal for them while women compete less (Niederle and Vesterlund 2007). Some studies find that competitiveness depends on the gender of the opponent(s) (Gneezy et al. 2003, Gneezy and Rustichini 2004a, Datta Gupta et al. 2005, Price 2008) whereas some find that women's competitiveness depend on the institutional framework (e.g., Niederle and Yestrumskas 2008, Niederle et al. 2009, Balafoutas and Sutter

2010). These gender differences have been suggested as a possible explanation for the gender gap in the labor market. The policy implications of a gender gap in competitiveness depend on the causes of the gap. Whether these gender differences are innate or acquired later in life remains unknown. Children therefore provide an interesting subject pool for the study of this distinction.

In this paper, we explore whether there are gender differences in competitiveness among children. Two previous studies also investigate this. Booth and Nolen (2009) look at willingness to compete in solving mazes among adolescent boys and girls from single sex schools and from mixed schools. Boys compete equally in both schools and more than girls do, whereas girls in single sex schools compete more than girls from mixed schools. In a field experiment looking at 9-10 year old Israeli children, Gneezy and Rustichini (2004a) find that boys, but not girls, respond to competition by running faster against another child than when running alone. Moreover, they find that the gender of the opponent matters only for girls, who compete less when running against another girl.

In this study we run a field experiment on 7-10 year old children in Sweden. The design is inspired by that of Gneezy and Rustichini (2004a), where the children compete in running. In addition, in our study the children also compete in skipping rope and dancing. The running task is included in order to have a comparison to previous work, while varying country (Israel vs Sweden), even though some parameters differ between the two studies. The other two tasks are included to study whether there are male and female areas of competition. If tasks are gendered, it is possible that this leads to gender differences in both motivation for, and payoffs from, competing. Most competitiveness studies build on tasks such as solving mazes and performing simple arithmetic, which are generally considered as male tasks. Several studies show that women perform worse on standardized tests when they are reminded of negative stereotypes about female math ability (Inzlicht and Ben-Zeev 2003, O'Brien and Crandall 2003, Shih et al. 1999, Steele 1997).[1] This kind of stereotype threat has been suggested as one rea-

[1] Interestingly when women are told that there are no differences between men and women in abstract math tests, women perform as well as men (Spencer et al. 1999).

son why women in mixed gender groups compete less than men in some of the tasks previously studied in this literature (Gneezy et al. 2003). There is mixed evidence on the role of the task on the gender gap in competitiveness. Günther et al. (2009) and Grosse and Reiner (2010) find a gender difference in performance change in a math task but not in a word task, whereas Wozniak et al (2010), using a maze task and the same word task, find no difference in the gender gap between the two tasks.[2] Thus, to explore competitiveness more generally than what has previously been done, in particular on children, we also look at what we consider more female tasks. Since our experiment is conducted with children, our inspiration comes from tasks that children perform. The tasks were chosen in agreement with the teachers.

Competitiveness is measured in the same way for all three tasks. First the children perform the task individually. Their performance is measured and they are then matched together in pairs of two depending on their result. Thereafter the children perform the task a second time in these matched pairs. Competitiveness is measured as the difference in performance between the individual and matched performance, and is thus considered as the reaction to competition. We have a control group of children who perform the task alone a second time, as in Gneezy and Rustichini (2004a). This allows us to control for unobservable factors that could cause differences in the outcome, such as, e.g., one gender getting tired faster than the other.

Given previous literature, we hypothesize that if there is a gender gap in running, boys will compete more than girls. We also hypothesize that if there is a gender gap in the female tasks it will be the opposite since, if anything, these tasks have positive stereotypes regarding female ability.

We find no evidence in support of our hypotheses. We find no gender differences in competitiveness among children in Sweden in any of the three tasks. Boys and girls increase their performance equally in the competitive setting for running and skipping rope, and there is no difference between the average increases. Regarding the dancing task, both boys and girls

[2] Wozniak et al. (2010) find no gender gap in performance change but find that men are more likely to self-select into competitions.

decrease their performance when competing, and this decrease in performance is not significantly different between the two genders. Our results also indicate that the gender of the opponent does not alter performance of either gender in any of the three tasks. Moreover, the findings from the control group indicate that our results are not driven by gender differences in factors such as tiredness.

This contradiction to earlier results by Gneezy and Rustichini (2004a) may be explained by context, such as culture. It has previously been shown that cultural factors such as gender norms may influence competitive behavior. Gneezy et al. (2009) compare a matrilineal society in India with a patriarchal society in Tanzania and find that women prefer the competitive setting more than men in the matrilineal society, whereas the inverse is found in the patriarchal society.[3] Our results suggest that cultural factors matter also among Western countries. Even though we cannot directly test this, we speculate that the difference between our results and those of Gneezy and Rustichini (2004a) may be due to differences in gender norms. Even though Sweden and Israel are both Western societies with high female labor force participation, Sweden usually performs higher on gender equality indices.[4]

Our paper is organized as follows. In section 2, we describe the experimental design of our field study. In section 3, we present our results. We conclude in section 4, where we also discuss the possible explanations for our findings as well as promising directions for future research.

2.2 Experimental design

The field experiment was conducted in 11 primary school classes in the Stockholm area during 2008 and 2009. We contacted all primary schools in

---

[3] The task at hand is the toss of a tennis ball into a bucket. Gneezy et al. (2009) are unaware of any resemblance between this task and some popular task in the cultures that are being studied, thus it is unlikely that the specific task had a certain gendered stereotype. In general, however, throwing objects could be considered more male in many cultures since men have typically been the hunters (e.g., men hunt through spear throwing).

[4] The Global Gender Gap Report 2009 lists Sweden as number four in the world in terms of gender equality. Israel ranks 45th out of 134 countries.

Stockholm with a letter explaining that we intended to study competitiveness among children. There was no mentioning of the gender dimension. All tasks were performed during physical education classes and the experiment was overseen by the teacher. The children, aged 7-10 years old, did not realize that they were participating in an experiment (as in Gneezy and Rustichini 2004a). The teachers did not mention the study to the children, and the tasks are standard in Swedish physical education classes. On two or three different occasions, the children competed in running, skipping rope and modern dance. These three tasks were carefully chosen. Running has previously been explored in Gneezy and Rustichini (2004a) and is part of physical education in Sweden. Skipping rope is a task that girls perform during school breaks throughout the world, including Sweden. Dancing is often considered female (Henschel-Pellet 2001), and during the Swedish school year it typically takes up one physical education class. The running task was administered by the teachers on a separate occasion (as in Gneezy and Rustichini 2004a), whereas the skipping rope was instructed and administered by the experimenters as an exercise complementary to the dancing, thus the experimenters were present at the occasion the dance competition. The dancing task was designed, instructed and scored by a professional dance teacher on one or two occasions depending on the length of the class. To avoid that teachers treated boys and girls differently in order to affect the results of the study, all teachers, including the dance teacher, were unaware of the gender dimension of the study. The children were given 40 minutes to practice the dancing task together with the whole class and the dance teacher, and 5 minutes to practice the skipping rope task prior to the start of the experiment.

In running, performance is measured by how fast the children ran 60 meters, the distance normally used for short distance running in Swedish schools. Note that this distance differs from what Gneezy and Rustichini (2004a) used, 40 meters. In skipping rope (where two individuals turn the rope while one child jumps), performance was measured as the number of jumps performed until the children missed. In dancing, the dance teacher scored the children based on how they performed compared to the set goal of the dance choreography. The dance choreography included ten distinct exercises and the children were awarded one point for each of these ten movements that they performed correctly.[5]

Each task consisted of two stages. At the first stage, the children performed the task by themselves and individual performance was measured. The teachers were aware of the setup of the study, whereas the children were unaware of the existence of a second stage when performing the task in the first stage in all three tasks.[6] At the second stage, the children performed the task in competition with another child. Matching started with the two children that had the best performance in the first stage in each task, and then continued down the list. If more than two children obtained the same result in the first stage, matching was done randomly (as in Gneezy and Rustichini 2004a).[7] In the case of dancing, both the individual performance and the competition occurred in a separate room where only the one or two children dancing and dance teacher were present. In all three tasks, the children knew that their competitor had achieved a similar score at the first stage. The dance teacher presented the tasks as competitive activities. The dance competition was presented as a "battle", somewhat in the spirit of a popular TV show.[8] In the skipping rope task, two ropes were put

[5] The dancing task consisted of a one minute long modern dance phrase. The choreography of the phrase was focusing on strength, coordination and balance rather than "feminine grace", in order to minimize subjectivity in the evaluation of dance. Since the dance teacher was not aware of the purpose of the study, any potential subjectivity is likely to be orthogonal to the gender of the child evaluated. The children were aware of how the task was scored.

[6] The teachers were aware of the two stages of each task, but did not inform the children about this. The experimenters gave oral instructions to the children about the setup of the study at the relevant stages.

[7] When an unequal number of children performed equally well, they were randomly paired. The remaining child was matched with the child with the next best result. If more than one child had the next best result, the remaining child with the higher score from the first matching was randomly matched with one of these children. During the competitive part of the experiment, the competing pairs participated in random order.

[8] The TV show "So you think you can dance" was aired on Swedish television before and during the time the study was performed. It has been pointed out to us that dancing is often

next to each other. The children were instructed to start jumping at the same time and were told that the winner was the child who performed the greatest number of jumps. All rules were explained by the dance teacher and the experimenters and no compensation was awarded apart from the intrinsic motivation that comes from winning, as in Gneezy and Rustichini (2004a). Our measure of competitiveness is the change in performance between the first and the second stage of the tasks.

2.3 Results

We test whether there is a gender gap in competitiveness among children in Sweden and whether the nature of the task affects the size and direction of the gender gap. We start by looking at gender differences in competitive behavior. Thereafter we address the effect of the gender composition in the competitive setting. We also present a robustness check and a survey on how boyish/girlish children perceive the explored tasks to be. For all tests in the analysis, we have performed a Mann-Whitney test, a two-sided t-test and used bootstrap techniques. Throughout the analysis we present only the p-value for the Mann-Whitney test.[9]

2.3.1 No significant gender differences in competitive behavior

In our study, 149 children participated in running, 143 in skipping rope, and 146 in dancing. The gender distribution in the three sports was 68 boys and 81 girls in running, 67 boys and 76 girls in skipping rope and 64 boys and 82 girls in dancing.[10] Consistent with sex-stereotypic expectations, we find

---

a cooperative or communal activity. We assume that the competitive element of the TV show decreased the cooperative or communal aspects of the dancing task.

[9] We present the Mann-Whitney test since none of our variables are normally distributed when using a skewness and kurtosis test. When there is a difference between the parametric and non parametric tests in terms of significance we also report the p-values for the t-test and the bootstrap-based critical values. We have also compared whether the distributions for each reported variable differ between men and women using a Kolmogorov-Smirnov test. The results are the same as those reported for mean values.

[10] Two subjects, one boy and one girl, were dropped from the sample due to physical disabilities. The differences in number of children between activities are due to the fact that we had different number of occasions depending on the structure of the physical education

that in the individual setting (stage 1) boys ran on average faster than girls (unlike in Gneezy and Rustichini 2004a), and girls skipped rope better compared to boys. In running and skipping rope, the p-value for a significant gender difference is 0.008, with boys performing better in the former ask and girls in the latter. In dancing, the non parametric test gives a p-value of 0.0478, whereas the difference is not significant with a t-test or a bootstratpped test.[11] When it comes to competitiveness, table 1 below shows that in all three tasks, and for both genders, average performance in the competitive setting differs significantly from average performance in the non-competitive setting, (p<0.01). Both genders improve their performance significantly in running and skipping rope in the competitive setting, but perform worse in dancing.[12]

Table 1: Average performance in stage 1 and in stage 2. Signrank (SR) test p-values of performance change for girls and boys separately.

	Running		**SR**	**Skipping**		**SR**	**Dancing**		**SR**
	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value
Girls	11.92	11.66	0.000	49.01	69.37	0.000	5.87	5.13	0.001
Boys	11.55	11.42	0.002	32.48	45.12	0.000	5.27	4.48	0.001

Figures 1-3 below show the distribution of the performance change in the different tasks. The three histograms show that there are no significant

classes in the different schools. There is no significant difference in performance change between school classes that had one occasion or school classes that had more occasions (ranksum: p=0.53).

[11] When we perform the tests on the inner quartile range (IQR, the distribution between the 25th and 75th percentile) the Mann-Whitney test is also insignificant.

[12] The other tests are not significant when it comes to the performance change of boys in running. However, when performing the tests on the IQR, all three tests are significant.

gender differences in any of the three tasks (running: p=0.47, skipping rope: p=0.24, dancing: p=0.85).[13]

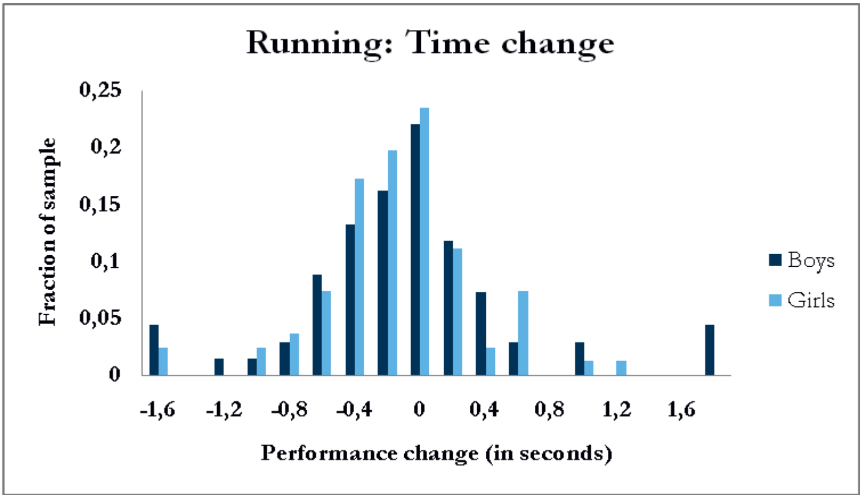

Figure 1. Distribution of change in running time (stage 2 - stage 1), by gender.

[13] To further investigate a possible gender difference in performance change we also performed quantile regressions for each task, controlling for gender of opponent (performed for quantile 0.1-1.0). Gender has an effect only in the top 10% of the performance change distribution in running and skipping rope. In this part of the distribution the performance change of boys is larger than girls in running and the opposite for skipping rope. There are however very few observations in the top 10% for each task. These results are therefore mere indications.

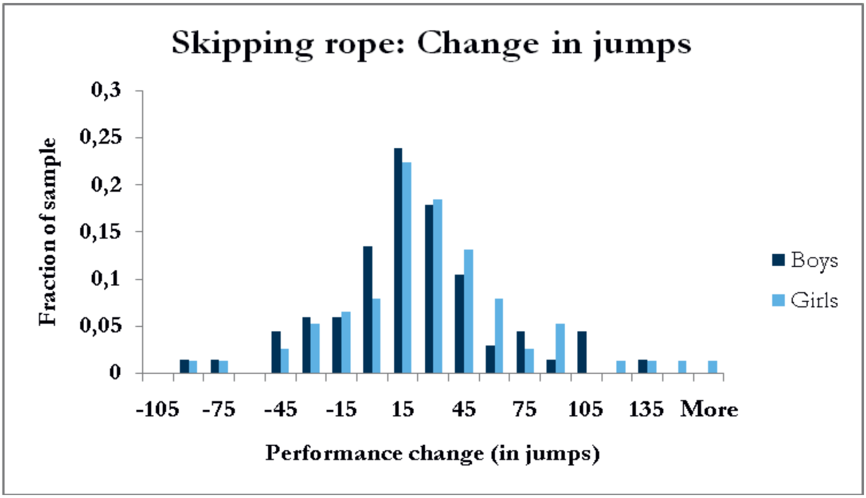

Figure 2. Distribution of change in jumps (stage 2 – stage 1), by gender.

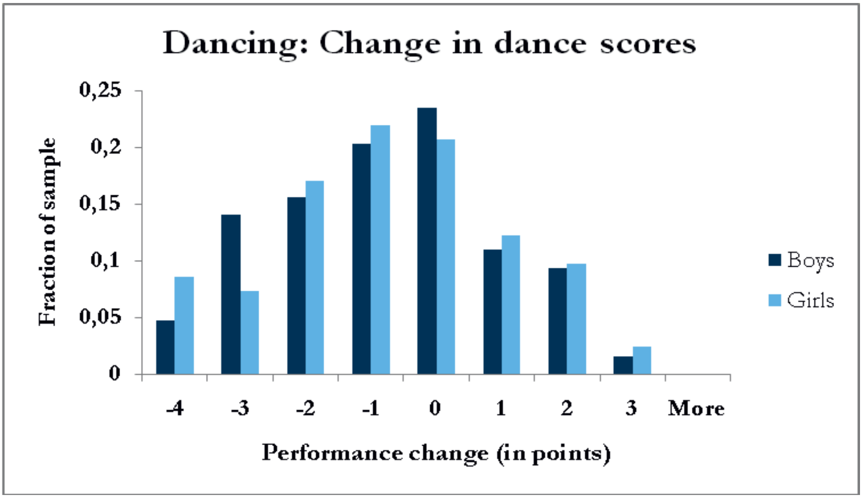

Figure 3. Distribution of change in dance scores (stage 2 – stage 1), by gender.

The pattern of gender similarities are displayed in an aggregated manner in figures 4-6 below. These plots show the average change in performance by each gender. In running, girls improve on average 0.26 seconds, or about 2.1%. This can be compared to the average decrease in running time of

0.13 seconds, or 1.1%, for boys.[14] The corresponding numbers for skipping rope is an increase of 20 versus 13 jumps, implying an improvement of 42% and 39% respectively. On average, girls' dance performance deteriorates by 0.73 points (13%) on average and boys' by 0.78 points (15%). As stated above, the difference in average change in performance between boys and girls is not statistically significant in any of the three cases.[15] These results also hold within all age groups in our sample.[16]

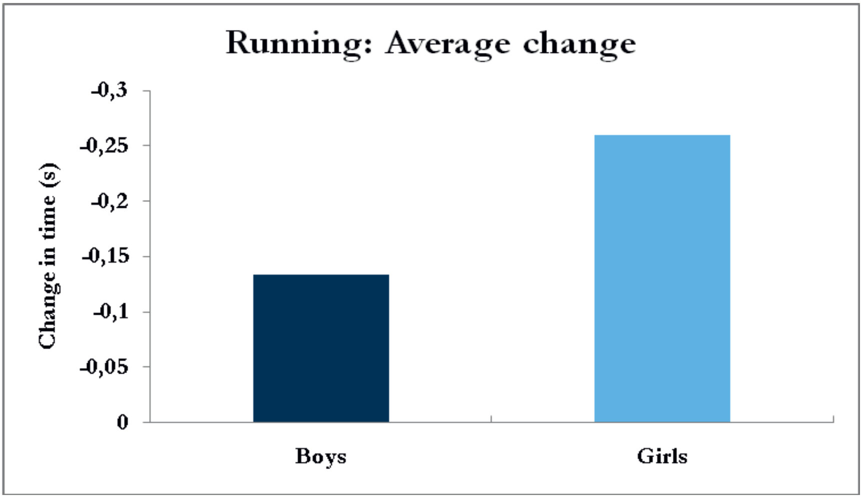

Figure 4. Average change in time (stage 2 – stage 1), by gender. 78 girls and 71 boys.

---

[14] For all three tasks, we conducted the same analysis with relative performance, where relative performance was defined as ((stage2-stage1)/stage1). This did not change any of our results. Our findings further remain stable when excluding outliers. An outlier is defined as an observation that lie more than two standard deviations away from the sample mean.

[15] A sample size analysis indicates that 1411, 965 and 38407 observations would be needed to obtain a significant result for the performance change in running, jumping and dancing respectively. The basis for the power calculation is a significance level of 5% and a power of 80%.

[16] In particular, when we restrict the running analysis to the same age group as studied in Gneezy and Rustichini (2004a), the gender gap among these 114 children aged 9-10 years old is still insignificant (p=0.47).

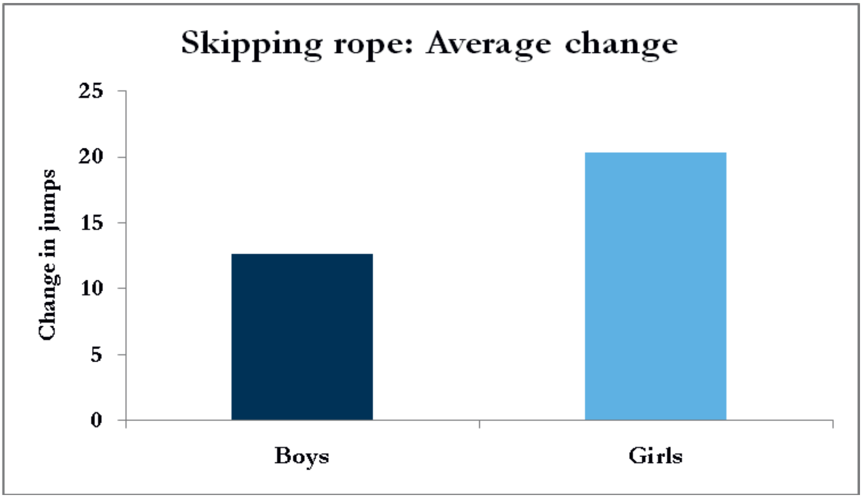

Figure 5: Average change in jumps (stage 2 – stage 1), by gender. 74 girls and 69 boys.

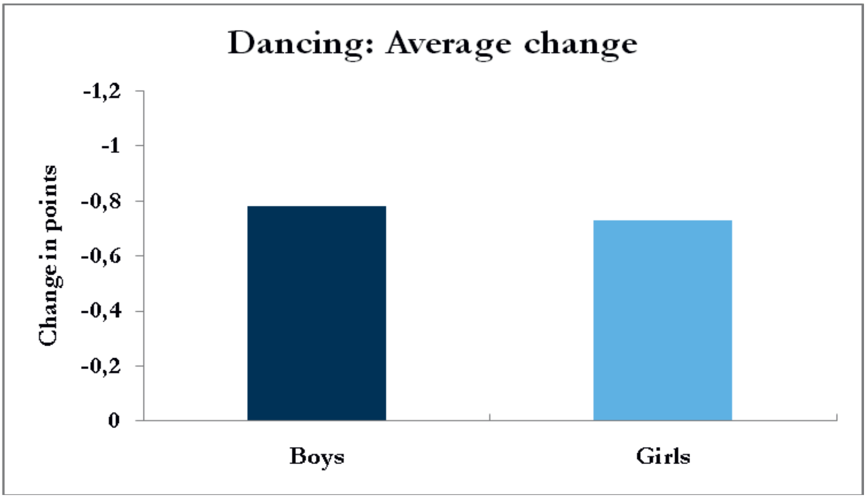

Figure 6. Average change in dance scores (stage 2 – stage 1), by gender. 82 girls and 64 boys.

2.3.2 Impact of opponent gender on competitive behavior

Some previous studies find that women compete more against women, and men more against men (e.g., Datta Gupta et al. 2005, Gneezy et al. 2003). On the contrary, Gneezy and Rustichini (2004a) find that boys are not affected by the gender composition but girls compete more against boys. Our results suggest that neither boys nor girls are influenced by the gender of their opponent. Table 2 gives an overall summary of our results for the different pair compositions in our study. In running, both girls and boys improve the most when running against a girl. However, the difference in competitive behavior when facing the same vs facing the opposite gender is statistically insignificant for girls (p=0.6221) and for boys (p=0.0701). In skipping rope and dancing, girls compete more fiercely against boys, but none of these results are significant (skipping rope: p=0.1864, dancing: p=0.4982). Boys on the other hand compete more against boys in skipping rope and more against girls in dancing, though also these differences are not significant (skipping rope: p=0.8401, dancing: p=0.4519).

Table 2. Performance change (stage 2 – stage 1) based on the gender composition of the competing pairs.

	Running			Skipping			Dancing		
Sample	n	Stage2 -stage1	p-value	n	Stage2 -stage1	p-value	n	Stage2 -stage1	p-value
Total	149	-0.20	0.000	143	17	0.000	146	-0.75	0.000
Girls with girls	47	-0.28	0.001	40	14	0.026	41	-0.83	0.002
Boys with boys	42	-0.13	0.175	30	15	0.014	27	-0.96	0.005
Girls mixed pairs	34	-0.24	0.001	36	27	0.001	41	-0.63	0.079
Boys mixed pairs	26	-0.14	0.001	37	10	0.127	37	-0.65	0.054

2.3.3 Robustness checks

We also let a separate group of children perform the task alone in the second stage, serving as a control group. We thereby control for unobservable factors that could cause differences in the outcome, such as one gender getting tired faster than the other. The control group includes 66 children in the running task (31 boys and 35 girls), 65 children in the skipping rope task (29 boys and 36 girls), and 49 children in the dancing task (19 boys and 30 girls). For running, both boys and girls perform worse in stage 2 compared to stage 1 ($p<0.001$). Importantly, however, there is no significant gender difference when we test performance change between boys and girls ($p=0.4878$). The fact that stage 2 performance in running is worse than stage 1 performance indicates an even greater reaction to competition in running for both boys and girls than if there would have been no performance change in the control. The absolute performance change between stage 2 and stage 1 in skipping rope and dancing is not significant (skipping rope: $p=0.1627$, dancing: $p=0.3206$). This indicates that when not competing against another child there is no significant improvement in performance in these two tasks. Moreover, there are no significant differences in these two tasks when we test performance change between boys and girls (skipping rope: $p=0.9106$, dancing: $p=0.9664$). See table 3 for more details on the results.

Table 3. Performance change (stage 2 – stage 1) in the control, and whether there is a gender difference in this performance change.

Control		**Running**			**Skipping**			**Dancing**	
Sample	n	Stage2-stage1	p-value	n	Stage2-stage1	p-value	n	Stage2-stage1	p-value
Total	66	0.35	0.001	65	6.77	0.163	49	-0.35	0.321
Gender difference	66	-0.20	0.488	65	-3.69	0.911	49	0.22	0.966

Even though we find no significant gender differences in mean change in performance in our main analysis, there may be differences in the variances of the performance distributions. We test this and find no significant differences in the variance of change in performance between boys and girls.[17]

Furthermore, we also perform a within subject analysis across tasks. We balance the sample by keeping only individuals that performed all three tasks (58 girls and 45 boys). We find no correlations between performance change in the different tasks for boys or girls (running and skipping rope: boys: p=0.5058, girls: p=0.3617; running and dancing: boys: p=0.4389, girls: p=0.9088; skipping rope and dancing: boys p=0.2710, girls: p=0.1089).[18] This suggests that in our sample there does not seem to be a general competitive type – some individuals perform better under competition in one task and not another.

2.3.4 Do children perceive the tasks to be gendered?

In a separate survey of children aged 9-10 years old, we asked how boyish/girlish they considered running, skipping rope and dancing to be. We also elicited perceptions of how boyish/girlish competing in these tasks was. The children were asked to use a scale where a lower number indicates rating the task as more boyish and a higher number as more girlish (1=very boyish, 2=boyish, 3=neutral, 4=girlish, 5=very girlish).

Table 4 shows that, on average, running is perceived to be more boyish than skipping rope and dancing. This is the case both in absolute and relative terms.

---

[17] The most common test for comparison of standard deviations, the F-test for the homogeneity of variances (sdtest), is very sensitive to the assumption that the data are drawn from an underlying normal distribution. Therefore we also performed a robust test (Levene's test with mean, median and 10% trimmed mean). None of these tests indicated significant differences in the variances.

[18] Performing this analysis on relative performance change does not alter the results qualitatively.

Table 4. Summary statistics of ratings.

Variable	Obs	Mean	Std Dev	Min	Max
Running	34	2.68	0.73	1	4
Skipping rope	35	4.17	0.79	3	5
Dancing	34	4.03	0.83	2	5
Competition running	35	2.29	0.83	1	4
Competition skipping rope	35	3.77	0.94	2	5
Competition dancing	35	4.03	0.82	3	5

Running is perceived as significantly more boyish than skipping rope ($p<0.001$) and dancing ($p<0.001$).[19] When comparing skipping rope and dancing there is no significant difference ($p=0.5432$). When it comes to the perceptions of how boyish/girlish it is to compete in these tasks, we observe the same pattern. Competing in running is rated as more boyish than competing in skipping rope and dancing

We also compare the rating of competing in a certain task with the general rating of the task. Competition in itself is rated as more boyish compared to the general rating for both running and skipping rope ($p=0.0315$ and $p=0.0211$), but not for dancing. For dancing there is no significant difference between competition and the general rating of the task ($p=1$). When merging these data, competition seems to be rated more boyish compared to the rating of the task in general ($p=0.0050$).

[19] Most of these variables are not normally distributed according to a skewness and kurtosis test. Thus, we perform a Mann-Whitney test for differences in distributions between the tasks.

2.3.4.1 Do boys and girls have different perceptions?

In table 5 we divide the ratings by gender. Girls tend to rate running as gender neutral and boys as more boyish (p=0.0021). Moreover, girls tend to rate dancing as more neutral, whereas boys rate it as more girlish (p=0.0430). Boys and girls give skipping rope a similar score. Regarding competition, there is no significant difference in the ratings for any of the tasks.

Table 5. Average ratings by gender.

	Running	Skipping	Dancing	Comp. running	Comp. skipping	Comp. dancing
Girls	3.06	4	3.81	2.53	3.65	3.88
Boys	2.31	4.35	4.35	2.06	3.82	4.18
Total	2.70	4.18	4.09	2.29	3.74	4.03

When merging the data on the three tasks, girls and boys rate competition in the same way in terms of how boyish/girlish it is (p=0.6993).[20]

2.4 Discussion

Previous literature on competitive behavior finds that men compete to a larger extent than women. This difference in behavior may explain part of the gender gap observed in many areas in society. In this literature, however, only a few tasks have been used to measure competitiveness, and these tasks can arguably be considered as more male than female. Three studies find that gender differences in competitiveness vary with the task at hand (Gneezy and Rustichini 2004b, Günther et al. 2009, Grosse and Reiner 2010), whereas another study find no difference in the gender gap between a maze task and a word task (Wozniak et al. 2010). Meanwhile, work in social psychology suggests that individual perceptions about rela-

[20] When we control for age in a tobit regression (upper limit 5 and lower limit 1), there is a gender difference in rating only for running, and age does not have a significant effect. It should be noted that the variation in age is very small. When controlling for age, boys and girls do not have different opinions concerning the rating of competition. It should be noted that the sample size is rather small.

tive performance, such as (over)confidence, and especially stereotypes may have important implications for actual performance (Steele 1997, Shih et al. 1999). Exploring more tasks than maze solving and simple arithmetic is thus important in order to increase our understanding about gender differences in competitiveness and the potential role of stereotypes.

In this paper we study how children compete in three distinct tasks. We let the children compete in running in order to create a comparison with previous literature. Moreover, we add two more female tasks to the competition; skipping rope and dancing. Competitiveness is measured by reaction to competition, i.e. as the child's increase in performance when competing against another child, compared to when the task is performed individually. We find no gender differences in competitive behavior in any of these tasks. Boys respond to competition, and so do girls. Contrary to previous literature (e.g., Datta Gupta et al. 2005, Gneezy et al. 2003, Gneezy and Rustichini 2004a) we also find that the gender of the opponent affects neither boys nor girls in any of the three tasks. The three performance measures we use here differ due to the difference in nature of the three tasks. This makes direct comparisons across tasks somewhat difficult, and we do find that there is actually an average decrease in performance when the children compete in dancing compared to the individual performance, unlike in running and skipping rope. However, in each of the three tasks we find no gender difference in performance change, and this is our main result.

One possible explanation to the difference between our running result and that of Gneezy and Rustichini (2004a) is culture. It has previously been shown that culture affects important economic decisions such as labor market participation and fertility (e.g., Fernández and Fogli 2006), and the institutional setting has been found to influence competitive behavior (e.g., Balafoutas and Sutter 2010, Gneezy et al. 2009, Niederle and Yestrumskas 2008, Cotton et al. 2009, Niederle et al. 2009, Wozniak et al. 2010). For example, the gender gap in self-selection has been shown to disappear with performance feedback (Wozniak et al. 2010) and the difference in performance change vanishes with repetition of the competition (Cotton et al. 2009). Women have also been found to compete more than men in a matrilineal society whereas men compete more than women in a patriarchal society (Gneezy et al. 2009). Even though our study only includes children in

Sweden, we can compare our running results to those of Gneezy and Rustichini (2004a).[21] Where we find no gender gap, Gneezy and Rustichini (2004a) instead find that among Israeli children only boys respond to competition in a running task. The specific mechanisms behind the different results in Sweden and Israel are unclear. It is possible that the more gender neutral culture in Sweden decreases the difference in competitive behavior between boys and girls in general, but also that it diminishes the degree to which tasks are gendered. If this is the case, this could explain why boys and girls compete equally in all tasks in our study.[22]

The results of two recent studies complicates this reasoning somewhat. Since we performed the study presented in this paper, there have been two other relevant studies. Sutter and Rützler (2010) look at willingness to compete among children aged 3 to 18 years old. Younger children are given the choice whether to compete or not in running 30 meters, whereas older children get the same choice for a math task. The authors find that boys are more competitive than girls in all age groups. In an even more recent study, Cárdenas et al. (2010) explore the gender gap in competitiveness and risk taking among 9-12 year olds in Colombia and Sweden. Boys and girls are equally competitive in all tasks and all measures in Colombia (including running), whereas the results in Sweden are mixed, with some indication of girls being more competitive than boys in skipping rope and math in terms of performance change, whereas boys are more likely to choose to compete

[21] Even though the two studies differ somewhat in their design. In our experiment, we look at three different tasks, not only running. Moreover, the children compete in running 60 meters, which differs from the 40 meters used in Gneezy and Rustichini (2004a). However, the setups are similar in many aspects: both setups explore competitiveness as the performance change when running against someone versus running alone, the children were not aware of participating in an experiment, the teachers administered the running task, the matching procedure of the competing pairs was the same, and there was only intrinsic motivation for winning. We also included a control group, as in Gneezy and Rustichini (2004a).

[22] Children in Sweden do not receive grades until year 8 (age 14) thus a higher motivation for both boys and girls to perform well due to grade concerns is not a plausible explanation to why boys and girls compete equally.

in general.[23] Cárdenas et al. (2010) also find that boys in both countries are more risk taking than girls, with a smaller gender gap in Sweden.

The absence of a gender gap in performance change in running in Austria and Colombia is surprising given the results in Israel. Both of these countries typically score as Israel on gender equality indices.[24] However, there are differences between the setup in Gneezy and Rustichini (2004a) and those in Cárdenas et al. (2010) and Sutter and Rützler (2010). An interesting avenue for future research would be to identify the specific components in explaining differences in the gender gap in competitiveness across a large number of countries using the exact same measures.

Making inferences about adult behavior from findings on children is not straightforward. Even though we do not find a gender gap among children in Sweden, it may be that male and female behavior change differently over time. Observing gender diversity in behavior among adults does not tell us the underlying reasons for these gender differences. For example, if a gender gap in behavior occurs during the teenage years, this could be caused by socialization or by the hormone surge that puberty brings along. More cross-cultural research and work on biological variables should also be of great interest. Thus far, studies looking at the importance of sex hormones to explain individual differences in competitiveness get mixed and inconclusive results. A study looking at competitiveness among men finds no relationship between self-selection into a tournament and current testosterone levels (Apicella et al. 2010). Buser (2009) finds that women are less likely to self-select into a tournament when progesterone and estrogen levels are high whereas Wozniak et al. (2010) find the opposite with women in the low-hormone phase being less competitive.[25] Meanwhile, Zethraeus et al.

---

[23] The difference between our results and those of Cárdenas et al. (2010) in skipping rope is perhaps due to the larger sample size in the latter study (520 children).

[24] The Global Gender Gap Report 2009 ranks Austria as 42nd and Colombia as 56th out of 133 countries on gender equality.

[25] Apicella et al. (2010) find that neither circulating testosterone, facial masculinity (considered a proxy of hormone exposure during puberty), nor digit ratios (considered a proxy of prenatal hormone exposure) correlate significantly with competitiveness in a sample of 98 young men. Moreover, Buser (2009) finds no effect of the cycle on competitiveness as measured by reaction to competition or risk preferences. This latter result contradict two

(2009) find that exogenously providing estrogen or testosterone to women does not affect their economic preferences, though the authors do not look at competitiveness specifically.[26] More work is thus needed to disentangle the importance of sex hormones in explaining gender differences in competitiveness and other economic preferences.

Our findings open up interesting directions for further research. If competitive behavior among boys and girls is cultural and/or task dependent, competitive behavior should be studied in a variety of tasks and cultural settings. Since we find no gender differences among children in Sweden, it would also be of great interest to see if there is a gender gap in competitiveness among Swedish adults, and if so at what age this first occurs. It would also be interesting to in future studies collect information about the cultural background of the participants in experiments, to explore cultural variation in that sense too. Moreover, we do not use any extrinsic incentives in this study. An interesting extension would be to test the robustness of our results to extrinsic rewards such as money or e.g. pens.[27] Once we have answers to these questions it will be possible to make more general claims about gender and competitiveness, and possibly how and if this relates to labor market outcomes.

---

studies that in turn also get opposing results when looking at competitive bidding/risk preferences. The first study finds that men and women who are menstruating (thus have low estrogen levels) act similarly (Chen et al. 2005), whereas a follow-up study finds that women menstruating or in the premenstrual part of the cycle act significantly different from men (Pearson and Schipper 2009).

[26] However, it could be the case that it is the long-term organizational effects of hormonal exposure that matter and not the effects from short-term exposure.

[27] Sutter and Rützler (2010) reward the children extrinsically when competing in running in Austria. Since they find no gender difference in performance change in running with this type of reward, this suggests that the lack of extrinsic reward is not necessarily what drives our results.

2.5 References

Apicella C.L., Dreber A., Gray P.B., Hoffman M., Little A .C., Campbell B.C., 2010, "Androgens and competitiveness in men", Forthcoming in *Journal of Neuroscience*, Psychology and Economics.

Balafoutas L., Sutter M., 2010, "Gender, competition and the efficiency of policy intervention", Working Papers in Economics 450, Göteborg University, Department of Economics.

Booth A., Nolen P., 2009, "Choosing to compete: the role of single-sex education", Mimeo.

Buser T., 2009. "The impact of female sex hormones on competitiveness", Mimeo.

Cárdenas J.C., Dreber A., von Essen E., Ranehill E., 2010, "Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden", Forthcoming in *Journal of Economic Behavior and Organization*.

Ceci S., Williams W., 2006, *Why aren't more women in science*, American Psychological Association.

Chen Y., Katuscak P., Ozdenoren, E., 2005, "Why can't a woman bid more like a man?", Mimeo.

Cotton C., McIntyre F., Price J., 2009, "Gender differences disappear with exposure to competition", Working Paper 2010-11, University of Miami, Department of Economics.

Datta Gupta N., Poulsen A., Villeval M.-C., 2005, "Male and female competitive behavior – experimental evidence", IZA Discussion Paper, No. 1833.

Fernández R., Fogli A., 2006, "Fertility: the role of culture and family experience", *Journal of the European Economic Association*, 4, 552-561.

Gneezy U., Rustichini A., 2004a, “Gender and competition at young age”, *American Economic Review Papers and Proceedings*, 118, 1049-1074.

Gneezy U., Rustichini A., 2004b, “Executives versus teachers”, Mimeo.

Gneezy U., Leonard K.L., List J.A., 2009, “Gender differences in competition: evidence from a matrilineal and a patriarchal society”, *Econometrica*, 77, 909-931.

Gneezy U., Niederle M., Rustichini A., 2003, “Performance in competitive environments: gender differences”, *Quarterly Journal of Economics*, 118, 1049 – 1074.

Günther C., Ekinici N.A., Schwieren C., Strobel M., 2009, “Women can’t jump? – An experiment on competitive attitudes and stereotype threat”, Working Paper, University of Graz.

Henschel-Pellet H.A., 2001, “Physical activity gender role stereotyping: an analysis of Children’s Literature”, *Research Quarterly*, 72.

Inzlicht M., Ben-Zeev T., 2003, ”Do high-achieving female students underperform in private? The implications of threatening environments on intellectual processing”, *Journal of Educational Psychology*, 95, 796-805.

Niederle M., Vesterlund L., 2007, “Do women shy away from competition? Do men compete too much?”, *Quarterly Journal of Economics*, 122, 1067-1101.

Niederle M., Segal C., Vesterlund L., 2009, “How costly is diversity? Affirmative action in light of gender differences in competitiveness”, Mimeo.

Niederle M., Yestrumskas A.H., 2008, “Gender differences in seeking challenges: The role of institutions”, Mimeo.

O'Brien L.T., Crandall C.S., 2003, “Stereotype threat and arousal: effects on women's math performance”, *Personality and Social Psychology Bulletin*, 29, 782-789.

Pearson M., Schipper B.C., 2009, "Menstrual cycle and competitive bidding", MPRA Paper.

Price J., 2008, "Gender difference in the response to competition", *Industrial and Labor Relations Review*, 61, 320-333.

Shih M., Pittinsky T.L., Ambady N., 1999, "Stereotype susceptibility: shifts in quantitative performance from sociocultural identification", *Psychological Science*, 10, 81-84.

Steele C.M., 1997, "A threat in the air. How stereotypes shape intellectual identity and performance", *American Psychologist*, 52, 613–629.

Sutter M., Rützler D., 2010, "Gender differences in competition emerge early in life", IZA Discussion Paper 5015.

Wozniak D., Harbaugh W., Mayr U., 2010, "Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback", MPRA Paper.

Zethraeus N., Kocoska-Maras L., Ellingsen T., von Schoultz B., Lindén Hirschberg A., Johannesson M., 2009, "A randomized trial of the effect of estrogen and testosterone on economic behavior", *Proceedings of the National Academy of Sciences of the United States of America*, 106, 6535-6538.

3 Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden

Juan-Camilo Cárdenas, Anna Dreber, Emma von Essen and Eva Ranehill

The authors are grateful for comments from Johan Almenberg, Thomas Buser, Alison Cool, Uri Gneezy, Bart Golsteyn, Moshe Hoffman, Magnus Johannesson, Astri Muren, David G. Rand and seminar participants at ESA Copenhagen 2010, Gruter's Squaw Valley Retreat on Law, Institutions and Human Behavior 2010, Harvard Kennedy School, MOVE Workshop on Gender Differences in Competitiveness and Risk Taking, Stockholm School of Economics, and Stockholm University. Anna Dreber is grateful for support from Iris Bohnet and the Women and Public Policy Program at Harvard Kennedy School. The field work in Colombia could not have been completed without the help of Adriana Molina and Gloria Rodriguez. Financial support from the Swedish Institute of Banking Research (Bankforskningsinsitutet) (A.D.), the Jan Wallander and Tom Hedelius Foundation and the Carl Silfvén Foundation (E.R.), the Swedish Council for Working Life and Social Research (FAS), and the Department of Economics at the Universidad de Los Andes is gratefully acknowledged.

ABSTRACT: We explore gender differences in preferences for competition and risk among children aged 9-12 in Colombia and Sweden, two countries differing in gender equality according to macro indices. We include four types of tasks that vary in gender stereotyping when looking at competitiveness: running, skipping rope, math and word search. We find that boys and girls are equally competitive in all tasks and all measures in Colombia. Unlike the consistent results in Colombia, the results in Sweden are mixed, with some indication of girls being more competitive than boys in some tasks in terms of performance change, whereas boys are more likely to choose to compete in general. Boys in both countries are more risk taking than girls, with a smaller gender gap in Sweden.

3.1 Introduction

Men typically occupy the majority of top positions in most sectors in most societies, whereas women in many western countries are at least as likely as men to pursue higher education and to participate in the labor market. One possible and suggested cause of gender differences in labor market outcomes is that men and women differ in terms of economic preferences. In particular, preferences for competition and risk, where women in general are found to be less competitive and less risk taking than men (see, e.g., Croson and Gneezy 2009 for an overview), might contribute to explaining the labor market gender gap. Competitiveness is typically measured as either the performance response to a competitive setting compared to a non-competitive setting, or as a preference for competition such as self-selecting into a competitive setting instead of a non-competitive setting. However, relatively little is known about how the gender gap in economic preferences varies with age, and to what extent cross-country differences in gender norms affect the gender gap. Studying children from different countries is one potential route to further this understanding.

In this paper we explore the gender gap in preferences for competition and risk among approximately 1200 children aged 9-12 in the two capitals Bogotá and Stockholm. Colombia and Sweden are two countries that differ in gender equality according to various macro-economic indices (e.g., Hausmann et al. 2010).[28] Our setup enables us to study to what extent there are systematic differences in the gender gap between Colombia and Sweden. We explore gender differences in competitiveness using four tasks: running, skipping rope, math and word search. These four tasks allow for the possibility that differences in gender stereotyping of the tasks influence the gender gap in competitiveness, i.e. there might be female and male areas of competition. We study competitiveness as the performance change between an individual setting and a forced competition in all four tasks, as well as the choice of whether to compete or not in math and word search. We also explore the gender gap in risk preferences by having the children choose between different incentivized gambles (using a measure adapted from Holt and Laury 2002).

There is some previous work on competitiveness and risk taking among children. In a field experiment on 9-10 year old children in Israel, Gneezy and Rustichini (2004a) find that boys react to competition by running faster against another child compared to an individual race, whereas girls do not change their performance. Contradictory to this finding, Dreber et al. (2009) find that 7-10 year old boys and girls in Sweden compete equally in running as well as in skipping rope and dancing.[29] Moreover, Booth and Nolen (2009a) explore how the gender gap in choosing to compete among 15 year old adolescents in the UK depends on whether they go to a single sex or mixed school. Girls in single sex schools, on the other hand, are more competitive than girls from mixed schools. Boys are found to be equally competitive in both types of schools, as well as more competitive than girls in both schools.

In parallel with our study, two other studies concerning gender differences in competitiveness among children have been conducted. Looking at running, Sutter and Rützler (2010) find that among 3-8 year old children in

[28] In this report, Colombia ranks 55th and Sweden 4th in terms of gender equality according to this index. As far as we know, there are no studies comparing adult behavior in competitiveness and risk taking in Colombia and Sweden.

[29] Dreber et al. (2009) find no impact of age on behavior. There are furthermore some differences between the setup of Gneezy and Rustichini (2004a) and that of Dreber et al. (2009).

Austria, boys are more likely than girls to choose to compete. Sutter and Rützler also look at 9-18 year old children competing in math and find similar results to those on younger children, i.e. no gender difference in performance change but boys are more likely to choose to compete than girls. Moreover, Andersen et al. (2010) compare competitiveness, measured as the choice to compete when throwing tennis balls, among children aged 7-15 in a matrilineal society (the Khasi) and a patriarchal society (the Kharbi) in India.[30] They find no significant gender difference in competitiveness in the matrilineal society, whereas in the patriarchal society a gender gap emerges in the age group 13-15, with boys being more competitive.

The type of competition task has also been shown to sometimes matter. Most of the literature focuses on math or maze tasks, tasks that are typically considered male, with a few exceptions.[31] Two studies comparing the gender gap in competitiveness between a maze task and a word task find that the gender gap is influenced by the task (Günther et al. 2009, Grosse and Riener 2010) whereas another study finds no difference between these tasks (Wozniak et al. 2010). Gneezy and Rustichini (2004b) find that the gender gap decreases when adult subjects can choose to compete in solving anagrams compared to shooting baskets, whereas Dreber et al. (2009) find no gender gap in performance change in running, skipping rope or dancing among children.

Previous literature on the gender gap in risk taking among children shows mixed results. Booth and Nolen (2009b) look at single sex and mixed schools and find that boys are more risk taking than girls in mixed schools but that there is no gender gap when comparing boys to girls from single sex schools. Girls are also more risk taking when assigned to all-girl groups than when assigned to mixed groups. Borghans et al. (2009) find a gender gap among 15-16 year old children in the Netherlands, with boys

---

[30] Matrilineal is a technical genealogical term, meaning that people trace descent through the mother's line. Patriarchal means that men have more power in society. These terms are not necessarily opposite: a society can for example be matrilineal (trace descent through the mother) and patriarchal (men have more power).

[31] The math task in this study is rated as being more boyish, see section 4f.

being more risk taking than girls.[32] However, unlike the latter two studies, Harbaugh et al. (2002) find no gender gap in risk taking among children aged 5-13 or among adolescents aged 14-20 in the US.

Moreover, evidence suggests that the gender gap in competitiveness and risk taking is influenced by the subject pool studied. Gneezy et al. (2009), in a study on adults, find that women compete more than men in a matrilineal society in India whereas the opposite is found in a patriarchal society in Tanzania. Moreover, the results of Booth and Nolen (2009a, 2009b), Andersen et al. (2010), and the differences between Gneezy and Rustichini (2004a), Dreber et al. (2009) and Sutter and Rützler (2010) also support the notion that the country or environment in which the study is performed matters. Since Colombia scores lower on gender equality indices than Sweden (Hausmann et al. 2010), we expect the gender gap to be bigger in Colombia in all four competition tasks as well as in risk taking compared to Sweden. We also expect the gender gap to be smaller (if there is any gap at all) in more feminine tasks such as skipping rope and word search compared to running and math in both countries.

We find little support for our hypotheses in Colombia, where boys and girls are equally competitive in all four tasks using both competitiveness measures. However, this is not the case in Sweden. Girls in Sweden increase their performance more than boys do when forced to compete in math, a traditionally male task, but there is also some indication of girls in Sweden being more competitive than boys in skipping rope, a traditionally female task. There is however no gender difference in reaction to competition in running or word search. Meanwhile, boys in Sweden choose to compete more than girls do when given the possibility. Boys and girls are thus consistently equally competitive in Colombia, whereas in Sweden boys are consistently more competitive in terms of choice and girls in terms of performance change. Our results suggest that tasks are only important for the gender gap in competitiveness in Sweden, but not in a uniform way. Risk taking, on the other hand, show results in line with our expectations; the gender gap is larger in Colombia than in Sweden. With this little sup-

[32] Borghans et al. (2009) also find that boys sometimes are more ambiguity averse than girls.

port for our hypotheses, however, we are agnostic to the specific variables that might drive our results.

The outline for our paper is the following. In section 2 we present the experimental setup. We give a summary of our hypotheses and results in section 3, and thereafter present these in more detail in section 4. We finish with a discussion in section 5.

3.2 Experimental setup

The study was divided into two parts: a physical education (PE) part and a classroom part. In the physical education part, the children competed in running and skipping rope, as well as participated in a cooperation task (the latter is described in Cárdenas et al. 2010).[33] Running and skipping rope each consisted of two stages. In stage 1, the children performed the task individually. In stage 2, the children performed the task in competition with another child. While performing the task in the first stage the children were unaware of the existence of a second stage. In the second stage, children were matched with someone who performed similarly to themselves in the first stage. If more than two children obtained the same result in stage one, the matching was random. The children were informed of the matching procedure. Performance in running was based on how fast the children ran 4*13 meters.[34] In the skipping rope task, children jumped with a long rope that one teacher or experimenter and one child turned. Performance was measured by the number of jumps. When competing in skipping rope, two ropes were put next to each other. The children were instructed to start jumping at the same time. Our measure of competitiveness during the physical education class is the absolute change in performance between the first and second stages, the most common measure of the reaction to competition. In the PE part, no compensation was awarded apart from the intrinsic motivation that comes from winning, as in Gneezy and Rustichini (2004a).

[33] In the physical education part, children performed the tasks in the presence of their classmates.

[34] Since this study was conducted indoors we were constrained by the size of a regular the PE class room.

In the classroom, the children competed in math or word search, participated in a risk task and answered a survey. In each class, half of the children were randomly chosen to solve math exercises, whereas the other half were given a word search task. The children did not get any feedback about their performance in any stage. In the first stage, a piece-rate scheme, the children were told that they had two minutes to solve as many exercises as possible, for which they would be given 3 points each. In the second stage, a tournament, the children were again told that they would get two minutes to solve exercises, but that they now would be randomly paired with someone in the class who solved the same type of task, and that if they solved more or the same amount of exercises as the other person, they would get 6 points per exercise, whereas if they solved fewer exercises than the other person they would get 0 points. In the third stage, the children were told that they were to solve exercises for another two minutes, and that they now could choose whether they wanted to be given points according to the piece-rate scheme or the tournament. Comparing performance in the second stage with performance in the first stage gives us a measure of competitiveness as absolute performance change or reaction to competition, whereas the choice in the third stage gives us a measure of competitiveness as a preference for competition. After the competitiveness task was over, we asked the children to guess how many children they believed had performed better than they had on the math task or the word task, for both the piece-rate scheme and the forced competition. This allows us to measure performance beliefs, or over- and underconfidence.

The risk task consisted of six Holt and Laury (2002) type of choices where the children could choose between a lottery in the form of a coin flip that gives 10 or 0 points with equal probability and a safe option where the certain amount increases successively in points (from 2 to 7.5 points). Our first measure of risk preferences relies on the unique switching point where the individual switches from preferring the lottery to preferring the safe option. Our main measure of risk preferences excludes inconsistent subjects, i.e. subjects with multiple switching points. Since some of our subjects are inconsistent we also analyze the number of times a person chooses the uncertain option compared to the safe option. This is our second measure of risk preferences.

After the risk task, a survey was included in order to measure beliefs concerning the different tasks, cooperation and competition, as well as to measure demographics.

In the end of the classroom part, points were converted into pens and erasers. Before the study started, the children were told that more points corresponded to more pens and erasers.

In sum, in this paper we analyze competitiveness as performance change in running, skipping rope, math and word search, competitiveness as choosing to compete or not in math and word search, and risk preferences through incentivized choices over lotteries and safe choices. We also look at additional measures such as overconfidence.

3.3 Summary of the results

Table 1 provides an overview of our hypotheses and results. Surprisingly, few of our hypotheses are supported. We discuss this more extensively in section 4 and 5.

Table 1. Summary of results.

Gender gap	Task	Hypothesis	Results	Hypothesis supported?
Colombia	Running – performance change	G<B	G=B	No
	Skipping rope – performance change	G<B	G=B	No
	Gender gap between tasks	R>S	R=S	No
	Math - performance change	G<B	G=B	No
	Word – performance change	G<B	G=B	No
	Gender gap between tasks	M>W	M=W	No
	Math – choice	G<B	G=B	No
	Word – choice	G<B	G=B	No
	Gender gap between tasks	M>W	M=W	No
	Risk	G<B	G<B	Yes

G=Girls, B=Boys, R=Running, S=Skipping rope, M=Math, W=Word, Col=Colombia, Swe=Sweden. In the results column, = indicates that the hypothesis of a difference could not be rejected.

Table 1 continued. Summary of results.

Gender gap	Task	Hypothesis	Results	Hypothesis supported?
Sweden	Running – performance change	G=B	G=B	Yes
	Skipping rope – performance change	G=B	G>B	No
	Gender gap between tasks	R=S	R<S	No
	Math - performance change	G=B	G>B	No
	Word – performance change	G=B	G=B	Yes
	Gender gap between tasks	M=W	M<W	No
	Math – choice	G=B	G<B	No
	Word – choice	G=B	G<B	No
	Gender gap between tasks	M=W	M=W	No
	Risk	G<B	G<B	Yes
Between countries	Running – performance change	Col>Swe	Col=Swe	No
	Skipping rope – performance change	Col>Swe	Col<Swe	No
	Math – performance change	Col>Swe	Col=Swe	No
	Word – performance change	Col>Swe	Col=Swe	No
	Math – choice	Col>Swe	Col<Swe	No
	Word – choice	Col>Swe	Col=Swe	No
	Risk	Col>Swe	Col>Swe	Yes

G=Girls, B=Boys, R=Running, S=Skipping rope, M=Math, W=Word, Col=Colombia, Swe=Sweden. In the results column, = indicates that the hypothesis of a difference could not be rejected.

3.4 Hypotheses and results

In this section we test whether there is a gender gap in competitiveness and risk taking among children in Colombia and Sweden and if the type of task matters for the size of the gender gap in competitive behavior within and between the countries.

We begin by looking at gender differences in competitiveness within and between the countries in the PE part and then continue by studying competitiveness in the classroom part. We also investigate whether the gender stereotype of a certain task affects the gender gap more in Colombia compared to Sweden. We thereafter look at the gender gap in risk taking within each country and between the countries, and explore how this relates to competitive behavior. Finally, we present some further analysis and robustness checks. All tests of the means are analyzed using the non-parametric Mann-Whitney test and a two-sided t-test. Only the p-values for the Mann-Whitney tests are displayed.[35] When the two tests display conflicting results this difference is usually due to outliers. When this occurs we therefore perform the two tests on the inner quartile range (IQR, the distribution between the 25th and the 75th percentile), and we again only present the p-values for the Mann-Whitney test, labeled IQR. In those cases, the p-values of the full sample are presented in a footnote. All regressions are OLS unless otherwise stated.

3.4.1 Basic statistics

The study was conducted on a total of 1240 children out of which 631 were in Colombia and 609 in Sweden.[36] In either country, approximately half of our sample consists of girls. We have a total of 54 primary classes in the years 3-5; 21 classes from the Bogotá region in Colombia and 33 classes

[35] We present the Mann-Whitney test since none of our variables are normally distributed when using a skewness and kurtosis test. When there is a difference between the tests in terms of significance we also report the p-values for the t-test. We have also compared whether the distributions for each reported variable differ between boys and girls using a Kolmogorov-Smirnov test. The results are similar to those reported for mean values.

[36] The data for Sweden was collected in parallel to the data collection in Colombia, hence the Swedish sample is not the same as in Dreber et al. (2009).

from the Stockholm region in Sweden. The classes were sampled during the fall of 2009 and spring of 2010. In each class, the study started with the PE part and continued with the classroom part either the same day or the same week. Both parts of the study were overseen by at least one teacher. A majority of the 1240 children completed all tasks except the math and word tasks where each child only participated in one of the two tasks.[37] Table 2 below provides summary statistics. For the set of variables used and variable descriptions, see appendix table A1.

Table 2. Summary statistics.

Variable	Mean	Sd	Median	N	Min	Max
Age	10.90	0.91	11	1120	8	15†
Class year	4.18	0.73	4	1240	3	5
Gender (boy=0, girl=1)*	0.48	0.50	0	1222	0	1
Country (Sweden=1, Colombia=0)*	0.49	0.50	0	1240	0	1

*(share between 0 and 1)
†There is one child who is 15 years old, two who are 14 years old, 20 that are 13 years old, and three that are 8 years old.

3.4.2 Competition PE part

In this section we explore competitiveness only as measured by absolute performance change in the PE part.

3.4.2.1 Hypotheses PE part

Previous studies indicate that the gender gap in competitiveness in running is influenced by the country in which the study is performed (Gneezy and Rustichini 2004a, Dreber et al. 2009, Sutter and Rützler 2010). Colombia typically scores lower than Sweden on gender equality indices, and our prior is that such indices capture the relevant factors influencing the gender gap

[37] Among those that did not participate in all the PE tasks this was either due to the different experimental parts (PE and class room parts) being run at separate occasions or to time constraints (in the PE part).

in competitiveness. We thus expect girls to be less competitive than boys in Colombia but not in Sweden, in both tasks. Moreover, Dreber et al. (2009) find no gender gap in Sweden in running and skipping rope, thus we expect no gender differences in Sweden in this sample.

Hypothesis 1: Girls are less competitive than boys in both running and in skipping rope in Colombia, whereas there is no gender gap in Sweden in these tasks.

In the current sample, the children rated skipping rope as more girlish and running as more boyish (see section 4f). In Dreber et al. (2009), the same finding did not influence behavior. We therefore expect the gender gap to be smaller in skipping rope than in running in Colombia, but that the task does not matter in Sweden.

Hypothesis 2: The gender gap in competitiveness is bigger in running than in skipping rope in Colombia, but not in Sweden.

3.4.2.2 Results -– performance change PE

Consistent with sex-stereotypic expectations, boys ran faster and girls skipped rope better on average in both stage 1 (individual performance) and in stage 2 (competition). This is the case in both Colombia and Sweden. Table 3 and table 4 show the average performances and p-values in both stages in Colombia and Sweden.[38]

Table 3. Average performance in stage 1 and in stage 2.

Columbia	**Running**		**SR**	**Skipping rope**		**SR**
	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value
Girls	16.62	15.79	<0.001	26.13	29.07	0.050
Boys	15.28	14.80	<0.001	19.77	22.96	0.203

A lower time for running indicates better performance. A higher number of jumps in skipping rope indicates better performance. Signrank (SR) test p-values of performance change for girls and boys separately in Colombia.

[38] Note that the children were not aware of the second stage when performing the first stage.

Table 4. Average performance in stage 1 and in stage 2.

Sweden	Running		SR	Skipping rope		SR
	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value
Girls	15.96	15.78	0.000	54.03	66.35	<0.001
Boys	15.46	15.32	0.006	24.34	31.58	<0.001

A lower time for running indicates better performance. A higher number of jumps in skipping rope indicate better performance. Signrank (SR) test p-values of performance change for girls and boys separately in Sweden.

With one exception, both boys and girls are competitive in terms of reacting to competition: they increase their performance when competing compared to performing the task individually in both Colombia and Sweden. When skipping rope, boys in Colombia are the only ones who don't increase their performance significantly when competing.

Testing whether there is a significant gender gap in competitiveness as measured by performance change in running, we find no gender gap in Colombia (IQR: p=0.236) or Sweden (p=0.875).[39] (see figure 1). The running result in Sweden is in line with what Dreber et al. (2009) found. In skipping rope, there is no gender gap in performance change in Colombia (p=0.379). In Sweden, there is some evidence that girls compete more than boys (IQR: p=0.014).[40] (see figure 2). This latter result differs from the result on skipping rope found in Sweden in Dreber et al. (2009). This is probably due to the larger sample size in this study, as indicated by the power test in Dreber et al. (2009). However, the gender gap in skipping rope disappears when using a relative measure of performance change, making this finding inconclusive (see the Appendix for further explanation of the relative measure).

39 Using the full sample in Colombia, the non-parametric test gives a significant gender difference (p=0.009) whereas the parametric test gives a borderline insignificant result (p=0.095).

40 Using the full sample in Sweden, the Mann-Whitney test gives a significant p-value (p=0.021) whereas the p-value from the t-test is insignificant (p=0.348).

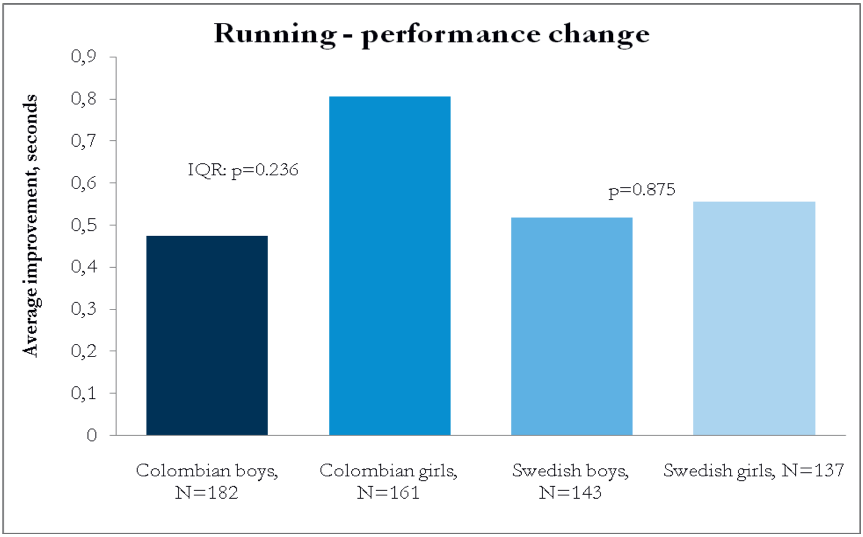

Figure 1. Average performance change in time (stage 2 – stage 1), by gender.

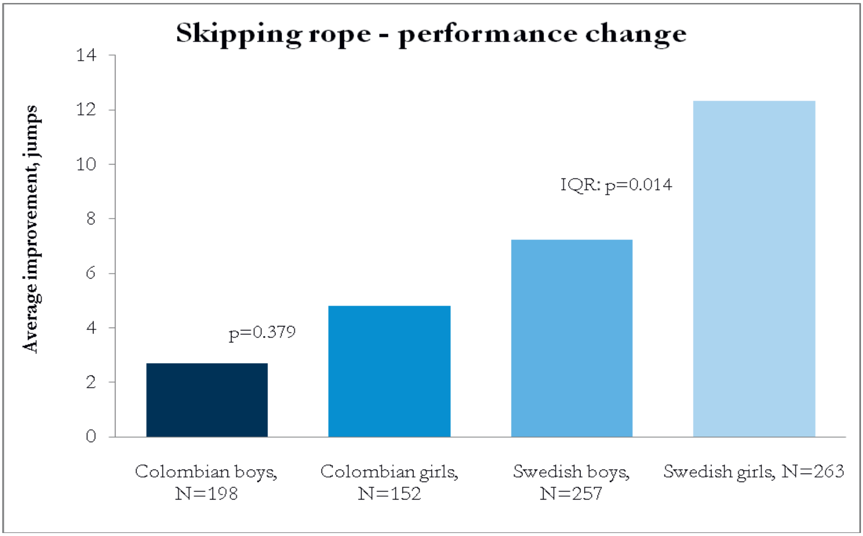

Figure 2. Average performance change in jumps (stage 2 – stage 1), by gender.

We also test whether the gender gaps differ between Colombia and Sweden in a regression analysis. Using the parametric tests we found no gender gap within each country, thus there are no significant differences in the regression analysis.[41] This result is not altered when adding control variables (see appendix tables A2-A3).[42]

Testing hypothesis 2, we look at whether the gender gap in competitiveness is bigger in running than in skipping rope in either country with a regression analysis. In order to be able to compare performance change between running and skipping rope we look at relative performance change rather than absolute performance change. See the first section of the Appendix for further analysis of relative performance change. We find no evidence of the gender gap being influenced by the task in neither Colombia nor Sweden (see appendix tables A4-A5).

We thus find no support for hypothesis 1 or for hypothesis 2. Boys and girls are equally competitive in running in both Colombia and Sweden; there is no gender gap in competitiveness in skipping rope in Colombia whereas there is some evidence of girls being more competitive than boys in skipping rope in Sweden. However the gender gaps in relative performance change display no significant differences between the two tasks.

The gender of the opponent is known in both running and skipping rope. There is some previous work suggesting that the gender of the opponent matters, but the results are mixed (see e.g., Croson and Gneezy 2009). In our sample the only opponent effects we find are that girls in Colombia run significantly faster when competing against another girl (p=0.001) and

[41] The gender gap in skipping rope becomes significantly larger in Sweden when using the other risk measure, see section 4d.

[42] When performing the regression analysis we compare the results from a regression with no control variables with regressions using two sets of controls. The first set of controls contain actual individual performance, expected individual performance (i.e. beliefs), age and risk preferences. These controls are included since previous work has shown that these are factors that play a role for both competitiveness measures. The second set of controls includes all variables from the first set plus four additional variables from the questionnaire that control for how gendered the children perceive the tasks to be and how important they consider competing to be. These four variables were included to control for motivational factors that may play a role in competitiveness.

boys in Sweden run significantly faster when competing against a girl (p=0.012).

Table 5. Differences in performance based on the gender composition in the competing pairs, p-values.

	Colombia				Sweden			
	Running		Skipping		Running		Skipping	
	n	p-value	n	p-value	n	p-value	n	p-value
Girls: boys vs girls	54/107	0.001	65/72	0.264	70/58	0.700	50/75	0.644
Boys: boys vs girls	126/56	0.039†	120/64	0.782	68/68	0.020	73/51	0.217

†This is not significant using a t-test (p=0.144) or with IQR (p=0.646).

3.4.3 Competition classroom

In this part we study competitiveness in math and word search as measured both by performance change as well as choosing to compete or not.

3.4.3.1 Hypotheses

There are no previous studies exploring the gender gap in different classroom tasks, such as math and word tasks, among children. Given the literature on performance change in the PE tasks among children we expect boys to be more competitive than girls in Colombia but not in Sweden. Since previous studies have found that competitiveness sometimes depends on the task for adults, we expect the gender gap to be bigger in math than in word search.

Hypothesis 3: Girls are less competitive than boys in Colombia in terms of performance change in both math and word search, whereas there is no gender gap in Sweden.

Hypothesis 4: The gender gap in competitiveness in terms of performance change will be bigger in the math task than in the word task in Colombia, but not in Sweden.

In the current sample, the children rated math as more boyish and word search as more girlish (see section f). Moreover, previous literature on adults show that men are more competitive when it comes to choosing to compete in math in western societies typically ranked less equal compared to Sweden, thus we expect girls to choose competition less than boys in Colombia but not in Sweden, for both tasks.[43] We also expect the gender gap to be bigger in math than in word search in Colombia but not in Sweden.

Hypothesis 5: Girls are less competitive than boys in Colombia in terms of choice in math and word tasks, whereas there is no gender gap in Sweden.

Hypothesis 6: The gender gap in competitiveness in terms of choice will be bigger in the math task than in the word task in Colombia but not in Sweden.

3.4.3.2 Results – performance change

When exploring performance in stage 1 (individual performance: piece-rate scheme), we find support for the math and word tasks being gendered in Sweden but not in Colombia. Performance in stage 1 differs between boys and girls in Sweden; boys perform better in the math task and girls perform better in the word task (Math: p=0.017, Word: p=0.043). In Colombia we find no gender differences in stage 1 (Math: p=0.746, Word: p=0.172). Tables 6 and 7 below display the average piece-rate performances and the average forced tournament performances.

Table 6. Average performance in stage 1 and in stage 2.

Colombia	**Math**		**SR**	**Word**		**SR**
	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value
Girls	6.61	7.15	0.163	3.36	4.22	<0.001
Boys	7.13	7.22	0.448	3.22	4.25	<0.001

Signrank (SR) test p-values of performance change for girls and boys separately in Colombia.

[43] E.g. Niederle and Vesterlund 2007 conduct their experiment on adults in the US. US is ranked 19th in the Global Gender Gap Report 2010 (Hausmann et al. 2010).

Table 7. Average performance in stage 1 and in stage 2.

Sweden	Math		SR	Word		SR
	Stage 1	Stage 2	p-value	Stage 1	Stage 2	p-value
Girls	9.60	10.73	0.001	9.41	9.81	0.303
Boys	11.22	11.11	0.378	8.28	8.34	0.705

Signrank (SR) test p-values of performance change for girls and boys separately in Sweden.

In Colombia, both boys and girls are competitive in word search in terms of reacting to competition, whereas this in not the case in math. In Sweden, only girls increase their performance significantly when forced to compete in the math task, but as for the result on skipping rope the gender difference disappears when we use a relative performance measure.

When we test whether there is a gender difference in competitiveness in Colombia and Sweden in either task, we find a gender gap only in Sweden and only in math: Girls in Sweden increase their performance in math significantly more than boys do (p=0.002). In Colombia however, there is no gender difference in performance change in the math task (p=0.747) or in the word task (p=0.172). In Sweden, there is no gender gap in competitiveness in the word task (p=0.555) (see figures 3 and 4).

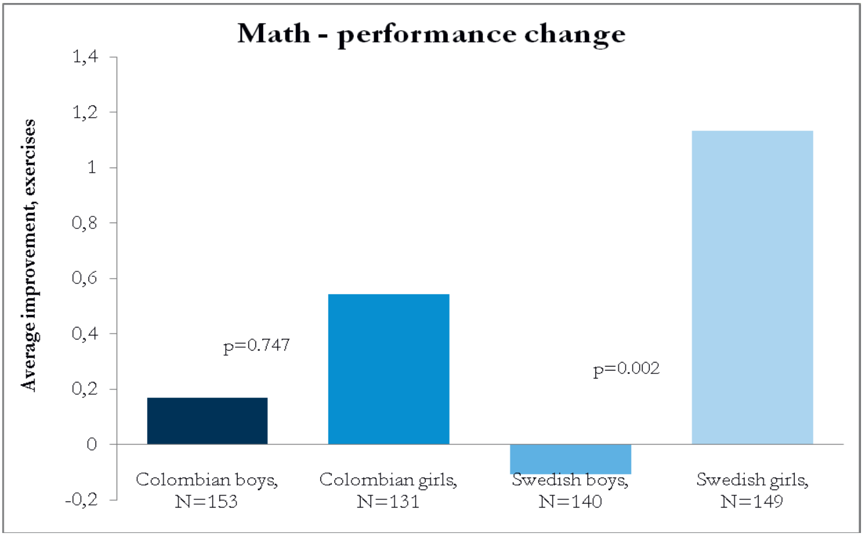

Figure 3. Average change in math exercises (stage 2 – stage 1), by gender.

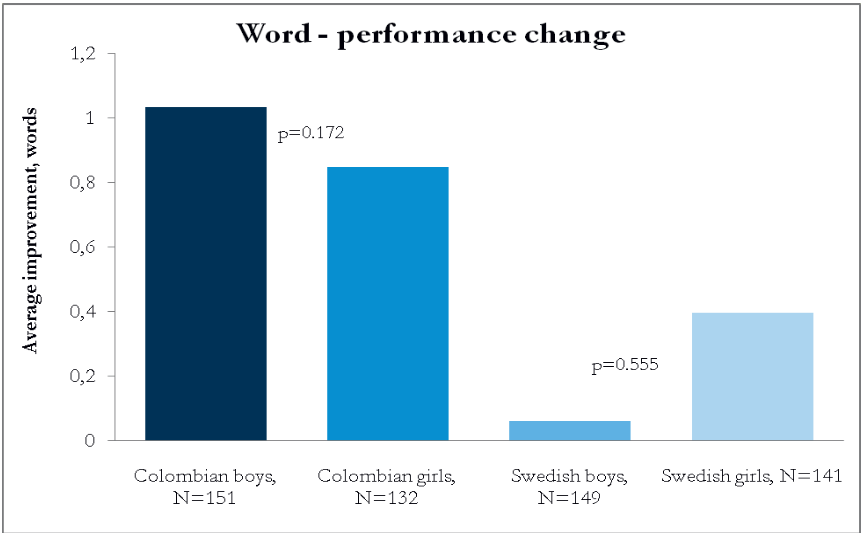

Figure 4. Average change in words found (stage 2 – stage 1), by gender.

In a regression analysis we find that the gender gap in performance change in math is not significantly bigger in Sweden than in Colombia (p=0.214).

When adding controls, the results remain similar (see appendix tables A6-A7). There is, as anticipated, also no significant difference in the gender gap in the word task between Colombia and Sweden (p=0.509).

We further test whether the gender gap in competitiveness in terms of relative performance change is bigger in math than in word search in either country. We find no evidence of this (see appendix tables A8-A9).

Little support is thus found for Hypotheses 3 and 4. There is no gender gap in competitiveness, as measured by performance change in Colombia in either task or in the word task in Sweden, whereas girls in Sweden are more competitive than boys in the math task. Yet, in a regression analysis of relative performance change, the gender gap does not seem to be influenced by the task.

3.4.3.3 Results – choice

In stage 3, when the children could choose whether or not to compete, we find that boys and girls in Colombia are equally likely to choose to compete in math and word search (Math: p=0.648, Word: p=0.610).[44] In Sweden, on the other hand, boys are significantly more likely to choose to compete both in math and in word search compared to girls: 44% of the boys and only 19% of the girls chose to compete in math (p<0.001), whereas in word search the corresponding numbers are 39% and 27% (p=0.041) (see Figures 5 and 6).

[44] Among Colombian children, 35% of the boys and 32 % of the girls chose to compete in math, with the corresponding numbers for word search being 26% resp. 29%.

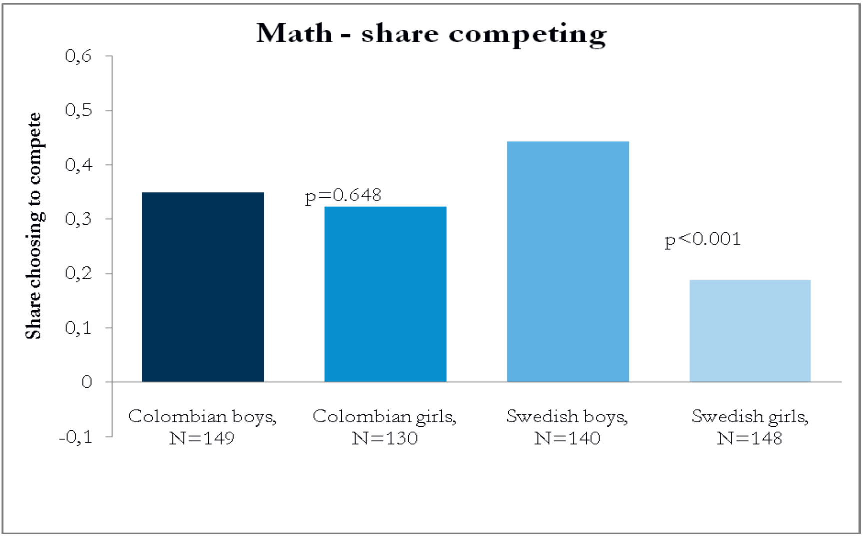

Figure 5. Share choosing to compete in math, by gender.

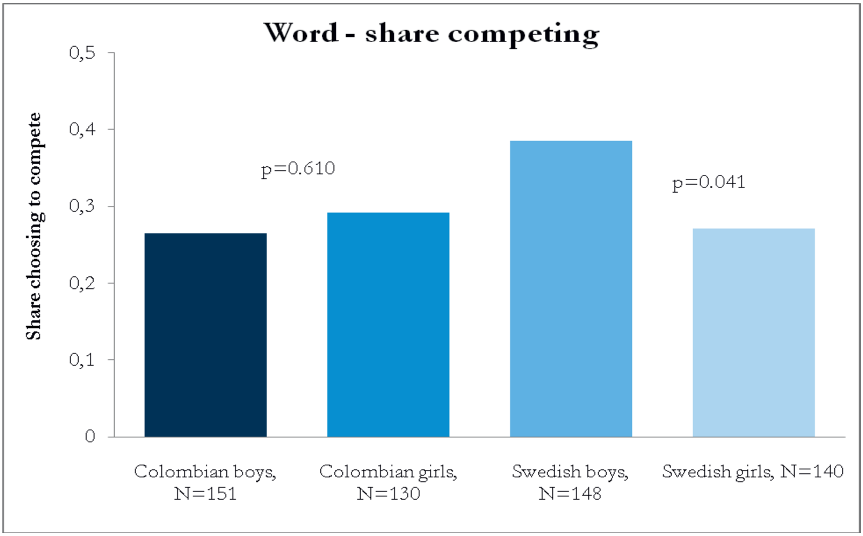

Figure 6. Share choosing to compete in word search, by gender.

Comparing the gender gap in choice between Colombia and Sweden, we find a significant difference in the math task. The gender gap in math is significantly larger in Sweden than the gender gap in Colombia (p=0.003). In word search we find a borderline insignificant gender gap between the two countries (p=0.068). When adding controls to the regression analysis (see footnote 11), the gender gap in competitiveness as measured by choice is significantly larger for both the math and the word task in Sweden (see appendix tables A10-A11).

Testing whether the gender gap in choice is bigger in math than in word search, we find some evidence of this being the case in Colombia (No controls: p=0.496, control Set 1: p=0.050, control Set 2: p=0.045) but not in Sweden.[45] (see appendix tables A12-A13).

We thus find no support of hypothesis 5. When it comes to competitiveness as measured by choice we find a gender gap in competitiveness in both tasks in Sweden but not in Colombia. It is however only the gender gap in math that is significantly different between the countries. Moreover, in Colombia, but not in Sweden, there is some support of hypothesis 6, with the gender gap in choice in math being somewhat bigger than in word search.

To summarize the section on competitiveness: when measuring competitiveness as a performance reaction to a competitive setting we find a some evidence of a gender gap only in Sweden where girls compete more in math. There is also some evidence of girls being more competitive in skipping rope in Sweden. When looking at the choice of competition we again find a gender gap only in Sweden, where boys choose to compete more often than girls in both math and word search (controlling for performance). Finally, there is only little evidence of the task being important for the gender gap in competitiveness. Though we find that girls in Sweden are more competitive than boys in terms of performance change in some instances, explicitly testing the gender gap in a regression analysis indicates that the only time the task matters is when it comes to competition choice in Colombia.

[45] The gender gap in choice reaches significance when adding controls in Colombia in this regression analysis, it disappears however when using the other risk measure. This is most likely due to the fact that we find a large gender difference in risk taking in Colombia.

3.4.4 Risk preferences

In this section we explore the gender gap in risk preferences measured from incentivized lotteries conducted in the class room.

3.4.4.1 Hypotheses

Previous work finds mixed results on the existence of a gender gap in risk taking among children and adolescents (Harbaugh et al. 2004, Booth and Nolen 2009b, Borghans et al. 2009). Among the studies that do find a gender gap, boys are found to be more risk taking than girls. We thus expect boys to take more risk in both countries, but given that Colombia scores lower on gender equality indices we expect the gap to be bigger in Colombia.

Hypothesis 7: Boys are more risk taking in both countries.

Hypothesis 8: The gender gap is greater in Colombia than in Sweden.

3.4.4.2 Results – risk

In the joint sample of children (including children in both Colombia and Sweden), 25% of the children were inconsistent in their choices of the safe option versus the lottery (coin flip). In general, the children are significantly more inconsistent in Colombia (29%) compared to Sweden (20%) ($p<0.001$).[46] There is however no gender difference in being inconsistent in either country (Colombia: $p=0.903$, Sweden: $p=0.205$). We also measure risk preferences in terms of the number of risky choices chosen, in order to not exclude inconsistent choices. Using this outcome measure the results are similar to those presented here.

---

[46] These shares are higher than what is typically found among adults, and could be an indication of a limited understanding of probabilities in this age group. Future research should take this into account.

Table 8. Summary table risk measures.

Variable	Mean	Sd	Median	N	Min	Max
Risk	3.99	2.22	3.5	875	1	8.75
Inconsistent answers	0.25	0.43	0	1166	0	1
Number of risky choices	2.54	1.66	3	1138	0	6

We find a gender gap in risk taking in both countries, with boys taking more risk. In Colombia, boys take 40% more risk than girls ($p<0.001$), with the corresponding number in Sweden being 15% ($p=0.001$) (see figure 7).

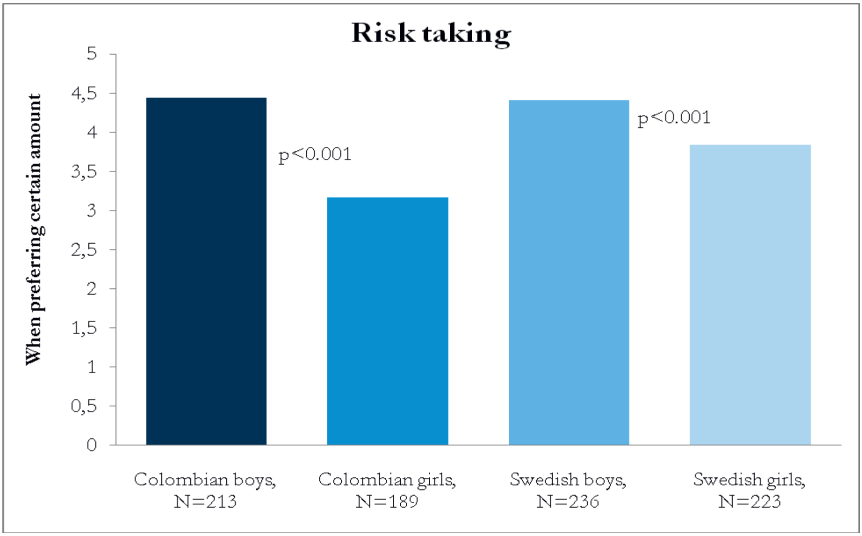

Figure 7. Risk taking, by gender.

Comparing Colombia and Sweden, we find that Colombian children take less risk than Swedish children ($p<0.001$). This result is driven by the difference between Colombian and Swedish girls, since boys are equally risk taking in the two countries. When testing the size of the gender gaps, we find a significantly larger gender gap in Colombia compared to Sweden ($p=0.015$).

Thus, hypotheses 5 and 6 are supported: boys take more risk in both countries, and the gender gap is greater in Colombia than in Sweden.

3.4.5 Competitiveness and risk preferences

We also explore the relationship between risk taking and competitiveness, since the two things often are related yet are two separate concepts, and there are strong gender differences in both preferences. We find a positive relationship between risk taking and choice of competition in Sweden ($p<0.001$), indicating that the children who choose to compete also tend to be more risk taking, and vice versa. In Colombia there seems to be no such relationship ($p=0.149$). Studying the sample split by gender within each country, both girls and boys display the same positive correlation pattern in Sweden (Girls: $p=0.017$, Boys: $p<0.001$). In Colombia neither boys nor girls display a positive pattern between choice of competition and risk taking behavior (Girls: $p=0.948$, Boys: $p=0.105$).[47]

Niederle and Vesterlund (2007) find that the gender gap in risk preferences only explains part of the gender gap in competitiveness as measured by choice among adults, and our results support this. Our results indicate that the cross-country factors in play seem to affect risk taking and competitiveness differently.

3.4.6 Further analysis and robustness checks

In this section we provide some further analysis of our findings. Additional tests and an analysis of differences in variance and relative performance can also be found in the first section of the appendix.

3.4.6.1 Performance beliefs

We asked the children to rank their believed performance in math and word search relative to their classmates. We create a variable that measures this discrepancy, thus both over- and underconfidence. Actual piece-rate performance differs significantly from the self-reported expected piece-rate

[47] The p-values come from testing equality of distribution of risk between those who chose competition to those who did not, using a Kolmogorov Smirnov test. This is the case for both indicators of risk preferences: the threshold children use for switching between a sure amount and a risky, or the number of risky choices they select out of all choices.

performance in both tasks and countries, except for the math task in Sweden. Children believe they perform better than they actually do in both tasks in both countries. We find no gender gap in this confidence measure when it comes to math or word search in either country. On average, the Colombian children are more overconfident than Swedish children ($p<0.001$). Using beliefs as a control variable does not alter any of our results. For relevant p-values please see appendix tables A14-A15.

It is surprising that we don't find that overconfidence, or a gender difference in beliefs about performance, explains part of the gender gap given that it has previously been shown to play an important role (e.g. Niederle and Vesterlund 2007). It is also surprising that there is no gender gap in overconfidence in either task in either country, since these results differ from those of Dahlbom et al. (2010), who find that among 14-year old children in Sweden, boys are overconfident and girls are underconfident in terms of math performance. Our results also differ from those of Jakobsson et al. (2010), who find that boys in El Salvador are overconfident and girls are underconfident in math whereas there is no gender gap in a more gender neutral task such as performance in social science, where both boys and girls are overconfident. The children in our study are younger than those in Dahlbom et al. (2010) and Jakobsson et al. (2010), and we ask a retrospective question whereas these other two studies ask the children about their expected performance on a math test that will be performed later. This may explain the discrepancy between our results.

3.4.6.2 Do the children perceive competing as important and tasks as gendered?

The final element in the classroom part is a survey where we elicit perceptions of how boyish/girlish the children considered running, skipping rope, math and word search to be. We further asked how boyish/girlish they considered competing in these tasks to be. We used a scale from 0 to 10 where a lower number indicates rating the task as more girlish and a higher number as more boyish (0=very girlish, 5=neutral, 10=very boyish).

In both countries, boys rate competition as more important compared to girls (Colombia: $p=0.009$, Sweden: $p<0.001$). In Colombia, both girls and boys believe that it is more important to compete against a boy than against a girl (Girls: $p=0.003$, Boys: $p<0.001$). Girls in Sweden rate competing against a boy as being more important compared to competing against a girl (p<0.001), whereas boys rate it as equally important (p=0.375). This does not correspond to what we observe in terms of the gender of opponent effect in performance change. For example, Swedish boys actually change their performance more when competing against a girl in running, see Table 4. Children in both Colombia and Sweden perceive math and running as being significantly more boyish (p<0.001 for both countries and both tasks) whereas skipping rope and word search are seen as being more girlish (p<0.001 for both countries and both tasks). Boys and girls tend to agree in these ratings, except that boys in both Colombia and Sweden perceive word search to be more girlish whereas girls perceive it to be more gender neutral (Colombia: Girls: p=0.111, Boys: p<0.001, Sweden: Girls: p=0.288, Boys: p<0.001). In Colombia, girls drive the results for skipping rope and word search and boys for running and math. The same holds for Sweden, except for skipping rope where boys and girls rate it as being equally girlish. Exact point estimates and p-values are found in table A16.[48]

3.5 Discussion

In studies on adults, men are typically more competitive, measured by both performance change in response to competition and the choice to compete, and more risk taking than women. This difference in behavior may explain part of the gender gap observed in many areas in society, including why men are more likely to be in top positions in most sectors. The foundations of the gender gap are currently being investigated in a number of ways. For example, some studies find that the type of task used to measure competitiveness matter (Gneezy and Rustichini 2004b, Günther et al. 2009, Grosse and Riener 2010), and influences the extent to which there is a gender gap in competitiveness, whereas other studies find no effect (Dreber et al. 2009, Wozniak et al. 2010). The gender gap in competitiveness among adults, as measured by choice, has been shown to disappear with performance feedback (Wozniak et al. 2010) and in setups where uncertainty about performance is minimized (Niederle and Yestrumskas 2008), and the gender dif-

[48] We have also performed a quantile regression analysis of competitiveness as measured by performance change.

ference in performance change vanishes with repetition of the competition (Cotton et al. 2009).

It has also been shown that the social and cultural environment in which the study is conducted plays an important role in explaining the gender gap in competitiveness (e.g. Gneezy and Rustichini 2004a, Dreber et al. 2009, Gneezy et al. 2009, Sutter and Rützler 2010). For example, Andersen et al. (2010) find that boys become more competitive than girls first around the age of 13-15 in a patriarchal society but not in a matrilineal society, where there is no gender gap in any age group. These discrepancies suggest that there is a need for more studies on this in a wide range of countries.

There are also studies that attempt to address the hormonal impact on the gender gap in preferences for competition and risk among adults (see Dreber and Hoffman 2010 for a review of this literature). These studies find conflicting results, while only looking at adults, on the impact of the menstrual cycle on competitiveness (Buser 2009, Wozniak et al. 2010) and on risk taking (Buser 2009, Chen et al. 2005, Pearson and Schipper 2009). The same is true for testosterone and risk taking (Apicella et al. 2008, Sapienza et al. 2009, Zethraeus et al. 2009), whereas the only study that we are aware of that looks at competitiveness find no hormonal correlates (Apicella et al. 2010). More work is thus needed in this field with inconclusive results, as well as studies looking at hormonal correlates among children and adolescents.

In this paper we study the gender gap in competitiveness and risk taking among children aged 9-12 in Colombia and Sweden. We consistently find no gender gap in competitiveness in Colombia, a country considered less gender equal than Sweden. In Sweden, we find clear evidence that boys choose competition more than girls in both math and word search. There is also some indication of girls being more competitive than boys in skipping rope and math when it comes to performance change in Sweden. Our hypotheses on competitiveness are thus not supported. Meanwhile, boys are more risk taking in both Colombia and Sweden, and the gender gap is greater in Colombia than in Sweden. This supports our hypotheses on risk preferences.

It is puzzling why our priors are not supported for competitiveness while they are supported for risk taking. Colombia and Sweden differ in many aspects, including the level of gender equality. Our results indicate that competitiveness and risk preferences pick up behaviors that are affected in a dissimilar way by these societal gender differences between the two countries. We hypothesized that the gender equality of the country would be a good proxy of the gender gap. Our sample of countries is obviously very small, but thus far the gender equality of the country seems to not be a good proxy of the gender gap in competitiveness. This should be elaborated further in more extensive studies. Moreover, focusing on identifying the specific components and how they relate to gender differences in competition, be it the country's educational gender gap, labor market gender gap, or political gender gap, is also a potentially fruitful avenue for future research.

Exploring the gender gap in preferences for competition and risk as we have done here contributes to further our understanding of the cultural impact on the gender gap in preferences as well as gives us more insights on what the gender gap in preferences looks like among children, which is not necessarily the same as among adults. It would be interesting to explore other age groups, including adults in a cross-cultural study, as well as to explore other types of preferences. This is an endeavor that will require collaborations among researchers across a wide range of countries, perhaps including other types of social and cognitive scientists for complementary perspectives of the gender gap and the development of preferences.

3.6 References

Andersen S., Ertac S., Gneezy U., List J., Maximiano S., 2010, ”Age and gender differences in competitiveness: evidence from matrilineal and patriarchal societies”, Mimeo.

Apicella C.L., Dreber A, Campbell B., Gray P.B., Hoffman M., Little A.C. 2008, “Testosterone and financial risk preferences”, *Evolution and Human Behavior*. 29: 384−390.

Apicella C.L., Dreber A, Gray P.B., Hoffman M., Little A.C. Campbell B., 2010, “Androgens and competitiveness in men”, Forthcoming in *Journal of Neuroscience, Psychology and Economics*.

Booth A.L. Nolen P.J., 2009a, ”Choosing to compete: how different are girls and boys?”, IZA Discussion Papers 4027.

Booth A.L. Nolen P.J., 2009b, “Gender differences in risk behaviour: does nurture matter?”, IZA Discussion Paper 4026.

Borghans L., Golsteyn B., Heckman J.J. Meijers H., 2009, “Gender differences in risk aversion and ambiguity aversion”, *Journal of the European Economic Association*.7(2-3): 649–658.

Buser T., 2009, “The impact of female sex hormones on competitiveness”, Mimeo.

Cárdenas J.C., Dreber A., von Essen E., Ranehill E., ”Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden”, Forthcoming in *Journal of Economic Behavior and Organization*.

Chen, Y., Katuscak P., Ozdenoren E., 2005, ”Why can't a woman bid more like a man?”, Mimeo.

Cotton C., McIntyre F. Price J., 2009, “Gender differences disappear with exposure to competition”, Working Paper 2010-11, University of Miami, Department of Economics.

Croson R., Gneezy U., 2009, "Gender differences in preferences", *Journal of Economic Literature*. 47(2): 1-27.

Dahlbom L., Jakobsson A., Jakobsson N., Kotsadam A., 2010, "Gender and overconfidence: are girls really overconfident?", Forthcoming in *Applied Economics Letters*.

Dohmen T., Falk A., Huffman D., Sunde U., Schupp J. Wagner G.G., 2010, "Individual risk attitudes: measurement, determinants and behavioral consequences", Forthcoming in *Journal of the European Economic Association*.

Dreber A., von Essen E., Ranehill, E., 2009, "Outrunning the gender gap: boys and girls compete equally", Forthcoming in *Experimental Economics*.

Dreber A., Hoffman, M., 2010, "Biological basis of sex differences in risk aversion and competitiveness", Mimeo.

Gneezy U., Rustichini A., 2004a, "Gender and competition at a young age", *American Economic Review*. 94(2): 377-381.

Gneezy U., Rustichini A., 2004b, "Executives verses teachers: gender, competition and self selection", Mimeo.

Gneezy U., Leonard K. List J., 2009, "Gender differences in competition: evidence from a matrilineal and a patriarchal society", *Econometrica*. 77(5): 1637-1664.

Grosse, N.D. Riener G., 2010, "Explaining gender differences in competitiveness: gender-task stereotypes", Jena Economic Research Papers 2010 – 017.

Günther C., Ekinici N.A., Schwieren C., Strobel M., 2009, "Women can't jump? – An experiment on competitive attitudes and stereotype threat", Working Paper, University of Graz

Harbaugh W.T., Krause K. Vesterlund L., 2002, "Risk attitudes of children and adults: choices over small and large probability gains and losses", *Experimental Economics*. 5:53–84.

Hausmann R., Tyson L., Zahidi S., 2010, "The global gender gap report", World Economic Forum. Geneve.

Holt C.A. Laury S.K.., 2002, "Risk aversion and incentive effects", *American Economic Review*. 92(5): 1644-1655.

Jakobsson N, Kotsadam A. Levin M., 2010, "Gender and overconfidence: effects of context, gender ttereotypes and peer groups", Mimeo.

Niederle M., Vesterlund L., 2007, "Do women shy away from competition? Do men compete too much?", *Quarterly Journal of Economics*. 122(3): 1067-1101.

Niederle M., Yestrumskas A.H., 2008, "Gender differences in seeking challenges: the role of institutions", Mimeo.

Pearson M., Schipper B.C., 2009, "Menstrual cycle and competitive bidding", MPRA Paper 16784.

Sapienza P., Zingales L., Maestripieri D., 2009, "Gender differences in financial risk aversion and career choices are affected by testosterone", *Proceedings of the National Academy of Sciences*. 106(36): 15268–15273.

Sutter M., Rützler D., 2010, "Gender differences in competition emerge early in life", IZA Discussion Paper 5015.

Wozniak D., Harbaugh E.T., Mayr U., 2010, "Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback", MPRA Paper 21097.

ZethraeusN., Kocoska-Maras L., Ellingsen T., von Schoultz B., Lindén Hirschberg A. Johannesson M., 2009, "A randomized trial of the effect of estrogen and testosterone on economic behavior", *Proceedings of the National Academy of Sciences*. 106(16): 6535-6538.

3.7 Appendix

3.7.1 Further analysis

3.7.1.1 Relative difference in performance

We also conduct the same analysis for performance with relative performance instead of absolute performance, where relative performance is defined as ((performance in stage 2 – performance in stage 1)/performance in stage 1). With this analysis the gender differences that we found using absolute performance change in skipping rope and math in Sweden disappear. Hence, we find no gender gap in competitiveness in neither Colombia nor Sweden in any task when it comes to relative performance change.

3.7.1.2 Variance

Studying gender differences in performance looking at gender differences in various parts of the performance distribution might provide further insight. Even though we find no significant gender differences in performance when looking at the mean, there may be differences in the variances of the performance distributions.[49]

The results when analyzing the variances in running and skipping rope are in line with what we find in the analysis of the means. In running in Colombia, there is no difference in variances when we look at the inner quartile range the gender difference is no longer significant (p=0.101). Sweden has no gender gap in the variance of the running performance distribution (p=0.487). In skipping rope, Colombian boys and girls have an equal variance (p=0.185), but in Sweden girls have a larger variance in skipping rope performance compared to boys (p=0.010), supporting our results in terms of mean differences. In Colombia, where we found no gender difference in mean performance, we also find no gender difference in the variance of

[49] The most common test for a comparison of standard deviations, the F-test for the homogeneity of variances (sdtest), is very sensitive to the assumption that that the data are drawn from an underlying normal distribution. Therefore we also performed a robust test (Levene's test with mean, median and 10% trimmed mean). None of these tests indicated significant differences in the variances. For simplicity we report only p-values from the non-parametric test using the mean.

math performance (p=0.255). In word search, however, where no gender difference in the mean was found, the non-parametric test displays an insignificant difference (p=0.522) whereas the parametric test indicates a significantly larger variance for boys (p<0.001) as does the test on the inner quartile range (p=0.029). In Sweden, the results for the mean analysis are supported, since we neither find a gender gap in the variance in math performance (p=0.426) nor a robust gender difference in the variance in word search performance (t-test: p=0.036, MW: p=0.086, IQR: p=0.142).

In sum, boys in Colombia have a larger variance in word performance, whereas in Sweden the girls have a larger variance in skipping rope.

Table A1. Set of variables used, variable description.

Variable	Variable description
Sweden	Dummy variable for country, Sweden=1
Female	Dummy variable for gender, girl=1
Female*Sweden	Interaction variable between gender and country
Individual performance	Performance in the non-competitive setting
Competitive performance	Performance in the competitive setting
Running	Dummy variable type of PE task, running=1
Math	Dummy variable type of lab task, math=1
Age	Age in years
Risk	Certainty equivalent in risk task
Risky choices	Number of risky choices in risk task
Performance beliefs	Participants' guessed rank
Importance winning female opponent	Importance of winning against a girl, scale 1-10
Importance winning male opponent	Importance of winningagainst a boy, , scale 1-10
Running gendered	How gendered running is, scale 1-10
Skipping gendered	How gendered skipping is, scale 1-10
Math gendered	How gendered math is, scale 1-10
Word gendered	How gendered word search is, scale 1-10

Table A2. Performance change running.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Sweden	0.323**	0.475***	0.490***
	(0.128)	(0.147)	(0.155)
Female	-0.288**	0.121	0.169
	(0.142)	(0.173)	(0.186)
Female*Sweden	0.247	-0.089	-0.163
	(0.185)	(0.211)	(0.223)
Individual performance		-0.279***	-0.278***
		(0.030)	(0.030)
Age		-0.0570	-0.047
		(0.060)	(0.060)
Risk		-0.0218	-0.025
		(0.024)	(0.025)
Importance winning female opponent			-0.027*
			(0.017)
Importance winning male opponent			-0.004
			(0.017)
Running gendered			0.014
			(0.024)
Observations	898	620	617
R-squared	0.029	0.153	0.159

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A3. Performance change skipping rope.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Sweden	4.547	3.700	5.408
	(3.647)	(4.442)	(4.561)
Female	2.129	7.085	8.916*
	(4.159)	(5.179)	(5.316)
Female*Sweden	2.953	10.30	8.064
	(5.361)	(6.488)	(6.683)
Individual performance		-0.364***	-0.369***
		(0.037)	(0.037)
Age		4.791***	4.256**
		(1.788)	(1.807)
Risk		0.334	0.156
		(0.757)	(0.756)
Importance winning female opponent			0.0311
			(0.503)
Importance winning male opponent			0.419
			(0.506)
Skipping gendered			0.865
			(0.666)
Observations	870	608	601
R-squared	0.009	0.151	0.157

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A4. Comparing the gender gap in performance change in running and skipping rope in Colombia.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Female	0.305	0.836	0.891
	(0.368)	(0.932)	(0.975)
Running	-0.741***	-0.850	-0.851
	(0.228)	(0.548)	(0.558)
Female*Running	-0.320	-0.986	-0.950
	(0.369)	(0.898)	(0.920)
Individual performance skipping		-0.020**	-0.020**
		(0.008)	(0.008)
Individual performance running		0.078	0.109
		(0.068)	(0.085)
Age		0.300**	0.230**
		(0.129)	(0.111)
Risk		-0.099	-0.098
		(0.066)	(0.060)
Importance winning female opponent			-0.024
			(0.070)
Importance winning male opponent			0.028
			(0.038)
Skipping gendered			0.096
			(0.077)
Running gendered			-0.013
			(0.044)
Constant	0.714***	-2.741	-2.818
	(0.228)	(1.704)	(1.932)
Observations	726	256	252
R-squared	0.035	0.084	0.090

Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A5. Comparing the gender gap in performance change in running and skipping rope in Sweden.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Female	-0.180	-0.051	-0.159
	(0.267)	(0.346)	(0.348)
Running	-1.122***	-1.223***	-1.234***
	(0.224)	(0.307)	(0.309)
Female*Running	0.177	0.290	0.367
	(0.267)	(0.348)	(0.345)
Individual performance skipping rope		-0.007***	-0.007***
		(0.001)	(0.001)
Individual performance running		0.053	0.056
		(0.105)	(0.103)
Age		0.346***	0.303***
		(0.120)	(0.116)
Risk		0.036	0.0118
		(0.060)	(0.066)
Importance winning female opponent			-0.019
			(0.026)
Importance winning male opponent			0.018
			(0.024)
Skipping gendered			0.093*
			(0.048)
Running gendered			-0.015
			(0.072)
Constant	1.114***	-3.449	-3.128
	(0.224)	(2.706)	(2.669)
Observations	1,042	720	714
R-squared	0.058	0.087	0.089

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A6. Performance change math and word search, control variables set 1.

VARIABLES	All	Math	Word
	(1)	(2)	(3)
Sweden	0.917**	0.662	1.442***
	(0.377)	(0.585)	(0.495)
Female	0.236	0.084	0.313
	(0.385)	(0.626)	(0.475)
Female*Sweden	0.468	0.765	0.484
	(0.497)	(0.782)	(0.634)
Individual performance	-0.374***	-0.340***	-0.468***
	(0.026)	(0.036)	(0.044)
Math	1.089***		
	(0.253)		
Age	0.515***	0.560**	0.648***
	(0.141)	(0.237)	(0.179)
Risk	0.0455	0.047	0.0477
	(0.060)	(0.095)	(0.075)
Performance beliefs	1.094*	1.789**	0.342
	(0.566)	(0.855)	(0.746)
Observations	767	375	392
R-squared	0.251	0.268	0.248

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A7. Performance change math and word search, control variables set 2.

VARIABLES	All	Math	Word
	(1)	(2)	(3)
Sweden	0.867**	0.444	1.362***
	(0.396)	(0.635)	(0.500)
Female	0.127	-0.282	0.435
	(0.412)	(0.685)	(0.500)
Female*Sweden	0.574	1.266	0.259
	(0.524)	(0.842)	(0.662)
Individual performance	-0.374***	-0.328***	-0.461***
	(0.026)	(0.037)	(0.044)
Math	1.099***		
	(0.256)		
Age	0.508***	0.578**	0.603***
	(0.143)	(0.246)	(0.178)
Risk	0.0313	0.067	0.033
	(0.061)	(0.099)	(0.075)
Performance beliefs	1.032*	1.668*	0.368
	(0.571)	(0.867)	(0.743)
Importance winning female opponent	-0.016	0.094	-0.095*
	(0.040)	(0.066)	(0.049)
Importance winning male opponent	0.016	0.014	-0.003
	(0.040)	(0.068)	(0.048)
Word gendered	-0.055	-0.032	-0.074
	(0.070)	(0.115)	(0.085)
Math gendered	-0.025	-0.171	0.058
	(0.072)	(0.123)	(0.085)
Observations	753	365	388
R-squared	0.251	0.274	0.262

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A8. Comparing the gender gap in performance change in math and word search in Colombia.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Female	-0.153	-0.170	-0.229
	(0.181)	(0.192)	(0.170)
Math	-0.153	-0.0321	-0.067
	(0.174)	(0.202)	(0.244)
Female*Math	-0.125	-0.256	-0.305
	(0.254)	(0.271)	(0.270)
Individual performance		-0.263***	-0.258***
		(0.019)	(0.075)
Competitive performance		0.279***	0.280***
		(0.021)	(0.083)
Age		-0.011	-0.016
		(0.075)	(0.050)
Risk		0.006	0.006
		(0.026)	(0.027)
Importance winning female opponent			-0.018
			(0.016)
Importance winning male opponent			-0.016
			(0.018)
Word gendered			-0.027
			(0.023)
Math gendered			-0.031
			(0.026)
Constant	0.730***	0.566	1.126*
	(0.124)	(0.825)	(0.601)
Observations	556	357	350
R-squared	0.011	0.416	0.423

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A9. Comparing the gender gap in performance change in math and word search in Sweden.

VARIABLES	No controls (1)	Set 1 (2)	Set 2 (3)
Female	0.012	0.027	0.030
	(0.091)	(0.065)	(0.066)
Math	0.065	0.150**	0.138*
	(0.091)	(0.065)	(0.071)
Female*Math	0.208	-0.028	-0.012
	(0.128)	(0.093)	(0.091)
Individual performance		-0.145***	-0.146***
		(0.006)	(0.011)
Competitive performance		0.126***	0.126***
		(0.007)	(0.008)
Age		-0.011	-0.016
		(0.028)	(0.034)
Risk		0.060***	0.063**
		(0.016)	(0.026)
Importance winning female opponent			0.010
			(0.009)
Importance winning male opponent			-0.006
			(0.009)
Word gendered			0.002
			(0.016)
Math gendered			-0.017
			(0.015)
Constant	0.125**	0.144	0.244
	(0.0632)	(0.318)	(0.348)
Observations	577	447	439
R-squared	0.022	0.575	0.575

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A10. Competition choice math and word search, control variables set 1.

VARIABLES	All	Math	Word
	(1)	(2)	(3)
Sweden	0.142***	0.068	0.233***
	(0.049)	(0.070)	(0.070)
Female	0.076	0.021	0.135**
	(0.050)	(0.075)	(0.067)
Female*Sweden	-0.244***	-0.190**	-0.284***
	(0.065)	(0.094)	(0.090)
Individual performance	0.002	0.0029	-0.001
	(0.003)	(0.004)	(0.006)
Math	0.034		
	(0.033)		
Age	-0.024	-0.018	-0.020
	(0.018)	(0.028)	(0.025)
Risk	0.040***	0.055***	0.024**
	(0.008)	(0.012)	(0.011)
Performance beliefs	0.383***	0.448***	0.309***
	(0.073)	(0.103)	(0.105)
Observations	767	374	393
R-squared	0.099	0.144	0.071

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A11. Competition choice math and word search, control variables set 2.

VARIABLES	All	Math	Word
	(1)	(2)	(3)
Sweden	0.150***	0.078	0.233***
	(0.052)	(0.077)	(0.072)
Female	0.094*	0.041	0.143**
	(0.054)	(0.083)	(0.072)
Female*Sweden	-0.248***	-0.202**	-0.274***
	(0.068)	(0.102)	(0.095)
Individual performance	0.002	0.002	-0.001
	(0.003)	(0.004)	(0.006)
Math	0.040		
	(0.033)		
Age	-0.021	-0.012	-0.022
	(0.019)	(0.030)	(0.026)
Risk	0.039***	0.055***	0.029**
	(0.008)	(0.012)	(0.011)
Performance beliefs	0.381***	0.457***	0.298***
	(0.074)	(0.104)	(0.106)
Importance winning female opponent	-0.002	-0.002	0.000
	(0.005)	(0.008)	(0.007)
Importance winning male opponent	0.008	0.007	0.008
	(0.005)	(0.008)	(0.007)
Word gendered	0.004	0.002	0.005
	(0.009)	(0.014)	(0.012)
Math gendered	-0.002	0.003	-0.007
	(0.009)	(0.015)	(0.012)
Observations	753	364	389
R-squared	0.098	0.145	0.068

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A12. Comparing the gender gap in choice in math and word search in Colombia.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Female	0.027	0.106	0.119*
	(0.055)	(0.065)	(0.066)
Math	0.084	0.174**	0.176**
	(0.053)	(0.069)	(0.068)
Female*Math	-0.053	-0.183**	-0.193**
	(0.078)	(0.093)	(0.096)
Individual performance		-0.002	-0.001
		(0.007)	(0.006)
Competitive performance		0.013*	0.014*
		(0.007)	(0.007)
Age		-0.055**	-0.057**
		(0.026)	(0.027)
Risk		0.013	0.013
		(0.009)	(0.009)
Importance winning female opponent			-0.005
			(0.007)
Importance winning male opponent			-0.001
			(0.007)
Word gendered			0.000
			(0.012)
Math gendered			0.002
			(0.012)
Constant	0.265***	0.659**	0.698**
	(0.038)	(0.283)	(0.309)
Observations	560	362	355
R-squared	0.005	0.060	0.064

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A13. Comparing the gender gap in choice in math and word search in Sweden.

VARIABLES	No controls	Set 1	Set 2
	(1)	(2)	(3)
Female	-0.114**	-0.152**	-0.119*
	(0.054)	(0.060)	(0.063)
Math	0.058	-0.023	-0.007
	(0.054)	(0.059)	(0.065)
Female*Math	-0.140*	-0.026	-0.032
	(0.076)	(0.084)	(0.085)
Individual performance		-0.001	-0.001
		(0.006)	(0.006)
Competitive performance		0.012**	0.012*
		(0.006)	(0.006)
Age		-0.034	-0.034
		(0.025)	(0.026)
Risk		0.086***	0.089***
		(0.014)	(0.014)
Importance winning female opponent			-0.002
			(0.009)
Importance winning male opponent			0.018**
			(0.009)
Word gendered			0.025*
			(0.013)
Math gendered			-0.007
			(0.014)
Constant	0.385***	0.307	0.122
	(0.0376)	(0.289)	(0.316)
Observations	576	448	440
R-squared	0.045	0.140	0.154

Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table A14. Actual and expected performance in math and in word search in Colombia

Columbia	Math			SR	Word search			SR
	Actual rank	Belief	Difference	p-value	Actual rank	Belief	Difference	p-value
Girls	0.526	0.627	0.101	0.002	0.517	0.564	0.047	0.061
Boys	0.539	0.603	0.064	0.201	0.548	0.582	0.034	0.313
p-value			0.208				0.629	

Table A15. Actual and expected performance in math and in word search in Sweden

Sweden	Math			SR	Word search			SR
	Actual rank	Belief	Difference	p-value	Actual rank	Belief	Difference	p-value
Girls	0.559	0.545	-0.015	0.890	0.513	0.515	0.000	0.001
Boys	0.544	0.597	0.053	0.145	0.592	0.619	0.027	0.381
p-value			0.163				0.141	

Table A16. How gendered boys and girls perceive the tasks

	Colombia		RS	Sweden		RS
	Boys	Girls	p-value	Boys	Girls	p-value
Running	6.852	4.722	0.000	6.754	5.392	0.000
Skipping	3.833	2.663	0.000	3.487	3.220	0.212
Math	6.088	4.549	0.000	5.456	5.012	0.000
Word	5.741	4.205	0.000	4.860	4.230	0.000

4 In Bloom: Gender Differences in Preferences among Adolescents

Anna Dreber, Emma von Essen, and Eva Ranehill

We are grateful for comments from Johan Almenberg, Uri Gneezy, Magnus Johannesson, Christoph Mathys, Astri Muren, Robert Östling, David G. Rand and seminar participants at Harvard Kennedy School, MOVE Workshop on Gender Differences in Competitiveness and Risk Taking, Stockholm School of Economics and Stockholm University. We further want to thank Aron Backström and Peter Gerlach for help with the data collection. Financial support from the Jan Wallander and Tom Hedelius Foundation and the Carl Silfvén Foundation (E.R) is gratefully acknowledged, as well as financial support from the Swedish Council for Working Life and Social Research (FAS).

ABSTRACT: We look at gender differences in preferences for altruism, risk and competition in math and word search among adolescents in Sweden. We find that girls are more altruistic and less risk taking than boys. We find no gender gap in performance change when comparing performance under non-competition with performance under competition. Boys and girls are equally likely to choose to compete in word search, but boys are significantly more likely to choose to compete in math. However, this gender gap diminishes and becomes non significant when we control for relative performance beliefs, indicating that some of the gender gap in our sample is not due to competition preferences per se.

4.1 Introduction

Women today are in many countries at least as likely as men to pursue higher education and to participate in the labor market. Yet, the wage gap and the segregation in the labor market persist. One possible cause of gender differences in labor market outcomes is that men and women differ in terms of economic preferences. In economic studies, men are typically found to be less altruistic and more risk taking than women, as well as more competitive (see, e.g., Eckel and Grossman 2008a, 2008b, Croson and Gneezy 2009, Bertrand 2010, Engel 2010).

There is some evidence suggesting that gender differences in social preferences, risk preferences and competitiveness may contribute to explaining the gender gap in labor market outcomes (see e.g. Bertrand 2010 for further discussion). For example, Fortin (2008) finds that four different non-cognitive traits among young employees, including a suggested measure of competitiveness and a suggested measure of altruism, can explain some of the gender gap in wages. Individuals that display more competitiveness and less altruism seem to earn more. Some previous studies illustrate the connection between risk preferences and the labor market. Manning and Swaffield (2008) explain some of the gender gap in earnings with a set of psychological factors, including attitudes toward risk and competition.[50] Bonin et al. (2007) find that individuals who self-report that they are less willing to take risks also work in occupations with more stable earnings, which tend to pay less on average due to compensating wage differentials. Though

[50] However, human capital factors are the most important variable explaining the gender wage gap.

quantitatively small, a related effect is found in Manning and Saidi (2010) who find that there are fewer women in occupations and establishments that use variable pay instead of fixed pay contracts in Britain.[51] Flory et al. (2010) test the relationship between labor market choices and competitiveness directly in a large scale, randomized field experiment. They find that women are less likely than men to choose to apply to jobs with competitive compensation regimes.

In this paper we explore the gender gap in preferences among adolescents. Little is known about the development of the gender gap in economic preferences, and to what extent adolescents exhibit the same type of gender differences in preferences as adults do. If gender differences in preferences can explain part of the gender gap in labor market outcomes, it is relevant to also assess gender differences before individuals enter the labor market. Many important decisions that have implications for labor market outcomes, such as education choices, are taken during adolescence. For this reason we set out to study gender differences in economic preferences among individuals aged 16-18 years in Sweden. We focus on preferences for altruism and risk, as well as competition.

A recent meta-analysis of dictator game giving among adults finds that women are more altruistic than men (Engel 2010). There are by now a number of studies on gender differences in altruism among children and adolescents. As with the adult literature, some of these studies find that girls are more altruistic (Harbaugh et al. 2003, Gummerum et al. 2010) whereas other studies find no gender gap (Benenson et al. 2007, Blake and Rand 2010) and one recent study finds that girls are less altruistic (Fehr et al. 2011).[52] Moreover, when it comes to social preferences, adolescent girls are generally found to be more inequality averse and boys more efficiency concerned (Almås et al. 2010, Sutter et al. 2010, Fehr et al. 2011, Martins-

[51] Dohmen et al. (2011) also find that risk preferences elicited from lab measures have a significant and positive, albeit low, predictive power for labor market behavior.

[52] However, Blake and Rand (2010) find that girls are more likely to give something compared to nothing than boys.

son et al. forthcoming).[53] In order to study altruism among adolescents, we have subjects play a dictator game where the recipient is a charity.

We also explore the gender gap in risk preferences. Among adults, women are typically found to be less risk taking than men. Previous literature on children finds either no gender gap (Harbaugh et al. 2002), or that boys are more risk taking than girls (Borghans et al. 2009, Sutter et al. 2010, Cárdenas et al. forthcoming).[54] However, context or sample selection also seems to influence the gender gap in risk taking. Booth and Nolen (2009b) look at single sex and mixed schools and find that in this sample of children around 15 years old, boys are more risk taking than girls in mixed schools but that there is no gender gap when comparing boys to girls from single sex schools. Girls are also more risk taking when assigned to all-girl groups than when assigned to mixed groups. When it comes to measuring risk preferences in our study, we use two measures. The main measure consists of six choices where individuals choose between a lottery in the form of a coin flip that gives SEK 100 or 0 with equal probability and a safe option where the certain monetary amount increases successively in (from SEK 20 to 75).[55] We also use a survey question where individuals are asked to self-report their general risk taking propensity. This measure has been shown to correlate with both risk taking in an incentivized experiment and with gender (Dohmen et al. 2011).

Competitiveness is typically measured as either the change in performance in a competitive setting compared to a non-competitive setting, or as a preference for competition, such as self-selecting into a tournament instead of a piece-rate payment scheme. Previous studies have found that if there is a gender gap in any of these measures, men and boys are more competitive (Gneezy et al. 2003, Gneezy and Rustichini 2004a, Datta Gupta et al. 2005, Niederle and Vesterlund 2007, Sutter and Rützler 2010). However, the gender gap can be influenced by both the task performed and the

---

[53] Sutter et al. (2010) actually find that maximin preferences become more important with age for girls.

[54] Harbaugh et al. (2002) have the smallest sample of the aforementioned studies on risk taking (129 children aged 5-13 and 58 children aged 14-20).

[55] When conducting the study 7 SEK corresponded to about 1 USD.

sample in which competitiveness is studied. For example, Gneezy and Rustichini (2004b), Grosse and Riener (2010), Günther et al. (2010) and Shurchkov (forthcoming) find that the gap in competitiveness varies or can even be reversed depending on the task performed and the time constraint, whereas Wozniak et al. (2010) find no effect of tasks. Yet other studies find that the existence of a gender gap in competitiveness varies with the sample studied (Booth and Nolen 2009a, Gneezy et al. 2009, Andersen et al. 2010, Zhang 2010). For example, Gneezy and Rustichini (2004a) find that boys but not girls are competitive when it comes to performance change in running, whereas Dreber et al. (2009) and Cárdenas et al. (forthcoming) find no gender gap with this measure in the same task. In this paper we explore the role of both performance change and self-selection into a competitive setting, in two different tasks: math and word search.

We find that adolescent girls are more altruistic and less risk taking than adolescent boys. We find no gender difference in performance change under a competitive setting in comparison to a non-competitive setting, in either math or word search. Boys and girls are equally likely to choose to compete in word search, but boys are significantly more likely to choose to compete in math. However, the gender gap in choosing to compete in math diminishes and is no longer significant when controlling for relative performance beliefs. This indicates that among adolescents, the gender gap in competitiveness is not always present, and when it is, it may largely be due to other factors than a gender gap in preferences for competition per se. We study adolescents in Sweden, a country which typically scores high on indices of gender equality, thus to what extent our findings are generalizable to other countries remains to be explored.

The outline for our paper is the following. We present the experimental setup in section 2, and move on to our results in section 3. We finish by a discussion in section 4.

4.2 Experimental setup

The study was conducted in 9 school classes in five high schools in the Stockholm area during the fall of 2009. We contacted all schools in the cities of Stockholm, Uppsala and Västerås. Though we may have some selection regarding which schools that decided to participate, participation at the student level was compulsory.[56] The school classes include a mix of different specializations.[57] A total of 216 adolescents in grades 10-12 participated in the study.[58] 50% of the participants are female.

The experiment consisted of three parts conducted in the classroom, measuring competitiveness, altruism and risk preferences. The subjects first competed in math and word search, then participated in a dictator game and finally participated in a risk task. They were informed that each of the three parts consisted of a chance to earn money. One of the three parts would be randomly selected for payment, and the amount of money they could earn depended on the outcome of the choices they made in this part. After completing all parts the subjects were given a survey with additional questions.

The competition in the classroom consisted of two tasks, math exercises and word search, and each task consisted of three stages. The order of math and word search was randomly chosen between classes. The subjects did not get any feedback about their performance in any stage. In the first stage, a piece-rate scheme, the subjects were told that they had two minutes to solve as many exercises as possible, for which they would be given SEK 3 each. In the second stage, a tournament, the subjects were again told that they would get two minutes to solve exercises, but that they now would be randomly paired with three other individuals in the class who solved the same type of task, and that if they solved more or the same amount of exercises as these other individuals, they would get SEK 12 per exercise, whereas if they solved fewer exercises they would get SEK 0. In the third stage, the subjects were told that they were to solve exercises for another

---

[56] The result for the gender gaps reported does not differ between schools, tested in a regression frame work.

[57] The Swedish high school is optional and the students can choose programs with different specializations. Specialization does not seem to explain our results when controlled for in a regression frame work.

[58] Grade 10-12 represents the Swedish "gymnasium". Among the participants, 56 attended the 10th grade, 95 the 11th and 50 the 12th. 15 students attended a mixed class with students from grade 10 and 11. For these students we have no information about which grade they actually attended at the moment of the study.

two minutes, and that they now could choose whether they wanted to be given points according to the piece-rate scheme or the tournament (where they would again compete against three random other individuals in their class). Our measure of reaction to competition is the absolute change in performance between the first and second stages. The choice in the third stage gives us a measure of competitiveness as a preference for competition. After the competitiveness task was over, we asked the subjects to guess where in the performance distribution of their class they believed themselves to be, for both the piece-rate scheme and the forced competition. This allows us to measure performance beliefs, or over-/under-confidence.

Next the subjects took part in a dictator game, where they were asked to distribute 50 SEK between themselves and a well known charity organization.[59] They were informed that if this part was selected for payment the money they gave to the charity would be sent by us to the charity at the end of the study. The amount that the subjects give to the charity is our measure of altruistic behavior.

The last part was a risk task consisting of six choices where the subjects could choose between a lottery in the form of a coin flip that gives SEK 100 or 0 with equal probability and a safe option where the certain amount increases successively in points (from SEK 20 to 75). Our first measure of risk preferences relies on the unique switching point where the individual switches from preferring the lottery to preferring the safe option. This measure excludes inconsistent subjects, i.e. subjects with multiple switching points.[60] To further analyze risk preferences we include a survey question where the subjects are asked to self-report their general risk taking propensity on a scale from 0 to 10, where 10 is "very risk taking" and 0 is "not risk taking at all". This second measure of risk preferences is not incentivized.

[59] The name of the charity organization was the Swedish section of "Save the children".

[60] 14 of our subjects are inconsistent. We therefore also analyze risk taking as the number of times a person chooses the risky option compared to the safe, in order to have a measure that includes the inconsistent subjects. Using this measure of risk preferences in our analysis does not change our results. There is no gender difference in the proportions of inconsistent subjects (p=0.102).

After the three parts of the study were conducted, a survey was included in order to measure beliefs concerning the different tasks, as well as demographics.[61] In the end, one part was randomly selected for payment and the money was handed out in cash to the subjects.

To summarize; we analyze competitiveness as performance change in math and word search, and as choosing to compete or not in math and word search; altruistic behavior via a dictator game; and risk preferences through incentivized choices over lotteries and safe options as well as self-reported risk taking. We further look at additional measures such as relative performance beliefs.

4.3 Results

This section consists of three parts, where we test whether there is a gender gap in altruism, risk and competitiveness. All tests of the means throughout the paper are analyzed using the non-parametric Mann-Whitney test and a two-sided t-test. Only the p-values for the Mann-Whitney tests are displayed.[62] We start by studying gender differences in altruism, followed by an analysis of risk preferences. We then explore competitiveness in the two tasks using the two measures of competitiveness. When exploring competitive preferences we control for risk preferences and relative performance beliefs.

4.3.1 Altruism

Girls are significantly more altruistic than boys in our sample of adolescents (p=0.014). Girls give on average SEK 29 and boys SEK 23 out of SEK 50

[61] We collected a variety of demographic variables, but age is the only demographic variable used in this paper. The sample of this study is too small to use all demographics in the analysis of the present paper. We nevertheless chose to include these in the questionnaire for the purpose of future research studies.

[62] We present the Mann-Whitney test since none of our variables are normally distributed when using a skewness and kurtosis test. When there is a difference between the tests in terms of significance we also report the p-values for the t-test. We have also compared whether the distributions for each reported variable differ between boys and girls using a Kolmogorov-Smirnov test. The results are similar to those reported for mean values.

to the charity organization that is the recipient in our dictator game (see table 1).[63]

Table 1. Altruism.

	n	Average donation
Boys	107	23.20
Girls	109	29.32
p-value	-	0.014

4.3.2 Risk preferences

In this section we explore the gender gap in risk preferences measured from incentivized lotteries and self-reported non-incentivized risk taking. Analyzing the incentivized risk task we corroborate most previous findings that boys are more risk taking than girls. The average certainty equivalent to the lottery with equal probabilities of winning 100 and 0 is 45.2 for boys. For girls the certainty equivalent is significantly lower, 37.1 (p=0.002).[64] Our second measure of risk taking, self-reported risk propensity, supports this pattern.[65] On a scale from 0 to 10, where 0 is not risk taking at all and 10 is very risk taking, boys rated their average risk propensity to 6.15, whereas girls averaged on 5.59 (p=0.026).[66]

[63] In a regression analysis, the coefficient on gender is not influenced by the inclusion of the additional control variables age and grades. The regression analysis is based on parametric assumptions that may not be fulfilled. A correlation analysis between all the behaviors we examine also shows that altruism is positively related to risk taking in the incentivized risk task (p<0.001), but not in the self-reported question. We also find no correlation between altruism and competitive choices (p=0.255 for math and p=0.479 for word).

[64] The result is qualitatively similar when analyzing the number of risky choices instead of the switching point in order to include inconsistent individuals (those that switch back and forth between the lottery and the safe points). Girls are still less risk taking compared to boys (p=0.007).

[65] There is no gender difference in the variance of incentivized risk taking (p=0.210).

[66] Our two risk measures are significantly correlated (Spearman's rho=0.219, p=0.002).

Table 2. Risk preferences.

	n*	Average certainty equivalent	n	General risk
Boys	103	45.22	107	6.15
Girls	98	37.12	109	5.59
p-value	-	0.002	-	0.026

*One girl did not participate in this part and 14 participants made inconsistent choices.

4.3.3 Competitiveness

In this section we explore competitiveness as measured by absolute performance change and as the choice whether to compete or not. All participants took part in both the math and the word search exercises.[67] We also include an analysis where we control for relative performance beliefs.

4.3.3.1 Performance and choice

Table 3 compares the performance between boys and girls in the first stage (a piece-rate scheme) and the second stage (a tournament). Studying performance in each stage separately, boys perform significantly better than girls in math in both stages, whereas there is no gender difference in performance in word search.

When it comes to absolute performance change, our first measure of competitiveness, we find no increase in performance under the competitive compensation scheme for either gender. In contrast to most previous literature measuring performance change, neither boys nor girls react to the competitive environment by increasing their performance comparing the second and the first stage. As a robustness check, we also analyze the relative performance change.[68] This does not alter our results. Thus, there is no significant gender gap in competitiveness with this measure in either task.

---

[67] Randomly chosen, half of the classes performed the math task first and half performed the word task. A regression analysis suggests that the order of the tasks does not influence our results.

[68] Relative performance change is defined as ((performance in stage 2 – performance in stage 1)/performance in stage 1). We also conduct a quantile regression on absolute per-

Table 3. Average performance, stage 1 and 2.

	n	Math, stage 1	Math, stage 2	p-value	Word, stage 1	Word, stage 2	p-value
Boys	107	8.79	8.82	0.948	8.79	8.57	0.546
Girls	109*	7.31	7.44	0.510	8.74	8.61	0.542
p-value		0.010	0.020	-	0.524	0.952	-

*One girl had to leave the class room and did not participate in the first part of the word task.

When it comes to the choice of competing or not, we find a significant gender gap in math but not in word search although the point estimate goes in the same direction for both tasks (see table 4).[69] In math, 38 (36%) of the boys choose to compete compared to 18 (17%) of the girls (p=0.001). The corresponding numbers in the word task are 34 (33%) and 29 (28%) respectively (p=0.356).[70] The difference in gender gaps between the two tasks is mainly due to girls choosing differently across the two tasks: the share of girls choosing to compete in the word task is significantly larger than the share choosing to compete in the math task (p=0.050), whereas the proportion of boys competing is stable across the two tasks (p=0.701).

Table 4. Percentage choosing to compete in stage 3.

Task\Gender	n	% competing math	n	% competing word	p-value
Boys	106	0.358	105	0.333	0.701
Girls	109	0.165	109	0.275	0.050
p-value	-	0.001	-	0.356	

formance change and find no gender gap in math or word search in any part of the performance distribution.

69 One subject did not choose payment scheme for the third stage in math, and two did not perform in this stage. In the word task, two participants did not choose payment scheme. When possible, these individuals are included in the analysis. Including or excluding these participants has no effect on the results.

70 A sample size analysis indicates that 1978 observations would be needed to obtain a significant result for the performance change in running, jumping and dancing respectively. The basis for the power calculation is a significance level of 5% and a power of 80%.

4.3.3.2 Relative performance beliefs

Due to the gender gap in performance in the math task in each of the two stages, all or part of the observed gender gap may be due to subjects correctly anticipating their probability of winning the tournament should they choose to participate. We thus control for individual performance in the second stage in a regression analysis (see Table 6 below). When doing so, however, the gender coefficient remains significant.

Gender differences in competitive choices may also be due to gender differences in performance in stage 3, if participants correctly anticipate this. However, there is no significant difference in performance increase between boys and girls in either task (math: $p=0.450$, word: $p=0.749$), nor is there a difference in performance increase dividing the sample based on their competitive choice.

Individual risk preferences as well as relative performance beliefs have previously been found to influence competitive choices (Niederle och Vesterlund 2007, Niederle and Yestrumskas 2008). Girls in our sample who self-select into competition are significantly more risk taking than other girls in both math ($p=0.049$) and the word task ($p=0.004$). For boys, there is a significant difference in risk taking between those that compete and those that do not only in math ($p=0.009$). However, exploring the self-reported risk measure, the only significant difference is when comparing boys choosing to compete or not in math ($p=0.006$).

Table 5 below report the number of correct guesses regarding relative performance, divided by task and gender. Relative to their performance, we find that girls are underconfident in terms of their performance beliefs (Math: $p<0.001$: Word: $p<0.001$), whereas there is some evidence that boys are underconfident in math but not word search (Math: $p=0.065$: Word: $p=0.659$).[71] When we compare boys and girls, girls are significantly more

[71] A t-test indicates that boys are significantly underconfident in math ($p=0.041$). Our measure of over/underconfidence is the difference between relative performance beliefs and actual relative performance, both in terms of quartile in the performance distribution. When assigning individuals to a quartile for actual relative performance, we divide each separate class into four equal groups (roughly equal groups when the class size cannot be divided by four). In some cases several individuals performed equally across groups. Those individuals

underconfident in word search (p<0.001), and there is some evidence that girls are more underconfident in math (p=0.097). This is interesting given that most studies on college students find that both boys and girls are overconfident.[72]

Table 5. Distribution of guessed ranks.

	Men			Women		
	Guessed rank	**Over-confi-dent**	**Under-confi-dent**	**Guesse d rank**	**Over-confi-dent**	**Under-confi-dent**
Math						
1. Best	9 (5)	4	-	3 (2)	1	-
2.	23 (9)	8	6	16 (4)	8	4
3.	30 (10)	7	13	31 (5)	10	16
4. Worst	18 (6)	-	12	41 (17)	-	24
Total	80			91		
Word						
1. Best	11 (3)	8	-	1 (1)	0	-
2.	25 (7)	11	7	22 (6)	4	12
3.	30 (13)	8	9	46 (13)	9	24
4. Worst	14 (6)	-	8	21 (10)	-	11
Total	80			90		

*Number of correct guessed in parenthesis.

are given an expected quartile. For example, if four individuals perform similarly, and two needs to be assigned to the worst quartile and two to the second to worst quartile, these individuals all received the expected quartile 3.5.

[72] However, in a study of confidence in math performance among 14-year old children in Sweden, Dahlbom et al. (forthcoming) find that boys are overconfident and girls are underconfident. While we ask our subjects about retrospective performance, Dahlbom et al. (2010) ask the children about their expected performance on a math test that will be performed later.

Conducting an OLS regression analysis[73] analyzing the gender gap in competitive choices we perform four regressions per task, stepwise including control variables as can be seen in table 6 (math) and 7 (word) below.[74] We analyze the full sample of individuals, however 45 participants (two classes) were not asked to state their performance beliefs regarding stage 2 performance. We thus also analyze a limited sample excluding these individuals and those for whom we don't have all control variables. The results are very similar. In math, we find that controlling for actual performance and class mean diminishes the size of the observed gender gap with 11% in the restricted sample (comparing the coefficients in regression 1 and regression 3 in Table 6). When comparing regression 1 with regression 5 in the restricted sample, i.e. also adding controls for relative performance beliefs and risk preferences, we see that the gender difference in competitive choice in math is no longer significant. The point estimate of the female coefficient is lower, but still negative. The four control variables account for about 56% of the gender gap found in regression 1. Performance beliefs account for about 33% of the observed gender gap and risk preferences for about 12%. This can be compared to the results reported in Niederle & Vesterlund (2007), who find that 27% of the gender gap in tournament entry in their sample can be attributed to differences in relative performance beliefs.[75] The gender gap in tournament choice in the word task is not significant, independent of whether we control for performance or not. Performance beliefs and risk taking are, as in math, positively related to choosing to compete in the word task.

[73] See Appendix table 1 for the same analysis using a logit regression.

[74] When performing the regression analyses we also use a specification including variables from the short survey that was distributed after the experiment. These variables measured for example how gendered the participants found the tasks to be, and how important it was to win dependent on the gender of the opponent. None of these variables were significant.

[75] The setups of our study and that of Niederle and Vesterlund (2007) differ in that they had participants submit a piece-rate performance to the tournament, whereas in our case subjects performed again when choosing to compete. Moreover, Niederle and Vesterlund find that gender differences in risk preferences only have a negligible effect on the choice to compete.

Table 6. Math: OLS regression controlling for performance, beliefs and risk behavior

	Math (restricted sample)					Math (full sample)		
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Female	-0.191	-0.212	-0.170	-0.107	-0.084	-0.193	-0.206	-0.162
	(0.069)**	(0.073)**	(0.072)*	(0.070)	(0.068)	(0.059)**	(0.060)**	(0.059)**
Class mean		-0.029	-0.042	-0.033	-0.037		-0.025	-0.042
		(0.029)	(0.028)	(0.027)	(0.026)		(0.025)	(0.023)
Performance			0.036	0.019	0.016			0.038
			(0.008)**	(0.009)*	(0.010)			(0.008)**
Beliefs				0.156	0.154			
				(0.042)**	(0.042)**			
Risk					0.059			
					(0.023)*			
Constant	0.380	0.630	0.413	0.133	0.027	0.358	0.565	0.377
	(0.055)**	(0.262)*	(0.265)	(0.253)	(0.236)	(0.047)**	(0.214)**	(0.211)
R-squared	0.045	0.050	0.144	0.219	0.254	0.049	0.052	0.149
Observations	169	169	169	169	169	215	215	215

* Robust standard errors in parentheses. *$p<0.05$, **$p<0.01$. 45 participants were not asked to state their beliefs about second stage performance, one participant did not answer the risk question and one did not choose payment scheme. Performance and beliefs in the regression analysis is based on performance and relative performance beliefs in the second stage (the tournament). Since 15 participants were inconsistent in their choices in the risk task, the risk measure included here is the number of risky choices the participants make.

Table 7. Word: OLS regression controlling for performance, beliefs and risk behavior

	Word (restricted sample)					Word (full sample)		
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Female	-0.056	-0.059	-0.062	0.021	0.039	-0.058	-0.063	-0.067
	(0.073)	(0.080)	(0.079)	(0.075)	(0.075)	(0.063)	(0.065)	(0.065)
Class mean		-0.005	-0.013	-0.020	-0.044		-0.008	-0.014
		(0.038)	(0.038)	(0.034)	(0.034)		(0.027)	(0.028)
Performance			0.017	-0.005	-0.003			0.010
			(0.010)	(0.010)	(0.009)			(0.008)
Beliefs				0.230	0.220			
				(0.042)**	(0.041)**			
Risk					0.067			
					(0.022)**			
Constant	0.359	0.404	0.330	0.025	0.028	0.333	0.401	0.369
	(0.055)**	(0.333)	(0.330)	(0.303)	(0.303)	(0.046)**	(0.248)	(0.249)
R-squared	0.004	0.004	0.021	0.152	0.191	0.004	0.004	0.011
Observations	167	167	167	167	167	214	214	214

* Robust standard errors in parentheses. *p<0.05, **p<0.01. 45 participants were not asked to state their beliefs about second stage performance, one participant did not answer the risk question and two did not choose payment scheme.

4.3.4 Summary of results

In sum, we find that among adolescents, girls give more in a dictator game where the recipient is a charity, and that boys are more risk taking than girls. We find no gender gap in performance change when our subjects are forced to compete. Boys and girls are equally likely to choose to compete in word search, but boys are significantly more likely to choose to compete in math. However, this gender difference diminishes and becomes insignificant when controlling for relative performance beliefs.

4.4 Discussion

The gender gap in preferences has recently been suggested as an explanation for the often observed gender differences in labor market outcomes. Studies on children and adults suggest that if there is any gender gap in economic preferences, it is the most robust for altruism, risk and competition; girls and women are, more altruistic, less risk taking and less competitive than boys and men. To what extent these findings are consistent across contexts, countries and age is something that is currently being investigated in a number of projects.

In this study we systematically explore the gender gap in these preferences among adolescents in Sweden. Relatively little is known about the development of the gender gap in economic preferences, and to what extent adolescents exhibit the same type of gender differences in preferences as adults do. For example, it is generally acknowledged that the teenage years are associated with increased risk seeking (Boyer 2006). At the same time, many important decisions during this period have lifelong consequences, for example pertaining to education and professional choice. It is therefore interesting to study the development of economic preferences during this period. In this study, we find that adolescent girls are more altruistic and less risk taking than adolescent boys, corroborating the general findings on adults as well as some previous findings among adolescents. When it comes to competitiveness, we find no gender gap in performance change in either math or word search, comparing performance in a forced tournament to performance in a piece-rate scheme. We also find no gender gap in the choice whether to compete or not in word search. In math, boys are more likely to choose to compete than girls, a finding in line with other studies on children (Sutter and Rützler 2010) and adults (e.g. Niederle and Vesterlund 2007). However, once we control for performance beliefs, this gap diminishes and becomes non-significant, indicating that a large part of the observed gender gap in our sample does not depend on a gender difference in preference for competition per se. Support for this is also given by studies on adults, which shows that the gender gap in competitiveness can be eliminated by performance feedback (Wozniak et al. 2010), repetition (Cotton et al. 2009) or in environments where uncertainty is minimized (Niederle and Yestrumskas 2008). Why we find no gender gap in selection into word search competition is puzzling given that the gender gap in underconfidence was even stronger for word search than for math. This lends some support for the importance of tasks when studying competitiveness.

Our results, in relation to previous literature, thus suggest that the gender gap in risk taking, and perhaps also in altruism, emerges before adulthood. The results on competitiveness are less conclusive, and the potential development of this gender gap remains to be explored further. Moreover, we study adolescents in Sweden, a country that typically scores among the highest on gender equality indices. To what extent our results are generalizable to adolescents in other countries and settings is thus not clear. In sum, more research is needed in order to establish when and why gender differences in preferences arise, as well as their exact implications for the labor market.

4.5 References

Almås I., Cappelen A.W., Sørensen E.Ø., Tungodden B., 2010, "Fairness and the development of inequality acceptance", *Science*, 328: 1176-1178.

Andersen S., Ertac S., Gneezy U., List J., Maximiano S., 2010, "Age and gender differences in competitiveness: evidence from matrilineal and patriarchal societies", Mimeo.

Benenson J.F., Pascoe J., Radmore N., 2007, "Children's altruistic behavior in the dictator game", *Evolution and Human Behavior*, 28(3): 168-175.

Blake P.R. Rand D.G., 2010, "Currency value moderates equity preference among young children", *Evolution and Human Behavior*, 31: 210-218.

Bonin H., Dohmen T., Falk A., Huffman D., Sunde U., 2007, "Cross-sectional earnings risk and occupational sorting: the role of risk attitudes", *Labour Economics*, 14(6): 926-937.

Bolton G. Katok E., 1995, "An experimental test for gender differences in beneficent behavior", *Economics Letters*, 48: 287-292.

Booth A.L., Nolen P., 2009a, "Gender differences in risk behaviour: does nurture matter?", Mimeo.

Booth A.L. Nolen and P.J., 2009b, "Choosing to compete: how different are girls and boys?", IZA Discussion Paper 4027.

Borghans L, Golsteyn B., Heckman J.J., Meijers H., 2009, "Gender differences in risk aversion and ambiguity aversion", NBER Working Paper 14713.

Boyer T., 2006, "The development of risk-taking: a multi-perspective view", *Developmental Review*, 26: 291–345.

Cárdenas, J.C., Dreber, A., von Essen, E., Ranehill, E., 2010, ”Gender differences in competitiveness and risk taking: comparing children in Colombia and Sweden”, Forthcoming in *Journal of Economic Behavior and Organization*.

Charness G. Rabin M., 2002, “Understanding social preferences with simple tests”, *Quarterly Journal of Economics*, 117(3): 817-869.

Cotton C. McIntyre F., Price J., 2009, “Gender differences disappear with exposure to competition”, Working Paper 2010-11, University of Miami, Department of Economics.

Croson R. Gneezy U., 2009, “Gender differences in preferences”, *Journal of Economic Literature*, 47(2): 1-27.

Dahlbom L., Jakobsson A., Jakobsson N., Kotsadam A., “*Gender and overconfidence: are girls really overconfident?*”, Fourthcoming in *Applied Economics Letters*.

Datta Gupta N., Poulsen A., Villeval M.C., 2005, ”Male and female competitive behavior – experimental evidence”, IZA Discussion Paper 1833.

Dohmen T., Falk A., Huffman D., Sunde U., Schupp J., Wagner G.G., “Individual risk attitudes: measurement, determinants and behavioral consequences”, Forthcoming in*Journal of the European Economic Association*.

Dreber A., von Essen E., Ranehill E., “Outrunning the gender gap: boys and girls compete equally”, Fortcoming in *Experimetal Economics*.

Eckel C., Grossman P., 2008a, “Men, women and risk aversion: experimental evidence”, *Handbook on Experimental Economics Results*, Volume 1, Edited by C. Plott and V. Smith. Chapter 113, 1063-1071. New York, Elsevier.

Eckel C., Grossman P., 2008b, “Differences in the economic decisions of men and women: experimental evidence”, *Handbook on Experimental Economics Results*, Volume 1, Edited by C. Plott and V. Smith. Chapter 57, 509-519. New York, Elsevier.

Engel C., 2010, “Dictator games: a meta study”, Working Paper 2010/07, Max Planck Institute for Research of Collective Goods.

Fehr E., Rützler D., Sutter M., 2011, ”The development of egalitarianism, altruism, spite and parochialism in childhood and adolescence”, IZA Discussion Paper 5530.

Flory J., Leibbrandt A., List J., 2010, “Do competitive work places deter female workers? A large-scale natural field experiment on gender differences in job-entry decisions”, NBER Working Paper Series w16546.

Fortin N.M., 2008, “The gender wage gap among young adults in the United States”, *Journal of Human Resources*, 884-918.

Gneezy U., Leonard K., List J., 2009, “Gender differences in competition: evidence from a matrilineal and a patriarchal society”, *Econometrica*, 77(5): 1637-1664.

Gneezy U., Niederle M., Rustichini A., 2003, “Performance in competitive environments: gender differences”, *Quarterly Journal of Economics*, 118(3): 1049-1074.

Gneezy U., Rustichini A., 2004a, “Gender and competition at a young age”, *American Economic Review Papers and Proceedings*, 94(2): 377-381.

Gneezy U., Rustichini A., 2004b, “Executives verses teachers: gender, competition and self selection”, Mimeo.

Grosse N. D., Reiner G., 2010, “Explaining gender differences in competitiveness: gender-task stereotypes”, Jena Economic Research Papers 2010 – 017.

Gummerum M., Hanoch Y., Keller M., Parsons K., Hummel A., 2010, “Preschoolers‘ allocations in the dictator game: the role of moral emotions”, *Journal of Economic Psychology*, 31(1): 25-34.

Günther C., Arslan Ekinici N., Schwieren C., Strobel M., 2009, "Women can't jump? – An experiment on competitive attitudes and stereotype threat", *Journal of Economic Behavior and Organization*, 75(3): 395-401.

Harbaugh W.T., Krause K.,. Vesterlund L., 2002, "Risk attitudes of children and adults: choices over small and large probability gains and losses", *Experimental Economics*, 5(1): 53–84.

Harbaugh W.T., Krause K., Liday, S., 2003, "Bargaining by children", Working Paper, University of Oregon.

Manning A., Farzad S., 2010, "Understanding the gender pay gap: what's competition got to do with it?", *Industrial and Labor Relations Review*, 63(4): 681-698.

Manning A., Swaffield J., 2008, "The gender gap in early-career wage growth", *Economic Journal*, 118: 983-1024.

Martinsson P., Nordblom K., Rützler D., Sutter M., 2011, "Social preferences during childhood and the role of gender and age – an experiment in Austria and Sweden", *Economics Letters*, 110(3), 248-251.

Niederle M., Vesterlund L., 2007, "Do women shy away from competition? do men compete too much?", *Quarterly Journal of Economics*, 122(3): 1067-1101.

Niederle M., Yestrumskas A.H., 2008, "Gender differences in seeking challenges: the role of institutions", Mimeo.

Shurchkov O., 2011, "Under pressure: gender differences in output quality and quantity under competition and time constraints", Forthcoming in *Journal of the European Economic Association*.

Sutter M., Rützler D., 2010, "Gender differences in competition emerge early in life", IZA Discussion Paper 5015.

Sutter M., Kocher M., Rützler D., Trautmann S.T., 2010, ”Impatience and uncertainty: experimental decisions predict adolescents' field behavior”, IZA Discussion Paper 5404.

Wozniak D., Harbaugh W.T., Mayr U., 2010, “Choices about competition: differences by gender and hormonal fluctuations, and the role of relative performance feedback”, MPRA Paper 21097.

Zhang J., 2010, “Do girls in china compete just as much as boys? evidence from an experiment that predicts educational choice”, Mimeo.

4.6 Appendix

Table A1. Math: Logit regression controlling for performance beliefs and risk preferences.

	Math (restricted sample)					Math (full sample)		
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Female	-0.967	-1.083	-0.963	-0.679	-0.574	-1.039	-1.108	-0.977
	(0.356)**	(0.383)**	(0.411)*	(0.426)	(0.433)	(0.329)**	(0.338)**	(0.358)**
Classmean		-0.152	-0.264	-0.209	-0.258		-0.134	-0.281
		(0.157)	(0.170)	(0.175)	(0.175)		(0.136)	(0.149)
Performance			0.200	0.121	0.101			0.220
			(0.053)**	(0.058)*	(0.056)			(0.051)**
Beliefs				0.863	0.939			
				(0.247)**	(0.269)**			
Risk					0.429			
					(0.169)*			
Constant	-0.491	0.817	-0.068	-1.871	-2.791	-0.582	0.537	-0.243
	(0.233)*	(1.367)	(1.535)	(1.651)	(1.603)	(0.203)**	(1.152)	(1.273)
Observations	169	169	169	169	169	215	215	215

* Robust standard errors in parentheses. *p<0.05, **p<0.01. 45 participants were never asked to state beliefs about second stage performance, one participant did not answer the risk question and one did not choose payment scheme.

Table A2. Word: Logit regression controlling for performance beliefs and risk preferences.

	Word (restricted sample)					Word (full sample)		
Variables	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Female	-0.967	-1.083	-0.963	-0.679	-0.574	-1.039	-1.108	-0.977
	(0.356)**	(0.383)**	(0.411)*	(0.426)	(0.433)	(0.329)**	(0.338)**	(0.358)**
Classmean		-0.152	-0.264	-0.209	-0.258		-0.134	-0.281
		(0.157)	(0.170)	(0.175)	(0.175)		(0.136)	(0.149)
Performance			0.200	0.121	0.101			0.220
			(0.053)**	(0.058)*	(0.056)			(0.051)**
Beliefs				0.863	0.939			
				(0.247)**	(0.269)**			
Risk					0.429			
					(0.169)*			
Constant	-0.491	0.817	-0.068	-1.871	-2.791	-0.582	0.537	-0.243
	(0.233)*	(1.367)	(1.535)	(1.651)	(1.603)	(0.203)**	(1.152)	(1.273)
Observations	167	167	167	167	167	214	214	214

* Robust standard errors in parentheses. *p<0.05, **p<0.01. 45 participants were never asked to state beliefs about second stage performance, one participant did not answer the risk question and two did not choose payment scheme.

5 The Effect of Competition on Physical Activity: A Randomized Trial

Magnus Johannesson, Robert Östling and Eva Ranehill

The authors thank Mia Hultberg, Anders Raustorp and three anonymous referees for helpful comments and Hanna Christiansson for editorial assistance. Financial support from the Jan Wallander and Tom Hedelius Foundation, the Swedish Research Council, and the Swedish Council for Working Life and Social Research is gratefully acknowledged

ABSTRACT: A recent literature in economics has highlighted that competition and symbolic awards can provide non-monetary incentives. In this paper, we report on a step contest that we carried out at a large Swedish workplace in order to test whether competition for symbolic awards can be used to promote physical exercise. Each individual was equipped with a pedometer and registered the number of steps daily during a four week period. Participants competed both in teams and individually and the winning team and individual received symbolic prizes. To evaluate the effect of the competition per se, we randomized teams into a control group and two treatment groups. We found that the step contest significantly increased both the fraction of subjects that completed the step contest and the number of steps. The number of steps was about 1,000 steps higher in the main treatment group than in the control group (an increase by about 10 percent). This is a conservative estimate as the dropouts on average walked fewer steps than individuals completing the study. In an additional treatment, we also included a daily step goal in the contest. The step goal had no additional significant effect on the number of steps, which may be due to the relatively low step goal used (7,000 steps per day).

5.1 Introduction

Given the increasing trend in obesity in Western countries (Cutler et al. 2003), it is of central importance to increase physical exercise. Both physical exercise and obesity are important determinants of health (Brown 1992; Miller et al. 1997; Lee et al. 1999; Speck and Looney 2001; Alevizos 2005; Guh et al. 2009; Whitlock et al. 2009). Physical exercise has both a direct effect on health, as well as an indirect effect through affecting important risk factors of diseases like obesity, blood pressure, and cholesterol levels (Brown 1992; Miller et al. 1997; Lee et al. 1999; Speck and Looney 2001; Alevizos 2005). The economic costs to society of obesity and inactivity are substantial (Keeler et al. 1999; Cox et al. 1998; Colditz 1999; Fiebelkorn et al. 2003; Branca et al. 2007). It is therefore important to test new and innovative approaches to promote exercise. A recent study by Charness and Gneezy (2009) showed that physical exercise can be increased by paying individuals to go to the gym, and another study showed some success in using cash incentives to reduce smoking (Volpp et al. 2009). The evidence is more mixed regarding financial rewards as a means of reducing obesity (Finkelstein et al. 2007; Volpp et al. 2008; Cawley and Price 2007). In this study we tested another incentive to increase exercise and health: a contest to win a symbolic award. We carried out a step contest at a major Swedish hospital in which employees, equipped with pedometers, competed (both individually and in teams) to win a symbolic prize.

Recent research in economics has shown that both competition per se and symbolic awards can act as non-monetary incentives. The introduction of a competitive element (for a given level of monetary incentives) has been shown to increase performance in a number of economic experiments performed on students and children (Gneezy, Niederle and Rustichini 2003; Gneezy and Rustichini 2004; Niederle and Vesterlund 2007; Booth and Nolen 2009; Niederle, Segal and Vesterlund 2010; Dreber, von Essen and Ranehill 2010; Sutter and Rützler 2010). These studies measure the impact of competition in a one-shot experimental setting, mainly using tasks such as solving mazes, math exercises, or running short distances (less than 100 meters). We contribute to this literature by testing if a competitive setting can also promote exercise. Compared to the above studies we also study behavior during a longer period and in a more representative sample.

In our experiment participants compete for a symbolic prize. Using the terminology of Frey (2006, 2007), the symbolic prize is a form of award, and Frey refers to such incentives as "non-material, extrinsic compensation".[76] One argument for why awards may improve performance is that it signals relative position, which is valuable if individuals care about their rank compared to other individuals (Besley and Ghatak 2008). There is a sizeable literature in economics on status and positional goods, see for instance Hirsch (1978) and Frank (1985). Another related argument is that awards are associated with increased social esteem that is directly valued by individuals (Frey 2006, 2007). For recent models of social esteem in economics, see Bénabou and Tirole (2006), Ellingsen and Johannesson (2008) and Andreoni and Bernheim (2009).[77] The widespread use of awards suggests that they are effective in enhancing performance. Although there is only limited empirical work testing this proposition, three recent studies suggest that awards can increase contributions to public goods (Neckerman and Frey 2007), increase the supply of blood donations (Lacetera and Macis

[76] As examples of awards Frey (2006) for instance includes Academy Awards (Oscars), prizes in sports competitions, the Pulitzer Prize, and awards in academia such as Nobel Prizes.
[77] Ariely et al. (2009) provides a recent empirical test of the social esteem model.

2008) and increase worker performance (Kosfeld and Neckermann 2010).[78] We contribute to this line of empirical research by testing whether competition for a symbolic prize can be used to increase physical exercise.

Pedometer use is a well-established way to increase physical activity.[79] A recent systematic review of studies testing the effects of using a pedometer showed that according to the randomized clinical trials of pedometer use, the use of a pedometer on average increased the number of steps per day by about 2,500; an increase by about 30 percent over the baseline activity level (Bravata et al. 2007). The corresponding effect based on observational studies was an increase by 2,183 steps per day. The effect of pedometer use was especially pronounced if a step goal was used, e.g. a step goal of at least 10,000 steps per day. According to the study by Bravata et al. (2007), the use of a pedometer also led to a significant reduction in body mass index (BMI) and systolic blood pressure.

The step contest studied in this paper was not invented by us, but is a common practice at many workplaces in Sweden. Out of a labor force of about 4.5 million, about 150,000 individuals participate in step contests every year.[80] To what extent these step contests increase physical activity is thus an important public health issue in Sweden. In a typical step contest the participants form teams and compete both individually and in teams. The participants in a step contest are equipped with a pedometer and then register the number of steps on a homepage created for the contest; on this homepage they can also see the number of steps of all other participants as well as their position in the contest. The winner is the individual and the team with the highest number of steps during the period of the contest

[78] See also the related work in psychology and management science about the effects of social recognition on performance (Stajkovic and Luthans 2003).

[79] The review article by Tudor-Locke et al. (2002) compares pedometers with other measures of physical activity and find strong correlations verifying that steps measured with a pedometer is a valid measure of physical activity.

[80] This is an estimate based on talking to some of the providers of step contests in Sweden (as the companies are unwilling to reveal their exact market size it is difficult to get the exact number).

(normally four weeks).[81] The reward is typically a symbolic award like a cup.

Step contests are a way of trying to promote increased exercise among the employees, but we know little about the actual effects on exercise and health. The studies evaluating the effects of the use of pedometers suggest that using a pedometer per se can have substantial positive effects (Bravata et al. 2007). In the present study we tested, in a randomized field experiment, if the competitive aspect of step contests has an additional positive effect on exercise.[82] Compared to the previous clinical trials included in the Bravata et al. (2007) review our study is the first to test the importance of symbolic rewards and competition. The study also investigates a "real world" intervention in a realistic setting.

The study was carried out at a major hospital in Sweden, which is similar to other workplaces that organize steps contests (a previous step contest had for instance been carried out at this very same hospital). Three experimental treatments were included. To be able to compare the effects of the contest per se we included a control group in the study in which individuals were equipped with a pedometer and registered their steps in the same way as in a step contest, but without being able to see the number of steps by other participants or their position in the contest. The second experimental group was a standard step contest, with the only difference compared to the control group being that individuals could see the number of steps of other participants and that the winning individual and team received a cup at the end of the four week period. The treatment we are looking at is the combination of an individual contest and a team contest to win a symbolic award. Individuals form teams which have a common goal and receive information about the number of steps both within and between teams to generate a competitive environment. Individuals in the control group share no information within or across teams. We also included a third experimental group that added a step goal to the experimental treatment. In this treatment the individuals had to walk at least 7,000 steps per

---

81 The winner of the team competition is the team with the highest number of steps per team member.

82 In the terminology used by Harrison and List (2004) the study is a framed field experiment.

day to participate in a random draw of prizes during the study (everyone in the other two groups were also randomly allocated prizes irrespective of their number of steps). With this third experimental group we tested the hypothesis that adding a step goal to a step contest will further increase physical activity. We chose a relatively modest step goal in order to provide the least physically active participants with an additional incentive to walk more.

Our design isolates the effect of the step contest per se, i.e. the effect over and above using a pedometer to monitor physical activity. A limitation of our design is that we cannot separate the effect of the symbolic reward (the award) from other aspects of the competition. Our estimate does also not separate the effect of the individual and the team contest, but accounts for the joint effect of these. With more experimental treatments it would have been possible to disentangle the effects of individual competition from team competition, but that was not within the scope of the present study. In principle a team contest could be either more or less effective than an individual contest. The team contest introduces possibilities to free ride which could reduce performance, but the effect of peer pressure would go in the other direction.

The design of the study is presented below, followed by a section about the hypotheses and statistical tests. Thereafter we present the results. The paper ends with a discussion of our findings.

5.2 Design of study

The study was carried out during the spring of 2009 at a major hospital in Sweden (Södersjukhuset with about 4,200 employees); we refer to Södersjukhuset as "the hospital" below. The step contest was handled by Select Wellness AB who handles a large number of step contests in Sweden (they have carried out step contests for more than 300,000 individuals).[83] In previous step contests organized by Select Wellnes AB the average number of steps has been 11,357 for hospitals and 10,674 for other employers. Al-

[83] Select Wellness AB charge the employers SEK 365 (approximately $50 based on the exchange rate at the time of the study) per participant in a step contest (including the cost of the pedometers). It is thus a relatively low cost intervention to increase physical exercise.

though the number of steps among hospital employees has been slightly higher than for average participants, these numbers still suggest that hospitals are relatively representative for the workplaces that typically participate in step contests.

The hospital offered all their employees to participate in the study and all employees received information about the study. The employees were told that the step contest was part of a scientific study conducted to evaluate the effects of step contests. They were also told that teams would be randomized to one of three groups and the three groups were briefly described. The information also stated that the study was voluntary and that individuals could discontinue the study at any point in time.[84]

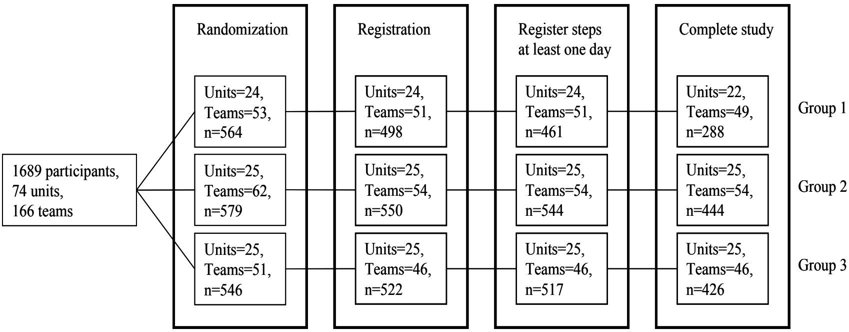

Figure 1. Design of the study

The project setup is illustrated in Figure 1. Teams consisting of 5–15 individuals could sign up for the study (individuals self select into teams). At the end of the entry period 166 teams with in total 1689 individuals had signed up for the study. This means that about 40 percent of the employees at the hospital signed up for the study. The population participating in the study is a much more unselected population than the patient populations typically taking part in the clinical trials on pedometer use included in the overview by Bravata et al. (2007). However, this does not imply

[84] The study was approved by the Regional Ethics Committee in Stockholm.

that the participants are representative of the general population in Sweden, nor of the employees at the hospital (the 40 percent selecting into the study can differ from the 60 percent not choosing to participate).

The teams were divided into 74 randomization units by the hospital so that teams in the same work unit were in the same randomization unit, i.e. a randomization unit consisted of all the teams in the same work unit (as defined by the hospital). This was done to make sure that everyone in the same work unit would face the same incentives and it was also a prerequisite by the hospital to carry out the study. The randomization was stratified with respect to the number of individuals in each randomization unit and the 74 randomization units were thereafter randomly allocated into the following three experimental groups:[85]

Group 1 (control group): Each participant received a pedometer and registered their number of steps daily during the four weeks of the study. The participants could not see the number of steps by other team members or other participants in the study.[86] The participants could win the prizes that were randomly allocated to all participants in the study.

Group 2 (the contest treatment group): Each participant received a pedometer and daily registered the number of steps during the four weeks of the study. The participant could observe the number of steps of all other participants, as well as see his/her position in the contest and the position of his/her team. Prizes were randomly allocated to all participants in the study. A symbolic prize (a cup) was given to the individual with the most steps and the team with most steps per team member.

Group 3 (the contest with a step goal): This treatment was the same as in group 2, with the difference that only participants with an average number of steps over 7,000 per day could win the prizes that were randomly allocated.[87]

---

[85] There were 44 randomization units with 22 or fewer individuals in each. Out of these, 14 were randomly selected to group 1 and 15 to group 2 and 3. Of the remaining 30 units with 23 or more individuals, 10 were randomly selected into each group.

[86] The steps of the control group were never revealed to any subject in any of the groups during the study.

[87] 7,000 steps corresponds to a distance of about five kilometers if each step is counted as 75 centimeter (which is a standard conversion used).

After the randomization all team leaders (each team had to have a team leader) received entry packages for all participants in the team and allocated one package to each team member. The package consisted of brief information about the study and the group to which the individual had been randomized, a pedometer, as well as information about how to register on the home page for the contest.[88] At the registration individuals were asked to fill out a questionnaire with some background information. They were also asked to register the number of steps they took on a daily basis.[89] After the four weeks the individuals were asked to fill out another questionnaire.[90]

Some prizes were randomly allocated among all participants (in group 3 only participants with over 7,000 steps per day could win prizes). Three weekly draws were conducted with 75 "winners" each time; each individual in these draws won a lottery ticket worth SEK 25.[91] After the end of the study, three SEK 1,000 gift certificates were also randomly allocated among all participants (but in group 3 only participants with over 7,000 steps per day could win).

The entry questionnaire consisted of a number of background questions (age, sex, education, marital status, if they had children, if they worked full time or not and if they had a desk job or not), questions on exercise habits (the hours of exercise per week in a typical week, the hours

---

[88] Participants received a step diary where they could write down the number of daily steps and they also registered the number of steps at a homepage. They were encouraged to register their steps every day on the home page, but it was also for instance possible to enter the homepage once a week and then register the steps from the step diary for each day in that week. There were no explicit instructions for what to do if a participant failed to use the pedometer a day. The pedometer is for measuring the activity of walking. For other physical activities like running individuals were told to convert 30 minutes of the activity to 2,700 steps.

[89] The number of daily steps was not verified by anyone; it was completely based on self-reported number of steps (which was also the standard in the clinical trials included in the Bravata et al. 2007 overview).

[90] It would have been ideal to continue to follow-up the participants after the end of the four week intervention period. Unfortunately this was not possible, and we have no data on the number of steps or other measures of physical activity or health after the end of the four week intervention period.

[91] The exchange rate at the time of the study was approximately $1 = SEK 7.5.

of exercise the previous week, and their self-assessed physical activity level compared to other individuals (below average, average, or above average), as well as questions on health status (length and weight to measure BMI; self-assessed health status: poor, fair, good, very good, or excellent).[92] The exit questionnaire only included questions about exercise habits and health status (but not height).

5.3 Hypotheses and tests

We tested two main hypotheses with this design. The first hypothesis was that the competition aspect of the step contest increases the levels of physical activity, i.e. that the number of steps would be higher in group 2 than in group 1. The second hypothesis was that the inclusion of a step goal would further increase the activity level, i.e. that the number of steps would be higher in group 3 than in group 2. As a secondary hypothesis we tested if the step contest led to a reduction in self-assessed BMI as an indicator of improved health.[93] It should be noted that self-assessed BMI is a noisy measure of overweight. It would have been preferable to measure BMI by measuring weight and height in the study. BMI as such is also an imperfect measure of overweight as muscle mass and bone structure can differ between individuals, and exercise can also lead to an increase in muscle mass. It is also unlikely to be able to detect a change in weight during just a four week intervention period. These limitations should be borne in mind with respect to our secondary hypothesis concerning BMI, but it is still of interest to report these results.

As some individuals dropped out of the study we evaluated our hypotheses both based on the individuals that completed the study (referred to as "completers" below) and on all the possible information about

---

[92] All employees at the hospital, including both nurses, physicians and other workers, were invited to participate in the study. Unfortunately, we did not collect data about occupation in the study apart from the variables about desk job or not and if the individuals worked full time or not, so we cannot directly control for occupation in the analyses (though occupation is likely to be correlated with years of education that we do control for).

[93] The BMI change was based only on changes in the reported weight during the study; the height was only measured in the beginning of the study.

dropouts (i.e. participants were included with the number of steps they walked before they dropped out; referred to as "all subjects" below).[94]

For "completers" the statistical tests were based on an OLS regression with dummy variables for the three experimental groups and the daily average number of steps as the dependent variable. To account for that the observations within teams and units may be correlated with each other, we estimated the OLS regression with clustered standard errors at the level of the randomization units (Wooldridge 2003). We carried out these tests with and without controlling for the background characteristics collected in the entry questionnaire (and to test if the background variables differed significantly between the groups we also used OLS regressions using clustered standard errors). The same statistical methodology was also used to test our secondary hypothesis about the effects on BMI as measured in the questionnaire responses (based on all individuals that filled out both the entry and the exit questionnaire).

In the estimation for "all subjects" we estimated the number of steps per individual during each day of the competition including all subjects that were still in the study; e.g. an individual that dropped out on day 11 would be included in the estimate of the number of steps per individual in days 1-10, but not in days 11-28. We did this estimation with and without controlling for background characteristics (when we controlled for background characteristics we ran a separate OLS regression for every day of the contest with the group dummy variables and the background characteristics as explanatory variables). As the "all subjects" sample (due to the dropouts) differed on different days in the study, it is not straightforward to carry out standard statistical tests of the difference in mean steps between

[94] An individual with zero steps in a day was defined as having dropped out of the study if the individual had not entered steps at any subsequent day. The dropout day is the day after the last day that an individual reported steps. If an individual for instance reported steps for days 1–10, but not for the subsequent days, the dropout day is day 11. Some individuals did not register steps for every day in the period, and any blank days were counted as zero steps during that day (for some individuals it was evident that they summed the steps for several days, e.g. a week, and entered the sum for a single day; but it was impossible to separate these observations from true missing values).

the groups. To carry out the statistical tests we therefore used bootstrap methods (Efron and Tibshirani 1993).[95]

Comparing the results for completers with the results for "all subjects" indicates in what direction dropouts biased the results (i.e. if dropouts prior to dropping out walked more or less steps than non-dropouts). As a further check of the impact of dropouts on our results we therefore compared the number of steps between dropouts and non-dropouts. We constructed two measures for each individual that dropped out. In the first, we estimated the difference between the average number of steps per day prior to dropping out and the average number of steps per day in the same period for the individuals that did not drop out at that day in that group (e.g. if an individual in group 3 that dropped out by day 11 had walked 11,500 steps per day in days 1-10 and the average number of steps in days 1-10 in group 3 for individuals that did not drop out on day 11 was 12,000 steps, this measure was equal to –500 (11,500 – 12,000)). This measure shows if individuals that dropped out on a specific day had walked fewer steps up until that day compared to individuals in their group that had not yet dropped out of the study. The second measure is similar, but with the difference that the steps per day prior to dropping out was now compared with the steps per day for completers in the same period (e.g. if an individual in group 3 that dropped out on day 11 had walked 11,500 steps per day in days 1-10 and the average number of steps in days 1-10 in group 3 for individuals that completed the study was 12,500 steps, this measure was equal to –1,000 (11,500 – 12,500)). This measure shows if individuals that dropped out on a specific day had walked fewer steps up until that day compared to individuals in their group that completed the study.

---

[95] The p-values were based on generating 1,099 bootstrap replications, which according to Davidson and MacKinnon (2000) should yield a relatively high precision in the estimated *p*-values. The reason for choosing numbers of replications ending with 99 is that Monte Carlo tests are exact only if $p(N + 1)$ is an integer, where N is the number of replications and p is the significance level (Davidson and MacKinnon 2000). In each bootstrap replication we randomly drew 74 randomization units with replacement and based on each draw we calculated the differences in mean number of steps between the three groups. The reported p-values were calculated under the assumption that the bootstrapped estimates were approximately normal (using the standard bootstrap procedure in STATA 10), but the p-values calculated directly from the 1,099 replications were highly similar.

These two measures give an indication of if individuals dropping out of the study were walking fewer or more steps than those that continued in the study. To statistically test if dropouts and non-dropouts differed in the number of steps based on the two above measures, we ran OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

5.4 Results

5.4.1 Subjects and dropouts

As can be seen in Figure 1, 1,689 individuals were initially randomized in the study. Out of these individuals 1,570 (93 percent) registered for the study and filled out the entry questionnaire. The dropout between randomization and registration was somewhat higher in group 1 (11.7 percent), than in groups 2 and 3 (5.0 percent in group 2 and 4.4 percent in group 3). Unfortunately we have no information about these individuals as the questionnaire was filled out at registration. The background characteristics of those individuals who filled out the questionnaire are shown in Table 1. As expected with the randomized design, the groups were similar in terms of background characteristics. In only four of the 33 pairwise comparisons were the differences statistically significant at the five percent level (there is a higher rate of married/cohabiting individuals in group 3 than in group 1 or 2; the hours of exercise are lower in group 3 than in group 2; and the self-assessed physical activity level is higher in group 1 than in group 3).

For the individuals that registered for the study we have at least some information about the number of steps for 97 percent of the individuals; i.e. they filled in the number of steps for at least one day. Also at this stage the dropout rate was larger in group 1 (7.4 percent of those who registered never reported any information about steps) than in groups 2 and 3 (1.1 percent in group 2 and 1.0 percent in group 3).[96]

[96] A possible explanation for the higher dropout rate in the control group between randomization and registration and between registration and registering any steps is that subjects became disappointed when they found out that they would not actually be part of a "real" contest. If the pattern in terms of the number of steps follows the pattern for subjects dropping out later in the study shown below, this effect would go towards underestimating the treatment effects.

Table 1. Background characteristics.*

	Group 1 (n=498)	Group 2 (n=550)	Group 3 (n=522)	p-value of difference**		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
Age (years)	46.29 (10.41)	44.31 (10.89)	43.97 (11.68)	0.130	0.133	0.836
Women (%)	86	89	90	0.434	0.330	0.783
Married/Cohabitant (%)	72	68	77	0.125	0.030	<0.001
Children (%)	73	70	72	0.487	0.888	0.627
Education (years)	14.27 (2.04)	14.31 (1.95)	14.57 (1.88)	0.827	0.062	0.157
Fulltime work (%)	71	77	68	0.279	0.630	0.101
Desk job (%)	33	31	30	0.856	0.768	0.887
BMI	24.36 (4.07)	24.17 (4.44)	24.46 (6.60)	0.547	0.795	0.468
Exercise (hours per week)	2.03 (1.47)	2.13 (1.47)	1.87 (1.41)	0.328	0.145	0.021
Self-assessed physi-				0.654	0.042	0.082
Less active than others (%)	26	24	30			
As active as others (%)	41	46	43			
More active than others (%)	33	30	27			
Self-assessed health status:				0.865	0.243	0.355
Poor or fair (%)	16	16	16			
Good (%)	43	43	46			
Very good (%)	31	29	28			
Excellent (%)	11	12	9			

* Standard deviations of continuous variables in parentheses.
** The *p*-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

The dropout rate during the study is shown in Figure 2 for the individuals that registered steps during the first day. It is evident that the dropout rate during the study was much higher in the control group than in the two other groups. In group 1, 37.5 percent of the individuals that started the study dropped out during the study, whereas these fractions were 18.4 percent in group 2 and 17.6 percent in group 3. The difference between group 1 and the other two groups was highly significant (the p-values are below 0.01 in both the comparison between group 1 and group 2 and in the comparison between group 1 and 3).[97] This means that in terms of the initially randomized individuals only 51.1 percent completed the study in the control group whereas 76.7 percent completed the study in group 2 and 78.0 percent completed the study in group 3. From these results it is evident that one important effect of the step contest was that it substantially increased the likelihood of completing the study; i.e. to continue using the pedometer and registering steps. Given that the use of a pedometer per se has been shown to increase physical activity and health in other studies (Bravata et al. 2007), this effect is important in itself.

We also tested if the probability of dropping out of the study was related to the observed characteristics in the study by running a linear probability model (with clustering on the randomization units as in the other regression models). In the regression we included all the individuals that registered in the study and filled out the entry questionnaire with background information. Only two of the observed characteristics were significantly related to the probability of dropping out at the five percent level; higher age and having a desk job significantly decreased the probability of dropping out of the study.[98]

5.4.2 Steps

Figure 3 shows the number of steps during the study in each of the three randomized groups using data for all individuals that were still in the study on each day (i.e. the “all subjects” sample). Figure 3 indicates that there are

---

[97] This statistical test as most of the other statistical comparisons across groups in the paper were based on an OLS regression with dummy variables for the experimental groups using clustered standard errors at the level of randomization units.

[98] The coefficient of age was –0.0065 and the coefficient of desk job was –0.074.

some day-of-the-week effects, with a drop in the number of steps during the weekend.[99] This indicates higher physical activity during weekdays, which is probably due to working per se involving physical activity (both at the workplace and when getting to and from work).[100] This pattern could be expected to vary between occupations.

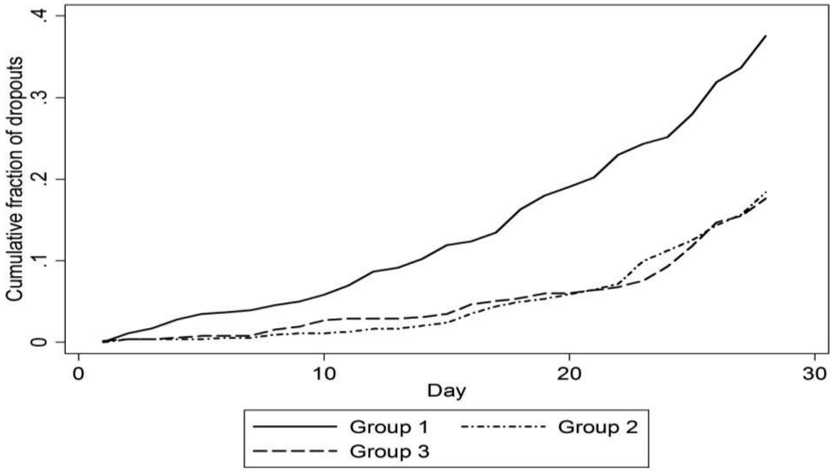

Figure 2. Dropouts during the study in each treatment group (estimated as a fraction of the subjects that registered steps on the first day).

---

[99] The day-of-the-week effects introduces some additional noise in the daily step measures, but we have refrained from controlling for these effects in our figures to show the actual pattern of steps across the four weeks of the study. Our estimate of the number of steps per day is based on the average number per day and will average out any day-of-the-week effects.

[100] One alternative explanation is that participants do not report steps during weekends. However, the day-of-the-week effects are highly similar also for those participants that reported positive number of steps during all days of the study.

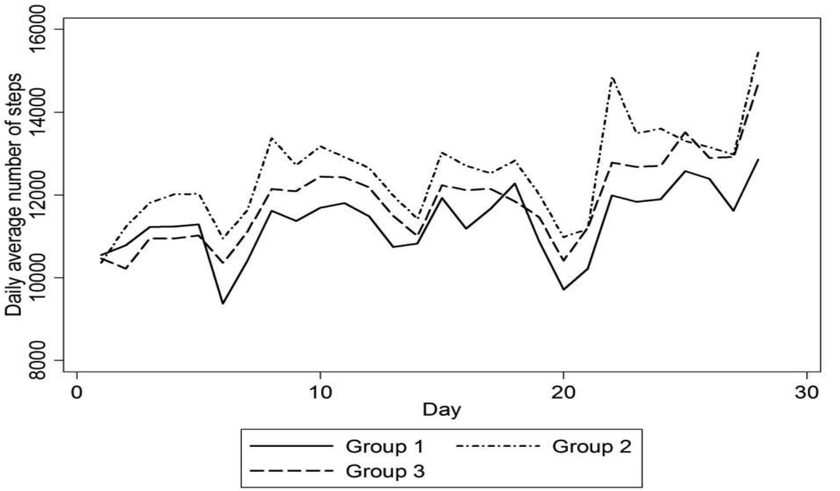

Figure 3. The number of steps in each treatment group during the study. The data includes all subjects that were still in the study each day (i.e. including those that subsequently dropped out).

Consistent with the first hypothesis, the number of steps was higher in group 2 than in the control group during the study. The average number of steps per day was 11,337 in the control group and 12,512 in group 2. The increase of 1,175 steps per day was statistically significant at the one percent level; see the results in Table 2. However, we failed to find support for our second hypothesis that a step goal would further increase physical activity. The number of steps in group 3, 11,872, was in between the two other groups and did not differ significantly from either of the two other groups.

In order to control for any differences in background characteristics between the groups we estimated the results controlling for the background characteristics collected in the study (as shown in Table 2 and in Figure 4). This reduced the difference in steps between group 1 and group 2 somewhat to 908 steps, but this difference was still significant at the one percent level. The difference between group 1 and group 3 increased somewhat to 643 steps, whereas the difference between groups 2 and 3 decreased to 265 steps.

Table 2. Step per day in the three experimental groups.

	Group 1	Group 2	Group 3	Difference in steps (p-value)**		
				Group 2 vs 1	Group 3 vs 1	Group 2 vs 3
All subjects*:						
No control for background variables	11337	12512	11872	1175 (0.008)	536 (0.131)	640 (0.199)
Control for background variables	11337	12244	11980	908 (0.003)	643 (0.027)	265 (0.482)
Subjects that completed the study:						
No control for background characteristics	11649	12701	12062	1052 (0.017)	413 (0.266)	639 (0.196)
Control for background characteritics	11649	12381	12150	731 (0.024)	501 (0.087)	230 (0.543)

* "All subjects" means that the estimate each day was based on all subjects that were still in the study that day.

** The p-values were estimated using bootstrap methods for the first two rows, whereas the p-values in the last two rows were based on OLS regressions with dummy variables for the three groups using clustered standard errors (at the level of randomization units).

An alternative way of analyzing the data is to only compare the number of steps between the subjects that completed the study. This is done in Figure 5 (without controlling for background factors) and in Figure 6 (controlling for background factors). This led to similar results as in the "all subjects" sample, although the difference between groups 1 and 2 decreased somewhat to 1,052 steps (without controlling for background characteristics) and 731 steps (with a control for background factors). As can be seen from Table 2, the difference was significant at the five percent level in both cases. The differences between the other groups were not significant at the five percent level.

Figure 5 and 6 indicate that there is an upward trend in the number of steps over time. We estimated a linear time trend in all groups. The time trend was significant (at the one percent level) in all groups with a point estimate of 46 steps per day in group 1, 82 steps per day in group 2, and 90 steps in group 3.

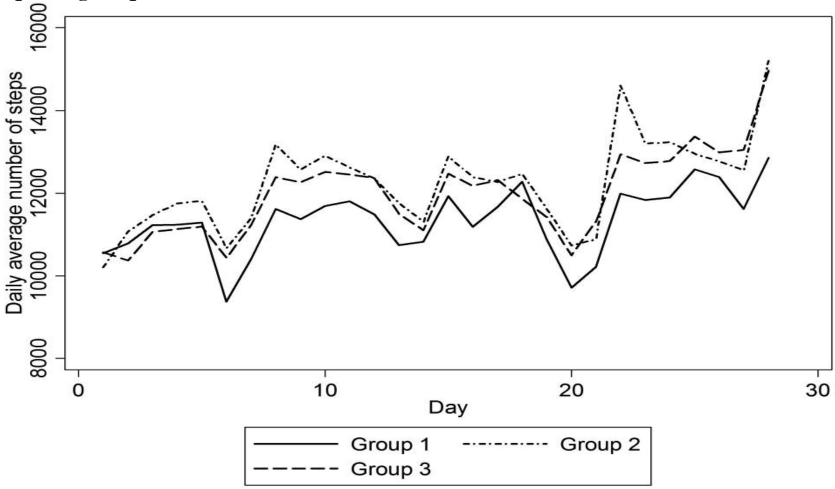

Figure 4. The number of steps in each treatment group during the study controlling for background characteristics. The data includes all subjects that were still in the study each day (i.e. including those that subsequently dropped out).

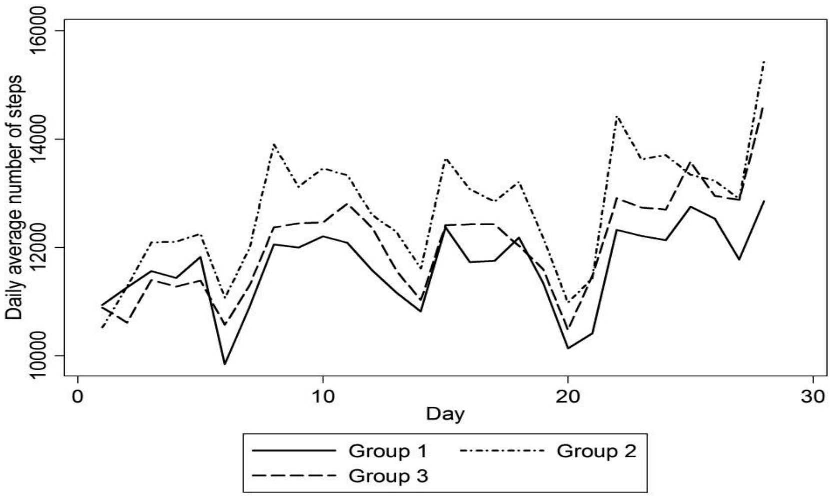

Figure 5. The number of steps in each treatment group during the study. The data includes only subjects that completed the study.

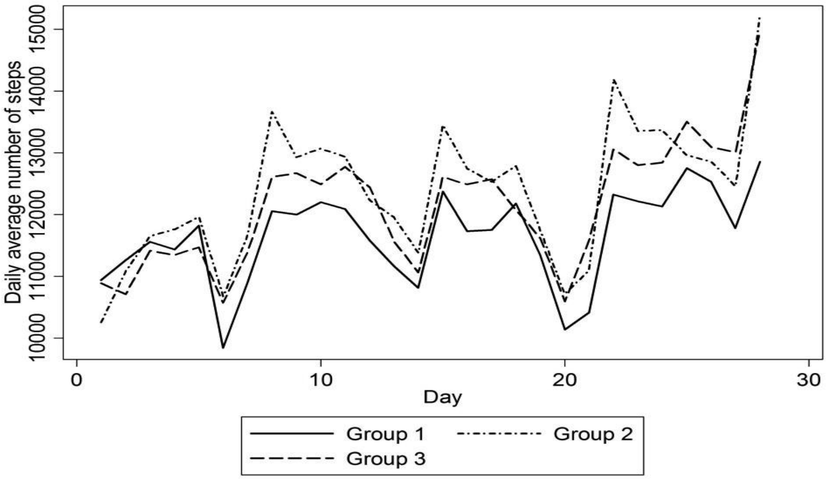

Figure 6. The number of steps in each treatment group during the study controlling for background characteristics. The data includes only subjects that completed the study.

The time trend was also significantly greater (at the five percent level) in groups 2 and 3 than in group 1, but did not differ significantly between groups 2 and 3. This implies that the difference in the number of daily steps between the two step contest groups and the control group increased over time in the study. A possible explanation for this is that the competition aspect spurs an increasing trend over time in groups 2 and 3 as individuals can observe the number of steps of other participants and teams. That there is a positive time trend also in the control group is consistent with the meta-analysis by Bravata et al. (2007); using a pedometer per se has a sizeable effect on the number of steps. It is likely that this increase occurs gradually over time after starting to use the pedometer.

Table 3 shows the full regression results for completers with all the background characteristics (corresponding to the last row of Table 2). Several background characteristics were significantly related to the number of steps reported during the study. Older and more educated participants took significantly fewer steps, as did participants with desk jobs and a high BMI. Full time workers, healthier and more physically active subjects as well as participants that exercise more tended to take more steps.

Table 3. OLS regression of the effects of background characteristics on the number of steps per day for subjects that completed the study.

Variable	Coefficient	Standard error*	p-value
Group 1 (baseline)			
Group 2	731.23	317.30	0.024
Group 3	500.84	288.15	0.087
Age (years)	-37.02	12.78	0.005
Woman	219.12	414.43	0.599
Married/Cohabitant	-441.62	268.80	0.105
Children	198.83	245.63	0.421
Education (years)	-182.26	60.11	0.003
Full time work	510.95	204.04	0.015
Desk job	-1011.08	237.37	0.000
BMI	-34.74	17.71	0.054
Exercise (hours per week)	807.22	89.26	0.000
Self-assessed physical activity:			
Less active than others (baseline)			
As active as others	597.62	290.32	0.043
More active than others	1023.10	343.24	0.004
Self-assessed health status:			
Poor or fair (baseline)			
Good	149.54	275.62	0.589
Very good	663.12	313.21	0.038
Excellent	263.15	377.57	0.488

*The standard errors were estimated with clustering at the level of randomization units.

We also carried out a sub-group analysis to investigate the effect of step contests in different sub-groups. These results are shown in Table 4. In one analysis we divided the subjects into desk workers and non-desk workers. In a second analysis we divided the sample into two BMI groups: above 25 (the definition of being overweight) and below 25. Finally, we divided the groups into different exercise levels: less than two hours per week and more than two hours per week. The effect of the step contest is weaker for desk workers, presumably because it is more difficult for desk workers to influence how many steps they take at work. The step contest seems to have a stronger effect for overweight people, but a weaker effect for people that exercise less than two hours per week. However, none of these differences between sub-groups are close to being statistically significant.[101]

Table 4. Step per day in different sub-groups (estimated for all subjects with controls for background variables).*

	Group 1	Group 2	Group 3	Difference in steps (p-value)**		
				Group 2 vs 1	Group 3 vs 1	Group 2 vs 3
Worker type:						
Desk workers	10787	11315	11411	528 (0.118)	623 (0.185)	-96 (0.820)
Non desk workers	11638	12744	12322	1106 (0.009)	684 (0.046)	423 (0.395)
BMI:						
BMI > 25	10356	11729	11123	1373 (0.005)	767 (0.042)	607 (0.276)
BMI ≤ 25	11957	12565	12481	608 (0.052)	523 (0.121)	84 (0.839)
Exercise:						
≤ 2 hour per week	10502	11113	10932	611 (0.112)	429 (0.159)	181 (0.673)
> 2 hour per week	12417	13638	13362	1221 (0.003)	945 (0.014)	276 (0.589)

* ”All subjects” means that the estimate each day was based on all subjects that were still in the study that day.
** The p-values were estimated using bootstrap methods.

[101] For each of the three sub-group analyses we compared the three estimates for the treatment effects in the two sub-samples using a standard Z test based on the bootstrapped estimates of the standard deviations. The lowest among the nine different p-values was 0.18.

Table 5. Comparison of steps between dropouts and non-dropouts.*

	Group 1	Group 2	Group 3	p-value of difference***		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
Dropouts vs non-drop-outs*	-738 (0.013)	-1192 (0.039)	-1697 (<0.001)	0.477	0.053	0.464
Dropouts vs com-pleters**	-1020 (0.001)	-1327 (0.022)	-1815 (<0.001)	0.629	0.109	0.482

* p-values within parentheses.
** Based on a comparison of the number of steps per day before the dropout day for subjects that dropped out a specific day versus subjects that did not drop out that day (but may have dropped out later in the study).
*** Based on a comparison of the number of steps per day before the dropout day for subjects that dropped out a specific day versus subjects that completed the study.
**** The p-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

The fact that the effect of the step contest decreased somewhat when we used "completers" instead of "all subjects", suggests that individuals that dropped out of the study were less physically active than other individuals. To analyze this issue further we compared the steps between "dropouts" and "non-dropouts". The data is summarized in Table 5. The dropouts on average walked fewer steps than the individuals that continued the study. The dropouts in group 1 walked 738 steps less per day compared to the non-dropouts. For groups 2 and 3, the dropouts walked 1,192 and 1,697 steps less, respectively. If non-dropouts were defined as completers these differences increased somewhat. Dropouts now walked 1,020 steps less per day in group 1 and the corresponding figures in groups 2 and 3 were 1,327 and 1,815, respectively.[102] We also tested whether the difference

[102] There was no significant difference between the groups in these estimates, but the point estimates goes in the direction of a larger difference between dropouts and non-dropouts in the step contest groups. It is possible that since there are benefits for people to stay in the step contest groups in terms of the potential to win the contest, only those individuals that

in the number of steps between dropouts and completers changed over time, but we could not reject the null hypothesis that this difference remained the same over time.[103]

That dropouts walked fewer steps than non-dropouts suggests that we have underestimated the effect of the step contest, as the rate of dropouts was substantially higher in group 1. The results in Table 5 can be used to make a rough adjustment to the step per day results in Table 2 for completers by imputing steps per day for individuals dropping out.[104] The adjustment for dropouts would imply that the difference in steps per day between group 1 and 2 would increase by about 200 steps and the difference between groups 1 and 3 would increase by about 100 steps and the difference between groups 2 and 3 would increase by about 100 steps. The effects on the estimates resulting from adjustment for dropouts would thus be relatively small, but they suggest that our estimates in Table 2 on the effect of the step contest (group 2 vs group 1) are conservative (lower bounds).

There is likely to be measurement error in the number of reported steps and the noise could potentially differ between the three treatment groups. For example, measurement errors could be larger in group 1 because participants in this group did not participate in the contest and therefore might have cared less about reporting accurately. To get an indication of whether measurement errors differed across groups, we calculated the standard deviation and coefficient of variation of the number of daily reported steps for each participant that completed the study. Both the stan-

---

really hate it will drop out. The benefits of staying in the study for the control group are lower, so the marginal person who drops out may be someone who is more active.

103 We ran a linear regression on all subjects dropping out testing if the effect (the difference in steps between dropouts and completers) was significantly related to either a linear time trend or a dummy variable for the second half of the study (controlling for the treatment groups). The time variable was not significant in either of these regressions ($p = 0.92$ with a linear time trend and $p = 0.69$ with a dummy variable for the second half of the study).

104 This can be done by imputing steps per day for dropouts based on the difference between dropouts and non-dropouts in Table 5 (i.e. individuals in group 1 that dropped out were assigned 1020 steps less per day during the study than completers in group 1; individuals in group 2 that dropped out were assigned 1327 steps less per day during the study than completers in group 2; and individuals in group 3 that dropped out were assigned 1815 steps less per day during the study than completers in group 3).

dard deviation and the coefficient of variation is the lowest in group 1, but the only difference between groups that is statistically significant at the 10 percent level is that the standard deviation is lower in group 1 compared to group 2 (the p-value is 0.01).[105]

5.4.3 BMI

In the questionnaire we included questions about height and weight to measure BMI. It should be emphasized that this data is based on self-reported data, which is an imperfect measure of actual BMI (Cawley 2000). In Table 6 we include data on the change in BMI during the study; this data is only for the individuals that completed the study and filled out the questionnaire at the end of the study. We also included data for two questions about exercise in Table 6. The first question was about the hours of exercise in a normal week and the second question was about hours of exercise in the preceding week. To measure a change in exercise due to the study, the second question is more appropriate. The question is phrased in terms of the hours of exercise that lead to a high pulse rate and sweating. As we can see from Table 6, there was no significant difference between the groups in the change in BMI or exercise. However, there was a tendency in all groups towards increased exercise and decreased BMI. When outliers were excluded the decrease in BMI during the study was significant in all groups. However, as the BMI measure was based on self-reported data it has to be interpreted very cautiously. To use a self-reported BMI measure may be particularly worrying in this case as people knew that they were part of an effort to get people to be more active and exercise more.

---

105 To test if the differences were statistically significant we ran OLS regressions with each measure of variation for each individual as the dependent variable and dummy variables for the three groups as explanatory variables. We then tested whether the group dummy estimates differed based on the clustered standard errors (at the level of randomization units).

Table 6. Change in BMI and exercise.*

	Group 1	Group 2	Group 3	p-value of difference**		
				Group 1 vs 2	Group 1 vs 3	Group 2 vs 3
Change in BMI	-0.022 (0.583)	-0.145 (0.237)	-0.144 (0.012)	0.337	0.077	0.994
Change in BMI (excluding >10 kg changes)	-0.064 (0.015)	-0.096 (<0.001)	-0.104 (0.004)	0.369	0.347	0.853
Change in BMI (excluding >5 kg changes)	-0.063 (0.009)	-0.068 (0.004)	-0.091 (0.016)	0.864	0.518	0.602
Change in exercise, hours per week	0.141 (0.087)	0.181 (<0.001)	0.233 (<0.001)	0.673	0.332	0.451
Change in exercise, hours the last week	0.468 (<0.001)	0.534 (<0.001)	0.540 (<0.001)	0.609	0.533	0.964

* p-values within parentheses. All changes were measured as the value at the end of the study minus the value at the beginning of the study and only data for subjects that completed the study were included. One participant reported 900 kg at the second occasion and was dropped in the comparison of BMI.

** These p-values were based on OLS regressions with dummy variables for the three experimental groups using clustered standard errors at the level of randomization units.

5.5 Discussion

Recent work has shown that cash incentives have the potential to increase physical exercise and promote health (Charness and Gneezy 2009). Our study complements those findings by showing that contests with symbolic rewards can also serve as an incentive to exercise. The study was designed to answer two main hypotheses about the effects of step contests. The first was if the competitive aspect per se increased the number of steps and thereby physical activity. We found support for this hypothesis. Our point estimates in our different estimations ranged between an effect of 731 and 1,175 steps depending on whether we controlled for background characteristics or not and depending on whether we incorporated any information about the steps of dropouts. If we included the available information about steps for dropouts the point estimates ranged between 908 and 1,175 steps per day.

The second hypothesis that we tested was whether adding a step goal would further increase the effect of the step contest. We failed to find support for this hypothesis. The point estimate of the number of steps in group 3 (with the 7,000 step per day step goal) was actually lower than in group 2, although the difference was not statistically significant. This is in contrast to the results of the meta-analysis by Bravata et al. (2007) who found that using a 10,000 step per day step goal significantly increased the number of steps. The reason for the lack of an effect of the step goal in our study is possibly that the step goal of 7,000 steps per day was too low; if we pool our data for all our three experimental groups only 6.3 percent of the subjects walked less than 7,000 steps per day on average.[106] At the same time it should be noted that the average baseline number of steps in the meta-analysis by Bravata et al. (2007) was 7,473. It is possible that the low step goal signaled to the participants that once they had reached this modest goal they had exercised sufficiently. Experimental work in psychology and economics suggest that reference points and framing may have important behavioral effects (Tversky and Kahneman 1974, 1981; McNeil et al. 1982; Ariely et al. 2003). The step goal provided a relatively low reference point and may therefore have decreased physical activity. To investigate if there was bunching of steps at or just above 7,000 steps in group 3, we plotted the distributions for the average number of steps during each week of the study (as the step goal was tied to the weekly lotteries), but there was no sign that the step goal caused people to report steps close to 7,000.[107] Our reason for the low step goal was that we wanted to try to provide an incentive for the most inactive individuals to increase their physical activity. An important lesson from our results is that if a step goal is used it should probably be set at a relatively high level like 10,000 steps per day.

---

[106] 1,425 of the 1,521 (93.7 percent) individuals who filled in any steps averaged above 7,000 steps and 1,110 of the 1,158 (95.9 percent) individuals that completed the study averaged above 7,000 steps.

[107] The fraction of individuals with 7,000 to 7,500 steps per day in group 3 was 4.69 percent in the first week, 1.88 percent in the second, 2.58 percent in the third and 1.64 percent in the final week of the study. The corresponding proportions in group 1 and 2 were slightly lower in the first week, but higher in the following three weeks.

Our estimates have to be interpreted cautiously due to the high rate of dropouts, especially in the control group. As a fraction of the individuals initially randomized to the study, the rate of dropout was almost 50 percent in the control group and about 20 percent in the two treatment groups. In their systematic review, Bravata et al. (2007) reported that nine out of the 26 studies on pedometer use had a zero dropout rate and that the average dropout rate in the remaining 17 studies was 20 percent. Those studies were, however, typically carried out on much more selected populations with small sample sizes (110 participants on average). Our study included a relatively unselected population compared to previous clinical trials (Bravata et al. 2007); out of the total number of employees at the hospital about 40 percent registered for the study. Our sample size of 1,689 is also much larger than the average sample size of 38 in the previous clinical trials (Bravata et al. 2007).

Our analysis of dropouts suggests that individuals that did not complete the study on average walked fewer steps than individuals that did. This suggests that our estimates of the effect of the step contest are lower bounds of the true effect. It should also be borne in mind that when we compared the estimates between dropouts and non-dropouts, it was based on the number of steps using a pedometer. It is likely that those individuals that dropped out of the study stopped using the pedometer and this may in itself have had a negative effect on physical activity (Bravata 2007). One important effect of the step contest was therefore that it greatly increased the likelihood that the individuals would continue in the study and thereby continue using the pedometer.

A limitation of our study is that it is based on the number of self-reported steps, i.e. that individuals honestly report their steps. We cannot directly verify that steps are honestly reported, but we investigated if the number of steps correlated with our measures of BMI, exercise (hours per week), and self-assessed physical activity relative to other individuals.[108] We found a significant correlation for all these measures, suggesting that the

[108] BMI and exercise were measured both at the beginning and at the end of the study, and self-assessed physical activity at the beginning of the study.

number of steps is related to physical activity.[109] The results of the meta-analysis by Bravata et al. (2007) also showed that the 30 percent increase in the number of self-reported steps from using a pedometer was associated with significant health improvements (a reduction in BMI and systolic blood pressure). This suggests that the number of self-reported steps is a valid and reliable measure of exercise. A difference compared to our study is that having a contest with a symbolic reward may introduce an additional incentive to misrepresent the number of steps. However, a growing literature in experimental economics suggest that many individuals have a psychic cost of lying (Ellingsen and Johannesson 2004; Gneezy 2005; Lundquist et al. 2009), which may counteract the incentive to lie about the number of steps.

The intervention in our study, the step contest, only lasted four weeks. This is a brief intervention period for a physical activity intervention and a limitation of our study. It is especially difficult to detect any effects on health measures such as BMI after such a short intervention period. It would thus have been ideal with a longer intervention period. Even if the intervention as such, the step contest, only lasted four weeks it would have been interesting to follow-up the physical activity and the BMI level in the groups after the end of the study. Charness and Gneezy (2009) for instance also studied a physical intervention that lasted for only one month, but continued to follow up the health effects after the end of the intervention period and detected significant improvements in BMI and other health measures four months after the end of the intervention period.

Our estimated effects of the step contest suggest that the step contest led to an increase in about 1,000 steps per day. To interpret the size of this effect it can be compared with the average effect of pedometer use in the meta-analysis by Bravata et al. (2007). They reported a mean increase of 2,491 steps for the 26 studies included in their overview. The effect of the step contest according to our study is thus about 40 percent of the effect of

---

109 The Pearson correlations between the average number of steps for completers and BMI before and after the study were -0.12 and -0.10 ($p < 0.001$). The Pearson correlations between number of steps and numbers of hours exercised where 0.38 (both before and after the study, $p < 0.001$). The Spearman correlation between the number of steps and self-assessed physical activity was 0.30 ($p < 0.001$).

using the pedometer per se. This is a sizeable additional effect on physical activity. The fraction of individuals that completed the study was also about 25 percentage units higher in the step contest groups compared to the control group. This is an important result in itself given that using a pedometer per se appears to be associated with a sizeable increase in physical activity (Bravata et al. 2007). Taken together these results suggest that step contests can potentially achieve important improvements in physical activity and health and be a useful public health tool. The potential for this as a desirable public health tool is furthered by the low cost per participant (≈$50). However, as in all studies using pedometers, it is unclear what the long term effects are on physical activity and health and this is important to investigate further.

5.6 References

Alevizos A., Lentzas J., Kokkoris S., Mariolis A., Korantzopoulos P., 2005, "Physical activity and stroke risk", *International Journal of Clinical Practice* 59, 922–930.

Andreoni J., Bernheim B. D., 2009, "Social image and the 50-50 norm: a theoretical and experimental analysis of audience effects", *Econometrica* 77, 1607–1636.

Ariely D., Bracha A., Meier S., 2009, "Doing good or doing well? Image motivation and monetary incentives in behaving prosocially", *American Economic Review* 99, 544–555.

Ariely D., Loewenstein G., Prelec D., 2003, "'Coherent arbitrariness': stable demand curves without stable preferences", *Quarterly Journal of Economics* 118, 73–105.

Bénabou R., Tirole J., 2006, "Incentives and prosocial behavior", *American Economic Review* 96, 1652–1678.

Besley T., Ghatak M., 2008, "Status incentives", *American Economic Review Papers and Proceedings* 98, 206–211.

Booth A. L., Nolen P., 2009, "Choosing to compete: how different are girls and boys?". University of Essex, Discussion Paper Series, No. 673.

Branca F., Nikogosian H., Lobstein T. (eds.), 2007, *The challenge of obesity in the WHO European Region and the strategies for response*. Copenhagen: WHO Regional Office for Europe.

Bravata D. M., Smith-Spangler C., Sundaram V., Gienger A. L., Lin N., Lewis R., Stave C. D., Olkin I., Sirard J. R., 2007, "Using pedometers to increase physical activity and improve health", *Journal of the American Medical Association* 298, 2296–2304.

Brown D. R., 1992, "Physical activity, ageing, and psychological well-being: an overview of the research", *Canadian Journal of Sport Science* 17, 185–193.

Cawley J., 2000, "Body weight and women's labor market outcomes", NBER Working Paper 7841.

Cawley J., Price J. A., 2009, "Outcomes in a program that offers financial rewards for weight loss", NBER Working Paper 14987.

Charness G., Gneezy U., 2009, "Incentives to exercise", *Econometrica* 77, 909–931.

Colditz G. A., 1999, "Economic costs of obesity and inactivity", *Medicine & Science in Sports & Exercise* 31, S663–S667.

Cox E.R., Kortt M.A., Langley P.C., 1998, "A review of cost-of-illness studies on obesity", *Clinical Therapeutics* 20, 772–779.

Cutler D. M., Glaeser E. L., Shapiro J. M., 2003, "Why have Americans become more obese?", *Journal of Economic Perspectives* 17, 93–118.

Davidson R., MacKinnon J. G., 2000, "Bootstrap tests: how many bootstraps?", *Econometric Reviews* 19, 55–68.

Dreber A., von Essen E., Ranehill E., "Outrunning the gender gap: boys and girls compete equally", Fortcoming in *Experimetal Economics*

.Efron B., Tibshirani R. J., 1993, "An introduction to the Bootstrap", *Monographs on Statistics and Applied Probability* 57. New York: Chapman and Hall.
Ellingsen T., Johannesson M., 2004. "Promises, threats and fairness", *Economic Journal* 114, 397–420.

Ellingsen T., Johannesson M., 2008, "Pride and prejudice: the human side of incentive theory", *American Economic Review* 98, 990–1008.

Fiebelkorn I. C., Finkelstein E. A., Wang G., 2003, "National medical spending attributable to overweight and obesity: how much, and who's paying?", *Health Affairs* W3, 219–226.

Finkelstein E. A., Linnan L. A., Tate D. F., Birken B. E., 2007, "A pilot study testing the effect of different levels of financial incentives on weight loss among overweight employees", *Journal of Occupational and Environmental Medicine* 49, 981–989.

Frank R. H., 1985, *Choosing the right pond: human behavior and the quest for status*. New York: Oxford University Press.

Frey B. S., 2006, "Giving and receiving awards", *Perspectives on Psychological Science* 1, 377–388.

Frey B. S., 2007, "Awards as compensation", *European Management Review* 4, 6–14.

Gneezy U., 2005, "Deception: the role of consequences", *American Economic Review* 95, 384–394.

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.

Guh D. P., Zhang W., Bansback N., Amarsi Z., Birmingham C. L., Anis A. H., 2009, "The incidence of co-morbidities related to obesity and overweight: a systematic review and meta-analysis", *BMC Public Health* 9:88.

Harrison G. W., List J. A., 2004, "Field experiments", *Journal of Economic Literature* 42, 1009–1055.

Hirsch F., 1976, *The social limits to growth*. London: Routledge & Kegan Paul.

Keeler E. B., Manning W. B., Newhouse J. P., Sloss E. M., Wasserman J., 1989, "The external costs of a sedentary life-style", *American Journal of Public Health* 79, 975–981.

Kosfeld, M., Neckermann, S., (2010), "Getting more work for nothing? Symbolic awards and worker performance", IZA Discussion Paper 5040.

Lacetera N., Macis M., 2008, "Social image concerns and pro-social behavior", IZA Discussion Paper 3771.

Lee C., Blair S., Jackson A., 1999, "Cardiorespiratory fitness, body composition, and all-cause and cardiovascular disease mortality in men", *American Journal of Clinical Nutrition* 69, 373–380.

Lundquist T., Ellingsen T., Gribbe E., Johannesson M., 2009, "The aversion to lying", *Journal of Economic Behavior & Organization* 70, 81–92.

Miller T. D., Balady G. J., Fletcher G. F., 1997, "Exercise and its role in the prevention and rehabilitation of cardiovascular disease", *Annals of Behavioral Medicine* 19, 220–229.

McNeil B., Pauker S. G., Sox H. Jr., Tversky A., 1982, "On the elicitation of preferences for alternative therapies", *New England Journal of Medicine* 306, 1259–1262.

Neckerman S., Frey B. S., 2007, "Awards as incentives". Institute for Empirical Research in Economics, University of Zurich, Working Paper No. 334.

Niederle M., Segal C., Vesterlund, L., 2010, "How costly is diversity? Affirmative action in light of gender differences in competitiveness". Mimeo, Stanford University.

Niederle M., Vesterlund L., 2007, "Do women shy away from competition? Do men compete too much?", *Quarterly Journal of Economics* 122, 1057–1101.

Speck B. J., Looney S. W., 2001, "Effects of a minimal intervention to increase physical activity in women: daily activity records", *Nursing Research* 50, 374–378.

Stajkovic A. D., Luthans F., 2003, "Behavioral management and task performance in organizations: conceptual background, meta-analysis, and test of alternative models", *Personnel Psychology* 56, 155–194.

Sutter M., Rutzler D., 2010, "Gender differences in competition emerge early in life". University of Innsbruck, Working Papers in Economics and Statistics, 2010–14.

Tudor-Locke C., Williams J. E., Reis J. P., Pluto D., 2002, "Utility of pedometers for assessing physical activity: convergent validity", *Sports Medicine* 32, 795–808.

Tversky A., Kahneman D., 1974, "Judgment under uncertainty: heuristics and biases", *Science* 185, 1124–1131.

Tversky A., Kahneman D., 1981, "The framing of decisions and the psychology of choice", *Science* 211, 453–458.

Volpp K. G., John J. K., Troxel A. B., Norton L., Fassbender J., Loewenstein G., 2008, "Financial incentive-based approaches for weight loss: A randomized trial", *Journal of the American Medical Association* 300, 2631–2637.

Volpp K. G., Troxel A. B., Pauly M. V., 2009, "A randomized, controlled trial of financial incentives for smoking cessation", *New England Journal of Medicine* 360, 699–709.

Whitlock G. et al., 2009, "Body-mass index and cause-specific mortality in 900 000 adults: collaborative analyses of 57 prospective studies", *Lancet* 373, 1083–1096.

Wooldridge J. M., 2003, "Cluster-sample methods in applied econometrics", *American Economic Review Papers and Proceedings* 93, 133–138.

6 Dominance and Submission: Social Status Biases Economic Sanctions

Emma von Essen, Eva Ranehill

We are grateful for comments from Konstanze Albrecht, Johan Almenberg, Anne Boschini, Anna Dreber, Tore Ellingsen, Armin Falk, Hannah Schildberg-Hörisch, Magnus Johannesson, Astri Muren and Mats Persson and seminar participants at Zürich University and CEE Copenhagen 2008. Financial support from the Jan Wallander and Tom Hedelius Foundation, as well as the Carl Silfvén Foundation (E.Ranehill) and Swedish council for working life and social research (FAS) is gratefully acknowledged (E. von Essen).

ABSTRACT: Social hierarchy is persistent in all almost all societies. Social norms and their enforcement are part of sustaining hierarchical systems. This paper combines social status and norm enforcement, by introducing status in a dictator game with third party punishment. Status is conveyed by surname; half of the third parties face a dictator with a noble name and half face a dictator with a common name. Receivers all have common names. We find that low status men are punished to a greater extent than low status women, high status men, or high status women. Interestingly, discrimination occurs only in male to male interaction. For offers below half, or close to half of the allocated resource, male third parties punish male dictators with common names almost twice as much as their noble counterparts. We find no support for female discrimination. This result suggests that social status has important implications for men's decisions to use economic punishment.

6.1 Introduction

In all societies, normative standards of behavior are enforced by formal and informal sanctions, and the importance of sanctioning possibilities for human economic interaction has been shown extensively (Ostrom, 2000, Fehr and Gächter, 2000, Carpenter and Matthews, 2005). However, despite the apparent social dimension of sanctioning behavior, most studies are performed in anonymous settings without social context. We thus still know little about how punishment is affected by social cues.

This study investigates how relative social status influences sanctioning behavior in a dictator game with third party punishment. The third party punishment game has been constructed and utilized to investigate norm enforcement (Fehr and Fischbacher, 2004). The status manipulation in this study pertains only to dictators. All participating third parties play against one of four confederates who participate as dictators in every session. The group of dictators consists of two men and two women. One of each gender has a noble name indicating high social status; the other has a common name indicating low status. We then compare the level and frequency of punishment among third parties facing dictators of different social status.

In society sanctions are often imposed by third parties, and the important role of people's actions as members of juries, committees and arbitrators has long been recognized. However, instead of sanctions being applied impartially, previous research finds that third party punishment is likely to be shaped by social context such as parochialism and the punishers' relation to the victim (Bernhard et al., 2006; Goette et al., 2006; Lieberman and Linke, 2007). Further, social status has been shown to affect the outcome of court proceedings in favour of higher status individuals (e.g. status of defendants see Sarnecki et al., 2006; Abrams et al., forthcoming, for status of the victim see Phillips 2009).

Status is often defined as the honor or prestige attached to one's position in society. General for many definitions of social status is that, apart from being a commonly recognized ranking of individuals in a given society, high status also implies favourable treatment and increased access to resources (Weiss and Ferschtman, 1998; Ball et al., 2001, Fershtman and Gneezy, 2001). In psychology and sociology, status hierarchies are analysed as a basis on which we construct our beliefs and behavior.[110] Social asymmetries thereby define perceptions of deservingness, implicit performance expectations and shape appropriate behavior (Weber, 1924; Cummins, 2000; Oxby, 2002). According to this literature, knowledge of relevant status relations is thus crucial to act successfully in social situations and relative status is likely to have a fundamental impact on human decision making.[111]

In this study we use noble names as a marker of high status; an indicator of possessing ascribed status through membership of the nobility. Using name as a status characteristic has many advantages. Most importantly, it is exogenous to the experimental setting and can be introduced without dra-

---

[110] For similar thoughts in economics see Akerlof and Kranton (2000) who discuss the impact of identity and social category on economic behavior.

[111] See for example the literature on Status Characteristics Theory (SCT), originally developed in Berger et al. (1966). It suggests that power and prestige rankings arise in interactional settings based on individual characteristics. Individuals possessing high status characteristics (such as being white, adult, male or tall) are judged as more able and better performing, as well as more deserving, independent of their actual performance (Hong and Bohnet, 2006). SCT thus predicts subordination and superordination due to a voluntary and partly unconscious process (Webster and Driskell, 1978). Though SCT has also been criticized, a large literature of empirical studies confirm that status characteristics have powerful and predictable effects on how we judge other individuals, and what we expect from them (Kalkhoff and Barnum, 2000; Ridgeway et al, 1998; Hong and Bohnet, 2006; Simpson and Walker, 2002).

wing any attention to the status manipulation in itself. Participants thus remain ignorant about the true aim of the study. Nobility has also been shown to have behavioral implications in the Swedish marriage market. Almenberg and Dreber (2009) find that noble individuals are more likely than commoners to find a partner from a higher wealth bracket than themselves. Further, possessing a noble name has no institutional meaning; the Swedish nobility lost its last formal privileges in the 19th century. However, the surnames remain as explicit hereditary status markers.[112]

Previous literature in economics indicates that social status does affect economic behavior. In laboratory studies, high social status seems to imply larger economic gains. Ball and Eckel (1996 and 1998) and Ball et al. (2001) investigate the economic benefit of experimentally manipulated status in an ultimatum game as well as a double oral auction game. Participants in the high status group receive better offers in both games, and this effect is persistent even in a treatment where allocation to the high status group is obviously random. In addition the authors test for the possibility of a status induced change in bidding behavior by rewarding status in private. In this treatment, where low status individuals are unaware of the status rewarded, the effect of status disappears. High social status thus appears to induce favourable treatment partly due to deference on behalf of lower status individuals. Preferential treatment of high status participants is also found by Glaeser et al. (2000) and Harbaugh et al. (2001) in the trust game. In both studies, participants with high status are found to elicit more trustworthy behavior on behalf of the trustees.[113] However, the studies by Ball and Eckel (1996, 1998), Ball et al. (2001) and Glaeser et al. (2000) investigate the

---

112 The last occasion a person was raised to the nobility in Sweden was in 1902, and the Swedish monarch has since then lost the right to ennoble. Surnames pertaining to a specific family benefit from stronger protection in the Swedish name law than more frequently occurring names.

113 Glaeser et al. (2000) investigate behavior in a sample of Harvard undergraduates. The participants meet their counterpart before they are separated again and play the trust game. As status variables the authors consider for example hours worked for pay, hours spent volunteering, father's education, number of close friends and proxies for popularity. Harbaugh et al. (2001) study the behavior of children aged 8, 11, 14 and 17. Age is here seen as a status marker.

impact of social status in a non-anonymous setting, and hence cannot distinguish between the effect of reputational or strategic concerns versus deference.

This study extends previous research in a number of ways. First, the setting is semi-anonymous. Semi-anonymity is achieved by revealing the names of the players participating in the dictator game to the third party, but at the same time keeping the identity of the third party unknown to all other players. This allows us to examine whether social status influence punishment in a transparent way, absent effects of potential future interactions. Second, by using the third party punishment game (Fehr and Fischbacher, 2004), we examine the effect of status through the sanction choices of individuals whose payoff is independent of the decisions previously taken. Third, earlier literature within economics has investigated preferential treatment in the reward domain. We here investigate whether social status also moderates sanctioning behavior.[114] The inclusion of one participant of each gender in each status category also allows us to investigate gender and status interactions. Gender itself is highly connected to social status and previous research indicates that men are more sensitive to social hierarchies (Campbell, 2002). For example, male groups develop steeper hierarchies and behave more competitively than female and mixed groups (Colarelli et al., 2006). In a laboratory experiment, Huberman et al. (2004) find that male participants sacrifice more resources to obtain social status than female. Fershtman and Gneezy (2001) also find an inter-male discrimination only when they study discrimination in the trust game among Jews of different ethnicity. They find discrimination on behalf of female participants.

Our results indicate that low status men are punished to a greater extent than low status women, high status men, or high status women. Interestingly, discrimination occurs only in male to male interaction. For offers below half, or almost half of the allocated resource, male third parties punish male dictators with common names almost twice as much as their noble counterparts. We find no support for female discrimination.

---

[114] Ball and Eckel 1996 perform ultimatum games with status manipulations, but they report no results regarding the responder behavior.

The differential treatment that we observe in this study can arise due to different reasons. If punishment evolved as a norm enforcing mechanism within relatively stable groups of individuals, altruistic behavior such as giving ans punishment, should primarily pertain to in-group members. If noble individuals are considered as out-group and commoners as in-group by the participating third parties we would then observe higher punishment of norm violations committed by common dictators in the third party punishment game and higher transfers to common recipients in for example a dictator game. If discrimination arises due to liking of noble individuals, this group would experience lower punishment in the third party punishment game and receive higher transfers in a dictator game. To separate these two explanations, we therefore also ran a dictator game. The confederates, who in the first game acted as dictators, are here featured as recipients.

The dictator game indicates no discrimination in transfers. We therefore find support neither for an in-group bias nor a difference in liking. We can therefore only speculate why discrimination arise in the punishment domain.

The rest of the paper is organized as follows. The subsequent section describes the experiment design of our study. In section three we present our results, before we conclude in section four where we discuss the possible explanations for our findings as well as future research.

6.2 Experimental design

The study consists of two separate economic games, a third party punishment game and a dictator game, each with the same status manipulation. Social status is differentiated via the participants' surname, as explained below. All participants in the third party punishment game received a show up fee of 50 SEK[115].

The third party punishment game features three participants, a dictator, a recipient and a third party. The implementation of the third party punishment game in this study is similar to that of Fehr and Fischbacher (2004),

[115] At the time of the experiment 1 USD corresponded to about 6 SEK, i.e. 100 SEK was about 15 USD.

where the dictator is endowed with twice the amount of the third party. In our setting the dictator received 100 SEK and the third party 50 SEK. The recipient gets no money and has no decision to make. The dictator can transfer money to the recipient in multiples of 10 SEK, with a maximum of 50 SEK. Thereafter the third party decides, according to the strategy method, on potential punishment of the dictator's payoff.[116] For each SEK that the third party chooses to punish the dictator's payoff is reduced by three. The payoffs (W_i) of the third party punishment game are thus:

Dictator: $W_d(x,p)=100-x-3*p$

Third party: $W_{tp}(p)=50-p$

Recipient: $W_r(x)=x$

where $x=\{10, 20, 30, 40, 50\}$ *is the sum the dictator transfers to the recipient and* $p=\{0,1,2,...,50\}$ *is the punishment imposed by the third party.*

Subjects were randomly allocated to the roles of recipients and third parties. For the position of dictator we recruited a group of four people; a woman and a man with the noble name von Essen, and a woman and a man with the common name Andersson.[117] This group participated as dic-

[116] The strategy method is elicitation of contingent responses. The third party made a sanctioning decision contingent on each possible transfer level before being informed of the dictator's decision. It is possible that this elicitation method induces different behaviors compared to a situation where the third party knows the dictators transfer decision (called the "specific response method"). Evidence from Cason and Mui (1998) and Brandts and Charness (2000) do not indicate that this is the case.

[117] These names are used as they are very strong indicators of nobility and vice versa. Von is a well known indicator of Swedish nobility and names ending with –sson are the most common names in Sweden. Andersson is the second most frequent surname. Swedish law awards intellectual property rights to surnames depending on how distinct they are. Names with the prefix von and other noble surnames are protected such that that a common person cannot add von to his or her surname (Statistics Sweden http://www.scb.se/Grupp/allmant/BE0801_2005K04_TI_10_A05ST0504.pdf, Access

tators in every session. Written instructions were distributed and read before making the decisions, and all third parties answered a set of control questions to ensure that they understood the consequences of their decisions. The recipients were asked to state their expectation concerning the third party punishment, also using the strategy method.

The status manipulation is apparent only to the third party, who sees the names of the two dictator game participants on top of the decision sheet. The experiment is thus semi-anonymous such that the third party knows the name of the dictator he or she punishes. In other aspects the game is anonymous and all players are aware of this. Apart from the names on the decision sheet, no reference was made to status or gender, and no subject indicated any interest in the names, nor in any aspect of status or gender. Each subject also answered a number of survey questions about age, gender, income, motives and beliefs about other player's income and wealth. This gives us an indication about possible mechanisms behind the observed result.

Based on the results from the third party punishment game only two of our confederates in the third party punishment game, the common and the noble man, were used as recipients in the dictator game. The dictator was endowed with 100 SEK and the recipient with was endowed with no money. Money could be transferred from the dictator to the recipient in multiples of 10 SEK, with a maximum of 50 SEK. The recipient had no decision to make.

The dictator game was run on a separate sample consisting only of male students. The status manipulation was semi-anonymous and implemented in the same way as in the third party punishment game. Each dictator saw the name of the recipient on top of his decision sheet. Apart from this, no reference was made to the other player or to status. In both games subjects

---

090122). Andersson was chosen since we had easy access to people with that name who could participate in the study. The noble name of one of the authors was never revealed to the participants. Due to Swedish tax regulations we had to collect the name and address of all participants after they had completed the experiment. We were therefore able to control for whether the third parties where noble or not. In the sample of third parties there were none with a noble name.

were placed at separate locations and each session took approximately 20 minutes.

6.3 Results

In this section, we start by presenting the general results regarding third party behavior. Thereafter we address the effect of the dictator's status category and gender on third parties decisions. Throughout the analysis we explore the proportion of punishment, i.e. the percentage of those who punished, as well as the level of punishment. To calculate the proportion of punishment we define the binary variable *punishment*, defined as a positive payment on behalf of the third party at any transfer level. Punishment level is simply the average punishment across the third party sample at a specific transfer level. We end by discussing the underlying mechanisms behind the results, by studying the behavior in the dictator game as well as the effect of third party beliefs regarding dictator wealth and income.

6.3.1 General results

In total, our sample consists of 132 observations of third party behavior, 63 male and 69 female.[118] 14 of these observations were removed from the sample. In a majority of the cases this was due to subjects indicating that they knew another participant, or incapacity to understand the experimental setup.[119]

[118] The participants came from three different Universities in Stockholm (Stockholm University, Stockholm School of Economics and Stockholm Royal School of Technology). We found no difference in punishment between the three schools.

[119] When running the experiment, we considered it important that participants understood the consequences of their actions. Further, we wanted to avoid participants with a previous personal relation that could influence their decisions. 10 of the 14 observations dropped were removed due to incapacity to correctly fill in the control questions before the actual experiment started or due to third parties indicating that they knew another participant. The remaining 4 third party observations were removed due to incomplete answers in the actual experiment. Of the 14 dropped third party observations 8 were males and 6 females. We also had two participating noble subjects as recipients. In these cases we removed the von in the names, which resulted in non noble names existing in Sweden.

We find that the majority of the third parties do punish, and the majority of the recipients also expect them to do so.[120] Figure 1a shows the average proportions of actual and expected punishment for each transfer level. The difference between expected and actual punishment and punishment level is only significant at the level of 50 (see table 1d in appendix). The proportion of punishers and the level of punishment we observe in our sample are in accordance with earlier studies (see for example Fehr and Fishbacher, 2004; Leibrandt and Lopéz-Peréz, 2008).[121]

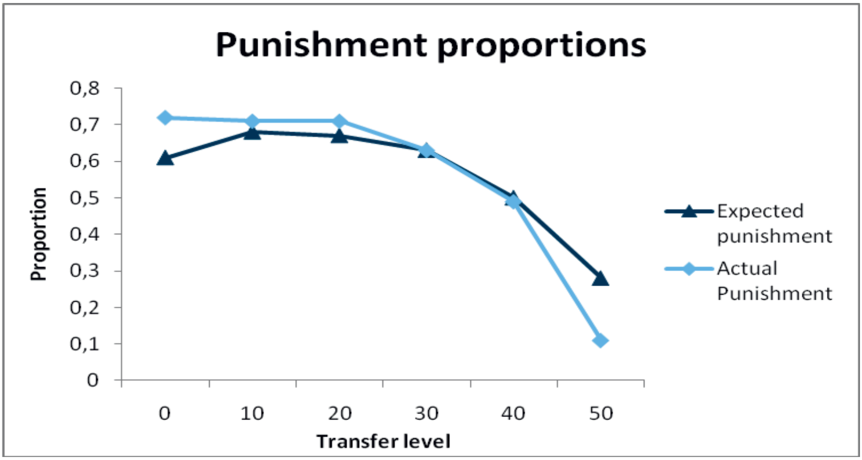

Figure 1a. Proportion of actual and expected punishment.

---

[120] None of our variables were normally distributed according to a skewness and kurtosis test. For all relevant tests in the analysis, we have therefore performed a Mann-Whitney test as well as a two-sided t-test. Throughout the analysis we refer only to the p-value for the Mann-Whitney test unless there are differences in significance between the measures (at the 5 % level). When testing the equality between proportions we have performed a chi square test and a parametric test of proportions. If nothing else is stated the chi square test is presented.

[121] We find close to significant gender differences in punishment expectations. 65% of the men and 79% of the women choose to punish ($p=0.09$). The corresponding numbers for expected punishment among male and female recipients is 60% and 76% ($p=0.06$). The level of punishment among third parties and the expected level of punishment among recipients do not vary by gender at any level of punishment.

For each 10 SEK reduction of transfer from the dictator, the third party imposes an average punishment of slightly more than 3 SEK, or a deduction of about 10 SEK. Thus, in expectance, the dictator is left with 50 SEK no matter what he or she chooses to do. The average punishment for a dictator who keeps the whole endowment was 17 SEK, deducing the dictator's income with 51 SEK.

In the next section we focus on the distribution of punishment based on two dictator characteristics; gender and nobility. All our third parties have a non noble name, implying that we compare the punishment decision of a common third party facing either a noble or a common dictator.

6.3.2 Punishment based on dictator nobility and gender

Variation in social status and gender entails four dictator categories in our experiment; noble women (NW), common women (CW), noble men (NM), and common men (CM). Among the common male dictators, 90% are punished at least at any level. The proportion of punishment in the other three categories lies between 60-70%. Testing for equality of punishment proportions between common men and the other groups we find a significant difference compared to noble and common women, and a close to significant difference in relation to the group noble men (p-values 0.05, 0.01, and 0.09 respectively, see table 1c in appendix). This indicaties that both gender and social status play a role for sanctioning behavior.

Figures 2a-e below illustrate the average level of punishment at each transfer level, separated by the two dimensions; gender and nobility.[122] Figure 2a shows the level of punishment of noble versus common dictators. The difference between the two categories is not significant, but the point estimate is consistently higher for common dictators. The following figure, 2b, shows the average punishment at each transfer level for female and male dictators separately. The total level of punishment and level of punishment at each transfer levels apart from 50 have p-values around 0.05, indicating that male dictators on average are punished harsher than female (see the corresponding p-values in table 2b in appendix).

---

[122] Subsequent Mann-Whitney p-values for test of equal averages in figures 2a-e are found in table 2 a-e in appendix. We report p-values for each transfer level, the total level and the total level excluding levels 40 and 50.

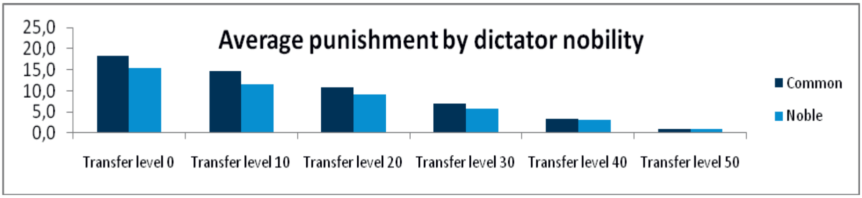

Figure 2a. Average level of punishment split by dictator nobility.

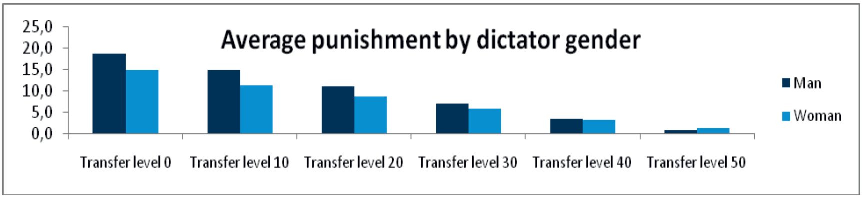

Figure 2b. Average level of punishment split by dictator gender.

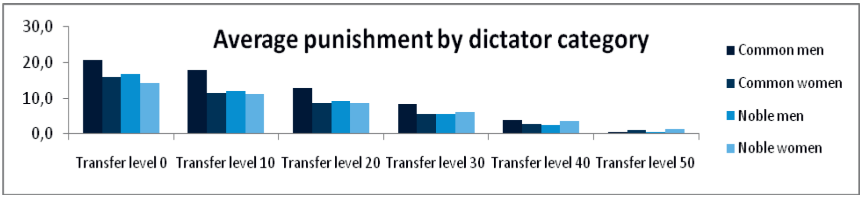

Figure 2c. Average level of punishment split by dictator nobility and gender.

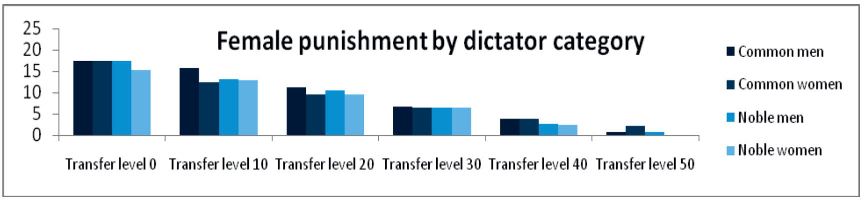

Figure 2d. Average level of female punishment split by dictator category.

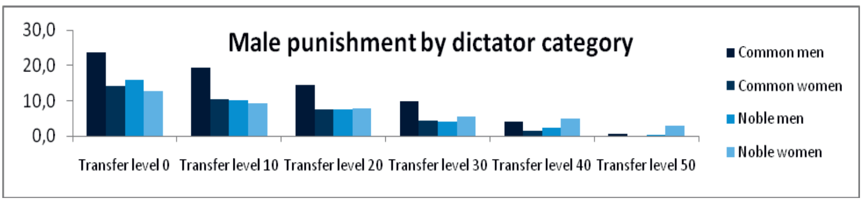

Figure 2e. Average level of male punishment split by dictator category.

Based on the results on punishment proportion we expect the group of common males to be punished the harshest. Figure 2c therefore shows a comparison of average punishment level of all four dictator categories, confirming this. Table 1 below states the p-values for comparison of punishment levels between common men and all other categories.

Table 1. Punishment by dictator category.

Transfer level	CM	NW	CW	NM	p-value	n
0	21.3	14.2	15.9	16.7	0.03, 0.06, 0.20	29, 32, 28, 29
10	17.8	11.3	11.5	12.0	0.01, 0.02, 0.03	29, 32, 28, 29
20	12.9	8.8	8.6	9.3	0.03, 0.02, 0.10	29, 32, 28, 29
30	8.4	6.0	5.6	5.4	0.07, 0.03, 0.08	29, 32, 28, 29
40	4.0	3.5	2.6	2.6	0.10, 0.02, 0.13	29, 32, 28, 29
50	0.7	1.3	1.1	0.6	0.39, 0.43, 0.74	29, 32, 28, 29
Total	65.1	45.1	45.4	46.6	0.02, 0.02, 0.08	29, 32, 28, 29
Total -50	64.5	43.8	44.3	46.0	0.05, 0.02, 0.07	29, 32, 28, 29
Total -40	60.5	40.3	41.6	43.4	0.02, 0.02, 0.08	29, 32, 28, 29

NW =Noble women, CW=common women, NM=noble men, and CM=common men. The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40. Sample size per group is reported in the order NW, CW, NM, CM.
P-values indicate the probability of equal punishment between common men and the other categories respectively (noble women, common women and noble men). When comparing the results from a two-sided ttest some of the differences in punishment between common men and noble and common women are insignificant, see appendix table 2c.

Based on the differences in punishment level between male and female third parties figure 2d and 2e report the same numbers as in table 1 split by gender of the third party. Female punishment level, depicted in figure 2d and in table 2d in appendix, shows small, insignificant and inconsistent differences in punishment across dictator categories. This result stands in stark contrast to the punishment by men (figure 2e and table 2 below). At transfer levels 0-30, where the dictator is most stingy, male third parties punish common male dictators almost twice as much as they punish noble ones.

At each level between 10 and 30, as well as the total level, this difference in punishment is significant.[123] For the levels 0-20, the difference in punishment is 7 SEK or more, implying an additional reduction of 20-30% of the common dictators' initial endowment. Thus our results indicate that sanctions in male to male interaction are influenced by relative status, as hypothesized in the introduction. High status males appear to be treated with leniency by other common male third parties.

Table 2. Male third party punishment by dictator category.

Transfer level	CM	NW	CW	NM	p-value	n
0	24.3	12.7	14.1	15.8	0.04, 0.05, 0.13	14, 13, 12, 16
10	19.4	9.2	10.5	10.3	0.02, 0.04, 0.02	14, 13, 12, 16
20	14.4	7.8	7.5	7.5	0.04, 0.03, 0.04	14, 13, 12, 16
30	9.9	5.6	4.5	4.2	0.06, 0.04, 0.04	14, 13, 12, 16
40	4.2	4.9	1.4	2.4	0.21, 0.03, 0.21	14, 13, 12, 16
50	0.6	3.0	0.0	0.4	0.42, 0.37, 0.88	14, 13, 12, 16
Total	72.9	43.2	37.9	40.6	0.04, 0.02, 0.05	14, 13, 12, 16
Total -50	72.3	40.2	37.9	40.2	0.04, 0.02, 0.05	14, 13, 12, 16
Total -40	68.1	35.4	36.5	37.8	0.03, 0.03, 0.04	14, 13, 12, 16

NW =Noble women, CW=common women, NM=noble men, and CM=common men. The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40. Sample size per group is reported in the order NW, CW, NM, CM. P-values indicate the probability of equal punishment between common men and the other categories respectively (noble women, common women and noble men). Some p-values for the group of common men compared to noble women are insignificant when running a two-sided t-test, see appendix table 2e.

[123] Punishment difference at the level of zero is not significant in our study due to a few male subjects motivating their punishment by either you give a lot or nothing.

In summary we find that the punishment decision of male third parties is affected by both the social status and the gender of the perpetrator. Female punishment decisions exhibit no consideration of social status.[124]

6.3.3 Giving in a dictator game

The discrimination in punishment that we observe can be due to at least two types of underlying mechanisms. First, in-group bias among common men would cause norm violations by in-group members to warrant harsher punishment than violations by other individuals, in order to successfully enforce norms within the group[125]. Second, common men might simply like noble men more than common men and therefore provide noble men with an economic premium.

In order to separate these mechanisms we also ran a dictator game. 59 males participated in the game; we assigned 31 subjects to face a recipient with a common name and 28 subjects to face a recipient with a noble name.[126] Our results indicate that giving in the dictator game does not depend on the status of the recipient. On average common men receive 34.2 SEK and noble men 36.8 SEK, but this difference is far from significant (p-value of 0.2809)[127]. Our results do thus neither support the explanation of in-group bias nor that of liking. They suggest that discrimination depending on status pertains only to the punishment domain, and not to giving.

---

[124] The analysis indicates that third party punishment behavior is dependent on the third party's relation to dictator characteristics. We therefore also studied whether third party discrimination was affected by gender composition of the dictator game participants. For example, we tested whether a male third party facing a male dictator punished differently depending on the gender of the recipient. We found no indication that this relation mattered. The sample sizes in each group are very small therefore this should be seen as mere indications.

[125] However previous empirical literature indicates that out-group members are punished harder when the victim is an in-group member (Bernhard et al., 2006; Lieberman and Linke, 2007).

[126] All participants were students at Stockholm University or Stockholm Royal School of Technology.

[127] A sample size analysis assuming an alpha of 0.05 and a power of 80% indicate that we would need a sample size of 2096 subjects in order to get significant results. Se appendix table 3a for descriptive statistics.

6.3.4 Further analysis

The discrimination based on social category that we observe might also be due not to social status per se but to variables correlated, or perceived to be correlated, with nobility. Nobility may for example influence beliefs about dictator wealth, income or education. We identify two potential sources of bias. First, if variables believed to be correlated with nobility have status implications in their own right, this leads us to overestimate the effect of nobility. Second, beliefs of high wealth or income may also reinforce mechanisms not directly linked to status such as inequality aversion (Bolton and Ockenfels, 2000, Croson and Gneezy, 2009, Fehr and Schmidt, 1999). If participants believe that nobility is positively correlated with wealth, inequality aversion would cause us to underestimate the effect of social status (nobility). As shown in Almenberg and Dreber, 2009 the Swedish nobility has a higher wealth than the common population. Previous literature has found women to be more inequality averse than men; a fact that could contribute to the gender differences in the discrimination we find (Andreoni and Vesterlund 2001, Dickinson and Tiefenthaler 2002, Selten and Ockenfels 1998).

All third parties in the third party punishment game were asked to state their beliefs regarding dictator wealth and income in relation to the average student and all dictators in the dictator game were asked to do the same with respect to the recipient. The number of third parties who indicated that they believed the dictator to have a higher wealth and income than the average student are roughly equal irrespective of whether they faced a noble or a non-noble dictator.[128]

Further, what subject's rate as being fair is not influenced by the status of the recipient in any of the two games. Further, 82 percent of the partici-

[128] Of those third parties facing a noble dictator, 17 (12) subjects thought their counterpart to had a higher wealth (income) than the average student. Among those facing a non-noble dictator, the corresponding number was 16 (14). Male third parties facing male noble dictators and assuming noble dictators to have a wealth above average, punish the least of all groups. Male third parties facing common dictators and assume common dictators to have a high wealth are punished the harshest. However, the sample sizes are too small for a relevant analysis.

pants consider an equal split fair. This corroborates the findings of Lieberman and Linke (2004), who find that even though third party punishment varies with social category, the third parties' moral judgement pertaining to the norm violation does not. It thus appears as if it is the scale of punishment that differs across status categories, not the norm in itself. These results indicate that the impact of social status on punishment is robust to controlling for beliefs of relative income and wealth. We also performed regression analysis which confirm our main results and indicate that third party beliefs concerning the wealth and income of the dictator as well as fairness have no significant effect on punishment.[129]

6.4 Discussion

Previous research in economics indicates that high social status conveys an economic premium (Ball and Eckel, 1996, 1998, Ball et al., 2001, Glaeser et al., 2000, Almenberg and Dreber, 2009). Our results illustrate that high social status induces lower punishment in a third party punishment game, though only in male to male interactions. Male third parties punish common male dictators almost twice as much as their noble counterparts, but no effect is found in all female or mixed interactions. We do not find a discriminatory effect of social status on altruistic behavior in a dictator game; male participants are not more generous to high status individuals. . This is in line with the results presented by Fershtman and Gneezy (2001), who they find difference in trust behavior depending on social category of the

[129] The OLS regression analysis is based on parametric assumptions that may not be fulfilled. The same set of control variables was included in the regression for each punishment level; beliefs about income, wealth and fairness, gender of all players, university and age of third party. The variables for beliefs regarding wealth and income were coded as 1 if third parties who believed the dictator to have a wealth (income) above the average student and 0 otherwise. The variable for fairness was coded to take the value 1 if subjects indicated an equal split of the initial dictator endowment as fair and 0 otherwise. We also conducted a regression pooling the data over all six decisions elicited by the strategy method for each third party, clustering on individual third parties. In this regression the male to male dummy is significant (p-value 0.033), and beliefs of wealth and income remain insignificant. The results from the regression analysis are available upon request.

trustee, but no differentiation based on recipient social category in a dictator game.

For several reasons, the effect of social status in the third party punishment game is surprisingly large. The Swedish nobility lost all its formal economic and political privileges in the 19th century. The status variable thus has no relation to the experimental context or to merit, and no references were made to nobility or social status during the experiment. The semi-anonymous design also diminishes reputation and strategic concerns. Despite this, the average punishment is significantly different between male status categories, and this difference represents 20-30 percent of the dictators' initial endowments. The effect remains stable when controlling for beliefs of wealth and income, and does not rely on fairness considerations being status dependent.

Our results corroborate previous research implying that men are more sensitive to social status in economic decision making tasks (Fershtman and Gneezy, 2001; Huberman et al, 2004). Many hierarchies in western society have throughout history been exclusive for men or male dominated. Nobility in Sweden, for example, is only hereditary on the male side; thus historically implying a larger value for men. If men were and are more likely to benefit from status, such as nobility, this explains the higher level of investment in status observed by men in comparison to women (Campbell, 2002; Huberman et al., 2004; Pawlowski et al., 2000). Men are also often found to be more competitive compared to women (see review by Croson and Gneezy, 2009). Since noble names are impossible to acquire, competition for status in the context of the present study is relevant only within the group of non noble names. An additional reason for the discrepancy in male and female behavior could be gender differences in inequality aversion. Previous literature has found women to be more inequality averse compared to men (Andreoni and Vesterlund, 2001; Dickinson and Teifenhaler, 2002; Selten and Ockenfels, 1998). If high social status is associated with other benefits, females should punish high status individuals more.

A few studies have explored in-group bias settings with third party punishment (Bernhard et al. 2008 and Götte et al. 2006). Contrary to what is found in this study, Bernhard et al. (2006) find that dictators from another group than the third party are punished harsher when facing a recipient from the same group as the third party. Lieberman and Linke (2004) found similar results in a hypothetical setting. This is the opposite of what in-group bias in a third party punishment game would predict if punishment serves as a norm enforcing device. All our third parties and recipients are from the same status group, individuals with non noble names. The divergence of our results in relation to the other studies may partly be explained by the fact that individuals with a common name can be considered a large and not well defined social group, or that status considerations override group belonging. Our results also differ from what Fehr et al. (2008) find in a trust game with third party punishment in India. They find that low caste participants punish less than high caste participants, but punishment in both groups is independent of the caste belonging of the norm violator. Several potential mechanisms might explain this variation in results. Norm enforcement could, for example, differ between norms pertaining to different behaviors. Fehr et al. (2008) propose that historically repressed groups have a lesser willingness to punish violators in general. Culture might create differences between the studies; India's cast system induces a more pronounced status hierarchy and may therefore repress the decision to punish by low status individuals more than the historical division between nobility and commoners in Sweden. Even though the Swedish nobility lost its privileges more than a decade ago one could speculate that the punishment specific discrimination is due to historic power inequalities. The decision to punish an individual of higher status may entail a larger risk of retaliation than the decision not to give. The result in Lieberman and Linke (2004) as well as the fairness judgements in our study supports the suggestion that the moral judgment of a norm transgression is stable across social categories ; however the scale of punishment differs.

Our study shows a surprisingly large effect of social status on punishment behavior, underlining the importance of social status as a modulator of behavior in male interactions. By guiding appropriate behavior, knowledge of status relations is an important key to successfully navigate in human societies. Future research is needed concerning various types of social status and its implications for economic decision making in different situations.

6.5 References

Abrams D., Bertrand M., Mullainathan S., 2008, "Do judges vary in their treatment of race?", Forthcoming in the *Journal of Legal Studies*.

Akerlof G. A., Kranton R. E., 2000, "Economics and identity", *The Quarterly Journal of Economics*, MIT Press, vol. 115:3, 715-753.

Almenberg J., Dreber A., 2009, "The lady and the trump: status and wealth in the marrige market", *Kyklos*, 62:2, 161-181.

Andreoni J. Vesterlund L., 2001, "Which is the fair sex? gender differences in altruism", *The Quarterly Journal of Economics*, 116, 293-312.

Ball S., Eckel C., 1996, "Buying status: experimental evidence on status in negotiation", *Psychology and Marketing*, 13, 381-405.

Ball S. Eckel C., 1998, "Stars upon thars", Working Paper. The University of Texas at Dallas.

Ball S., Eckel C., Grossman P. Zame W., 2001, "Status in markets", *The Quarterly Journal of Economics*, 116:1, 161-188.

Berger J., Cohen B. P. Zelditch M. (Eds.), 1966, Status characteristics and expectation states. sociological theorie in progress, 1. Houghton Mifflin. Boston.

Bernhard H., Fischbacher U. Fehr E., 2006, "Parochial altruism in humans", *Nature* 442:7105, 912-915.

Bernhard H., Fehr E., Rockenbach B., 2006, "Egalitarianism in young children", *Nature* 454:28, 1079-1084.

Bolton G. Ockenfels A., 2000, "ERC: a theory of equity, reciprocity and competition", *American Economic Review*, 90, 166-193.

Brandts J., Charness G., 2000, "Hot vs. cold: sequential responses and preference stability in experimental games", *Experimental Economics*, 2:3, 227-238.

Campbell A., 2002, A mind of her own. Oxford University Press, Oxford.

Carpenter J., Matthews P.H., 2005, "Norm enforcement: anger, indignation or reciprocity?", Middlebury College Working Paper Series 0503, Middlebury College, Department of Economics.

Cason T., Mui V.L., 1998, "Social influence in the sequential dictator game", *Journal of Mathematical Psychology*, 42, 248-265.

Colarelli S. M., Spranger J. L., Hechanova, M. R., 2006, "Women, power, and sex composition in small groups: an evolutionary perspective", *Journal of Organizational Behavior*, 27, 163–184.

Croson R., Gneezy U., 2009 "Gender differences in preferences", *Journal of Economic Literature*, 47:2, 448–74.

Cummins, D., 2005, "Dominance, status, and social hierarchies", In D. M. Buss (Ed.), The Handbook of evolutionary psychology, 676–697. Hoboken, NJ: John Wiley and Sons.

Dickinson D., Teifenhaler J., 2002, "What is fair?" *Experimental Evidence. Southern Economic Journal*, 69, 414-428.

Fehr E., Hoff K., Kshetramade M., 2008, "Spite and development", *American Economic Review Papers and Proceedings*, 98:2, 494–99.

Fehr E., Fischbacher U., 2004, "Third-party punishment and social norms", *Evolution and Human Behavior*, 25, 63-87.

Fehr E., Gächter S., 2000, "Cooperation and punishment in public goods experiments", *American Economic Review*, 90:4, 980-994.

Fehr E., Schmidt K., 1999, "A theory of fairness, competition and cooperation", *The Quarterly Journal of Economics*, 114:3, 817-868.

Fershtman C., Gneezy U., 2001, "Discrimination in a segmented society: an experimental approach", *The Quarterly Journal of Economics*, 116:1, 3511-3377.

Glaeser E.L., Laibson D.I., Scheinkman J.A., Soutter C.L., 2000, "Measuring trust", *The Quarterly Journal of Economics*, 115, 811-846.

Goette L., Huffman D., Stephan M., 2006, "The impact of group membership on cooperation and norm enforcement: Evidence using random assignment to real social groups", Institute for the Study of Labor, IZA Discussion paper, 2020.

Harbaugh W.T., Krause K., Liday S., Vesterlund Lise. 2001, "Trust in children", University of Oregon Economics Department Working Papers 2002-10, University of Oregon Economics Department, revised 25.

Hong K., Bohnet I., 2006, "Status and distrust: the relevance of inequality and betrayal aversion", *The Journal of Economic Psychology*, 28, 197-213.

Huberman B., Loch C., Öncüler A., 2004, "Status as a valued resource", *Social Psychology Quarterly*, 67:1, 103-114.

Kalkhoff W., Barnum C., 2000, "The effects of status-organizing and social identity processes on patterns of social influence", *Social Psychology Quarterly* 63, 95-115.

Leibrandt A., Lopéz-Peréz R., 2008, "The envious punisher: understanding third and second party punishment with simple games", Institute for Empirical Research in Economics, University of Zurich, Working Paper Series, Working Paper No. 373.

Lieberman D., Linke L., 2007, "The effect of social category on third party punishment", *Evolutionary Psychology*, 5, 289-305.

Ostrom E., 2000, "Collective action and the evolution of social norms", *Journal of Economic Perspectives*, 14(3), 137-58.

Phillips S., 2009, "Status disparities in the capital of capital punishment", *Law and Society Review*, 43:4.

Ridgeway C., Boyle E., Kuipers K., Robinson R., 1998, "How do status beliefs develop? The role of resources and interaction", *American Sociological Review*, 63, 331-350.

Sarnecki J., 2006, "Är rättvisan rättvis? Tio perspektiv på diskriminering av etniska och religiösa minoriteter inom rättssystemet", SOU 2006:30.

Selten R. Ockenfels A., 1998, "An experimental solidarity game", *Journal of Economic Behavior Organization*, 34, 517-539.

Simpson B. Walker H.A., 2002, "Status characteristics and performance expectations: a reformulation", *Sociological Theory*, 20, 24-39.

Weber M., 1924, *From Max Weber: essays in sociology*. (trans, eds) Gerth H. H., Wright Mills C., London, Routledge and K. Paul, 1970.

Webster M., Driskell J.E., 1978, "Status generalization: a review and some new data", *American Sociological Review* 43, 220 - 36.

Weiss Y., Ferschtman C., 1998, "Social status and economic performance: a survey", *European Economic Review*, 42, 801-820.

6.6 Appendix 1

Table 1a. Descriptive statistics.

Variable	n	Mean	p50	Sd	Min	Max
Female TP	118	0.53	1	0.50	0	1
Female R	118	0.56	1	0.50	0	1
Female D	118	0.51	1	0.50	0	1
Nobel women	118	0.27	0	0.45	0	1
Nobel men	118	0.25	0	0.43	0	1
Common women	118	0.24	0	0.43	0	1
Common men	118	0.25	0	0.43	0	1
Level 0	118	16.94	16.5	14.29	0	50
Level 10	118	13.11	13	11.55	0	40
Level 20	118	9.88	10	9.19	0	40
Level 30	118	6.38	5	7.38	0	40
Level 40	118	3.22	0	5.08	0	30
Level 50	118	0.93	0	4.06	0	30
Punishment	118	0.73	1	0.45	0	1
Fairness	109	0.82	1	0.39	0	1
Age	118	25.88	24	8.33	17	71

Descriptive statistics of all variables included in our analysis. Level 0-50 indicates the actual level of punishment at each transfer level. Punishment and Justice are dummy variables; Punishment takes the value 1 when punishment was exerted at any level and 0 otherwise, and Justice takes the value 1 if a 50/50 split of the initial money was indicated as the fair division and 0 otherwise.

Table 1b. Sample size and attrition.

Sample	TP	R
Full sample	132	132
Males	59	52
Females	73	73
Attrition	14	11
Males	7	5
Females	7	6
Total	118	121

Recipient attrition refers to the case of recipients who were not able to answer either the control questions correctly, or who did not state beliefs for all alternatives elicited through the strategy method. In these cases we used their names on the third parties decision sheets, but dropped them when estimating expected proportion and level of punishment.

Table 1c. Proportion of punishment.

Punishment proportions	All	Men	Women
Nw	0.69	0.57	0.78
Cw	0.61	0.54	0.67
Nm	0.72	0.58	0.82
Cm	0.90	0.88	0.92
Punishment proportions	p-value (Ch2)	p-value (prtest)	
Cm vs Nw	0.05	0.05	
Cm vs Cw	0.01	0.01	
Cm vs Nm	0.09	0.09	

Proportion of actual punishment by dictator characteristics and third party gender. NW=noble female dictators, CW=common male dictators, NM=noble male dictators and CM=common male dictators.

Table 1d. Expected vs. actual punishment (proportion and level)

Punishment proportions	Expected	Actual	p-value ch2	p-value prtest
0	0.61	0.72	0.10	0.07
10	0.68	0.71	0.58	0.57
20	0.67	0.71	0.49	0.48
30	0.63	0.63	1.00	0.99
40	0.50	0.49	1.00	0.95
50	0.28	0.11	<0.01	<0.01

Punishment level	Expeted	Actual	p-value MW	p-value ttest
0	14.8	16.9	0.19	0.27
10	13.3	13.1	0.93	0.89
20	11.2	9.9	0.70	0.32
30	9.3	6.4	0.13	0.02
40	6.9	3.2	0.94	<0.01
50	5.2	0.9	<0.01	<0.01

Table 2a. Punishment level of noble and common dictators.

Transfer level	Av. Noble	Av. common	Differ-ence	p-value (MW)	p-value (ttest)	n
0	15.4	18.6	3.3	0.23	0.22	61, 57
10	11.6	14.7	3.1	0.15	0.14	61, 57
20	9.0	10.8	1.8	0.26	0.29	61, 57
30	5.8	7.0	1.3	0.36	0.35	61, 57
40	3.1	3.4	0.3	0.65	0.76	61, 57
50	1.0	0.9	-0.1	0.85	0.89	61, 57
Total	45.8	55.4	9.6	0.24	0.26	61, 57
Total -50	44.8	54.5	9.7	0.24	0.24	61, 57
total -40	41.8	51.2	9.4	0.22	0.21	61, 57

The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40.

Table 2b. Punishment level of female and male dictators.

Transfer level	Av. Women	Av. Men	Differ-ence	p-value (MW)	p-value (ttest)	n
0	15.0	19.0	4.0	0.04	0.13	60,58
10	11.4	14.9	3.5	0.03	0.10	60,58
20	8.7	11.1	2.4	0.03	0.15	60,58
30	5.8	6.9	1.1	0.05	0.41	60,58
40	3.1	3.3	0.2	0.04	0.85	60,58
50	1.2	0.6	-0.6	0.15	0.44	60,58
Total	45.2	55.9	10.7	0.03	0.21	60,58
Total -50	44.0	55.2	11.2	0.03	0.17	60,58
Total -40	40.9	51.9	11.1	0.03	0.14	60,58

The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40.

Table 2c. Punishment by dictator category.

Transfer level	CM	NW	CW	NM	p-value (MW)	p-value (ttest)	n
0	21.3	14.2	15.9	16.7	0.03, 0.06, 0.20	0.05, 0.16, 0.18	29, 32, 28, 29
10	17.8	11.3	11.5	12.0	0.01, 0.02, 0.03	0.03, 0.04, 0.03	29, 32, 28, 29
20	12.9	8.8	8.6	9.3	0.03, 0.02, 0.10	0.08, 0.08, 0.09	29, 32, 28, 29
30	8.4	6.0	5.6	5.4	0.07, 0.03, 0.08	0.22, 0.17, 0.06	29, 32, 28, 29
40	4.0	3.5	2.6	2.6	0.10, 0.02, 0.13	0.71, 0.32, 0.12	29, 32, 28, 29
50	0.7	1.3	1.1	0.6	0.39, 0.43, 0.74	0.53, 0.69, 0.94	29, 32, 28, 29
Total	65.1	45.1	45.4	46.6	0.02, 0.02, 0.08	0.02, 0.12, 0.07	29, 32, 28, 29
Total -50	64.5	43.8	44.3	46.0	0.05, 0.02, 0.07	0.02, 0.10, 0.07	29, 32, 28, 29
Total -40	60.5	40.3	41.6	43.4	0.02, 0.02, 0.08	0.02, 0.09, 0.07	29, 32, 28, 29

NW =Noble women, CW=common women, NM=noble men, and CM=common men. The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40. Sample size per group is reported in the order NW, CW, NM, CM. P-values indicate the probability of equal punishment between common men and the other categories respectively (noble women, common women and noble men).

Table 2d. Female third party punishment by dictator category.

Transfer level	CM	NW	CW	NM	p-value (MW)	p-value (ttest)	n
0	17.5	15.3	17.5	17.3	all >0.2	all >0.2	18, 15, 17, 13
10	15.8	12.9	12.3	13.2	all >0.2	all >0.2	18, 15, 17, 13
20	11.2	9.5	9.6	10.6	all >0.2	all >0.2	18, 15, 17, 13
30	6.6	6.4	6.5	6.4	all >0.2	all >0.2	18, 15, 17, 13
40	3.8	2.5	3.8	2.7	all >0.2	all >0.2	18, 15, 17, 13
50	0.7	0	2.1	0.8	all >0.2*	all >0.2*	18, 15, 17, 13
Total	55.6	46.6	51.8	50.9	all >0.2	all >0.2	18, 15, 17, 13
Total - 50	54.9	46.6	49.7	50.1	all >0.2	all >0.2	18, 15, 17, 13
Total - 40	51.1	44.1	45.9	47.4	all >0.2	all >0.2	18, 15, 17, 13

*The difference between the punishment of common men and noble women is significant at the level of 50 (MW: p=0.04, ttest: p=0.07). However, given the number of tests ran, we would expect some false positives to occur. NW=noble female dictators, CW=common male dictators, NM=noble male dictators and CM=common male dictators. Sample size per group is reported in the order NW, CW, NM, CM. P-values indicate the probability of equal punishment between common men and the other categories respectively (noble women, common women and noble men).

Table 2e. Male third party punishment by dictator category.

Transfer level	CM	NW	CW	NM	P-value (MW)	P-value (ttest)	N
0	24.3	12.7	14.1	15.8	0.04, 0.05, 0.13	0.02, 0.06, 0.10	14, 13, 12, 16
10	19.4	9.2	10.5	10.3	0.02, 0.04, 0.02	0.02, 0.05, 0.03	14, 13, 12, 16
20	14.4	7.8	7.5	7.5	0.04, 0.03, 0.04	0.06, 0.04, 0.03	14, 13, 12, 16
30	9.9	5.6	4.5	4.2	0.06, 0.04, 0.04	0.19, 0.04, 0.02	14, 13, 12, 16
40	4.2	4.9	1.4	2.4	0.21, 0.03, 0.21	0.79, 0.04, 0.19	14, 13, 12, 16
50	0.6	3.0	0.0	0.4	0.42, 0.37, 0.88	0.26, 0.38, 0.80	14, 13, 12, 16
Total	72.9	43.2	37.9	40.6	0.04, 0.02, 0.05	0.11, 0.04, 0.04	14, 13, 12, 16
Total -50	72.3	40.2	37.9	40.2	0.04, 0.02, 0.05	0.06, 0.04, 0.03	14, 13, 12, 16
Total -40	68.1	35.4	36.5	37.8	0.03, 0.03, 0.04	0.03, 0.04, 0.03	14, 13, 12, 16

NW =Noble women, CW=common women, NM=noble men, and CM=common men. The variable "Total -50"compounds punishment on all levels except level 50. The variable "Total -40"compounds punishment on all levels except level 50 and level 40. Sample size per group is reported in the order NW, CW, NM, CM. P-values indicate the probability of equal punishment between common men and the other categories respectively (noble women, common women and noble men).

Table 3a. Descriptive statistics dictator game.

Variable	N	Mean	Median	Sd	Min	Max
Common name	31	65.8	50	20.6	50	100
Noble name	28	63.2	50	21.6	50	100
Total	59	64.6	50	21.0	50	100

Table 3b. Giving by recipient category.

	Av. Noble	Av. common	p-value (MW)	p-value (ttest)	n
Level of giving	63.2	65.8	0.28	0.64	28,31
Proportion of giving	.77	.78	0.58	0.92	28,31

6.7 Appendix 2: Experiment instructions

Instructions for participant A.

Welcome to this study in economics!

Please read the following instructions carefully. Depending on how you and your counterparts decide, you can earn money in addition to the 50 SEK you earn by participating. Therefore, it is important that you read and follow the instructions.

<u>Please do not talk during the study.</u> If you have any questions, raise your hand and we will come to answer your question.

Throughout this study you will use Swedish crowns. The study comprises three types of participants: Participants A, participants B and participants C. **You are a participant A**. During the study, you will interact with one randomly assigned participant B and one randomly assigned participant C.

Specific Instructions for the Experiment Procedure

Stage one
In the first stage, participants A are the sole decision-makers. As a participant A you have got an endowment of 100 SEK. Participant C gets 50 SEK, and participant B gets no endowment. We ask you to decide how many of the 100 SEK that you wish to assign to participant B. You can give participant B a number of SEK between 0 and 50 in a multiple of tens, i.e. 0, 10, 20, 30, 40, or 50 SEK. If, for example, you grant participant B 40 SEK, your income at the end of stage one will amount to 60 SEK, and participant B's income will amount to 40 SEK. If you accord her/him 10 SEK, your income will be 90 SEK, and the income of participant B will be 10 SEK at the end of stage one. If you grant B 0 SEK, your income at the end of stage one will amount to 100 SEK while participant B's equals 0 SEK.

Stage two

In stage two, only participants C have a decision to make. Participant C can pay to deduct money from your payoff. Each SEK charged to you as participant A diminishes your income by 3 SEK, and participant C's income is reduced by 1 SEK. Participant C can deduct a number of SEK between 0 and 50. Suppose participant C deducts 2 SEK: your income will then be reduced by 6 SEK while participant C's income will be reduced by 2 SEK. If participant C deducts 19 SEK to you, your income diminishes by 57 SEK and participant C's income is reduced by 19 SEK. Participant C takes her or his decision before knowing your decision, and hence answers how they would like to allocate their money for every possible decision you can make. Neither you nor participant B knows the identity of participant C.

This is how we calculate participants A's, B's, and C's respective incomes:

<u>Participant A's income amounts to</u>

+ 100 SEK (participant A's endowment)
- number of SEK assigned to participant B by participant A
- 3 times the number of deduction SEK transferred to participant A by participant C

<u>Participant B's income amounts to</u>

+ number of SEK assigned to participant B by participant A

<u>Participant C's income amounts to</u>

+ 50 (participant C's endowment)
- number of deduction SEK charged participant A by participant C

Please note that your earnings may be negative, in which case the SEK will be deducted from your participation payment.

Control Questions

It is important that all participants have understood the rules of the game. Therefore we ask you to answer the following control questions. When you have finished, signal to us by raising your hand.

A. Participant A assigns 0 SEK to participant B.

a) Participant C charges participant with 0 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

b) Participant C charges participant A with 30 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

2. Participant A assigns 40 SEK to participant B.

a) Participant C charges participant A with 0 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

b) Participant C charges participant A with15 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

Decision sheet participant A.

Below we ask you to decide how much you want to transfer to participant B. We also ask you to give us your best estimation of C's decision. We want you to make an estimate for every possible decision that you can make, i.e. what decision do you believe participant C will take if you transfer 0, 10, 20, 30, 40, or 50 SEK .

Your decision

You may transfer 0, 0, 20, 30, 40, or 50 SEK to participant B

How many SEK do you want to transfer? _______________.

How do you believe is participant C going to decide?

Number of SEK you grant B	**Number of SEK participant C transfers to you (prior to the sum being tripled)**
0	
10	
20	
30	
40	
50	

When you have taken your decisions, please turn the page and answer a few survey questions.

Survey questions for participant A

1. What are you studying? ___________________________________

2. I am a *woman* ☐ *man* ☐

3. I am __________ *years old.*

4. What do you think is a fair allocation of the 100 SEK?

5. What is the reason behind the decision you made?

Instructions for participant B.

Welcome to this study in economics!

Please read the following instructions carefully. Depending on how you and your counterparts decide, you can earn money in addition to the 50 SEK you earn by participating. Therefore, it is important that you read and follow the instructions.

<u>Please do not talk during the study.</u> If you have any questions, raise your hand and we will come to answer your question.

Throughout this study you will use Swedish crowns. The study comprises three types of participants: Participants A, participants B and participants C. **You are a participant B**. During the study, you will interact with one randomly assigned participant A and one randomly assigned participant C.

Description of the two parts of the study

Stage one
In the first stage, participants A are the sole decision-makers. At the beginning of stage one, participants A get an endowment of 100 SEK. Participants C get an endowment of 50 SEK, whereas you as a participant B get no endowment. Participant A must decide how many of her/his 100 SEK s/he wishes to assign to you. S/he can transfer to you a number of SEK between 0 and 50 in a multiple of tens, i.e. 0, 10, 20, 30, 40, or 50 SEK. If, for example, participant A grants you 40 SEK, her/his income at the end of stage one will amount to 60 SEK, and your income will amount to 40 SEK. If s/he accords you 10 SEK, her/his income will be 90 SEK, and your own income will be 10 SEK. If s/he grants you 0 SEK, her/his income, at the end of stage one will result in 100 SEK, and your own income will result in 0 SEK.

Stage two
In stage two, only the participants C have a decision to make. Participant C can pay to deduct money from A. Each SEK charged to participant A diminishes A's income by 3 SEK, and participant C's income is reduced by 1

SEK. Participant C can deduct a number of SEK between 0 and 50. Suppose participant C charges 2 SEK: A's income will then be reduced by 6 SEK while participant C's income will be reduced by 2 SEK. If participant C deducts 19 SEK to A, A's income diminishes by 57 SEK and participant C's income is reduced by 19 SEK. Participant C takes her or his decision before knowing A's decision, and hence answers how they would like to allocate their money for every possible decision A can make. Neither you nor participant A knows the identity of participant C.

This is how we calculate participants A's, B's, and C's respective incomes:

<u>Participant A's income amounts to</u>

\+ 100 SEK (participant A's endowment)
\- number of SEK assigned to participant B by participant A
\- 3 times the number of deduction SEK transferred to participant A by participant C

<u>Participant B's income amounts to</u>

\+ number of SEK assigned to participant B by participant A

<u>Participant</u> C's <u>income amounts to</u>

\+ 50 (participant C's endowment)
\- number of deduction SEK charged participant A by participant C

Control Questions

It is important that all participants have understood the rules of the game. Therefore we ask you to answer the following control questions. When you have finished, signal to us by raising your hand.

A. Participant A assigns 0 SEK to participant B.

a) Participant C charges participant with 0 SEK deduction.

What is participant A's income? _______________

What is participant B's income? _______________

What is participant C's income? _______________

b) Participant C charges participant A with 30 SEK deduction.

What is participant A's income? _______________

What is participant B's income? _______________

What is participant C's income? _______________

2. Participant A assigns 40 SEK to participant B.

a) Participant C charges participant A with 0 SEK deduction.

What is participant A's income? _______________

What is participant B's income? _______________

What is participant C's income? _______________

b) Participant C charges participant A with 15 SEK deduction.

What is participant A's income? _______________

What is participant B's income? _______________

What is participant C's income? _______________

Decision sheet participant B.

Below we ask you to estimate the other participants' decisions. This estimate is to be made for every possible decision A and C can make. Enter your estimates below on this sheet. Estimate the amount you believe A will give to you in the first stage, and how much you believe participant C will deduct. We want you to make an estimate for every possible decision that participant A can make. In the box to the right of the number 0, you enter the number of SEK you believe participant C transfers to A in the event that A grants participant you 0 SEK. In the box beside the number 10 you enter the number of SEK you believe participant C transfers to A if A chooses to grant you 10 SEK, and so on.

Your estimation

A may transfer 0, 10, 20, 30, 40, or 50 SEK to you
How many SEK do you think A transfers? ________________..

How do you believe is participant C going to decide?

Number of SEK participant A grants you	Number of SEK deducted by participant C
0	
10	
20	
30	
40	
50	

When you have estimated the other participants' behavior, please turn the page and respond to a few survey questions.

Survey questions

1. What are you studying? ___________________________________

2. I am a *woman* ☐ *man* ☐

3. I am __________ *years old.*

4. How do you think that your average monthly income relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

5. How do you think that your wealth relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

5. What do you think is a fair allocation of the 100 SEK?

Instructions for participant C

Welcome to this study in economics!

Please read the following instructions carefully. Depending on how you and your counterparts decide, you can earn money in addition to the 50 SEK you earn by participating. Therefore, it is important that you read and follow the instructions.

Please do not talk during the study. If you have any questions, raise your hand and we will come to answer your question.

Throughout this study you will use Swedish crowns. The study comprises three types of participants: Participants A, participants B and participants C. **You are a participant C**. During the study, you will interact with one randomly assigned participant B and one randomly assigned participant C.

Description of the two parts of the study

Stage one

In the first stage, participants A are the sole decision-makers. At the beginning of the stage, participant A gets an endowment of 100 SEK. You as a participant C get an endowment of 50 SEK. Participant B gets no endowment. Participant A must decide how many of her/his 100 SEK s/he wishes to assign to participant B. S/he can transfer to participant B a number of SEK between 0 and 50 in a multiple of tens, i.e. 0, 10, 20, 30, 40, or 50 SEK. If, for example, participant A grants participant B 40 SEK, her/his income at the end of stage one will amount to 60 SEK, and participant B's income will amount to 40 SEK. If s/he grants participant B 10 SEK, her/his income will be 90 SEK, and participant B's income will be 10 SEK at the end of stage one. If s/he grants participant B 0 SEK, her/his income, at the end of stage one, will result in 100 SEK, and participant B's own income will result in 0 SEK.

Stage two

In stage two, you, as participant C, are the only one to make a decision. You can pay to deduct SEK from participant A. Each SEK you deduct from participant A diminishes your income by 1 SEK and participant A's income by 3 SEK. You can assign any number of SEK between 0 and 50. Suppose you deduct 2 SEK to participant A, your income will be reduced by 2 SEK, and participant A's income will be reduced by 6 SEK. If you assign 19 SEK to participant A, your income is diminished by 19 SEK and participant A's income is reduced by 57 SEK. Neither participant A nor participant B knows your identity.

This is how we calculate participants A's, B's, and C's respective incomes:

Participant A's income amounts to

+ 100 SEK (participant A's endowment prior to stage 1)
- number of SEK assigned to participant B by participant A
- 3 times the number of deduction SEK transferred to participant A by participant C

Participant B's income amounts to

+ number of SEK assigned to participant B by participant A

Your income amounts to

+ 50 (your endowment)
- number of deduction SEK charged participant A by participant C

Control Questions

<u>It is important that all participants have understood the rules of the game. Therefore we ask you to answer the following control questions. When you have finished, signal to us by raising your hand.</u>

A. Participant A assigns 0 SEK to participant B.

a) Participant C charges participant with 0 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

b) Participant C charges participant A with 30 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

2. Participant A assigns 40 SEK to participant B.

a) Participant C charges participant A with 0 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

b) Participant C charges participant A with15 SEK deduction.
What is participant A's income? _______________
What is participant B's income? _______________
What is participant C's income? _______________

A: ______________________________

B: ______________________________

Decision sheet participant C.

In the first part participant A decides how many SEK to give to B. In the second part you decide how many SEK to deduct. We ask you to state your decision for **every possible decision** that A may make.

Your decision

Number of SEK participant A grants participant B	**Number of SEK you transfer to participant A (prior to the sum being tripled)**
0	
10	
20	
30	
40	
50	

When you have taken your decision, please turn the page and respond to a few survey questions.

Survey questions participant C

1. What are you studying? ___________________________________

2. I am a *woman* ☐ *man* ☐

3. I am __________ *years old.*

4. How do you think that participant A's income relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

5. How do you think that participant A's wealth relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

6. How do you think that your average monthly income relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

7. How do you think that your wealth relates to the average student?

_ Below average _ Close to average _ Above average _ Don't know

8. What do you think is a fair allocation of the 100 SEK?

9. What is the reason behind the decision you made?

10. Did you know any of the other participants you interacted with?

7 Pubertal timing and educational outcomes

Anna Dreber, Emma von Essen, and Eva Ranehill

We are very grateful for comments from Johan Almenberg and Magnus Johannesson. Financial support from the Jan Wallander and Tom Hedelius Foundation and the Carl Silfvén Foundation (E.R) is gratefully acknowledged, as well as financial support from the Swedish Council for Working Life and Social Research (FAS).

ABSTRACT: Education has important short and long run implications for individual outcomes. In this paper we explore the association between puberty and educational outcomes in a sample of Swedish girls. Previous research suggests that girls that mature earlier perform worse in school compared to girls that mature later. To test if this is also true among Swedish girls, we investigate the association between pubertal development and grades, educational aspirations and educational choice. We also investigate whether changes in risk attitudes, time preferences and priorities concerning school versus friends mediate this potential correlation. We confirm that earlier maturing girls have lower grades and lower educational aspirations, but find that they make educational choices similar to those of later maturing girls. Furthermore, we do not find that these differences in grades and aspirations are mediated by risk attitudes, time preferences or priorities.

7.1 Introduction

Educational outcomes have important impacts on the individual, in the short run as well as in the long run. For example, secondary school outcomes correlate with important subsequent outcomes in life through their effect on college enrollment, childbearing, income, health and mortality (Angrist and Krueger 1991, Pallas 2000). From a policy perspective, it is therefore important to study the determinants of the individual variation in educational outcomes.

In economics, the literature studying educational outcomes has mainly focused on family characteristics, such as parental education, income and occupation, as well as individual characteristics, such as gender and birth month (Mehgir and Palme 2004, Fredriksson and Öckert 2009, Björklund et al. 2010). Meanwhile, a number of studies in sociology and developmental psychology have pointed to the importance of pubertal development for educational outcomes. Puberty typically occurs around the ages 10-14 among girls, and is a period of major physical, hormonal, psychological and behavioral change. Some studies find that, on average, girls that mature earlier have lower grades (Simmons and Blyth 1987, Dubas et al. 1991, Cavanagh et al. 2007), lower academic goals (Graber et al. 1997), and a higher probability of dropping out of school early (Cavanagh et al. 2007). However, the relationship between early puberty and educational outcomes among girls is not always found (Stattin and Magnusson 1990, Dubas et al. 1991, Graber et al. 1997, Koivusiltay and Rimpelä 2004).[130] For boys, the relationship appears to be the opposite; earlier maturing boys typically perform better than later maturing boys. In this paper we study how pubertal development in girls is associated with educational outcomes among a sample of 344 adolescents in Sweden.[131]

Pubertal development could affect educational outcomes through various channels. A potential channel is through changes in risk attitudes and time preferences. One of the most salient characteristics of adolescence is an increase in behaviors with inherently risky and impulsive elements, such as drinking, smoking, and engaging in unprotected sex (Arnett 1999, Boyer 2006, Steinberg 2010). These behaviors have previously also been linked to low academic achievement among adolescents.[132]

The onset of pubertal development occurs through hormone signals from the brain to the reproductive system, which thereafter produces hormones that affect the brain and other organs (Ellison 2001). There is some, albeit mixed, evidence of correlations between hormones, risk and time preferences (e.g. Takahashi et al. 2007, Apicella et al. 2008, Sapienza et al.

---

[130] Nevertheless, even in studies where there is no correlation between puberty and grades, such as in a previous Swedish study by Stattin and Magnusson (1990), early maturing girls experience school as more negative and play truant to a larger extent than later girls. Previous literature also shows that other incidences of negative consequences in relation to puberty among girls are the largest among those that mature early, when it comes to for example anxiety, depression, eating disorders, and substance abuse (see Mendle et al. 2007 for review). Early maturing girls are physically different from their same age peers and this may also lead to a negative self-appraisal.

[131] Adolescence is often referred to as the psychosocial transition between childhood and adulthood, and puberty, in a strict sense, refers to the physical sexual maturation. Adolescence overlaps somewhat with puberty, where the former is often roughly considered to be the period between 13 and 19 years of age.

[132] For a review of this literature see http://aspe.hhs.gov/hsp/riskybehav01/index.htm, accessed the 15 of April 2011.

2009, Zethraeus et al. 2009). It is thus possible that these hormonal changes during puberty affect risk and time preferences.

A second channel through which pubertal development could affect educational outcomes is priorities regarding school work vs. friends and romantic interests. Changes in priorities could be influenced by changes in preferences for these activities, or by differential treatment in the social environment, where early maturing girls stand out and are given different attention by e.g. boys and parents compared to later maturing girls. Support for this reasoning is given by studies showing that girls who mature earlier are more likely to select into, and to be selected into, peer groups with older boys and girls that are characterized by riskier behavior and lower academic achievement (Stattin and Magnusson 1990, Haynie 2003).[133]

In our study, 344 girls are sampled in the 9th grade when they are 15-16 years old. This is the last year of compulsory education in Sweden and the year when students make their choice of secondary education. The educational outcomes we measure are grade point average (GPA), educational aspirations and educational choice, where the latter is indicated by the choice of vocational or academic track in secondary education. We further use three measures of self-reported pubertal development. Pubertal timing, the age when menarche occurs for girls, is our first measure of pubertal development.[134] Our second measure is a compound variable of relative pubertal development in five areas of physical change, where the participants rate their development in relation to other girls of the same age. Our third measure focuses on one of these areas, namely relative breast development.

We find that girls that mature earlier, measured through pubertal timing, have lower grades as well as lower educational aspirations. Conversely, the compound measure of relative pubertal development is consistently insignificant. Relative breast development, however, is significant as an explanato-

---

[133] If early maturing girls have more peers with riskier behavior, and value friends more than school, this further supports the importance of looking at risk preferences since the peer network might reflect underlying similarities in preferences rather than peer effects.

[134] For a discussion regarding the onset of menarche as a measure of pubertal development see Dorn et al (2010).

ry variable in the regressions with GPA and educational choice as dependent variables.

We also attempt to understand whether the two different channels that we propose can mediate the relationship between pubertal development and educational outcomes. We thus measure attitudes for risk as well as time preferences (i.e. patience). To our knowledge, this has previously not been explored in the literature linking pubertal development to educational outcomes. We further measure the subjective importance of school versus peers. In line with previous literature, we find that patience has positive implications for educational outcomes. In contrast, unlike previous literature we find that a high risk taking propensity has negative implications for educational outcomes. However, we do not find any evidence that risk attitudes, time preferences or changes in priorities regarding school versus friends mediate the relation between puberty and educational outcome. There is no correlation between the potential mediating factors and pubertal development.

Moreover, age at menarche is partly heritable (Ellison 2001), and the impact of early pubertal development on educational outcomes may thus be overestimated through the impact of socio-economic background on educational outcomes. However, there is no clear evidence of a correlation between pubertal development and socio-economic background, and the effect of pubertal development on educational outcomes is generally robust to controlling for the parents' socio-economic background.[135] Here we thus include a set of demographic variables as additional control variables: age, parental education and the number of siblings. We find that our results are robust to the inclusion of these controls.

There is further some relevant literature in economics, e.g. Pekkarinen (2005) presents suggestive evidence of the impact of puberty on educational choice. Investigating the effect of a change in the tracking age to secondary education from 11-12 to 15-16 it is found that this favors girls compared to

---

[135] For example Windham et al. (2009) and Obeidallah (2000) find that higher socio-economic status is correlated with lower age at menarche, Semiz et al. 2009 find no relationship, and Short and Rosenthal (2008) as well as Semiz et al. (2009) find that in-family stress such as disease, conflict or absent fathers is associated with lower age at menarche.

boys. The author argues that girls at the age of 15 or 16 have reached the end of puberty, whereas boys are in the middle of it, and that being in the middle of puberty has adverse effects on educational aspirations. We find that pubertal development measured by pubertal timing is of importance for educational outcomes in girls.

In sum, our findings suggest that absolute pubertal timing matters for GPA and for educational aspirations, whereas relative breast development correlates with GPA and educational choice. Later maturing girls thus appear to have higher GPA and educational aspirations and to be more likely to choose an academic track for secondary education. Risk attitudes and time preferences seem to influence educational outcomes, but do not appear to mediate the relationship between pubertal development and educational outcomes. An important caveat is that our study does not allow us to infer causality. Our results should be seen as a first step investigating the relationship between pubertal timing and educational choice from an economic perspective. Future studies should attempt to further investigate the mechanisms behind the correlation between pubertal timing and educational outcomes, preferably in a large longitudinal sample.

The outline for the paper is the following. In section 2, we present the survey design. Section 3 presents our results, and section 4 concludes with a discussion of our findings.

7.2 Design of study

7.2.1 Survey description

All relevant schools in the Swedish cities Stockholm and Malmö that had contact information on their webpage were contacted via email.[136] 11 schools agreed to participate in the study. Though we have selection at the school level that we cannot control for, all students present at the day of

[136] The principal and the study and career advisor of all schools with grade 9 were contacted. Schools with a particular religious or pedagogical focus were not contacted. A comparison of the data collected in each city reveals that whereas GPA is somewhat higher in the Stockholm sample (p=0.093) the educational aspirations are somewhat lower (p=0.099). Further, the Stockholm sample is slightly less risk taking (p=0.012) and have 0.3 more siblings (p=0.087).

the survey participated. The study was conducted in April and May 2009 and 2010, just after the students had made their choices of specialization to secondary education. We thus have data on two different cohorts. The survey was introduced as part of the school curricula during a regular school class, headed by a teacher, the school nurse, and/or a study and career advisor depending on the preference of the school.[137] The survey consisted of four parts.

The first part included hypothetical measures of risk attitudes and time preferences. Risk attitudes are measured by a question where the subjects are asked to self-report their general risk taking propensity on a scale from 1 to 10, where 10 is "very risk taking" and 1 is "not risk taking at all". This measure was used in Dohmen et al. (2011) where it was found to predict incentivized risk taking as well as risk taking in other domains. Time preferences, i.e. patience, were assessed through a set of questions where participants had to choose between hypothetical money "now" or hypothetical money "later". The amount of money ”later" was fixed whereas money "now" increased for each pair of alternatives. 23% of the participants provided inconsistent answers (i.e. switched between money now and money later multiple times). We therefore used the number of choices for money later, of 19 possible pairs of alternatives, as our variable for patience.

The second part inquired about school related variables. From this part we created three outcome variables; GPA, educational aspirations and the choice of secondary education. In Sweden, GPA is specified every semester from the 8th grade and onwards, consisting of an average of the grade in each course weighted by size of the course (in number of hours).[138] To some extent the GPA obtained correlates with subsequent educational choices, as some popular specializations require a higher GPA. It is not, however, the case that an academic specialization always requires a higher GPA than a vocational educational choice. In order to measure educational

[137] Every school in Sweden has a career advisor in order to inform students about alternative future educational options.

[138] The grading scale has four levels: "fail", "pass", "pass with distinction" and "excellent" (the authors' translation), where fail corresponds to 0 points, pass corresponds to 10 points, pass with distinction corresponds to 15 points and excellent corresponds to 20 points.

aspiration, we asked the participants to state the highest type of diploma they wished to obtain. This variable consisted of four categories where higher numbers implied higher diplomas (1 implies diploma from compulsory school, 2 diploma from gymnasium, 3 diploma from tertiary education, excluding university, and 4 diploma from university). We also included a variable for educational choice, indicating whether the student had chosen an academic specialization or a vocational specialization as secondary education. Sweden has 9 years of compulsory schooling, starting the year a child turns 7. In the 9th grade, the large majority of students choose a specialization for secondary education.[139] At the time of the study there were 17 possible different specializations; 15 vocational and 2 academic. All specializations comply with the minimal standards for access to tertiary education. However, most higher education requires complementary studies unless students have attended one of the two academic specializations. The choice consists of up to three ranked pairs of schools and specializations. We focus on the first pair; the participant's favored choice, creating a binary choice variable.

In the second part we also included additional school related measurements such as the importance of friends in relation to the GPA obtained, time spent studying and time spent with friends, as well as parents' educational aspirations. The questions pertaining to the importance of school and friends both assessed answers on a scale from 1-10. We divided the answer on importance of school with the answer on importance of friends to measure the relative value the participants' placed on school to friends.

The third part of the survey investigated puberty and health related outcomes. In the literature puberty is measured in a number of ways, ranging from invasive and non-invasive clinical examinations, to self-reported measurements. In the context of the present study only the latter approach was considered appropriate.[140] We have three measures of pubertal develop-

[139] There are only 9 students in our sample dropping out of school after the 9th grade. Hence we cannot use drop out or not as an outcome variable.

140 In our sample most female participants are at a later stage of their pubertal development. Clinical measurements would probably have required repeated measurements or a measurement at a point earlier in time to provide the required variation. Self-reported mea-

ment; one measure of pubertal timing and two measures of relative pubertal development. In order to measure pubertal timing we ask the girls to state the year and the semester when they received their first menstruation. Pubertal timing is simply the age of menarche. We also include a set of questions on relative pubertal development. These are based on the Adolescent Scale (AS), a self-reported measure created by Kaiser and Gruzelier (1999). This is a brief scale that intends to measure pubertal development relative to others of the same sex and age.[141] The AS consists of a set of questions, asking the respondent to rate their status of physical pubertal maturation relative to other girls of the same age based on six criteria; breast development, growth spurt, body hair, skin changes, menstruation, and general development. Participants were asked to pin down the ratings on a 5 degree scale, where 1 corresponded to "much earlier than other girls", 2 to "somewhat earlier than other girls", 3 to "about the same as other girls", 4 to "somewhat later than other girls", and 5 to "much later than other girls".[142] From the AS we created two variables of relative pubertal development. First we use five of these questions (all but the question on general development) to create a compound variable that we refer to as relative pubertal development.[143] Second, since breast growth is arguably the most parable change to others we used this question also as a separate variable.[144]

---

surements may also more accurately reflect the individual perception of relative pubertal development, which is partly what we are interested in investigating,

[141] The most widely used self-report measurement of relative pubertal timing for pubertal development is the Pubertal Development Scale, PDS (Petersen et al. 1988). This measure is suitable for longitudinal studies. The AS is better suited for the type of study that we do.

[142] The exact question read "For each question, category a-f below, please indicate how you think your development in this area corresponds to other girls your age by ticking the alternative that you think describes you the best".

[143] The compound variable is highly correlated with self-reported general development (coefficient 0.698 and $p<0.001$).

[144] This part of the survey also included questions on height, weight, exercise, life satisfaction, "locus of control", the importance of having a partner and the importance of being good looking. We did not however use these variables in this paper since our sample is too small to use all variables in the analysis. We nevertheless chose to include these in the questionnaire for the purpose of future research studies.

The last part of the survey included demographic questions such as year and month of birth, number and sex of siblings and parental education.[145] Parental education was measured similarly to the educational aspiration level, though we differed between theoretic and vocational secondary education implying that this variable has 5 categories where, as before, higher numbers pertain to higher diplomas. Table A1 in the appendix provides a list of all variables included in the survey.[146]

Our dataset allows us to identify correlations only, and though participating schools come from areas with different socio-economic background, generalizations should be made with great caution. Further, a longitudinal approach also including clinical measurements of pubertal development would of course have increased the quality of our data. This study should therefore be seen as a first attempt to study the influence of puberty and its mediating mechanisms on educational choices.

7.2.2 Hypotheses and tests

In accordance with most previous literature we hypothesized that early maturing girls would obtain lower grades and have lower educational aspirations than their later maturing peers. In addition, previous literature finds that early girls are more exposed to older peers and deviant behavior. They have also been found to experience school more negatively. We thus also hypothesized that early girls would exhibit lower motivation for studies and therefore have lower educational aspirations and be more likely to choose vocational tracks than their later maturing peers. Moreover, the measures of relative pubertal development further allow us to explore whether girls' assessments of their relative development matter as much as pubertal timing, which is more about absolute timing. If this is the case, it could suggest an important role of feedback from the environment.

---

[145] This part also included questions about parental occupation, the respondent's origin and religiosity. The question about parental occupation was an open question and unfortunately the quality of the data was too bad to be included in the analysis. Origin, sex of siblings and religiosity were also not used in the analysis.

[146] We also collected the corresponding data for boys in the surveyed classes. However, among boys partial attrition was much larger. In this paper we thus only focus on girls.

We further had some expectations regarding mediating variables. We hypothesized that early girls would be more risk taking and impatient, as well as more prone to rate the importance of school to friends lower than their later peers. In turn, we hypothesized that risk taking and impatience would be negatively correlated with educational outcomes whereas the correlation between the importance of school to friends would be positively correlated.

7.3 Results

7.3.1 Descriptive statistics

A total of 344 girls participated in the survey. Table 1 below presents descriptive statistics and attrition for the variables used in the analysis. At the time of the study the participating girls are on average 15.9 years old, and reached menarche at the average age of 12.8.[147] The compound variable of relative pubertal development shows that girls on average find their pubertal development as well as their breast development to be about the same as other girls. The median girl has a grade point average of 236 on a scale ranging up to 320. In terms of educational aspiration, 68% of the girls in our sample aspire to get a university education whereas 73% of the girls chose an academic specialization when it comes to secondary education.[148]

7.3.2 Regression analysis

We look at how pubertal timing, relative pubertal development and breast development correlate with grades, educational aspirations and educational choice in separate regressions.[149] Our main analysis is a regression analysis, based on OLS regressions.[150]

---

[147] The true mean is somewhat higher as 10 girls stating that they had not yet reached menarche were excluded from this statistic.

[148] Only 2.6% of our sample did not indicate a choice of secondary education.

[149] For each pubertal development variable, all regressions are run with a sample not including partial attrition so that we can compare significance and effect sizes with and without controls.

[150] To control that our results are not dependent on specification, functional form, or regression method we further conducted a logit regression for the binary outcome variable of

Table 1. Descriptive statistics, 344 respondents

Variable	mean	p50	N	sd	se(mean)	min	Max
Timing	12.75	12.71	324	1.07	.06	10.67	16
Puberty 5	2.86	3	338	.58	.03	1	5
Breast	2.87	3	338	.88	.05	1	5
Continue studying	.97	1	343	.16	.01	1	1
GPA	236	240	301	46.51	2.68	95	320
Aspirations	3.49	4	297	.83	.05	1	4
Educational choice	.73	1	334	.44	.02	0	1
Risk	5.89	6	336	1.70	.09	1	10
Patience	10.01	9	344	5.86	.32	0	19
School to friends	1.06	1	341	0.53	0.03	0.22	7
Age	15.87	15.92	337	.34	.02	14.33	16.83
Education father	4.08	5	283	1.51	.09	1	6
Education mother	4.09	4	287	1.32	.08	1	6
# siblings	1.63	1	338	1.05	.06	0	5

7 participants did either omit the year or the month they were born and 2 girls did not answer whether they had reached menarche. An additional 3 girls answered that they had not yet had their menses, thus these girls were excluded in the main analysis. In the remaining sample, 8 girls omitted information about which school year they reached menarche. For the 14 girls that did not state which term they got their first menstruation, we assigned the timing to be the average of that school year (between fall and spring semester).

We conduct three types of regressions for each pubertal and educational variable. First we study the educational variables only including each of the three different pubertal development variables separately. Second we add the three variables we expect to be mediating the effect of puberty on educational outcomes. Third we include a set of demographic control variables that could be important in understanding educational outcomes in each

---

educational choice, and a tobit regressions for the truncated outcome variable educational aspirations, see table A4 and A5 in the appendix.

separate regression. This provides us with nine separate regressions per educational choice variable.

7.3.2.1 Grades

Table 2 below shows that pubertal timing appears to be of some importance for grades. When we only use pubertal timing as a regressor, we find that it is significantly positively correlated with grades (p=0.030). Everything else equal, reaching menarche one year later corresponds to an increase in grades of about 7 points in our sample, i.e. an improvement of about 0.15 standard deviations.[151] When we add the potential mediating variables, we find that effect of pubertal timing remains about the same in size and significance (p=0.027).[152] Adding the demographic control variables age, parental education and number of siblings, the effect of pubertal timing decreases to about 6 points and becomes marginally significant (p=0.057).

We do not find any evidence of relative pubertal development being significantly related to GPA. When including relative breast development by itself, however, it is positively and significantly related to grades (p=0.020). The result is similar when we include the potential mediating factors (p=0.036), and as for pubertal timing, the effect is lower and marginally significant when we add the other controls (p=0.058). Since a higher value on breast development corresponds to later development, this supports our hypothesis.

Time preferences appear to influence GPA; patience correlates positively with grades. There is also some evidence of a negative relationship with risk taking.[153] We find no evidence of a correlation between grades and priorities regarding how much girls' value school relative to friends. Among the socio-demographic variables, only the father's education is marginally significantly correlated with GPA.

---

[151] When running regressions without control variables pubertal timing has a larger effect size, 6.94 GPA.

[152] The results from a correlation analysis also confirms that the puberty variables are uncorrelated with the mediating variables risk attitudes, time preferences and priorities regarding the importance of friends to school.

[153] This relationship is not dependent on what pubertal development variables we use.

7.3.2.2 Educational aspirations

We next turn to the impact of the puberty variables on educational aspirations. Our results indicate that pubertal timing is positively related to educational aspirations when not controlling for anything else (p=0.002), when controlling for the potential mediators (p=0.002) and when including the other socio-demographic variables (p=0.009).[154] We find no significant relationship between educational aspirations and either relative pubertal development or relative breast development.

When it comes to risk attitudes and time preferences, we find that risk taking is significantly negatively related to educational aspirations, where risk taking individuals appear to be more likely to aspire to quit school early. There is some evidence of a positive correlation between patience and aspirations, when using the relative pubertal development variables. The father's education is positive and marginally significant in all specifications. Priorities regarding how much girls value school relative to friends and love interests seem to have no impact on aspirations.

7.3.2.3 Educational choices

Analyzing the impact of puberty on the choice of specialization, we find no relationship between educational choice and either pubertal timing or the compound relative pubertal development variable.[155] Relative breast development, however, is positively and significantly or marginally significantly related to educational choice (p=0.034, p=0.044, or p=0.076). Even if the significance level changes the coefficient remains rather stable across regressions. This indicates that girls that develop breasts relatively late are more likely to choose an academic track in secondary school. We also find that patience is positively significantly correlated with choosing the academic track across regressions.

---

[154] Running the same set of regressions while controlling for GPA diminishes the effect of pubertal timing on educational aspirations to about half (see table A2). Further, the coefficient is only significant at the 10% level in the first two specifications, and not at all in the third.

[155] Similarly to the previous case, running the same set of regressions controlling for GPA diminishes the effect of relative breast development to about half the size noted in table 4 (see table A3). Further the coefficient is no longer significant.

Table 2. Pubertal development and grades

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	6.940**	6.562**	5.826*						
	(2.179)	(2.226)	(1.916)						
Puberty 5				3.411	2.383	1.447			
				(0.618)	(0.464)	(0.277)			
Breast							8.331**	7.271**	6.654*
							(2.338)	(2.108)	(1.908)
Risk		-3.282*	-3.375*		-3.299*	-3.337*		-2.773	-2.864*
		(-1.875)	(-1.958)		(-1.874)	(-1.930)		(-1.590)	(-1.658)
Patience		2.126***	2.051***		2.149***	2.063***		2.125***	2.047***
		(4.061)	(4.032)		(4.162)	(4.084)		(4.161)	(4.086)
Priorities		4.683	6.650		3.767	5.988		4.535	6.542
		(0.738)	(1.094)		(0.600)	(1.015)		(0.695)	(1.061)
Age			-1.527			-0.363			0.919
			(-0.167)			(-0.0414)			(0.108)
Educ father			4.786*			5.019*			4.866*
			(1.677)			(1.780)			(1.716)
Educ mother			1.270			1.978			1.814
			(0.431)			(0.677)			(0.623)
# siblings			-3.914			-3.961			-3.925
			(-1.315)			(-1.359)			(-1.321)
Observations	235	235	235	239	239	239	239	239	239
R-squared	0.025	0.132	0.167	0.002	0.107	0.150	0.026	0.126	0.166

Robust t-statistics in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3. Pubertal development and educational aspirations

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	0.150***	0.146***	0.124***						
	(3.153)	(3.125)	(2.630)						
Puberty 5				-0.028	-0.042	-0.049			
				(-0.373)	(-0.573)	(-0.712)			
Breast							-0.001	-0.028	-0.033
							(-0.0124)	(-0.499)	(-0.592)
Risk		-0.076**	-0.080**		-0.076**	-0.079**		-0.078**	-0.082**
		(-2.209)	(-2.313)		(-2.164)	(-2.260)		(-2.204)	(-2.308)
Patience		0.014	0.013		0.018**	0.017*		0.018**	0.017*
		(1.552)	(1.489)		(1.993)	(1.960)		(1.972)	(1.935)
Priorities		-0.147	-0.120		-0.167	-0.138		-0.169	-0.140
		(-1.206)	(-1.117)		(-1.417)	(-1.359)		(-1.430)	(-1.380)
Age			0.112			0.180			0.180
			(0.795)			(1.301)			(1.304)
Educ father			0.081*			0.082*			0.081*
			(1.709)			(1.728)			(1.732)
Educ mother			0.074			0.084			0.085
			(1.355)			(1.503)			(1.506)
# siblings			-0.044			-0.036			-0.036
			(-0.802)			(-0.663)			(-0.675)
Observations	224	224	224	230	230	230	230	230	230
R-squared	0.042	0.080	0.137	0.000	0.047	0.116	0.000	0.047	0.117

Robust t-statistics in parentheses. *** p<0.01, ** p<0.05, * p<0.1

Table 4. Pubertal development and educational choice

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	0.019	0.017	0.015						
	(0.718)	(0.687)	(0.564)						
Puberty 5				0.044	0.038	0.029			
				(0.969)	(0.849)	(0.620)			
Breast							0.060**	0.057**	0.052*
							(2.128)	(2.028)	(1.785)
Risk		0.003	0.004		-0.002	-0.001		0.002	0.003
		(0.173)	(0.216)		(-0.111)	(-0.0612)		(0.139)	(0.152)
Patience		0.014***	0.014***		0.014***	0.013***		0.014***	0.013***
		(3.074)	(2.940)		(2.999)	(2.825)		(2.999)	(2.837)
Priorities		0.048*	0.057**		0.037	0.050*		0.042	0.053*
		(1.728)	(2.044)		(1.417)	(1.833)		(1.548)	(1.915)
Age			-0.021			-0.022			-0.013
			(-0.288)			(-0.319)			(-0.184)
Educ father			0.003			0.005			0.005
			(0.115)			(0.220)			(0.205)
Educ mother			0.029			0.036			0.034
			(1.139)			(1.407)			(1.328)
# siblings			0.004			0.004			0.004
			(0.154)			(0.151)			(0.143)
Observations	253	253	253	258	258	258	258	258	258
R-squared	0.002	0.050	0.059	0.004	0.047	0.062	0.016	0.059	0.072

Robust t-statistics in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

7.4 Discussion

Given the short and long run impacts of educational outcomes for the individual, it is important to understand the determinants of e.g. GPA, educational aspirations and educational choice. In this study we replicated the common finding that girls that mature early perform worse in school and have lower educational aspirations with a sample of 15-16 year old girls in Sweden. We also found some results suggesting that later maturing girls, when it comes to relative breast development, are more likely to choose the academic track in high school. Moreover, we explored possible mediating factors in order to explain this relationship. We hypothesized that changes in risk attitudes and time preferences, perhaps associated with hormonal changes during puberty, were one channel through which puberty could affect educational outcomes. We also hypothesized that changes in priorities, where earlier girls would put less emphasis on school relative to friends, would be another mediating factor. We found no evidence of any of these variables mediating the correlation between early pubertal development and educational outcomes. However, this study should be seen as an exploratory attempt, and not as a conclusive study on the role of these mediating factors.

Puberty is typically related to an increase in behaviors that are associated with risk taking and impulsivity. However, when it comes to comparing different age groups in studies in economics and developmental psychology, most of the focus has been on adolescence rather than puberty. Some studies find that adolescents are more risk taking and less patient than other groups (e.g. Green et al. 1994, 1997, 1999, Steinberg et al. 2008, Burnett et al. 2010), whereas other studies find a linear decrease in risk behavior from childhood to adulthood (Bettinger and Slonim 2007, van Leijenhorst et al. 2008), and yet others do not find a difference across age groups (Harbaugh et al. 2002, Crone et al. 2010, Sutter et al. 2010). However, puberty and adolescence only overlap partially, thus it would be of interest to focus on whether boys and girls at different stages of puberty, and not just adolescents, act differently than other groups. Moreover, it is not clear whether experimentally elicited preferences for risk and time correspond easily to the propensity to engage in the risky and impulsive behaviors that typically are associated with puberty (see e.g. Sutter et al. 2010). This might be one explanation for why we do not find risk attitudes and time preferences to be mediating the relationship between pubertal development and educational outcomes.

Another reason we do not find any mediating effects of risk attitudes and time preferences could be due to the fact that we measure these at a point in time where most girls have reached a more advanced pubertal status. Potentially, differences in these preferences are larger when pubertal discrepancies are more important. However, our results suggest that risk attitudes and time preferences correlate with educational outcomes. Patience correlates positively with educational outcomes and risk taking negatively. A handful of studies have previously related educational outcomes to experimentally elicited preferences for risk and time. Benjamin et al. (2006) find that risk taking is positively correlated with standardized test scores, whereas Sutter et al. (2010) find no correlation between risk preferences and GPA. Patience has been found to correlate negatively with deviant behavior in school (Castillo et al. 2008) and positively with GPA (Kirby et al. 2005, Benjamin et al. 2006, Sutter et al. 2010). However, studying patience, cognitive capacity and imaginative powers Borghans and Goldsteyn (2004) find a slightly more complicated picture. In their study, individuals with high time discounting (impatience) have lower grades but stay longer in college since they also have lower ability to imagine the future. Even if we use hypothetical measures of time preferences compared to other studies in economics our results are similar. When it comes to risk attitudes, however, our results differ somewhat from what previous economic literature has found.

Puberty could further affect decision making pertaining to education through both its hormonal effects on the brain, and through its effect on how one is treated by the social environment. The latter is partly what we aimed to capture by the relative pubertal development measures we used, and in particular what we had in mind when we looked at the relative breast development variable, since this is arguably the most parable physical change of sexual character for girls during this period. Support of this reasoning comes from a study where relatively early maturing girls are shown to be treated differently by their peers, for example by boys (Stattin and Magnusson 1990). This might cause early maturing girls to change their behavior. With reliable measures of pubertal development and complete information on peer groups this could be explored further. Another natural extension is also to study same-sex schools, in order to see whether the effects are similar in those schools compared to mixed schools.

Our study highlights the importance of including pubertal development measures in studies regarding educations outcomes. Nevertheless, a number of caveats should be kept in mind when interpreting the results. Though we have no selection into our study at the student level we have selection at the school level, and the sample included in this study is unlikely to be representative of the population in Sweden as a whole. Sampling schools at the end of the last semester of the compulsory school, when schools with less advantaged students work hard to get as many students to pass as possible, may have led to a selection towards the upper end of the spectra of socio-economic status. Further, sampling schools in two of the biggest cities in Sweden probably exacerbated this. This is also apparent when we look at the data. For example, our sample has a higher GPA than the national average, even if we compare with the average in big cities. One can only speculate whether the impact of puberty would have been greater or not had we had access to a different sample. However our estimates are not likely to be an overestimation of the true effects. For example, with respect to educational choice, where we find the weakest results, it is for example worth noting that all (50) of the students in one of the participating schools chose an academic specialization. Pubertal timing may have a larger impact on educational outcomes among students from less affluent areas. Further, we only investigated girls and only relied on self-reported answers related to pubertal timing and relative pubertal development. For future research, larger longitudinal studies with more objective measures should be used to explore potential impact of gender differences regarding pubertal development and educational outcomes.

7.5 References

Aksglaede, L., Sørensen, K., Petersen, JH., Skakkebaek, NE., Juul, A., 2009, "Recent decline in age at breast development: the Copenhagen Puberty Study", *Pediatrics,* 123(5):932-9.

Angrist, J.D., Krueger, A.B., 1991, "Does Compulsory School Attendance Affect Schooling and Earnings?", *The Quarterly Journal of Economics,* 106(4): 979-1014.

Apicella, C.L., Dreber, A., Campbell, B.C., Gray, P.B., Hoffman, M., Little, A.C., "Testosterone and Financial Risk Preferences", *Evolution and Human Behavior*, 29(6), 384-390.

Arnett, J., 1999, "Adolescent storm and stress, reconsidered", *American Psychologist,* 54:317–26.

Benjamin, D.J., Brown, S.A., Shapiro, J.M., 2006, "Who is Behavioral? Cognitive Ability and Anomalous Preferences", Mimeo. Harvard University.

Bettinger, E., Slonim, R., 2007, "Patience among Children", *Journal of Public Economics* 91:343–363.

Björklund, A., Lindahl, L., and Lindquist, M. J., 2011, "What more than parental income, education and occupation? An exploration of what Swedish siblings get from their parents", *The B.E. Journal of Economic Analysis & Policy*, 10 (1).

Blumenthal, H., Leen-Feldner, E.W., Trainor, C.D., Babson, K.A., Bunaciu, L., 2009, "Interactive roles of pubertal timing and peer relations in predicting social anxiety symptoms among youth", *Journal of Adolescent Health*, 44:401-403.

Borghans, L., Golsteyn, B., 2005, "Imagination, Time Discounting and Human Capital Investments", Mimeo.

Boyer, T., 2006, "The Development of Risk-Taking: A Multi-Perspective View", *Developmental Review*, 26: 291–345.

Burnett, S., Bault, N., Coricelli, G., Blakemore, S.J., 2010, "Adolescents' heightened risk-seeking in a probabilistic gambling task", *Cognitive Development.* 25(2), 183–196.

Cavanagh, S., Riegle-Crumb, C., Crosnoe R., 2007, "Puberty and the education of Girls", *Social Psychology Quarterly*, 70:2, 186-198.

Castillo, M., Ferraro, P., Jordan, J., & Petrie, R, 2008, "The today and tomorrow of kids", Working Paper, Georgia State University.

Crone, E.A., Bullens, L., Van Der Plas, E.A.A., Kijkuit, E.J., Zelazo, P.D., 2008, "Developmental changes and individual differences in risk and perspective taking in adolescence", *Development and Psychopathology,* 20, 1213–1229.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp J., Wagner. G.G., 2011, Forthcoming. "Individual Risk Attitudes: Measurement, Determinants and Behavioral Consequences", *Journal of the European Economic Association.*

Dubas, Semon J., Graber, J.A., Petersen, A.C., 1991, "The Effects of Pubertal Development on Achievement during Adolescence", *American Journal of Education*, 99(4):444-460.

Duke, P. M., Carlsmith, M., Jennings, D., Martin J. A., Dornbusch, S. M., Gross, R. T., Siegel-Gorelick, B., 1982, "Educational Correlates of Early and Late Sexual Maturation in Adolescence", *The Journal of Pediatrics,* 100 (4).

Ellison, P.T., 2001, *On Fertile Ground*, Harvard University Press, Cambridge, MA.

Fredriksson, P., Öckert, B., 2005, "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance", IZA DP No. 1659.

Graber, J. A., Lewinsohn, P. M., Seeley J. R., Brooks-Gunn, J.,1997, "Is Psychopathology Associated With the Timing of Pubertal Development?", *Journal of the American Academy if Child and Adolescent Psychiatry,* 36 (12):1768-76.

Haynie, D.L., 2003, "Contexts of Risk? Explaining the Link between Girls Pubertal Development and Their Delinquency Involvement", *Social Forces*, 82:355-397.

Harbaugh, W.T., Krause, K., Vesterlund, L., 2002, "Risk Attitudes of Children and Adults: Choices over Small and Large Probability Gains and Losses", *Experimental Economics*, 5(1): 53–84.

Kaiser, J., Gruzelier, J.H., 1999, "The adolescence scale (AS-ICSM): a tool for the retrospective assessment of puberty milestones", *Acta Paediatrics* 88:64–68.

Kaltiala- Heino, R., Kosunen, E., Rimpela, M., 2003, "Pubertal timing, sexual behaviour and self- reported depression in middle adolescence", *Journal of Adolescence*, 26:531-545.

Kirby, C., Winston, G., Santiesteban, M., 2005,. "Impatience and Grades: Delay-Discount Rates Correlate Negatively with College GPA", *Learning and Individual Differences*, 15(3), 213-222.

Koivusiltay, L., Rimpelä. A., 2004, "Pubertal timing and educational careers: a longitudinal study", *Annals of Human Biology*, 31(4):446–465.

Lanza, S.T., Collins, L.M., 2002, "Pubertal Timing and the Onset if Substance Use in Females During Early Adolescence", *Prevention Science* 3(1):69-82.

Van Leijenhorst, L., Westenberg. P.M., Crone, E.A., 2008, "A Developmental Study of Risky Decisions on the Cake Gambling Task: Age and Gender Analyses of Probability Estimation and Reward Evaluation", Developmental Neuropsychology, 33:2,179-196

Mendle, J., Turkheimer, E., Emery, R., 2007, "Detrimental psychological outcomes associated with early pubertal timing in adolescent girls", *Developmental review,* 27:151–171.

Obeidallah, D., Brennan, RT., Brooks-Gunn, J., Earls, F., 2004 "Links between pubertal timing and neighbourhood contents: implications for girls' violent behaviour", *Journal of the American Academy of Child and Adolescent Psychiatry*, 43:1460-1468.

Pallas, A.M., 2000, The Effects of Schooling on Individual Lives, in *Handbook of the Sociology of Education*. Kluwer Academic/Plenum Publishers, New York.

Palme, M., Meghir, C., 2005, "Educational Reform, Ability and Parental Background", *American Economic Review*, 95(1), 414-424.

Pekkarinen, T., 2008, "Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment", *Scandinavian Journal of Economics*, 110 (4), 807-825, 2008.

Sapienza, P., Zingales, L., Maestripieri, D., 2009, "Gender Differences in Financial Risk Aversion and Career Choices are Affected by Testosterone", *Proceedings of the National Academy of Sciences*, 106(36): 15268-15273.

Simmons, R.G., Blyth, D.A., 1987, *Moving into Adolescence.* Hawthorne, N.Y.: Aldine De Gruyter.

Short, M.B., Rosenthal, S.L., 2008, "Psychosocial Development and Puberty", *Annals of the New York Academy of Scienes*, 1135: 36–42.

Stattin, Hakan and David Magnusson, 1990, *Pubertal Maturation in Female Development*, Hillsdale: Lawrence Erlbaum Associates.

Steinberg, L., 2010, "Commentary: A Behavioral Scientist Looks at the Science of Adolescent Brain Development", *Brain Cognition*, 72(1): 160-164.

Sutter, M., Kocher, M.G., Rützler, D., Trautmann, S.T., 2010, "Impatience and Uncertainty: Experimental Decisions Predict Adolecents' Field Behavior", Discussion Papers in Economics 12114, University of Munich, Department of Economics.

Takahashi, T., Sakaguchi, K., Oki, M., Hasegawa, T., 2008. "Sex hormonal modulation of hyperbolic discount factor in men", *Journal of Neuroscience, Psychology, and Economics*, 1(1):7–16.

Tanner, J.M., 1990, *Foetus into Man: Physical Growth from Conception to Maturity.* Harvard University Press.

Windham, G.C., Zhang, L., Longnecker, M.P., Klebanoff, M., 2008, "Maternal smoking, demographic and lifestyle factors in relation to daughter's age at menarche", *Paediatric and Perinatal Epidemiology.*, 22(6):551-61.

Zehr, J.L., Culbert, K.M., Sisk, C.L., Klump, K.L. 2007, "An Association of Early Puberty with Disordered Eating and Anxiety in a Population of Undergraduate Women and Men", *Hormones and Behaviour 52*

Zethraeus, N., Kocoska-Maras, L., Ellingsen, T., von Schoultz, B., Lindén Hirschberg, A., Johannesson, M., 2009, "A Randomized trial of the effect of estrogen and testosterone on economic behavior", *Proceedings of the National Academy of Sciences of the United States of America*, 106:6535-6538.

7.6 Appendix

Table A1. Variable description

Variable	Variable description
GPA	Grade Point Average
Aspiration	Highest diploma aimed four, 1= diploma compulsory school (9 years), 2= Diploma secondary education (12 years), 3= diploma tertiary education excluding university, 4= diploma from university
Educational choice	Choice of academic or vocational track in secondary education, 1= academic track as first choice
Pubertal timing	Age ate menarche, measured in years
Puberty 5	Average of five self-estimated ratings on pubertal progress in relation to same aged peers. The scale ranged from 1= much earlier than other girls, to 5= much later than other girls. The five estimations pertained to breast development, growth spurt, body hair, skin problems and menstruation
Breast	The self-estimated relative breast development used in puberty 5 on its own
Risk	Self reported general risk taking propensity, reported on a scale from 1= "not risk taking at all" to 10="very risk taking"
Patience	The number of patient choices in a hypothetical question involving a choice between money now and later. The later amount was consistently 200 SEK, whereas the mount to be obtained today ranged from 20 SEK to 200 SEK in brackets of 10 SEK
Priorities	The ration between a question asking participants to state the importance of getting good grades from 1-10, where 10 corresponded to very important, and a similar scale asking about the importance of friends
Age	Age in years
Educ Father	Father's education, 1=diploma from compulsory school, 2= diploma from vocational secondary education, 3= diploma from academic secondary education, 4= diploma from tertiary education excluding university, 5= diploma from university
Educ Mother	Mother's education, 1=diploma from compulsory school, 2= diploma from vocational secondary education, 3= diploma from academic secondary education, 4= diploma from tertiary education excluding university, 5= diploma from university
# siblings	Number of siblings.

Table A2. Educational aspiration controlling for grades

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	0.090*	0.086*	0.076						
	(1.867)	(1.776)	(1.540)						
Puberty 5				-0.080	-0.088	-0.081			
				(-1.174)	(-1.317)	(-1.254)			
Breast							-0.053	-0.076	-0.070
							(-0.950)	(-1.337)	(-1.269)
GPA	0.005***	0.005***	0.005***	0.006***	0.006***	0.005***	0.006***	0.006***	0.005***
	(4.429)	(4.048)	(3.707)	(4.782)	(4.291)	(3.878)	(4.860)	(4.481)	(4.052)
Risk		-0.076**	-0.077**		-0.072**	-0.072**		-0.078**	-0.078**
		(-2.119)	(-2.169)		(-2.019)	(-2.037)		(-2.163)	(-2.160)
Patience		0.000	0.000		0.004	0.004		0.004	0.004
		(0.0389)	(0.0010)		(0.481)	(0.467)		(0.422)	(0.408)
Priorities		-0.198**	-0.164**		-0.214**	-0.181**		-0.221***	-0.187**
		(-2.183)	(-2.028)		(-2.535)	(-2.405)		(-2.685)	(-2.559)
Age			0.105			0.168			0.166
			(0.716)			(1.168)			(1.175)
Educ father			0.051			0.047			0.047
			(1.116)			(1.029)			(1.022)
Educ mother			0.093*			0.095*			0.095*
			(1.659)			(1.689)			(1.681)
# siblings			-0.042			-0.033			-0.032
			(-0.824)			(-0.646)			(-0.644)
Observations	208	208	208	212	212	212	212	212	212
R-squared	0.135	0.167	0.214	0.123	0.155	0.206	0.123	0.158	0.209

Robust t-statistics in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A3. Educational choice controlling for grades

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	0.005	0.005	0.003						
	(0.194)	(0.198)	(0.132)						
Puberty 5				0.036	0.034	0.032			
				(0.853)	(0.795)	(0.727)			
Breast							0.031	0.033	0.031
							(1.182)	(1.246)	(1.167)
GPA	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***	0.003***
	(5.237)	(4.686)	(4.579)	(5.262)	(4.711)	(4.594)	(5.108)	(4.581)	(4.477)
Risk		0.006	0.006		0.006	0.006		0.008	0.008
		(0.357)	(0.369)		(0.349)	(0.358)		(0.477)	(0.477)
Patience		0.006	0.005		0.005	0.005		0.006	0.005
		(1.143)	(1.046)		(1.106)	(1.021)		(1.164)	(1.079)
Priorities		0.020	0.027		0.021	0.027		0.024	0.030
		(0.690)	(0.902)		(0.740)	(0.926)		(0.844)	(1.019)
Age			-0.010			0.000			0.001
			(-0.144)			(0.00419)			(0.0227)
Educ father			-0.001			-0.001			-0.001
			(-0.0378)			(-0.0630)			(-0.0555)
Educ mother			0.022			0.021			0.021
			(0.871)			(0.848)			(0.823)
# siblings			0.011			0.011			0.010
			(0.464)			(0.478)			(0.447)
Observations	234	234	234	234	234	234	234	234	234
R-squared	0.143	0.150	0.155	0.145	0.152	0.157	0.147	0.155	0.159

Robust t-statistics in parentheses, *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table A4. Tobit regression educational aspirations

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ageatmenarche	0.525***	0.527***	0.484***						
	(2.902)	(3.018)	(2.779)						
averagepuberty5				0.100	0.044	0.011			
				(0.319)	(0.143)	(0.0366)			
relative_puberty_breast							0.018	-0.054	-0.076
							(0.0881)	(-0.273)	(-0.389)
risk_general		-0.268**	-0.262**		-0.226**	-0.221**		-0.230**	-0.226**
		(-2.395)	(-2.379)		(-2.027)	(-2.027)		(-2.043)	(-2.059)
patientchoices		0.037	0.039		0.051	0.054*		0.051	0.054*
		(1.166)	(1.204)		(1.588)	(1.683)		(1.608)	(1.691)
schooloverfriends		-0.480*	-0.414		-0.477*	-0.425		-0.483*	-0.427
		(-1.754)	(-1.507)		(-1.720)	(-1.532)		(-1.739)	(-1.543)
age			0.126			0.468			0.451
			(0.241)			(0.912)			(0.879)
education_father			0.098			0.094			0.098
			(0.753)			(0.726)			(0.761)
education_mother			0.123			0.143			0.145
			(0.845)			(1.007)			(1.017)
numbersiblings			-0.135			-0.097			-0.098
			(-0.818)			(-0.589)			(-0.601)
Observations	214	214	214	216	216	216	216	216	216

Robust t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table A5. Logit regression educational choice

VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Timing	0.130	0.122	0.129						
	(0.809)	(0.797)	(0.803)						
Puberty 5				0.345	0.353	0.332			
				(1.277)	(1.264)	(1.149)			
Breast							0.385**	0.399**	0.385**
							(2.229)	(2.223)	(2.096)
Risk		-0.015	-0.010		-0.030	-0.028		-0.001	-0.002
		(-0.155)	(-0.0992)		(-0.299)	(-0.276)		(-0.0142)	(-0.0226)
Patience		0.073**	0.070**		0.068**	0.066**		0.071**	0.069**
		(2.522)	(2.406)		(2.345)	(2.261)		(2.438)	(2.355)
Priorities		0.474	0.518		0.461	0.480		0.497	0.502
		(1.063)	(1.188)		(1.056)	(1.141)		(1.059)	(1.115)
Age			-0.257			-0.145			-0.087
			(-0.539)			(-0.310)			(-0.183)
Educ father			-0.028			-0.022			-0.026
			(-0.218)			(-0.173)			(-0.199)
Educ mother			0.131			0.136			0.123
			(0.946)			(0.972)			(0.855)
#siblings			0.074			0.074			0.082
			(0.471)			(0.470)			(0.493)
Observations	216	216	216	215	215	215	215	215	215

Robust t-statistics in parentheses, *** p<0.01, ** p<0.05, * p<0.1

The Stockholm School of Economics

A complete publication list can be found at *www.hhs.se/research/publications*. Books and dissertations are published in the language indicated by the title and can be ordered via e-mail: *efi.publications@hhs.se*.

A selection of our recent publications

Books

Barinaga, Ester (2010). *Powerful dichotomies.*

Ericsson, Daniel (2010). *Den odöda musiken.*

Ericsson, Daniel (2010). *Scripting Creativity.*

Holmquist, Carin (2011). *Kvinnors företagande – kan och bör det öka?*

Lundeberg, Mats (2011). *Improving business performance: a first introduction.*

Melén, Sara (2010). *Globala från start: småföretag med världen som marknad.* Forskning i Fickformat.

Mårtensson, Pär & Mähring, Magnus (2010). *Mönster som ger avtryck: perspektiv på verksamhetsutveckling..*

Sjöström, Emma (2010). *Ansiktslösa men ansvarsfulla.* Forskning i fickformat.

Wijkström, Filip (2010). *Civilsamhällets många ansikten.*

Engwall, Lars (2009). *Mercury meets Minerva: business studies and higher education: the Swedish case.*

Dissertations

Alexandersson, Gunnar (2010). *The accidental deregulation : essays on reforms in the Swedish bus and raillway industries.*

Bohman, Claes (2010). *Attraction: a new driver of learning and innovation.*

Buturak, Gökhan (2010). *Choice deferral, status quo bias, and matching.*

Ejenäs, Markus (2010). *Ledning av kunskapsintegration – förutsättningar och hinder: en studie av en fusion mellan IT- och managementkonsulter.*

Engvall, Anders (2010). *Poverty and conflict in Southeast Asia.*
Glassér, Charlotte (2010). *The fountainhead of innovation health: a conceptualization & investigation.*
Hemrit, Maetinee (2011). *Beyond the Bamboo network: the internationalization process of Thai family business groups.*
Juks, Reimo (2010). *Corporate governance and the firm´s behaviour towards stakeholders.*
Lundvall, Henrik (2010). *Poverty and the dynamics of equilibrium unemployment: [essays on the economics of job search, skills, and savings].*
Lychnell, Lars-Olof (2010). *IT- relaterad verksamhetsförändring : processer som formar växelspelat mellan utveckling och användning.*
Magnusson Bernard, Kristin (2010). R*emittances, regions and risk sharing.*
Mohlin, Erik (2010). *Essays on belif formatic and pro-sociality.*
Monsenego, Jérôme (2011). T*axation of foreign business income within the European internal market: an analysis of the conflict between the objective of achievement of the European internal market and the principles of territoriality and worldwide taxation.*
Nakatani, Tomoaki (2010). *Four essays on building conditional correlation GARCH models.*
Nelson, Maria (2010). *Utflyttning av aktiebolag: en analys i ljuset av den internationella skatterätten och EU-rätten.*
Siming, Linus (2010). *Private equity and advisors in mergers and acquisitions.*
Sjöqvist Rafiqui, Pernilla (2010). *Evolving economic landscapes: institutions and localized economies in time and space.*
Stepanok, Ignat (2011). *Essays on international trade and foreign direct investment.*
Strid, Ingvar (2010). *Computational methods for Bayesian inference in macroeconomic models.*
Sunesson, T. Daniel (2010). *School networks and active investors.*
Tolstoy, Daniel (2010). *International entrepreneurship in networks: the impact of network knowledge combination on SMEs' business creation in foreign markets.*
Öhman, Niclas (2010). *Considering intentions.*