CESifo WORKING PAPERS

10779
2023

November 2023

The Persistent Effect of Competition on Prosociality

Fabian Kosse, Ranjita Rajan, Michela Tincani

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
https://www.cesifo.org/en/wp
An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: https://www.cesifo.org/en/wp

CESifo *Working Paper No. 10779*

The Persistent Effect of Competition on Prosociality

Abstract

We present the first causal evidence on the persistent impact of enduring competition on prosociality. Inspired by the literature on tournaments within firms, which shows that competitive compensation schemes reduce cooperation in the short-run, we explore if enduring exposure to a competitive environment persistently attenuates prosociality. Based on a large-scale randomized intervention in the education context, we find lower levels of prosociality for students who just experienced a 2-year competition period. 4-year follow-up data indicate that the effect persists and generalizes, suggesting a change in traits and not only in behavior.

JEL-Codes: D640, C900.

Keywords: prosociality, competition, cooperation, social skills, socio-emotional skills, tournaments, comparative pay, incentive schemes.

*Fabian Kosse**
University of Würzburg
Sanderring 2
Germany – 97070 Würzburg
fabian.kosse@uni-wuerzburg.de

Ranjita Rajan
The Karta Initiative

Michela Tincani
University College London / United Kingdom
m.tincani@ucl.ac.uk

*corresponding author

November 2023
We thank Teodora Boneva, Luca Braghieri, Thomas Buser, Armin Falk, Ximeng Fang, Michele Fioretti, Lorenz Goette, Johannes Haushofer, Matt Lowe, Katherina Kosse, Enrico Miglino, Rubén Poblete-Cazenav, Chris Roth, Matthias Sutter, Florian Zimmermann and many seminar and conference participants for fruitful discussions. Financial support through the Jacobs Foundation, the R24 pilot grant program, the ESRC (Grant ES/N015622/1) and the German Research Foundation (CRC TR 190) is gratefully acknowledged. The pre-analysis plan can be found in the AEA RCT Registry (AEARCTR-0002288). This research received approval from the UCL Research Ethics Committee (Project ID 10515/002) and from the UCL Data Protection Office (Z6364106/2017/06/101).

1 Introduction[1]

Elements of prosociality such as reciprocity, altruism and trust are particularly important aspects of human personality and affect a wide range of economic decisions and outcomes. At the level of groups and societies, they are essential for the functioning of markets and the well-being of societies,[2] as they work as a contract enforcement device (Fehr, Gächter, and Kirchsteiger, 1997) and drive economic exchange (Guiso, Sapienza, and Zingales, 2009). At the level of individuals, recent evidence indicates the benefits of prosociality and social skills regarding, e.g., labor market and health outcomes (Deming, 2017; Fang et al., 2022), which is in line with the large body of literature on the returns to non-cognitive or socio-emotional skills (e.g., Heckman, Stixrud, and Urzua, 2006).

Despite the fundamental importance of prosociality for the well-being of societies and individuals, little is known about its determinants at the individual level. Inspired by the literature on the cooperation-reducing effect of competitive compensation schemes in firms (e.g., Lazear, 1989) and on the decline of ethical conduct under market competition (Shleifer, 2004; Tergiman and Villeval, 2022), we explore the role of competitive environments in determining prosociality. While competition is pervasive in economic, political and education systems and there is growing laboratory experimental evidence that individuals behave less prosocially in short-lasting competitive situations (e.g., Chowdhury and Gürtler, 2015), the real-world effects of enduring competition on prosociality are largely unknown. Equally unknown is whether any effects are short-lived, reflecting short-term behavioral responses, or long-lasting, potentially reflecting changes in prosociality as a trait.

To answer these questions, we take advantage of a large-scale, randomized education intervention in Chile. We elicited comprehensive measures of prosociality based on validated survey items at the end of the competition period and four years later. We compare the levels of prosociality of students who were randomized into a more competitive environment for the last two years of high school to that of a control group. Our sample consists of students from low socioeconomic status families and the intervention implemented was a percent plan policy offering university admission for the top 15% of students within each high school. Compared to the control group, the policy increases the level of competition within schools as it increases the importance of the relative standing within school.

[1]This research received approval from the UCL Research Ethics Committee (Project ID 10515/002) and from the UCL Data Protection Office (Z6364106/2017/06/101).

[2]Examples for the importance of prosociality on the level of groups and societies are shown in Knack and Keefer (1997) and Fang et al. (2022).

Our results show that experiencing a more competitive environment in the last two years of high school leads to lower prosociality. This finding holds for a composite measure of prosociality, as well as for the three facets, reciprocity ($p < 0.01$), altruism ($p < 0.1$) and trust ($p < 0.05$), separately. We show that these findings are not driven by selective attrition and are robust to corrections for multiple hypothesis testing. Analyzing heterogeneous treatment effects reveals a stronger treatment effect for male students compared to females. The 4-year follow-up data indicate that the prosociality reducing effect is persistent and general, i.e., the effect is not only directed towards former contestants but also towards individuals not related to the competition.

The contribution of this paper is fourfold. First, our findings contribute to the literature on the role of the social environment for the formation of prosociality. Building on studies that point to the development of prosociality in childhood and adolescence (e.g. Sutter and Kocher, 2007; Fehr, Bernhard, and Rockenbach, 2008; Fehr, Glätzle-Rützler, and Sutter, 2013; Falk et al., 2021), recent evidence indicates that enriching the social environment in childhood in form of, e.g., desegregation, preschool attendance or mentoring program participation persistently increases individuals' prosociality (Rao, 2019; Cappelen et al., 2020; Kosse et al., 2020). We complement this literature by providing the first causal evidence on environmental factors that attenuate the formation of prosociality. Thereby, we also contribute to the study of formative periods for preferences and skills, see e.g., Kautz et al. (2014), by providing further evidence on their malleability in late adolescence.

Second, we add to the literature on the effects of competition, comparative payment schemes and tournament-style promotion mechanisms. Previous theoretical and empirical studies have shown that when individuals are rewarded according to their performance relative to others, they behave less cooperatively than when they are rewarded for their performance in absolute terms. Contestants are inclined to not help their rivals, and instead to cheat or sabotage to improve their relative standing (e.g., Lazear, 1989; Chowdhury and Gürtler, 2015). Most empirical evidence comes from relatively short-lasting laboratory or online experiments, where individuals are exposed to short competitive situations, and where their prosocial behavior is observed within or directly (i.e. a few minutes) after the competition period (e.g., Falk, Fehr, and Huffman, 2008; Harbring and Irlenbusch, 2011; Charness, Masclet, and Villeval, 2014; Buser and Dreber, 2016; Huber et al., 2023). However, it is vital to understand the effects of longer-lasting competitions, as this is what happens in real-world organizations and societies: for instance, bonuses are usually paid on a yearly basis and decisions about promotions are often based on even longer time horizons. Entirely open questions in these contexts are 1.) how enduring exposure to a competitive environment, in contrast to a short-lasting competitive situation, affects perceptions and reactions of individuals and 2.) how persistent these effects are. If habit formation takes place, reduced cooperation will become automatic and persistent by its repeated execution (e.g., Lally et al., 2010). In contrast, if habituation takes place, the effect of the stimulus will vanish after its prolonged presentation (e.g., Thompson and Spencer, 1966). Therefore, it is an empirical question if the previously documented phenomenon of reduced cooperation also holds in long-lasting competitions and if the effect persists and generalizes across situations and environments. Our unique setting allows us to answer both questions for the first time: we exploit a real-world policy that introduced a 2-year competition period and we measure prosociality across different contexts, at the end of the competition period, and 4-years later, providing the longest-run evidence on the impacts of competition of prosociality to date.

Third, by showing that the prosociality reducing effect is not only directed towards former contestants but also towards individuals not related to the competition, we also contribute to the experimental literature on behavioral spillovers across contexts (e.g., Bednar et al., 2012; Cason, Savikhin, and Sheremeta, 2012; Cason and Gangadharan, 2013; Buser and Dreber, 2016). Our findings complement existing studies by showing that the phenomena of behavioral spillovers is not limited to laboratory environments but is also observable in a long-run real-world field context.

Fourth, we contribute to a better understanding of the effects of percent plans, which are becoming more and more popular in education policy, for example the states of Texas, California and Florida use percent plans for college admissions (Horn and Flores, 2003). Therefore, our findings are important from a policy perspective as they highlight that percent plans can have unintended side-effects which should be taken into account in cost-benefit analyses.

2 Study design

This section introduces the design of the study. We first outline the societal and institutional background in Chile. Next, we describe the intervention and the corresponding incentive schemes. Finally, we discuss the data sources and measures used in the analyses.

2.1 Societal and institutional background

Education in Chile is compulsory from ages 6 to 18 and (theoretically) all high school degrees qualify for university. However, only very few students from low socioeconomic status (SES) backgrounds progress to university.[3] This low intergenerational mobility regarding education outcomes is combined with a high education earnings premium on the labor market (Hastings, Neilson, and Zimmerman, 2013; OECD, 2016) and led to calls for more equality of opportunity and a series of social unrests. In response, the Chilean government extended the tuition fee waiver program to all students from the bottom 60% of the income distribution. However, students from low SES families kept enrolling into selective universities[4] only at low rates, as to obtain admission, students must perform well on a standardized university admission exam.[5] In 2018 more than 65% of students from low SES families took the admission exam, but only about 8% enrolled in university.[6]

2.2 Intervention and randomization

To increase the admission chances of low SES students, in 2014 the freshly elected government led by Michelle Bachelet introduced a policy called PACE (*Programa de Acompañamiento y Acceso a la Educación Superior*). The policy targets high schools serving disadvantaged students. Eligibility is based on a school-level vulnerability index called IVE (*Indice de Vulnerabilidad Escolar*), based on the socioeconomic characteristics of students. The key feature of PACE is a percent plan that offers university admission to the top 15% of students within each high school based on the grade point average[7] in grades 9 to 12.[8] The PACE seats are supernumerary: they

[3]Students whose parents have tertiary education have a more than six times (77% vs. 12%) higher probability of attaining tertiary education compared to children whose parents have below secondary education (OECD, 2016).

[4]These are universities participating in the centralized admission system. They offer 5-year (and longer) programs of an academic nature. They include the 23 public and private not-for-profit colleges that are part of the Council of Rectors of Chilean Universities (CRUCH) and 14 additional private colleges. Higher education institutions outside of this system do not have minimum admission requirements, and typically provide vocational and shorter degrees.

[5]At the time of our study the test was called PSU (*Prueba de Selección Universitaria*).

[6]The numbers refer to our control group (see below). For details also see Tincani, Kosse, and Miglino (2023).

[7]To be precise, the relevant ranking refers to a score called Puntaje Ranking de Notas (PRN). The Pearson's correlation coefficient between the unadjusted 4-year grade average and the PRN score is 0.974. The adjustment considers the historical performance of high schools. This adjustment is known to the students and is not affecting the incentives.

[8]Moreover, like in the Texas Top Ten plan, light-touch orientation classes (2 hours per month on average) are offered in PACE high schools, for details see Cooper et al. (2019, 2022). The classes covered topics such as learning techniques and practicality of application processes but nothing related to determinants of prosociality.

do not replace regular seats but instead are offered in addition to them. Therefore, PACE did not make it mechanically harder to be admitted through the regular channel. If a student does not accept a PACE admission, that PACE seat remains vacant.

In 2014, PACE was introduced in 69 disadvantaged high schools and later expanded to further schools. The expansion step of PACE in 2016 was conducted with evaluation as one of the key purposes. The government defined a new eligibility threshold based on the IVE which made 221 new schools eligible for PACE participation. 64 of these eligible schools were randomly chosen to become part of PACE.[9] Operating within the constraints of a limited research budget, we could not collect survey data from all elegible schools. Therefore, we sampled all 64 treatment schools and randomly chose 64 control schools (out of 157). In section 2.4.1, we present evidence for the baseline balance of this sample of 128 schools.

2.3 Incentives, timing and perception of the treatment

With the backdrop of the high education premium in Chile and the relatively low chances of students from eligible schools to gain admission to universities through the regular admission channel, PACE creates strong incentives to graduate among the top 15%: While PACE offers admission to all students of the top 15% in a given school in the treatment group, in the control group (empirically) only 32.8% of the top 15% and only 7.0% of the bottom 85% got admitted.[10] In this context, we outline below the incentive structures in treatment (PACE) and control (no PACE) schools.

Control group: To do well in the standardized university admission exam and gain admission to university, students must accumulate high absolute levels of human capital. With every hour studying, students marginally increase their human capital and thus increase their individual probability of admission to university. It is immaterial if schoolmates accumulate more or less human capital. We refer to the control group environment in which the relative ranking within school is irrelevant, as less competitive environment (see also e.g. Gneezy, Niederle, and Rustichini, 2003; Benistant, Galeotti, and Villeval, 2022).[11]

[9]The (not stratified) randomization code was written by PNUD - Chile (United Nations Development Program).

[10]For details on the effects on educational outcomes see Tincani, Kosse, and Miglino (2023).

[11]Please note that the control group environment also includes elements of competition: University admission is based on a nation-wide ranking of performance in the standardized university admission exam. However, this does not give relevance to the relative ranking within a school.

Treatment group: As in the control group, students in PACE schools can also attempt admission to university through the regular channel (where chances are relatively low), but can additionally earn a guaranteed slot at university if they graduate among the top 15% in their school. Therefore, the absolute level of performance is less important whereas the relative standing within school is of major importance. In line with the (laboratory) experimental literature we refer to the treatment group environment in which the relative ranking within group is more important as more competitive environment (e.g., Gneezy, Niederle, and Rustichini, 2003; Benistant, Galeotti, and Villeval, 2022).

Timing of the competition period: In this study we focus on the cohort of students who just entered grade 11 when PACE was implemented.[12] At the start of the school year, students who were enrolled in treatment schools were informed that their schools were now part of PACE. The announcement was made after the deadline for school enrollment, consequently ruling out strategic enrollment into treatment schools.[13] While the GPAs from grades 9 and 10 were already determined at this time, there were still two years of schooling to improve their relative standing in the grade distribution.

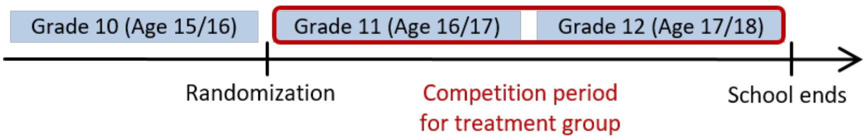

Figure 1: Timing of the competition period for the cohort under study.

Perceptions of the schooling environment: As displayed in figure 1, the randomized implementation of PACE at the beginning of grade 11 creates the following situation: Compared to students in control schools, students in treatment schools experience a two-year *competition period* created by the pronounced focus of the relative comparison within schools. This aligns with teachers' and students' perceptions of the schooling environment. Using survey data collected by the Chilean Ministry of Education at the beginning of grade 12 we show that teachers and students in treatment schools perceive the schooling environment as more competitive

[12]Only this cohort allows for a valid treatment-control comparison. Considering grade 12 was not possible as, when PACE enters a school, only the grades 11 and younger are treated. Considering grades 9 and 10 was not possible, as for those cohorts the government added additional schools from the control group to the treatment group in a non-random way.

[13]Eligibility for a guaranteed slot requires enrollment in the same school for the last two years of schooling. Therefore, there were no incentives to change schools at a later time. In order to focus on eligible students, we restrict the sample to students who were in the same school for the last two high school years.

compared with teachers and students in control schools. For details see sections B.1 and B.2 in the Appendix. The results indicate, e.g., that in treatment schools compared with control schools, about 40% more teachers "agree" or "strongly agree" with the statement "There is a lot of competition to get the best grades in my course" (see table B1).

2.4 Data and measures

In the following we describe the data and measures used in the analysis. For the baseline we rely on administrative data, for the endline and the 4-year follow-up we conducted our own tailor-made surveys. We link all data sources through a unique student identifier, which is available for every student in Chile.[14]

2.4.1 Baseline

Our baseline measures come from a large administrative data collection called SIMCE (*Sistema de Medición de la Calidad de la Educación*), which was conducted at the end of grade 10. This data set contains information on a standardized achievement test and a large set of socioeconomic variables such as gender, age and parental education and income.[15] Based on these data, table A1 presents balancing tests for the 128 schools that we sampled. The results indicate that students in treatment and control schools do not significantly differ at baseline with regard to gender, age, parental socioeconomic status (SES) and academic performance. Moreover, treatment and control schools do not differ by cohort size and location.

2.4.2 Endline

Our endline data comes from a survey that we implemented at the end of grade 12. The data collections in the 128 treatment and control schools were conducted by trained field-workers in cooperation with a Chilean survey company during regular schooling hours. The data collection took place in form of a self-administered paper pencil survey. While completing the survey, students were seated in a standardized seating order (see figure A1) and were not allowed to talk in order to prevent interactions. To minimize experimenter demand effects, e.g. triggered by thankfulness

[14]We thank the Chilean Ministry of Education for performing the merges and delivering to us anonymized student-level data.

[15]As the SIMCE data contain many missings on parental education and income (as they were collected in form of a parental survey), we also use an additional source of administrative data (*Subvención Escolar Preferencial*), which give us a dummy variable indicating very low SES students (*prioritario* student) for all students in the sample.

in the treatment group, the survey did not contain any reference to the Ministry of Educational or the Chilean government in general, and described the study as independent research by the authors of this paper

We surveyed 6,094 students, constituting nearly 70 percent of the students enrolled in the 128 sample schools. Our response rate compares favorably with surveys conducted by the Ministry of Education (e.g., MinEduc, 2017), and it reflects a natural drop-out in the final weeks of the last high school year. Due to missing responses on one or more prosociality items (see below) our endline sample consists of 5,371 students. The results in table A2 indicate that attrition is neither significantly related to treatment status nor to the interaction of treatment status and baseline characteristics as achievement, gender and parental SES. Nevertheless, in the analysis we also provide results based on Inverse Probability Weights (IPW) and Lee (2009) bounds as robustness checks.

Prosociality measure (endline): We refer to prosociality as positive other-regarding behaviors and beliefs. To measure it, we closely follow the procedure proposed and validated by Kosse and Tincani (2020). In order to yield a comprehensive measure of students' prosociality we combine measures of three main facets: altruism, reciprocity, and trust. Altruism reflects an individual's willingness to benefit others (without expecting anything in return), (positive) reciprocity reflects an individual's willingness to reward kind behavior and trust indicates prosocial beliefs in the actions of others. To measure these facets we use the Chilean Spanish version of the qualitative items of the Global Preference Survey (GPS, Falk et al., 2022).[16] Positive reciprocity was measured by the item "When someone does me a favor I am willing to return it." Trust was measured by the item "I assume that people have only the best intentions." To measure altruism we used the slightly age-adapted item "How willing are you to help others without expecting anything in return?".[17] All items were rated on an 11-point Likert scale. We aggregate the three (standardized) facets of prosociality using principal component analysis (PCA). In the analyses we use the resulting (standardized) first principal component as our measure of prosociality.[18]

[16]These items were selected in an ex ante experimental validation procedure among large sets of items in order to exhibit the highest predictive power for corresponding incentivized behavioral measures. We did not use the quantitative items as they are hard to implement in paper-pencil surveys.

[17]The original GPS item is "How willing are you to give to good causes without expecting anything in return?" Our focus group pre-tests indicated that "giving to good causes" is very unusual among low SES Chilean students. To measure meaningful variations we decided to slightly age-adapt the item.

[18]For a detailed discussion on this procedure see Kosse and Tincani (2020).

2.4.3 4-year follow up

Four years after the end of school, in summer 2021, we collected a follow-up survey in order to explore if the effects found at endline are persistent and are general, i.e., translate to situations which are independent of the school environment. At endline, in preparation for the follow-up, we asked the students for permission to re-contact them in order to conduct a follow-up survey. About 90% percent of the students in the endline sample gave their permission to be contacted for a follow-up interview and provided email addresses and/or phone numbers. The follow-up data collection was implemented in cooperation with the same survey company as the endline data collection and it was conducted in form of an online survey with links distributed via email and phone numbers. As at endline, the survey did not contain any reference to the Ministry of Education or the Chilean government in general and was described as an independent research study by the authors of this paper.

Unfortunately, many of the originally provided email addresses and phone numbers were not in use anymore after four years. Nevertheless and despite the questionnaire being online, we received complete survey answers from 1,018 students, constituting approximately 20 percent of our endline sample.[19] While such a drop in observations is not unusual when switching from a central-location in-person to an online interview mode (as inevitable in this situation), it raises two potential concerns for the analysis of the data and the interpretation of the results: selective attrition and lack of statistical power.

To explore the potential issue of selective attrition empirically, as in section 2.4.2, we make use of our rich baseline data. We explore if attrition differs between treatment and control group, and if baseline measures differentially predict attrition in the treatment and control group. The results in table A3 indicate that attrition is neither significantly related to treatment status nor to the interaction of treatment status and baseline characteristics as achievement, gender and parental SES. This indicates the absence of selective attrition based on observables.[20] Nevertheless, in the analysis we also provide results based on IPW and Lee (2009) bounds as robustness checks.

While the statistical power at endline is satisfactory (post-hoc power of 0.82 ($\alpha = 0.1$) and 0.72 ($\alpha = 0.05$)), it is not much lower in the four-year follow-up

[19]For administrative reasons we were not able to offer monetary incentives for survey participation. However, at endline and follow-up, we raffled off, respectively, one and three iPads among all participants.

[20]Selective attrition, i.e. different drivers of attrition in treatment and control group, would have been indicated by significant interaction effects. For a discussion of interaction effects see, e.g., chapter 3.1.4 in Angrist and Pischke (2009).

(post-hoc power of 0.78 ($\alpha = 0.1$) and 0.67 ($\alpha = 0.05$)).[21] There are three reasons why we maintain relatively high levels of power despite losing many observations. First, to counteract the expected loss in observations, we increased the precision of the prosociality measure by collecting more items for its identification (for details see below). Second, given the cluster structure of our data, we did not lose independent observations but observations within clusters.[22] Third, the within-cluster correlation decreased from 0.033 to 0.001. This is not surprising, as at follow-up, participants are no longer exposed to the same (school) environment as at endline.

Prosociality measures (follow-up): At the 4-year follow-up we collected three sets of items. 1.) The same set of items which we collected at endline. All of them have a general frame and are not referring to a specific group of people ("When *someone*...", "I assume that *people* have...", "How willing are you to help *others* ..."). 2.) In order to reduce measurement error, we collected a second set of validated generally framed items, one additional item per facet. For positive reciprocity, we used "I go out of my way to help somebody who has been kind to me in the past" (Dohmen et al., 2009). For trust, we used "In general, one can trust other people" (Fehr et al., 2002). For altruism, we used "Imagine the following situation: Today you unexpectedly received 128.000 Pesos. How much of this amount would you donate to a good cause?" (Falk et al., 2022). 3.) To be able to distinguish between general behavior and behavior specific towards the former contestants, we also collected items on all three facets explicitly focusing on former high school peers. The items are adaptations of the generally framed items and read as follows: "Please think about your former high school class peers. How willing are you today to ..." "Return a favor to them?" (reciprocity), "Trust them?" (trust), "Help them without expecting anything in return?" (altruism). All items were rated on an 11-point Likert scale, except the generally framed altruism item, which was measured in Pesos. Based on these items we construct three prosociality measures by aggregating (sub-)sets of them using PCA:

[21]The power calculations for cluster randomized controlled trails were conducted as described in Hemming and Marsh (2013). The calculations take into account: the means in treatment and control groups, the intra cluster correlation, the average cluster size and the number of clusters per arm. Instead of calculating the ex-post power, the same information can be used to calculate minimum detectable differences (MDDs). At endline, e.g., the MDD for power = 0.8 and $\alpha = 0.05$ is 0.12, the MDD for power = 0.7 and $\alpha = 0.1$ is 0.09. At the 4-year follow-up, e.g., the MDD for power = 0.8 and $\alpha = 0.05$ is 0.18, the MDD for power = 0.7 and $\alpha = 0.1$ is 0.14. Please note that the presented values are conservative, as they are not taking into account the power gain by using baseline controls.

[22]In the endline sample the average cluster size is 42 students, in the follow-up sample it is 8 students.

- Joint measure of prosociality: using all items
- General prosociality: using only generally framed items
- Prosociality towards former schoolmates: using only items with a focus on former schoolmates

The 'Joint measure of prosociality' minimizes measurement error and maximizes power. The comparison of the second and the third measure allows to answer the question of whether the effect on prosociality is specific to former contestants or also applies to individuals and situations independent of the former school environment.

3 Results

In the following we present the results of our analyses. In section 3.1, we answer the question of whether lasting competitive environments attenuate prosociality and in section 3.3 we explore the persistence and generalization of the effect.

3.1 The causal effect of competition on prosociality

The experimental setup makes our empirical strategy straightforward. Equation 1 shows our main empirical model:

$$PS_{is} \quad = \quad \beta_0 + \beta_{Treat} Treat_s + X_i' \beta + u_{is} \qquad (1)$$

where PS_{is} is the standardized prosociality measure of student i in school s. The treatment indicator $Treat_s$ takes the value one if a school was randomly selected to be part of PACE and zero otherwise. The vector X includes baseline measures of achievement, gender and parental SES to increase precision of the estimates. u_{is} represents the error term of the model. We estimate equation (1) using OLS and cluster standard errors at the school level.

Table 1 shows our main result at endline. In column 1 we show the raw treatment effect without further controls and in column 2 we show the treatment effect controlling for baseline measures of achievement, gender and parental SES. The results indicate a significant negative treatment effect on students' prosociality of about 11 to 12% of a standard deviation. This shows that living in a more competitive environment for two years during adolescence attenuates the formation of prosociality. The effect is sizable and compares to a one standard deviation increase in achievement or the difference between female and male students (see column 2).

As already discussed in section 2.4.2, there are no signs of selective attrition (see also table A2). Nevertheless, in column 3 of table 1 we apply inverse probability weighting to counteract minor imbalances.[23] Compared to column 2 the results are largely unchanged, if anything, the treatment effect is slightly bigger in column 3. In addition, we estimate treatment effects using the trimming procedure suggested by Lee (2009). Instead of correcting point estimates based on observables, this approach yields non-parametric bounds of effect sizes on the basis of extreme assumptions about selection. The estimates of the bounds shown in table A4 confirm the negative treatment effect.

In table A5 we show treatment effect estimates for the respective facets of prosociality. The results indicate negative and significant treatment effects for altruism, trust and reciprocity, irrespectively of considering original or multiple hypothesis corrected p-values. The results indicate similar sized effects on altruism and trust, with the strongest effect on reciprocity. This pattern indicates that living in a more competitive environment comprehensively attenuates the formation of various facets of prosociality.

	Std. Prosociality (endline)		
	(1)	(2)	(3)
Treatment dummy	-0.107**	-0.120***	-0.122***
	(0.053)	(0.044)	(0.045)
Achievement (at baseline, standardized)		0.107***	0.108***
		(0.016)	(0.016)
Female		0.123***	0.129***
		(0.034)	(0.034)
Very low SES dummy		0.022	0.025
		(0.031)	(0.031)
Weights	No	No	IPW
Observations	5,343	5,343	5,343

Table 1: Treatment effect on prosociality at endline. Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. For details on the control variables see section 2.4.1. Column 3 applies inverse probability weights which account for potential selective attrition and are estimated from a probit model of a binary selection indicator (indicating whether the prosociality measure is available at endline) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A2 column 3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

[23] Weights are estimated from a probit model of a binary selection indicator (indicating whether the prosociality measure is available for endline) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A2 column 3.

The laboratory experimental literature indicates gender differences in reaction towards competition (e.g., Gneezy, Niederle, and Rustichini, 2003; Niederle and Vesterlund, 2007; Buser, 2016). Most related to our study, Buser and Dreber (2016) find a stronger cooperation reducing effect for males than for females after being randomized into a competitive payment scheme in an online experiment. Table A6 shows the treatment effects of living in a more competitive environment on prosociality separately for female and male students. While female students are generally more prosocial than male students (table 1, column 2), the results indicate a more than 60% stronger treatment effect for males than for females (-0.075 vs. -0.127). This is in line with the findings of Buser and Dreber (2016) and indicates that this gender-specific reaction does not only hold for short-lasting competitions but also in high-stake long-run competitions.

3.2 Discussion of alternative mechanisms and explanations

We argue that the prosociality-reducing effect of participation in PACE is driven by prolonged exposure to a more competitive environment (see sections 2.3, B.1 and B.2). In the following section, we discuss two potential alternative mechanisms and explanations: cognitive load-induced heuristics and behavioral changes of school principals and teachers.

3.2.1 Cognitive load-induced use of heuristics

The laboratory experimental literature on the relation of competition and prosocial behavior discusses cognitive load as a potential mechanism (e.g., Buser and Dreber, 2016): Given the strategic concerns and the added uncertainty, competition is a more complex situation that might lead people to rely on simple heuristics when making decisions. While the direction of the effect of cognitive load on prosocial behavior is not clear, we test for the presence of differential answering behavior across the treatment and the control groups. We use two complementary analyses. First, as shown in table A7 we directly test for differential use of the common heuristic "always ticking the same response". Our results indicate that the probability of always giving the same response when answering the three prosociality items is significantly negatively related to achievement at baseline. This suggests that the use of this heuristic is indeed negatively related to cognitive resources. However, independently of the specifications of the analysis, the use of this heuristic is not related to treatment status ($p > 0.690$). Second, we compare the variances of prosociality between treatment and control group to test more generally for treatment differences in answering patterns. Levene's tests for equality of variances (Levene, 1960) does not indicate differential variances of the prosociality measure across treatment and control group ($W = 0.182, p = 0.670$). Taken together, these results indicate the absence of differential use of answering heuristics in treatment and control group. Therefore, cognitive load is unlikely to be a mechanism driving our results.

3.2.2 Behavioral changes of school principals and teachers

Principals and teachers could potentially respond to the treatment in various ways and thereby directly affect the students' prosociality. In a companion paper, based on teachers' and principals' surveys conducted in the same sample of schools, Tincani, Kosse, and Miglino (2023) explore the behavior of teachers and principals. Based on these analyses,[24] in the following, we discuss potential treatment responses of principals and teachers:

- Grading: Teacher could potentially change their grading pattern in response to the treatment and thereby, e.g., create additional envyness and resentment. However, the empirical evidence does not support this channel. Table G3 in Tincani, Kosse, and Miglino (2023) indicates that the mapping between standardized achievement scores and grades does not differ between treated and control schools and table G4 indicates that the school principals report the same grading practices across treatment groups.

- Teachers' effort and focus of instruction: Teachers could potentially change their focus of instruction (e.g., focusing teaching only on top students), or they could change their own effort (preparation hours and absence days) as an effect of the treatment. However, the empirical evidence does not support this channel. Table A12 in Tincani, Kosse, and Miglino (2023) indicates that there is no evidence for such behavioral responses to the treatment.

- Principals and school policies: While the curriculum is not a possible margin of treatment response because the Ministry of Education mandates it, school principals in treated schools might potentially change the amount of support classes for students. However, table G4 in Tincani, Kosse, and Miglino (2023) indicates that this is not the case. Treated schools do not differ from control schools regarding the support offered to students. Moreover, principals might also change the assignment of students to classrooms, e.g., by grouping students according to their ability, and thereby affect the social behavior of

[24] For details see sections 5.2 and E.1 in Tincani, Kosse, and Miglino (2023).

the students. However, the empirical evidence does not support this channel either. Table G5 in Tincani, Kosse, and Miglino (2023) indicates that there is no evidence for differential classroom assignment between treatment and control schools.

In sum, based on detailed teachers' and principals' surveys, Tincani, Kosse, and Miglino (2023) find no effects on the behaviors of school principals and teachers in response to the treatment. This suggests that changes in the behavior of teachers and school policies are unlikely to be a mechanism driving our results.

3.3 Persistence and generalization of the effect

Table 2 shows treatment effects on prosociality in the 4-year follow-up.[25] In column 1 we show the raw treatment effect without further controls, in column 2 we show the treatment effect controlling for baseline measures and in column 3 we additionally apply IPW.

In Panel A, the dependent variable is the joint measure of prosociality. It combines general prosociality and prosociality towards the former contestants. The measure makes use of all survey items, it minimizes measurement error and maximizes statistical power. The coefficients shown in columns 1 to 3 demonstrate a robust, negative and significant effect of the tournament schema in high school on prosociality four years later. Table A8 shows Lee (2009) bounds and also confirms the negative treatment effect.

As the prosociality measures in Panel A of table 2 and table 1 differ, a direct comparison of the magnitudes of the effects is not possible, unless one makes strong distributional assumptions. However, as both analyses, at endline and at the 4-year follow-up, show significant negative effects, the results clearly indicate persistence of the effect of the tournament schema. This shows unambiguously that rank-based reward schemes can have persistent negative effects on prosociality.

To get a sense if the treatment effect is decreasing or increasing over time, we also rebuild the endline measure of prosociality at follow-up and use it to analyze the dynamics of the treatment effect in a panel framework. The results in table A9 indicate that the treatment effect regarding this measure does not differ between endline and follow-up (interaction effect of treatment and time: 0.003, $p = 0.971$). In table A10 we show the results for the facets altruism, trust and reciprocity separately.

[25] All dependent variables are standardized using the distribution in the 4-year follow-up sample. Using the distribution in the endline sample would largely simplify a comparison of the effects over time, but this is not possible as the follow-up measures (see section 2.4.3) rely on items that are not available at endline.

As for the composite measure, there is no significant change in the treatment effects of the facets between endline and follow-up. This suggests that the size of the treatment effect on prosociality is pretty stable over time.

	Std. Prosociality (4-year follow-up)		
	(1)	(2)	(3)
Panel A: Joint measure of prosociality			
Treatment dummy	-0.152**	-0.146**	-0.149**
	(0.067)	(0.063)	(0.068)
Panel B: General prosociality			
Treatment dummy	-0.126*	-0.120*	-0.128*
	(0.072)	(0.071)	(0.074)
Panel C: PS towards former schoolmates			
Treatment dummy	-0.132**	-0.128**	-0.130*
	(0.065)	(0.061)	(0.066)
Baseline controls	No	Yes	Yes
Weights	No	No	IPW
Observations	1,018	1,018	1,018

Table 2: Treatment effect on prosociality four years after the end of the treatment. Coefficients are OLS estimates. Standard errors clustered at school level are shown in parentheses. For details on the control variables see section 2.4.1. All dependent variables are standardized using the distribution in the 4-year follow-up sample. Column 3 applies inverse probability weights which account for potential selective attrition and are estimated from a probit of a binary selection indicator (indicating whether the prosociality measure is available for the 4-year follow-up wave) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A3 column 3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Potentially, the long-run effects might be affected by the fact that a higher share of treated students enter college and, thereby, experience a more academic and potentially more competitive social environment also after finishing high school. To explore if this is the case, we make use of the empirical result that, for the subsample of students in the bottom 85% of the baseline GPA ranking of their school, the treatment did not significantly affect the probability of college admission and attendance.[26] Therefore, in this subsample, treated students did not experience a more academic and potentially more competitive social environment after finishing high school. Hence, in table A11, we repeat the previous panel analysis for the subsample of students in the bottom 85% at baseline. The results for this subgroup, which is unaffected by potential treatment-induced college experiences, are in line with the previous finding and indicate that the treatment effect does not differ between endline and follow-up (interaction effect of treatment and time: 0.028, $p =$

[26]The treatment effect on the probability to get a college admission for this subgroup is 0.011 ($p = 0.255$). For details see tables A4 and A7 in Tincani, Kosse, and Miglino (2023).

0.740). This suggests that the documented long-term effects are driven by changes in the high school environment, and not by different after high school experiences which go along with the treatment.[27] Thereby, the empirical pattern points to a persistent effect of a more competitive environment at high school and a change of prosociality as a trait.

Finally, in the endline analysis (table 1), it was not possible to distinguish between effects on general prosociality or prosociality towards former contestants. While the items used a general framing, for high school students, it is reasonable to expect that schoolmates constitute a large share of their social environment. Therefore, behavior towards "others" and behavior towards schoolmates has necessarily a large overlap and is hard to distinguish. Fours years later, driven by heterogeneous life choices, this overlap between general "others" and former schoolmates is much weaker and a distinction is possible.[28] In table 2, Panel B, we show treatment effects on general prosociality (towards "others"), in Panel C, we show treatment effects on prosociality towards former schoolmates, i.e. former contestants. While a direct comparison of the coefficients in Panel B and C, would, again, require strong assumptions about the underlying distributions, the fact that all coefficients in Panel B and C are significantly negative shows that the effect of competition on prosociality generalizes outside of the pool of immediate (former) competitors. The effect is not only present for behavior towards former contestants, but also towards individuals who were not involved in the competition.

4 Conclusion

We combined a large-scale randomized education intervention in Chile with a series of tailor-made data collections and show that enduring exposure to a more competitive environment persistently and generally attenuates prosociality. Based on a 4-year follow-up, we have shown, for the first time, that enduring competition does not only change situation- and context-specific behaviors, but also future behaviors in situations independent of the past competition. Therefore, our results suggest that competitive environments might not only affect prosocial behavior but also prosociality as a trait, and thus affect the lives of people at large.

We connect two important scientific and practical debates. First, we contribute to a better understanding of the formation of prosociality by providing the first

[27] For a detailed analysis of the treatment effects on education outcomes see Tincani, Kosse, and Miglino (2023).

[28] At the 4-year follow-up, the correlation between 'General prosociality' and 'Prosociality towards former schoolmates' is 0.475.

causal evidence on attenuating environmental factors and by showing that the late adolescence constitutes a formative period. Second, we extend the theoretical and laboratory-experimental literature on the effects of tournament compensation schemes by presenting the first real-world effects of enduring competition on prosociality. Notably, a recent meta-study based on 45 short-term, online experiments on the effect of competition and moral behavior reports a quite similar effect size of about 0.09 standard deviations (Huber et al., 2023). Our study complements these findings by showing the effects of a persistent competitive environment can be long-lasting and independent of the specific context. This indicates that prosociality, as a trait, can change in response to competitive environments. These findings are of great practical relevance as competition among individuals is a pervasive fact of life: relative comparisons often determine success in firms, politics, education, sports, and many other contexts.

Moreover, our findings are of concrete policy relevance as they indicate that percent plan policies can have unintended side-effects by lowering prosociality.[29] This pattern might counteract other positive effects of percent plan policies as prosociality is generally positively related to e.g., labor market and health outcomes (Deming, 2017; Fang et al., 2022). These arguments are of particular importance in the context under study, as Kosse and Tincani (2020) have shown a relatively strong relation between prosociality and labor markets success in Chile.

Future research could study how negative effects of percent plan policies on individuals' prosociality might be avoided or at least reduced. One possibility would be to shift the level of competition. The rules could be adapted in a way such that not the ranking within a certain school but, e.g., the ranking within all low SES students from a certain region decides the allocation of preferential university slots. In such a regime, the level of within-school competition is lower, as the relative ranking within a school is less important. Moreover, creating between-school competitions, e.g. in form of school-level incentives for performance, might even increase within-school cooperation (Bornstein, Gneezy, and Nagel, 2002; Goette et al., 2012; Lowe, 2021).

[29]For a related study on spillover effects of affirmative action on competitiveness and unethical behavior, see Banerjee, Gupta, and Villeval (2018).

References

Angrist, Joshua D and Jörn-Steffen Pischke. 2009. *Mostly harmless econometrics: An empiricist's companion.* Princeton university press.

Banerjee, Ritwik, Nabanita Datta Gupta, and Marie Claire Villeval. 2018. "The spillover effects of affirmative action on competitiveness and unethical behavior." *European Economic Review* 101:567–604.

Bednar, Jenna, Yan Chen, Tracy Xiao Liu, and Scott Page. 2012. "Behavioral spillovers and cognitive load in multiple games: An experimental study." *Games and Economic Behavior* 74 (1):12–31.

Benistant, Julien, Fabio Galeotti, and Marie Claire Villeval. 2022. "Competition, information, and the erosion of morals." *Journal of Economic Behavior & Organization* 204:148–163.

Bornstein, Gary, Uri Gneezy, and Rosmarie Nagel. 2002. "The effect of intergroup competition on group coordination: An experimental study." *Games and Economic Behavior* 41 (1):1–25.

Buser, Thomas. 2016. "The impact of losing in a competition on the willingness to seek further challenges." *Management Science* 62 (12):3439–3449.

Buser, Thomas and Anna Dreber. 2016. "The flipside of comparative payment schemes." *Management Science* 62 (9):2626–2638.

Cappelen, Alexander, John List, Anya Samek, and Bertil Tungodden. 2020. "The Effect of Early-Childhood Education on Social Preferences." *Journal of Political Economy* 128 (7):2739–2758.

Cason, Timothy N and Lata Gangadharan. 2013. "Cooperation spillovers and price competition in experimental markets." *Economic Inquiry* 51 (3):1715–1730.

Cason, Timothy N, Anya C Savikhin, and Roman M Sheremeta. 2012. "Behavioral spillovers in coordination games." *European Economic Review* 56 (2):233–245.

Charness, Gary, David Masclet, and Marie Claire Villeval. 2014. "The dark side of competition for status." *Management Science* 60 (1):38–55.

Chowdhury, Subhasish M and Oliver Gürtler. 2015. "Sabotage in contests: a survey." *Public Choice* 164 (1-2):135–155.

Cooper, Ryan, Javier Guevara, James Kinder, Mario Rivera, Antonia Sanhueza, and Michela M. Tincani. 2022. "The impacts of preferential college admissions for the disadvantaged: experimental evidence from the PACE programme in Chile." IFS Working Paper 22/19.

Cooper, Ryan, Javier Guevara, Mario Rivera, Antonia Sanhueza, and Michela M. Tincani. 2019. "Evaluación de Impacto del Programa PACE." Report of the Chilean Ministry of Education.

Deming, David J. 2017. "The Growing Importance of Social Skills in the Labor Market." *The Quarterly Journal of Economics* 132 (4):1593–1640.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde. 2009. "Homo Reciprocans: Survey Evidence on Behavioural Outcomes." *The Economic Journal* 119 (536):592–612.

Falk, Armin, Anke Becker, Thomas Dohmen, David Huffman, and Uwe Sunde. 2022. "The preference survey module: A validated instrument for measuring risk, time, and social preferences." *Management Science* .

Falk, Armin, Ernst Fehr, and David Huffman. 2008. "The power and limits of tournament incentives." *Working paper, University of Bonn* .

Falk, Armin, Fabian Kosse, Pia Pinger, Hannah Schildberg-Hörisch, and Thomas Deckers. 2021. "Socioeconomic status and inequalities in children's IQ and economic preferences." *Journal of Political Economy* 129 (9):2504–2545.

Fang, Ximeng, Timo Freyer, Chui-Yee Ho, Zihua Chen, and Lorenz Goette. 2022. "Prosociality predicts individual behavior and collective outcomes in the COVID-19 pandemic." *Social Science & Medicine* 308:115192.

Fehr, Ernst, Helen Bernhard, and Bettina Rockenbach. 2008. "Egalitarianism in Young Children." *Nature* 454 (7208):1079–1083.

Fehr, Ernst, Urs Fischbacher, Bernhard von Rosenbladt, Jürgen Schupp, and Gert G. Wagner. 2002. "A Nation-Wide Laboratory." *Journal of Applied Social Science Studies* 122:519–542.

Fehr, Ernst, Simon Gächter, and Georg Kirchsteiger. 1997. "Reciprocity as a Contract Enforcement Device: Experimental Evidence." *Econometrica* 65 (4):833–860.

Fehr, Ernst, Daniela Glätzle-Rützler, and Matthias Sutter. 2013. "The Development of Egalitarianism, Altruism, Spite and Parochialism in Childhood and Adolescence." *European Economic Review* 64 (1):369–383.

Gneezy, Uri, Muriel Niederle, and Aldo Rustichini. 2003. "Performance in Competitive Environments: Gender Differences." *The Quarterly Journal of Economics* 118 (3):1049–1074.

Goette, Lorenz, David Huffman, Stephan Meier, and Matthias Sutter. 2012. "Competition Between Organizational Groups: Its Impact on Altruistic and Antisocial Motivations." *Management Science* 58 (5):948–960.

Guiso, Luigi, Paola Sapienza, and Luigi Zingales. 2009. "Cultural Biases in Economic Exchange?" *The Quarterly Journal of Economics* 124 (3):1095–1131.

Harbring, Christine and Bernd Irlenbusch. 2011. "Sabotage in tournaments: Evidence from a laboratory experiment." *Management Science* 57 (4):611–627.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman. 2013. "Are some degrees worth more than others? Evidence from college admission cutoffs in Chile." NBER Working Paper No. 19241.

Heckman, James J., Jora Stixrud, and Sergio Urzua. 2006. "The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24 (3):411–482.

Hemming, Karla and Jen Marsh. 2013. "A menu-driven facility for sample-size calculations in cluster randomized controlled trials." *The Stata Journal* 13 (1):114–135.

Horn, Catherine L. and Stella M. Flores. 2003. "Percent Plans in College Admissions: A Comparative Analysis of Three States' Experiences. Cambridge, MA: Civil Rights Project at Harvard University."

Huber, Christoph, Anna Dreber, Jürgen Huber, Magnus Johannesson, Michael Kirchler, Utz Weitzel, Miguel Abellán, Xeniya Adayeva, Fehime Ceren Ay, Kai Barron, Zachariah Berry, Werner Bönte, Katharina Brütt, Muhammed Bulutay, Pol Campos-Mercade, Eric Cardella, Maria Almudena Claassen, Gert Cornelissen, Ian G. J. Dawson, Joyce Delnoij, Elif E. Demiral, Eugen Dimant, Johannes Theodor Doerflinger, Malte Dold, Cécile Emery, Lenka Fiala, Susann Fiedler, Eleonora Freddi, Tilman Fries, Agata Gasiorowska, Ulrich Glogowsky, Paul M. Gorny, Jeremy David Gretton, Antonia Grohmann, Sebastian Hafenbrädl, Michel Handgraaf, Yaniv Hanoch, Einav Hart, Max Hennig, Stanton Hudja, Mandy Hütter, Kyle Hyndman, Konstantinos Ioannidis, Ozan Isler, Sabrina Jeworrek, Daniel Jolles, Marie Juanchich, Raghabendra Pratap KC, Menusch Khadjavi, Tamar Kugler, Shuwen Li, Brian Lucas, Vincent Mak, Mario Mechtel, Christoph Merkle, Ethan Andrew Meyers, Johanna Mollerstrom, Alexander Nesterov, Levent Neyse, Petra Nieken, Anne-Marie Nussberger, Helena Palumbo, Kim Peters, Angelo Pirrone, Xiangdong Qin, Rima Maria Rahal, Holger Rau, Johannes Rincke, Piero Ronzani, Yefim Roth, Ali Seyhun Saral, Jan Schmitz, Florian Schneider, Arthur Schram, Simeon Schudy, Maurice E. Schweitzer, Christiane Schwieren, Irene Scopelliti, Miroslav Sirota, Joep Sonnemans, Ivan Soraperra, Lisa Spantig, Ivo Steimanis, Janina Steinmetz, Sigrid Suetens, Andriana Theodoropoulou, Diemo Urbig, Tobias Vorlaufer, Joschka Waibel, Daniel Woods, Ofir Yakobi, Onurcan Yilmaz, Tomasz Zaleskiewicz, Stefan Zeisberger, and Felix Holzmeister. 2023. "Competition and moral behavior: A meta-analysis of forty-five crowd-sourced experimental designs." *Proceedings of the National Academy of Sciences* 120 (23):e2215572120.

Kautz, Tim, James J. Heckman, Ron Diris, Bas Ter Weel, and Lex Borghans. 2014. "Fostering and Measuring Skills: Improving Cognitive and Non-cognitive Skills to Promote Lifetime Success." OECD Education Working Paper No. 110.

Knack, Stephen and Philip Keefer. 1997. "Does Social Capital Have an Economic Payoff? A Cross-Country Investigation." *The Quarterly Journal of Economics* 112 (4):1251–1288.

Kosse, Fabian, Thomas Deckers, Pia Pinger, Hannah Schildberg-Hörisch, and Armin Falk. 2020. "The Formation of Prosociality: Causal Evidence on the Role of Social Environment." *Journal of Political Economy* 128 (2):434–467.

Kosse, Fabian and Michela M Tincani. 2020. "Prosociality predicts labor market success around the world." *Nature Communications* 11 (1):1–6.

Lally, Phillippa, Cornelia HM Van Jaarsveld, Henry WW Potts, and Jane Wardle. 2010. "How are habits formed: Modelling habit formation in the real world." *European Journal of Social Psychology* 40 (6):998–1009.

Lazear, Edward P. 1989. "Pay equality and industrial politics." *Journal of Political Economy* 97 (3):561–580.

Lee, David S. 2009. "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *The Review of Economic Studies* 76 (3):1071–1102.

Levene, Howard. 1960. "Robust tests for equality of variance." In *Contributions to Probability and Statistics*, edited by I. Olkin, S.G. Ghurye, W. Hoeffeling, W.G. Madow, and H.B. Mann. Stanford University Press: Stanford, CA, 278–292.

Lowe, Matt. 2021. "Types of Contact: A Field Experiment on Collaborative and Adversarial Caste Integration." *American Economic Review* 111 (6):1807–44.

MinEduc. 2017. "Levantamiento de información para el seguimiento a la implementación del PACE." *Report, MinEduc* .

Niederle, Muriel and Lise Vesterlund. 2007. "Do women shy away from competition? Do men compete too much?" *The Quarterly Journal of Economics* 122 (3):1067–1101.

OECD. 2016. "Education at a Glance 2016: OECD Indicators." OECD Publishing, Paris.

Rao, Gautam. 2019. "Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools." *American Economic Review* 109 (3):774–809.

Romano, Joseph P. and Michael Wolf. 2016. "Efficient Computation of Adjusted p-values for Resampling-based Stepdown Multiple Testing." *Statistics & Probability Letters* 113 (1):38–40.

Shleifer, Andrei. 2004. "Does competition destroy ethical behavior?" *American Economic Review* 94 (2):414–418.

Sutter, Matthias and Martin G. Kocher. 2007. "Trust and Trustworthiness across Different Age Groups." *Games and Economic Behavior* 59 (2):364–382.

Tergiman, Chloe and Marie Claire Villeval. 2022. "The way people lie in markets: Detectable vs. deniable lies." *Management Science* .

Thompson, Richard F and William A Spencer. 1966. "Habituation: a model phenomenon for the study of neuronal substrates of behavior." *Psychological Review* 73 (1):16.

Tincani, Michela M, Fabian Kosse, and Enrico Miglino. 2023. "College Access When Preparedness Matters: New Evidence from Large Advantages in College Admissions." CEPR Discussion Paper No. DP18564.

Appendix

A Additional Tables and Figures

Baseline variables	Mean of Control group	Difference Treatment - Control	N
Student characteristics			
Female	0.476	0.001 (0.054)	9,006
Age	17.54	0.031 (0.052)	9,006
SIMCE score	221.4	7.600 (5.256)	8,944
GPA (grade 10)	5.438	0.013 (0.035)	8,944
Very low SES	0.602	0.014 (0.020)	9,006
Mother's education (years)	9.553	0.081 (0.168)	6,000
Father's education (years)	9.320	0.115 (0.178)	5,722
HH income (in 1000 CLP)	284.0	14.33 (12.79)	6,018
School characteristics			
Cohort size	66.11	-8.50 (10.63)	128
Rural area	0.109	0.016 (0.054)	128

Table A1: Checks for baseline balance. The first column shows means of the baseline variables for the control group. The second column shows the difference in means between treatment and control group. Standard errors of the differences clustered at school level are shown in parentheses. None of the differences is significantly different from zero at any conventional level. Data on parental education and income were collected in form of a parental survey and contain some missings. For details on variables and data sources, see section 2.3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Available at endline (= 1 if yes)		
	(1)	(2)	(3)
Treatment dummy	-0.044	-0.051	-0.011
	(0.036)	(0.036)	(0.038)
Achievement (at baseline, std)		0.048***	0.067***
		(0.013)	(0.010)
Female		-0.027	0.011
		(0.023)	(0.025)
Very low SES dummy		-0.031**	-0.025
		(0.012)	(0.017)
Treatment x achievement			-0.031
			(0.022)
Treatment x female			-0.070
			(0.043)
Treatment x very low SES dummy			-0.013
			(0.024)
Observations	9,006	8,944	8,944
R-squared	0.002	0.013	0.015

Table A2: Test for selective attrition at endline. Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. The dependent variable is one if a student endline measure of prosociality is available, and zero otherwise. All independent variables were collected before the treatment assignment took place. Neither the treatment dummy, nor the interaction effects are significantly different from zero at any conventional level in any specification. The number of observations slightly differs as baseline measures are missing in the registry data for a few students. For details on variables and data sources, see section 2.3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Available at follow-up (= 1 if yes)		
	(1)	(2)	(3)
Treatment dummy	0.010	0.003	0.013
	(0.012)	(0.012)	(0.018)
Achievement (at baseline, std)		0.039***	0.044***
		(0.006)	(0.007)
Female		0.016**	0.015*
		(0.007)	(0.009)
Very low SES dummy		-0.028***	-0.018
		(0.008)	(0.013)
Treatment x achievement			-0.008
			(0.011)
Treatment x female			0.003
			(0.015)
Treatment x very low SES dummy			-0.019
			(0.016)
Observations	9,006	8,944	8,944
R-squared	0.000	0.019	0.019

Table A3: Test for selective attrition at 4-year follow-up. Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. The dependent variable is one if a student endline measure of prosociality is available, and zero otherwise. All independent variables were collected before the treatment assignment took place. Neither the treatment dummy, nor the interaction effects are significantly different from zero at any conventional level in any specification. The number of observations slightly differs as baseline measures are missing in the registry data for a few students. For details on variables and data sources, see section 2.3.***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Standardized prosociality	
	Lower bound	Upper bound
Treatment dummy	-0.292	-0.006
Number of observations	8,944	8,944
Number of selected obs.	5,343	5,343

Table A4: Lee bounds of the treatment effect at endline. The bounds are estimated using the trimming procedure suggested by Lee (2009). For the estimation, to tighten the bounds, the sample is split by baseline achievement (median split).

	Standardized Altruism (1)	Standardized Trust (2)	Standardized Reciprocity (3)
Treatment dummy	-0.073	-0.071	-0.121
Original p-values	(0.093)	(0.013)	(0.003)
Romano-Wolf p-values	[0.094]	[0.032]	[0.009]
Observations	5,343	5,343	5,343

Table A5: Treatment effect on facets of prosociality. Coefficients are ordinary least squares estimates. All regressions use the standard baseline controls (achievement, female, very low SES) as in table 1 column 2. Original p-values are shown in (parentheses), Romano-Wolf p-values are shown in [square brackets], for details on the stepdown adjusted p-values, robust to multiple hypothesis testing, see Romano and Wolf (2016).

	Standardized Prosociality					
	Females (1)	Males (2)	Females (3)	Males (4)	Females (5)	Males (6)
Treatment dummy	-0.075 (0.066)	-0.127** (0.057)	-0.094 (0.061)	-0.143*** (0.051)	-0.096 (0.062)	-0.144*** (0.052)
Baseline controls	No	No	Yes	Yes	Yes	Yes
Weights	No	No	No	No	IPW	IPW
Observations	2,507	2,836	2,507	2,836	2,507	2,836

Table A6: Heterogeneous treatment effects by gender. Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. The table repeats the analyses shown in table 1 separately for female and males students. Columns 1, 3 and 5 show the treatment effects for female students. Columns 2, 4 and 6 show the treatment effects for male students. For details on baseline controls and inverse probability weights, see table 1. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Using the heuristic (= 1 if yes) "always ticking the same response"		
	(1)	(2)	(3)
Treatment dummy	0.001	0.003	0.002
	(0.009)	(0.009)	(0.009)
Achievement (at baseline, std.)		-0.017***	-0.017***
		(0.004)	(0.004)
Female		-0.006	-0.006
		(0.009)	(0.009)
Very low SES dummy		0.012	0.013
		(0.008)	(0.008)
Weights	No	No	IPW
Observations	5,343	5,343	5,343

Table A7: Treatment effect on the use answering heuristics. Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. The dependent variable is one if the participant gave the same response (on 11-point Likert scales) to all three prosociality items and zero else. For details on the control variables see section 2.4.1. Column 3 applies inverse probability weights. For details on baseline controls and inverse probability weights, see table 1. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Standardized prosociality	
	Lower bound	Upper bound
Treatment dummy	-0.232	-0.027
Number of observations	8,944	8,944
Number of selected obs.	1,018	1,018

Table A8: Lee bounds of the treatment effect at 4-year follow-up. The bounds are estimated using the trimming procedure suggested by Lee (2009). For the estimation, to tighten the bounds, the sample is split by baseline achievement (median split).

	Standardized Prosociality (Full sample)	
	(1)	(2)
Treatment x Time	0.003 (0.071)	0.011 (0.080)
Fixed effects	Individual & Time	Individual & Time
Weights	No	IPW
Observations	2,036	2,036

Table A9: Treatment effect over time in the full sample. Coefficients are two-way fixed effects estimates. The regressions include fixed effects for individuals and time. Standard errors clustered at school level are shown in parentheses. The panel covers two points in time: endline and 4-year follow up. The prosociality measure (at both points in time) corresponds to the endline measure as described in section 2.4.2. To construct the measure at follow-up, we use the same three items as at endline, use the endline distribution for the standardization, and apply the PCA weights from the PCA at endline. Therefore, we use the same measure (with the same metric) at endline and at follow-up. Column 2 applies inverse probability weights that account for potential selective attrition and are estimated from a probit model of a binary selection indicator (indicating whether an individual is available for this analysis) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A2 column 3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Standardized Altruism (1)	Standardized Trust (2)	Standardized Reciprocity (2)
Treatment x Time	-0.022 (0.077)	-0.047 (0.078)	0.074 (0.070)
Fixed effects	Individual & Time	Individual & Time	Individual & Time
Weights	IPW	IPW	IPW
Observations	2,036	2,036	2,036

Table A10: Treatment effect over time: altruism, trust reciprocity . Coefficients are two-way fixed effects estimates. The regressions include fixed effects for individuals and time. Standard errors clustered at school level are shown in parentheses. The panel covers two points in time: endline and 4-year follow up. The respective measure (at both points in time) corresponds to the endline measure as described in section 2.4.2. To construct the measure at follow-up, we use the endline distribution for the standardization. Therefore, we use the same measure (with the same metric) at endline and at follow-up. All regressions apply inverse probability weights that account for potential selective attrition and are estimated from a probit model of a binary selection indicator (indicating whether an individual is available for this analysis) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A2 column 3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

	Standardized Prosociality (Sample: Bottom 85% according to GPA ranking at baseline)	
	(1)	(2)
Treatment x Time	0.028	0.049
	(0.083)	(0.087)
Fixed effects	Individual & Time	Individual & Time
Weights	No	IPW
Observations	1,408	1,408

Table A11: Treatment effect over time in the sample: Bottom 85% according to GPA ranking at baseline. Coefficients are two-way fixed effects estimates. The regressions include fixed effects for individuals and time. Standard errors clustered at school level are shown in parentheses. The panel covers two points in time: endline and 4-year follow up. The prosociality measure (at both points in time) corresponds to the endline measure as described in section 2.4.2. To construct the measure at follow-up, we use the same three items as at endline, use the endline distribution for the standardization, and apply the PCA weights from the PCA at endline. Therefore, we use the same measure (with the same metric) at endline and at follow-up. Column 2 applies inverse probability weights that account for potential selective attrition and are estimated from a probit model of a binary selection indicator (indicating whether an individual is available for this analysis) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. See also table A2 column 3. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

Figure A1: Example of the data collection setup. The data collections were conducted by trained field-workers during regular schooling hours. The data collection took place in form of a self-administered paper pencil survey. During the interview students were sitting in a standardized seating order and were not allowed to talk in order to prevent interactions.

B Additional Analyses

B.1 Perceptions of the schooling environment by teachers

At the beginning of grade 12 the Chilean Ministry of Education conducted a survey among class teachers in treatment and control schools. Class teachers are regular teachers assigned to a specific classroom, who have additional tutoring responsibilities (meeting with students, holding orientation classes, meeting with parents). Data are available for 110 schools (54 control, 56 treatment). The survey includes the item "There is a lot of competition to get the best grades in my course". Teachers indicated how much they agree on a 5-point Likert scale.

Table B1 shows the differences in perceived competition in schools rated by teachers in treatment and control schools. Teachers in treatment schools perceive a 26.4% of a standard deviation higher level of competition for the best grades in their classes (see column 1). While in control group schools only 26.2% of teacher "agree" or "strongly agree" with the statement above, column 2 indicates that about 40% more teachers do so in the treatment group (see column 2: 0.109/0.262 = 41.6%).

	Competition for grades (rated by teachers)	
	Standardized	Binary
	(1)	(2)
Treatment dummy	0.264*	0.109
	(0.158)	(0.076)
Observations	165	165

Table B1: Treatment increases perceived competition (rated by teachers). Column 1 shows an OLS coefficient. Column 2 shows an average marginal effect after Probit estimation. Standard errors clustered at school level are shown in parentheses. The binary measure in column 2 takes the values one if teachers rated "agree" or "strongly agree" and zero else. The estimations include controls for the following student and teacher characteristics: shares of female and very low SES students, school average achievement scores (at baseline), teachers' gender and age. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.

B.2 Perceptions of the schooling environment by students

At the beginning of grade 12 the Chilean Ministry of Education conducted a survey among students in treatment and control schools. Data are available for more than 80% of the students in our main sample. The survey includes the following three items regarding aspects of a competitive schooling environment: "I feel support from my classmates when I have trouble learning or understanding something (reversed)", "We fight a lot when we do group work.", "I think it is good that teachers promote the participation of all students in class (reversed)". Students responded using a 5-point Likert scale to indicate their level of agreement. To create a joint measure of the perceived level of competition in the schooling environment, we aggregate the three standardized measures using principal component analysis (PCA). In the analysis below we use the resulting standardized first principal component as our measure of the level of competition. While each of the three items just captures specific aspects of a competitive environment, the idea of our approach is that the source of the joint variation, represented by the principal component, serves as an indicator of the perceived level of competition.

Table B2 indicates that there is a robust and significant positive effect of the treatment on the level of competition in the schooling environment as perceived by students.

	Competition in school score (rated by students, standardized)	
	(1)	(2)
Treatment dummy	0.090**	0.092**
	(0.045)	(0.045)
Weights	No	IPW
Observations	4,246	4,246

Table B2: Treatment increases perceived competition (rated by students). Coefficients are ordinary least squares estimates. Standard errors clustered at school level are shown in parentheses. The regressions include controls for achievement (at baseline), gender and SES. Column 2 applies inverse probability weights which account for potential selective attrition and are estimated from a probit model of a binary selection indicator (indicating whether the competition score is available) regressed on baseline measures of achievement, gender and SES, the treatment dummy and the interactions of baseline measures and the treatment dummy. ***, **, * indicate significance at the 1%, 5% and 10% level, respectively.