

Swiss Leading House

Economics of Education • Firm Behaviour • Training Policies

Working Paper No. 236

How Selectivity Shapes Selection

Claudio Schilter



Universität Zürich
IBW – Institut für Betriebswirtschaftslehre

u^b

^b
UNIVERSITÄT
BERN

Working Paper No. 236

How Selectivity Shapes Selection

Claudio Schilter

March 2025

Please cite as:
"How Selectivity Shapes Selection." Swiss Leading House "Economics of Education"
Working Paper No. 236, 2025. By Claudio Schilter.

Die Discussion Papers dienen einer möglichst schnellen Verbreitung von neueren Forschungsarbeiten des Leading Houses und seiner Konferenzen und Workshops. Die Beiträge liegen in alleiniger Verantwortung der Autoren und stellen nicht notwendigerweise die Meinung des Leading House dar.

Discussion Papers are intended to make results of the Leading House research or its conferences and workshops promptly available to other economists in order to encourage discussion and suggestions for revisions. The authors are solely responsible for the contents which do not necessarily represent the opinion of the Leading House.

The Swiss Leading House on Economics of Education, Firm Behavior and Training Policies is a Research Program of the Swiss State Secretariat for Education, Research, and Innovation (SERI).

www.economics-of-education.ch

How Selectivity Shapes Selection

Claudio Schilter*

University of Bern

March 17, 2025

Keywords: Occupational Choice, Diversity, Gender, HR Policies

JEL Codes: C93, J21, J24, M14

Abstract

This field experiment investigates how stressing selectivity at career information events affects the diversity, size, and quality of the applicant pool. While the total number of applications remains unaffected by stressing selectivity, it reduces the share of female participants and children of migrants in the applicant pool. A key mechanism driving this effect is that treated participants perceive their (also treated) peers as more competitive during the event. Leveraging treatment timing, I find that exposure to such peer behavior significantly contributes to the gender-specific effect of stressing selectivity. Moreover, further analysis reveals that stressing selectivity deters high-quality female participants from applying and attracts low-quality male participants. The results point to de-emphasizing selectivity as a simple way of boosting diversity, particularly when potential applicants interact with one another.

*claudio.schilter@unibe.ch

I thank Uschi Backes-Gellner, Eric Bettinger, Anne Brenøe, Enzo Brox, Friedrich Geiecke, Daniel Goller, Guido Schwerdt, Carmit Segal, Melanie Wasserman, Stefan Wolter, Martina Zanella, and the seminar participants at ESPE, the University of Zurich, the Personnel Economics and Economics of Education Workshop, the VPET-ECON Leading House Retreat, the SERI Conference on VET Research, and the 3rd Berlin Workshop on Empirical Public Economics for their helpful inputs to this project. I also thank Zürcher Kantonalbank and Carmen Manso for their partnership and their full support in running the study and Christina Arn and Lukas Diethelm for their excellent research assistance. Finally, I am grateful to the Swiss State Secretariat for Education, Research and Innovation for the financial support provided through its Leading House on the Economics of Education. The project has been pre-registered at the AEA RCT registry with the ID AEARCTR-0010893.

1 Introduction

Firms strive to build a workforce that is both qualified and diverse, with several recent studies investigating how firms can achieve this goal (e.g., Flory et al., 2021, 2023; Mocanu, 2023; Del Carpio and Guadalupe, 2022; Delfino, 2024). Market analysts predict that annual global spending on diversity programs will exceed \$15 billion by 2026 (Duchene, 2024). At the same time, many firms emphasize selectivity in their hiring processes, e.g., Microsoft seeks “top-performing graduates,” and McKinsey looks for “exceptional people.”¹ Indeed, in a simple economic model, stressing selectivity appears reasonable: it can serve as a signal for discouraging lesser-qualified individuals from applying, thereby saving time and effort, while possibly also attracting higher-qualified ones. However, if women and other minorities are more easily deterred from applying, then stressing selectivity contradicts the goals that firms aim to achieve through their substantial investments in diversity.

The literature on application behavior is inconclusive as to whether stressing selectivity reduces diversity. Recent studies show that providing more selectivity-related information in job listings, such as the current number of applicants (Gee, 2019) or a longer list of necessary qualifications (Abraham et al., 2024), can attract qualified women and boost diversity. In contrast, studies building on gender differences in competitiveness (Niederle and Vesterlund, 2007) show that qualified women are deterred from applying to positions that are more challenging (Coffman et al., 2024) or that have more competitive compensation schemes (Flory et al., 2015; Miller et al., 2024; Samek, 2019).

However, while these studies provide invaluable insights into specific policies associated with selectivity, they often face data limitations. For example, some studies observe only those who apply (e.g., Flory et al., 2015), while others have only sparse data on the (potential) applicants and their beliefs (e.g., Gee, 2019). These limitations both prevent any detailed analysis of how a firm’s perceived selectivity influences application decisions and constrain researchers’ ability to explore the mechanisms driving the effects of

¹See Microsoft (2025); McKinsey (2024).

selectivity-related policies. These limitations, together with firms' emphasis on selectivity and substantial investments in diversity programs, requires asking and answering the following two questions: Does stressing selectivity reduce diversity among applicants and new hires? When firms decide to emphasize selectivity, is there a trade-off between quality and diversity?

Using a randomized field experiment at a firm's career information events, a data source yielding highly detailed information on both applicants and potential applicants, this paper analyzes how a firm's stressing selectivity affects application behavior. Specifically, 28 events targeting students about to finish schooling were organized by the third largest Swiss bank in Zurich, Switzerland. A key advantage of this population is their nearly identical CVs, apart from measurable differences in their academic transcripts. For nearly all event participants, participation in the field experiment was the default option with an opt-out alternative, thereby ensuring high coverage. Combining rich survey data from questionnaires completed by participants before and after each event with firm-level data on their applications, their behavior during the event, and their performance in the hiring process, this setting constitutes a unique and valuable dataset.

All banks in Switzerland are selective by virtue of the high prestige of working for a Swiss bank and usually stress selectivity in their hiring communications. The government-owned partner firm in my study differs in their communication from privately owned Swiss banks in that it explicitly de-emphasizes selectivity—for example, by not requiring applicants to have been in their school's highest academic track. The field experiment allows me to study the differences between de-emphasizing selectivity and the common status quo of emphasizing it. The treatment—administered to half of the participants—is aimed at counteracting this de-emphasis by mimicking the behavior of most privately owned banks.

My analysis consists of five main steps. First, I compare the application behavior of the treatment and the control group to obtain my main results. Second, I examine

how treatment affects participants being offered a position, the participants' beliefs, and their exposure to potentially different peer behavior. I can study peer behavior because I randomize stressing selectivity at the group level. Therefore, the treated (control) participants interact only with other treated (control) participants during the information event. Third, I disentangle the effects of information and peer interaction. As I also randomize the timing when a key part of the treatment is delivered during the career information event, I can use this timing as an instrument for the change in peer behavior.

Fourth, I conduct an explorative analysis who is attracted and who is deterred by stressing selectivity. I investigate this using a causal forest (Athey et al., 2019; Wager and Athey, 2018), grouping participants into terciles of their conditional average treatment effect (CATE), and describing each tercile. I also specifically investigate the effect of stressing selectivity on applicant quality. Finally, I examine the characteristics driving gender differences in treatment response. One explanation is that the female dummy serves as a proxy for gender-typical characteristics. Therefore, I analyze to what extent the female-treatment interaction loses explanatory power when the regression also includes interactions between these characteristics and treatment.

I find that while stressing selectivity encourages local male participants to apply, it discourages female participants from applying, with the effect strongest among those whose parents were born in relatively poor countries. As a result, stressing selectivity does not affect the total number of applications. Moreover, the relative treatment effect—having a lower female share in the pool of applicants—is more pronounced (18 percentage points) than the absolute effect of discouraging female participants from applying (10 percentage points). Adjusting for spillovers amplifies these effects. Furthermore, the effect carries through to the pool of successful applicants, implying that neither meaningful affirmative action on the part of the firm nor self-selection of participants offsets the effect on applications.

Moreover, I find participant interactions to be an important mechanism driving these

effects. Treated participants perceive their (also treated) peers as being more competitive. In line with this result, the firm's staff observed more active and engaged participants in the treated groups—an effect driven entirely by male participants. Therefore, treated participants both received the information about selectivity and experienced their peers' changed behavior. Exploiting the randomized timing of treatment and conducting a back-of-the-envelope calculation suggests that changed peer behavior accounts for approximately half the gender difference in the treatment effect.

Another candidate mechanism is participants' perception of success. Stressing selectivity influences participants' perceptions of general success probabilities, bringing their estimates of how many out of 50 applicants are hired closer to the actual figures. However, it does not affect participants' subjective assessment of their own chances of success. This combination of results suggests that although participants' perceived success probability is not a relevant mechanism, the treatment effectively conveyed an increased sense of selectivity. Moreover, the additional finding that participants overestimate the number of hired applicants—with treatment reducing the magnitude of these overestimates—provides evidence that the partner firm indeed de-emphasizes its selectivity.

Grouping participants into terciles of their CATE reveals negative treatment effects for those in the lowest tercile—indicating that these participants are discouraged from applying—and positive effects for those in the highest tercile. Participants in the lowest tercile not only are more likely to be female but also have higher grades, have parents with higher levels of education, and are more likely to be under-confident about their math and communication skills. Particularly for female participants, stressing selectivity disproportionately discourages those with high grades but low confidence. Moreover, and contrary to the predictions of a simple economic signaling model, my analyses find no evidence that stressing selectivity improves overall applicant quality. Instead, I find that stressing selectivity predominantly discourages high-quality female participants from applying while encouraging low-quality male participants to apply.

Analyzing gender as a potential proxy variable reveals that the gender difference in treatment effects is generally very robust. However, its size is nearly halved and becomes statistically insignificant when the treatment is interacted with measures for the participants being both social and under-confident about their communication skills—a combination typically associated with female participants (correlation coefficient: 0.22). This finding is in line with my earlier finding of the importance of participant interactions.

This paper offers both practical and research contributions to the literature on application behavior. At the practical level, its findings are particularly valuable for firms that collectively spend billions of dollars on diversity initiatives, often investing in strategies with questionable effectiveness (Dobbin and Kalev, 2016). My results demonstrate that de-emphasizing selectivity—a free intervention—can effectively enhance diversity without sacrificing applicant quality or overburdening hiring processes with additional applications.

At the research level, this paper contributes to the growing literature analyzing how the supply side of the labor market can attract a pool of both diverse and qualified applicants. While previous studies use quotas (Niederle et al., 2013), role models (Del Carpio and Guadalupe, 2022), reduced ambiguity in qualifications (Coffman et al., 2024), information about the current number of applicants (Gee, 2019), information on the share of high-performing workers (Delfino, 2024), and firms signaling their desire for diversity (Flory et al., 2021, 2023), my treatment directly examines stressing selectivity by (a) providing the precise number of positions and prior year’s number of applications² and (b) stating that the firm only hires the “most suitable” people. My paper is the first to use highly detailed data to investigate who is attracted to and who is deterred from applying when firms stress selectivity, with clear findings showing that stressing selectivity actually reduces diversity in both applications and hires, and that it predominantly repels high-quality

²Gee (2019) randomly providing the current number of applicants is possibly most related to the treatment in my paper. However, different to my treatment, subjects are informed about the current (not expected) number of people that have applied, with information provided by a third party (LinkedIn), and there is no interaction among potential applicants.

female participants while attracting low-quality male participants. Moreover, my design uniquely allows for examining how job seekers with a shared interest in an employer or a position interact with one another, and I demonstrate and quantify its significance.

In every field experiment, external validity is an important consideration. The participants in this study are potential applicants for apprenticeships, which they typically start at around 16 years of age. However, in the Swiss system, these apprenticeships mark the direct start of careers, comparable in status and trajectory to university pathways. Therefore, it is a high-stakes decision where to apply for apprenticeships. Swiss apprenticeships involve substantial on-the-job experience, with apprentices working in the firm roughly four days a week, interacting with customers, and functioning as regular employees. Banking apprenticeships in particular target students who are mature for their age, are ready to enter the workforce, and serve as a pipeline for the bank's future managers.³ These characteristics, along with the bank's international importance, suggest that my findings may extend beyond this specific context to other structured career-entry programs and broader hiring practices.

The remainder of this paper is organized as follows: Section 2 explains the institutional context, Section 3 outlines the experimental design, and Section 4 presents the main results. Section 5 analyzes the mechanisms underlying these results. Section 6 concludes and discusses the hiring process implications for practitioners.

2 Setting

The study takes place in Zurich, Switzerland, in cooperation with Zürcher Kantonalbank (ZKB). ZKB is the third largest bank in Switzerland by assets, the 40th largest in Europe, and comparable in size to top-20 banks in the United States (Jimenea et al., 2024). It is also consistently ranked as one of the safest banks in the world (currently second, see Global

³Given the partner firm's focus on addressing low diversity in managerial positions, these apprenticeships play a crucial role in tackling the issue at its root.

Finance, 2024), making it a globally relevant player in financial stability, if not in overall banking.

Switzerland has an apprenticeship-focused education system: around two thirds of the population complete a three- to four-year apprenticeship after nine years of schooling. Apprenticeships can lead to prestigious careers with the previous CEO of ZKB and the current CEO of the largest Swiss bank (UBS) being just two examples of successful apprenticeship graduates. Aside from different levels (level A being more challenging than level B),⁴ students have virtually identical courses and therefore also highly comparably CVs—except for their transcripts—when they choose an apprenticeship. Students generally choose their target professions and employers in eighth grade and then apply roughly in early ninth grade. They apply on the free and open market, with each employer having their own application process.

ZKB offers apprenticeships in four professions. They host information events about these apprenticeships (and the careers thereafter) in the second half of eighth grade (see Figure 1). The events take half a day and students are allowed to skip school for such information events. ZKB aims to inform broadly, hence the only requirements on signing up are to be in eighth grade and fast enough to secure a spot before the events are fully booked. Visiting the information events is not necessary for applying to an apprenticeship. However, although not communicated to the students, ZKB “greenlists” students that make a good impression during the information events. All students receive an email after the event encouraging them to apply, but greenlisted participants receive an additional phone call from ZKB on top of that. Moreover, their green status can be beneficial in the assessment process. The assessment process involves first the screening of the application package and then two separate assessment days (if the first day was successful).

ZKB is a state-owned firm and, contrary to privately owned banks in Zurich, they

⁴Level C is the easiest, but none of the participants in this study visit level C. There is also the academic school type. These students do not usually apply for apprenticeships. Only one participant visits an academic school. This person has been pooled into level A.

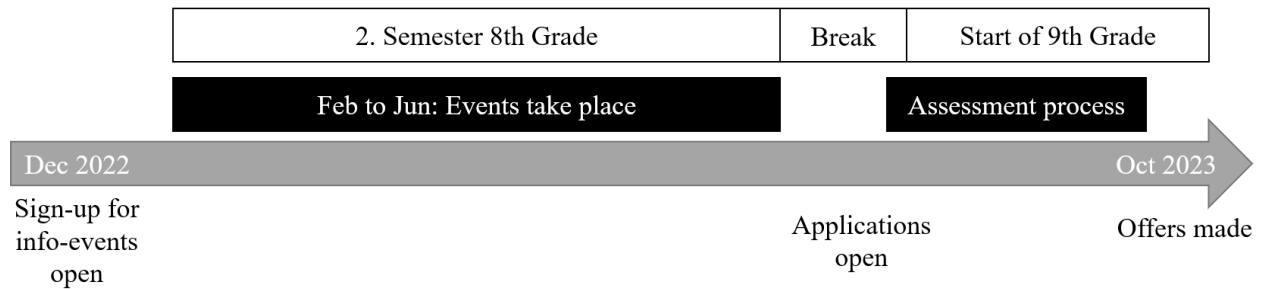


Figure 1: Timeline

de-emphasize their selectivity. They are the only bank not requiring applicants to be in the highest school track; they explicitly write on their website that they are not looking for the best, but motivated and curious people; and they are the only bank to mention the necessary grade-cutoff online—and it is relatively low. Since an apprenticeship in a bank is popular and prestigious, they are still only employing a small fraction of their applicants. Thus, the key difference to privately owned banks is in the communication of selectivity rather than the selectivity itself. ZKB also aims to provide career-information to a wide audience—beyond a strict profit motif. Therefore, they host more information events than privately owned banks, making them an attractive partner for research in terms of sample size considerations. They also do not stress their selectivity during these events (absent of treatment).

The information events begin and end with the participants watching a presentation by a supervisor that introduces ZKB to them (at the beginning) and highlights the requirements, perks, and application deadline (at the end). In between, they attend workshops led by current apprentices that are aimed to provide some insight to each of the apprenticeship years. Additionally, one workshop is held by a supervisor and includes a part where common questions are discussed (students can send in questions for this). For the most popular apprenticeship, the commercial apprenticeship, approximately 40 students visit each information event. They are divided into four groups to attend the workshops, which are held as parallel sessions. For the other apprenticeship occupations, IT and two

occupations that sit between IT and the commercial apprenticeship,⁵ the information events have fewer participants and all workshops are attended together.

3 Experimental Design

The experiment contains three surveys and a treatment with four parts. Having a multi-part treatment helps ensure that students receive the information despite potential lapses in attention. The students are exposed to new information over the course of four hours and they are not used to interacting with a potential employer, making it easy to overlook one or two brief messages.

3.1 Sequence and Treatment

3.1.1 Baseline survey

Upon arrival at the information event (or shortly before that), all participating students were asked to take a baseline survey, which includes questions about basic demographic characteristics, the students’ educational and parental background, perceptions of own skills, and baseline beliefs regarding ZKB’s selectivity. The first component of the treatment is embedded in the treatment group’s survey:

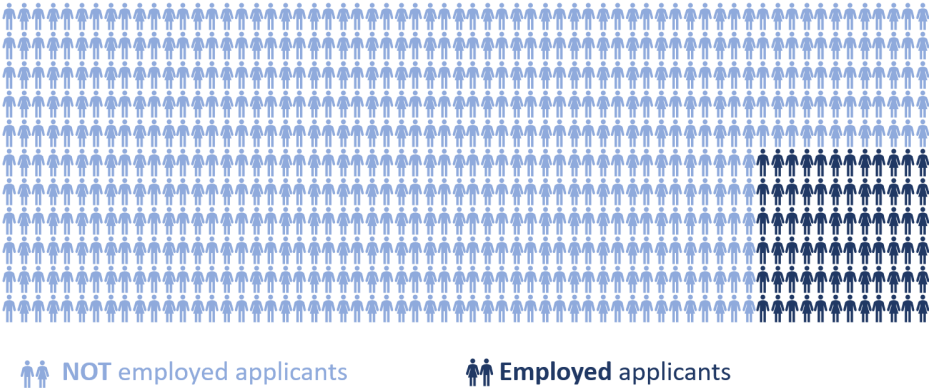


Figure 2: Illustration shown to Treated Participants

⁵Namely “Mediamatics” and “Digital Business Development.”

Did you know that ZKB is a very selective employer? Every year around X students apply for an apprenticeship in occupation Y at ZKB. However, only Z start their apprenticeship. The competition is fierce and ZKB only looks for the most suitable applicants

The placeholder Y is the relevant apprenticeship occupation of the information event in question, and the placeholders X and Z are the respective numbers of applicants and filled positions. Along with this, the numbers (X and Z) are illustrated as shown in Figure 2. The control group received information on how each week is divided between school days and work days during the apprenticeship instead. This information is publicly accessible and generally mentioned later during the information event.

3.1.2 Q&A session

The treatment group was always exposed to two questions concerning selectivity (among other questions): “How many apprenticeship applications does ZKB receive in occupation Y?” and “How many apprenticeship positions does ZKB have in occupation Y?”. The supervisor in charge of this session asked the participants, provided the answers, and emphasized ZKB’s selectivity. The control group only saw other questions. Since the students do not send in many questions (and none on selectivity were sent in during the study period), it is common to use questions sent in by previous cohorts. This is likely the most impactful part of the treatment, as it is delivered in person by a senior ZKB employee.

3.1.3 Follow-up survey

At the end of the information session, the students were asked to take a follow-up survey with questions about their experiences during the day, impressions of ZKB, knowledge of ZKB’s selectivity, and beliefs regarding their chances of being offered an apprenticeship if they apply. The third part of the treatment is embedded in this survey. Both treatment and

control students were asked how many out of 50 applicants actually start an apprenticeship at ZKB. After answering, treatment students received the correct answer, again along with an illustration as in Figure 2. The control group received instead information on the information sessions held by ZKB, the timeline of when students apply, and when applicants hear back from ZKB. This information is publicly available and previously communicated during the day.

3.1.4 Follow-up email

Shortly after the event, each student received a personalized email with links to information about the apprenticeships at ZKB, where to apply, and when the call for applications opens. In the email to treated students, the part of the email that says “We would be very happy if you apply with us this summer!” (for the control group) was replaced by “We only consider the most suitable candidates. If you think you are a fit, we would very happy if you apply with us this summer!”.

3.1.5 Endline survey

In late November 2023, i.e., after offers were made and either accepted or rejected, students were invited to complete a brief endline survey online. This survey mostly contains questions regarding the occupations and firms to which the students have applied and how these other firms compare to ZKB in terms of competitiveness and selectivity.

3.2 Randomization

I first randomly assigned individual students attending the information events on the commercial apprenticeship to either treatment or control, stratifying by gender. Next, I divided both treated and control students into two groups each, without stratification. Finally, I randomly determined the order in which these groups attended the workshop with the Q&A session, placing them first, second, third, or fourth on the event schedule.

For students attending any of the other three occupations, who were not divided into groups, I conducted randomization at the event level. I paired the first event with the second, the third with the fourth, and so on. Within each pair, I randomly assigned one event to treatment and the other to control, ensuring that each pair had one event in each condition.

Spillovers are an important concern in this design. First, for events on the commercial apprenticeship, the design does technically not rule out that participants ask about selectivity in the plenum sessions at the beginning and end of the event (where both treatment and control groups are present). However, the head organizers were specifically instructed to notice any such questions and there were no discussions concerning selection in any of the events. Moreover, all measures that are elicited during the event cannot be influenced by spillovers. Therefore, the mechanisms discussed in Section 5.2 and 5.3 are not affected. Finally, I address spillovers on the application behavior using information of the participants' school and class in Sections 4 and 5.1.

	Control			Treatment			Diff
	n	mean	sd	n	mean	sd	
Female	224	0.46	0.50	287	0.41	0.49	-0.044
Age	224	14.19	0.47	287	14.23	0.46	0.038
School Type B	224	0.08	0.27	287	0.05	0.22	-0.027
Occ: Commercial	224	0.75	0.43	287	0.69	0.46	-0.057
Occ: IT	224	0.14	0.35	287	0.18	0.39	0.046
Occ: Mediamatics	224	0.08	0.27	287	0.06	0.24	-0.017
Occ: Digital Business	224	0.04	0.19	287	0.06	0.24	0.027
Has Baseline Survey	224	0.69	0.46	287	0.67	0.47	-0.019
Has Follow-Up Survey	224	0.71	0.45	287	0.75	0.43	0.039
Parents born low-GDP C.	150	0.29	0.45	206	0.29	0.46	0.005

Note: School Type B is the lower of two relevant school types, a low-GDP country is defined to be a country with a GDP per capita lower than the EU-average.

Table 1: Balance Table

The main dataset consists of 511 observations⁶ from 28 information events. Table

⁶If a participant attends an information event about the same occupation and in the same treatment status, I only keep the first observation. If they experienced both treatment and control, I only keep the observation where treatment occurred. Moreover, I drop 87 additional observations from the main dataset

1 shows that the randomization is balanced with no statistically significant differences between treatment and control. Whether the applicants' parents are born in a low-GDP country is a pre-registered dimension of heterogeneity. However, this information is only available from the baseline survey and therefore unavailable from those who have not completed this survey at least up to the relevant question.

4 Main Results: Effect on Application Behavior

The two pre-registered hypotheses both concern the participants' application behavior. Hypothesis 1 states that the treatment deters students from applying to the position. Hypothesis 2 states that the treatment will deter minorities from applying; specifically, girls and children of migrants from countries with a GDP per capita below the EU average. Following my pre-registration, I am testing this by interacting the treatment with the participants' gender and their parents' origin respectively.

The results depicted in Table 2's columns (1) and (2) address Hypothesis 1 and clearly refute it: The treatment did not lead to fewer applications in total.⁷

The remaining columns evaluate Hypothesis 2, examining differential treatment effects by gender and parental origin. For girls and children with parents from low-GDP countries, the difference in treatment effect ranges from 16 to 18 percentage points (approximately 0.3 SD). The sum of the coefficients for treatment and its interaction with gender is significant (due to the one-sided pre-registration), indicating that the treatment discourages girls from applying, with an effect size of 10 percentage points (0.2 SD). While the sums of the respective coefficients are also negative in columns (4) and (5), they are not statistically significant, likely due to the smaller sample size. Column (6) reveals a sig-

where key information is missing (age, date, and/or whether or not they have applied). However, I include these 87 observations to identify classes where spillovers may have occurred.

⁷Throughout this paper, I cluster standard errors at the level of randomization, meaning that observations from information events on the commercial apprenticeship are clustered at the observation-level and observations from the other information events are clustered at the event level (see Section 3.2).

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.004 (0.051)	0.049 (0.065)	0.080 (0.067)	0.133 (0.081)	0.105 (0.077)	0.180** (0.088)
Treated*Female			-0.178** (0.092)	-0.185** (0.112)		-0.174* (0.112)
Tr*Par. born low-GDP C.					-0.173* (0.122)	-0.160* (0.124)
Female	-0.055 (0.047)	-0.044 (0.058)	0.047 (0.067)	0.065 (0.085)	-0.046 (0.058)	0.058 (0.085)
Parents born low-GDP C.					0.114 (0.093)	0.103 (0.093)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared	0.090	0.090	0.097	0.097	0.095	0.102
N	511	356	511	356	356	356
Control Mean of Dep. Var.	0.509	0.520	0.509	0.520	0.520	0.520
Sum Tr. + Interacation(s)			-0.099* (.071)	-0.052 (.091)	-0.068 (.104)	-0.155* (.117)

Note: The dependent variable is if a participant has applied. Columns (2) and (4) replicate the specification in columns (1) and (3) respectively, but with the same sample that also contains information on migration background (as in columns 5 and 6). Standard errors are clustered at the level of randomization. One-sided p-values are used for the interactions in (3), (4), (5), and (6) due to one-sided pre-registration (including the sums of the interactions with the treatment coefficients). Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Application Behavior

nificant and substantial effect: Girls with migrant parents are discouraged from applying by 15 percentage points (0.3 SD), while boys with local or high-GDP-origin parents are encouraged to apply, with an effect size of 18 percentage points (0.4 SD).

In sum, the results in Table 2 show that the treatment reduces diversity in the pool of applicants while not changing its size. Table A1 in the appendix leads to the same qualitative conclusions using probit estimation instead of OLS. The results are also robust to adding grades and non-cognitive characteristics (sociability, risk-aversion, competitiveness, shyness) as control variables. However, doing so makes columns (1) and (3) obsolete

since these characteristics are only available for participants who completed the baseline survey.

The results in Table 2 do not account for information spillovers occurring after the information event but before the application deadline. The most relevant peer group of secondary school students is arguably their classmates. I use baseline-information about the participants' school and class to run a "spillover-proof" regression by excluding any classes that include both treated and control participants. Classes with only one participant are retained, while classes with two participants are excluded with a 50% probability. Classes with three participants are excluded with a 75% probability et cetera.⁸ Finally, I apply probability weights to account for the exclusion probability depending on the number of participants per class.

Table A2 in the appendix presents the results from this "spillover-proof" regression. The point estimates are considerably higher, consistent with classic spillovers that mitigate the treatment effect. For example, the effect of treatment to discourage girls from applying in column (3) increases from 10 to 17 percentage points. These higher magnitudes also lead to the coefficient on treatment in columns (3), (4), and (5)—i.e., the treatment's effect on boys and locals respectively—to be statistically significant.

5 Mechanisms

5.1 Effect on Offered Positions

A first important question regarding mechanisms is whether discouraged female participants were those who were unlikely to receive an offer regardless. Table 3 shows that the results presented in Table 2 extend to offered positions, underscoring the treatment

⁸For the commercial apprenticeship, participants are randomized individually, so this always hold. The other occupations are randomized on the event level. If classmates coordinate to attend these events together, the probabilities mentioned above need not be true. However, only three pairs of two people from the same class attend the same event that is randomized at the event-level. This could likely occur even absent of any coordination. The results also do not change meaningfully by excluding these 6 participants.

	Everybody		Spillover-Proof		Cond. on Application	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.037 (0.034)	0.082* (0.046)	0.111 (0.081)	0.303*** (0.094)	0.059 (0.068)	0.106 (0.083)
Treated*Female		-0.096* (0.062)		-0.444*** (0.134)		-0.107 (0.122)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared	0.080	0.084	0.281	0.324	0.126	0.129
N	508	508	253	253	264	264
C. Mean of Dep. Var.	0.11	0.11	0.12	0.12	0.22	0.22

Note: The dependent variable is if a participant has received an job offer. For columns (1) to (4), this is the combination of participants applying and succeeding in their application. Columns (5) and (6) focus on the subsample of those participants that apply. Columns (3) and (4) use the reduced sample from school classes that either have only control or only treated participants, and p-weights are used to reflect the probability of remaining in the sample. Standard errors are clustered at the level of randomization. One-sided p-values are used for the interactions in Columns (2), (4), and (6) in line with Table 2. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Offered Positions Depending on Treatment

effect’s importance beyond applications alone.⁹ ZKB did not save the treatment status after treatment was completed and this information was therefore not available in the hiring process.

In the first four columns, the dependent variable indicates that a participant both applied and received an offer from ZKB. The last two columns focus on applicants only, examining ZKB’s decision in isolation. The one-sided p-value for the female-treatment interaction in column (6) is 0.19 (0.17 for probit). Although this result is not statistically significant, the negative coefficient—combined with the results in the first four columns and the fact that the coefficients of a spillover-proof version of column (6) are highly significant ($p=0.02$)—challenges theories suggesting that affirmative action could meaningfully

⁹A probit estimation reaches again the same conclusion (see Table A3 in the appendix).

mitigate the gender-specific treatment effect of emphasizing selectivity.

5.2 Interest, Effort Provision, and Behavior during the Event

Treated participants receive information about ZKB’s selectivity through the surveys and directly from ZKB. Economic theory suggests that this treatment can affect participants in at least three ways. First, it may signal aspects of the work environment at ZKB, such as its competitive nature (Connelly et al., 2024). Second, the information can shape participants’ perceived probability of success if they apply—both by influencing their subjective assessment and by providing previously unknown objective data on past hiring rates (Lazear et al., 2018). Third, because treated participants are grouped together, they may experience a different social dynamic, potentially interacting with peers who behave more competitively or are perceived as more competitive, which may deter female participants from applying (Niederle and Vesterlund, 2007). The follow-up survey includes targeted questions to assess each of these potential effects.

Column (1) in Table 4 shows that the treatment had no effect on how competitive ZKB’s work environment is perceived by the participants. Treatment also had no effect on how friendly participants believe the work environment to be or whether participants indicated they would like to work in that environment. Furthermore, there is no statistically significant gender difference in the treatment effect for the regression in column (1), nor for any other regression in Table 4 (results not shown).

Columns (2) and (3) examine participants’ perceptions of the general probability of success when applying to ZKB for an apprenticeship. The treatment directly provides information on the fraction of applicants who secure an apprenticeship. At baseline, participants overestimated this number, consistent with ZKB de-emphasizing its selectivity. The treatment helped correcting this misperception by causing a downward revision of these assessments.¹⁰ However, as shown in column (4), the treatment did not alter

¹⁰The participants’ answers are rather noisy though, with some presumably not willing to do maths

	Comp WE	How many Start		Can Get Job	Competitive Peers	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.139 (0.223)	-3.015*** (0.910)	-1.998** (0.937)	0.024 (0.256)	0.451** (0.227)	0.131** (0.063)
Baseline Value			0.353*** (0.055)			
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared		0.253	0.366			0.071
N	366	300	237	365	366	366
C. Mean of Dep. Var.		12.94	12.36			0.37
Method	ologit	ols	ols	ologit	ologit	ols

Note: The dependent variables are if the participants perceive the work environment at ZKB (1) or the other participants (5 & 6) to be competitive, how many out of 50 applicants they believe to start the apprenticeship at ZKB (2 & 3), and how likely participants think it is that they are hired if they apply (4). Columns (1), (4) & (5) are measured with 5-item Likert scales. Column (6) reports a dummy-dependent variable that takes the value 1 if the participants answered “Rather Competitive” or “Very Competitive”. For columns (2) and (3), outliers were dropped that indicated that more than half of the applicants would start an apprenticeship at ZKB. Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type. The abbreviation “ologit” stands for ordered logit.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: Immediate Self-Reported Effects of Treatment

participants’ perceptions of their own likelihood of receiving an offer from ZKB.

Furthermore, treated participants perceived the other participants in their group to behave more competitively (see column 5). The magnitude of this effect is considerable: an OLS-regression reveals that treated participants were 13.1 percentage points (0.3 SD) more likely to perceive others to be “Rather Competitive” or “Very Competitive” (see column 6).

This change in behavior during the information event is evident not only in reports from the follow-up survey but also in observations by ZKB’s staff. The event organizers—typically one or two senior HR representatives along with about eight of ZKB’s

at the end of an intense information event or accidentally answering how many out of 50 were declined. Estimates in Table 4 therefore use a sample that drops the 17% of outliers that stated that more than half of the applicants would start the apprenticeship.

apprentices—meet after each event to identify (“greenlist”) promising applicants. They assess candidates based on their engagement, such as asking questions, participating in discussions, and responding to organizers’ inquiries. ZKB does not inform participants about this practice. While the senior representative delivering the treatment during the Q&A session is necessarily aware of the treatment status, the apprentices—who play a key role in greenlisting as the interactive nature of the other sessions provides better opportunities to assess participants—are generally blind to it.¹¹

For most participants, these information events represent one of their first interactions with an actual firm, and the majority tends to be reserved. The behaviors associated with greenlisting—actively engaging, asking questions, and participating in discussions—deviate from this general tendency and require effort. Moreover, by definition, being greenlisted indicates that a participant has demonstrated interest in the position.

The treatment makes participants considerably more likely to be greenlisted, as shown in Table 5’s columns (1) and (2). The effect size of between 6 and 9 percentage points corresponds to between 0.2 and 0.3 SD. As all findings in Table 5, they are robust to using probit instead of OLS (see Table A4 in the appendix).

This response to the treatment differs significantly by gender. Boys are strongly motivated by the treatment, and their treatment effect is between 0.4 and 0.5 SD. However, the gender-treatment interaction is equally large and negative, implying a zero-effect of the treatment on girls’ behavior during the day (see columns 3, 4, and 6). Participants with parents born in low-GDP countries also respond less to treatment, but this interaction is not statistically significant. At the same time, the sum of the main effect with the interaction effect of migrant parents (columns 5 and 6) or both interaction effects (column 6) is statistically not significantly different from 0.

¹¹The expectation that the apprentices are blind to the participants’ treatment status is arguably more realistic when treatment is provided to only a fraction of the participants at once in a separate room away from all but one apprentice, i.e., at information events about the commercial apprenticeship (majority of the sample). Using just this sub-sample for the regressions in Table 5 leads to very similar results (even with slightly higher point estimates and lower p-values for the coefficients on the interaction of treatment with gender).

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.064** (0.032)	0.092** (0.045)	0.118*** (0.044)	0.152*** (0.052)	0.124** (0.050)	0.178*** (0.055)
Treated*Female			-0.116* (0.061)	-0.132* (0.073)		-0.126* (0.072)
Tr*Par. born low-GDP C.					-0.100 (0.085)	-0.091 (0.084)
Female	0.015 (0.031)	0.019 (0.038)	0.081* (0.042)	0.097* (0.054)	0.017 (0.038)	0.092* (0.053)
Parents born low-GDP C.					0.078 (0.064)	0.071 (0.062)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared	0.065	0.103	0.072	0.111	0.107	0.114
N	509	356	509	356	356	356
Control Mean of Dep. Var.	0.090	0.107	0.090	0.107	0.107	0.107
Sum Tr. + Interaction(s)			.002 (.044)	.02 (.064)	.024 (.078)	-.038 (.086)

Note: The dependent variable is if participants have been noted by ZKB's staff to have behaved sufficiently interested and competent such that they encourage their application in a phone call after the event. Columns (2) and (4) replicate the specification in columns (1) and (3) respectively, but with the same sample that also contains information on migration background (as in columns 5 and 6). Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 5: Noted as Desirable Applicant

5.3 Interaction vs. Information

As a consequence of the changed behavior during the event, treated participant experience both a direct effect from receiving information about selectivity and an indirect effect from interacting with peers who also respond to this information. I can analyze the impact of this peer behavior change by leveraging the randomized timing of the Q&A sessions at the information events for the commercial apprenticeship.

In treated groups, participants who experience the Q&A session receive a key part of the treatment—the part delivered directly by firm staff—earlier in the information event. This timing means that the behavioral effect should occur sooner and be more noticeable

compared to participants with later Q&A sessions. Column (1) of Table 6 shows that this effect indeed exists.¹²

In columns (3) and (4) of Table 6, I regress whether a participant applied on the share of greenlisted peers in their group. I instrument this share and the share interacted with gender with both the timing of the Q&A session and gender interacted with this timing. The first stages are strong, as reported in columns (1) and (2). In columns (3) and (4), the rk LM test rejects underidentification with a p-value of 0.00. The Hansen J test for overidentification yields a p-value of 0.72. While the 2SLS regression in column (3) falls short of Stock-Yogo thresholds for strength, the LIML regression in column (4) meets this criterion. A placebo first-stage regression for the control group yields insignificant coefficients near zero, underscoring that the session order only influences outcomes when a treatment message is involved.

The results indicate that in groups where the treatment induced more participants to engage and become greenlisted, a gender gap in application rates emerges. A back of the envelope calculation reveals that this effect is substantial. For this calculation, I use the regressions reported in column (1) of Table 5 and in column (3) of Table 2, but I use the subsample of participants who attended events for the commercial apprenticeship. The former regression results in a point estimate of 5.7 percentage points more participants becoming greenlisted due to the treatment. The latter regression has a point estimate of -16.1 percentage points for the gender difference in applications due to the treatment. Multiplying the 5.7 percentage points with the point estimate of -1.40 (-1.35) from column (4) (column 3) of Table 6 yields a gender difference in applications due to exposure to greenlisted peers of -8.0 (-7.7) percentage points—almost exactly half of -16.1 percentage points.¹³

¹²More specifically, the results in column (1) show that the primary effect stems from receiving treatment immediately in the first workshop (possibly making the treatment particularly salient).

¹³Using the regressions reported in column (2) of Table 5 and in column (4) of Table 2 instead of column (1) and (3) respectively yields the similar result that exposure to greenlisted contributes approximately 60% to the total treatment effect.

	First Stages		IV Reg.	
	(1)	(2)	(3)	(4)
Share noted desirable			0.872 (0.734)	0.905 (0.767)
Share noted desi.*Female			-1.348* (0.805)	-1.397* (0.834)
Q&A second	-0.198*** (0.031)			
Q&A third	-0.080*** (0.018)			
Q&A fourth	-0.097*** (0.025)			
Q&A second*Female		-0.216*** (0.038)		
Q&A third*Female		-0.174*** (0.035)		
Q&A fourth*Female		-0.139*** (0.039)		
All prev. FE & Controls	✓	✓	✓	✓
Dep. Var.	Share Desi.	S. Desi.*Fem.	Appl.	Appl.
R-squared	0.67	0.63	0.07	0.07
N	199	199	199	199
Mean of Dep. Var.	0.14	0.07	0.51	0.51
Method	ols	ols	2sls	liml
Joint Sign. Instr. (F)	14.73	11.77		
Underident. Test (rk LM)			25.74	25.74
Weak Ident. Test (C-D Wald F)			7.96	7.96
Overident. Test (Hansen J)			2.10	2.10

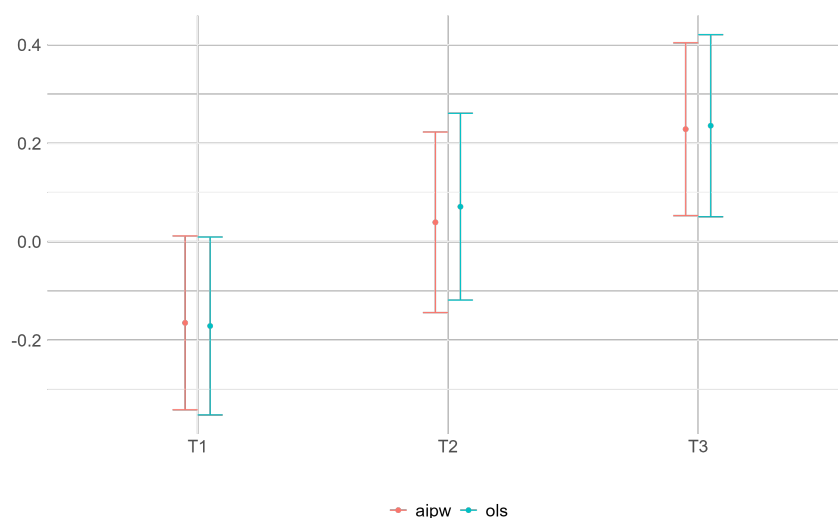
Note: The sample is reduced to treated participants in those information events where participants were further randomized into four groups at each event. The dependent variables in the first stages are the group-share of participants noted as active and engaged (“desirable”) and this same variable interacted with gender respectively. A key part of the treatment was delivered during Q&A sessions, which groups attended in random order. The groups attending the Q&A session first serve as the reference group. The dependent variable in the IV regressions is whether or not participants applied. Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Treatment Effect of Changed Peer Interaction

5.4 Who is Attracted and Who is Deterred by Treatment?

The results in Section 4 confirm the pre-registered hypothesis that stressing selectivity deters girls and children of migrants from relatively low-income countries more strongly from applying. In this section, I examine in greater detail who is attracted or deterred from applying due to the treatment. To do so, I conduct an exploratory heterogeneity analysis using a method based on causal forests, closely following Golub Capital Social Impact Lab (2023). Causal forests are ensembles of decision trees that segment groups to maximize differences in conditional average treatment effects (CATE), assigning participants different average treatment effects based on their individual characteristics (see Athey et al., 2019; Wager and Athey, 2018, for further details). After estimating the CATEs with the causal forest, I categorize participants into terciles according to their CATEs.

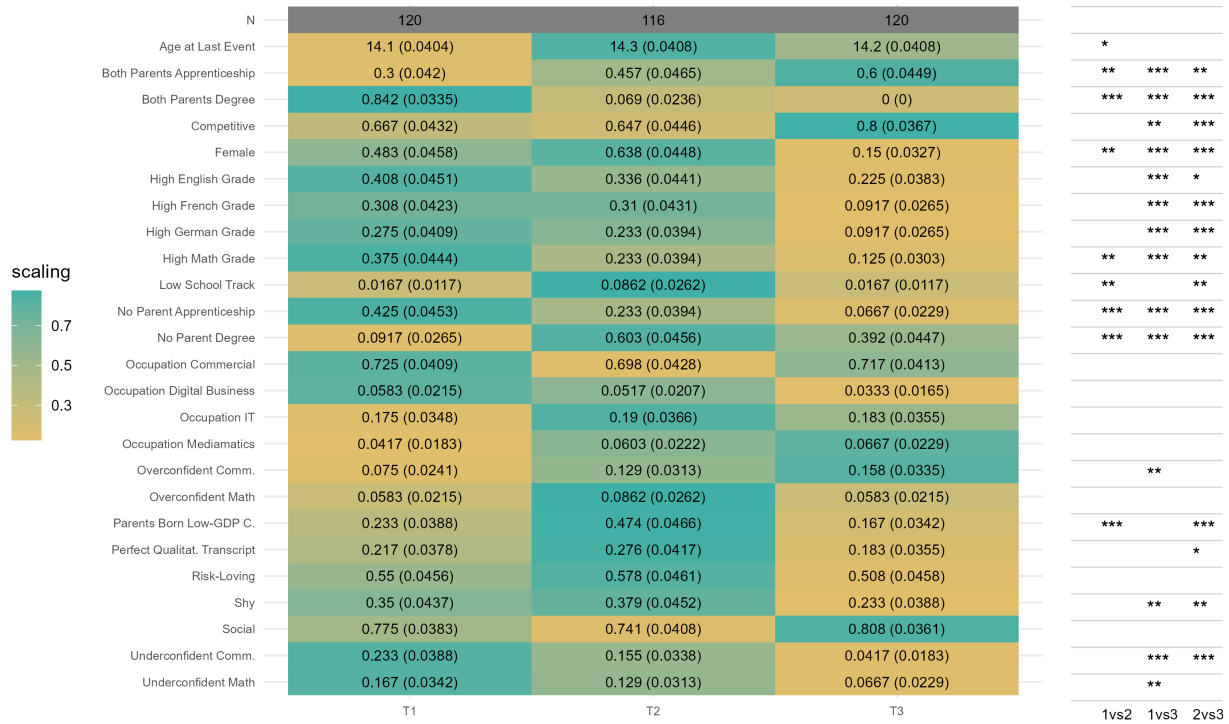


Note: The figure displays tercile averages with their 95% confidence intervals. The abbreviation “aipw” stands for augmented inverse probability weighting. In line with Figure 4, the sample is restricted to those observation with a complete baseline survey.

Figure 3: Conditional Average Treatment Effect Terciles

Figure 3 illustrates that participants in the first tercile experience a negative treatment effect, meaning they are deterred by the emphasis on selectivity (the 95% confidence intervals for both OLS and augmented inverse probability weighting (aipw) include low positive values, though the 90% confidence interval does not). In contrast, participants

in the third tercile exhibit a positive treatment effect, indicating that the emphasis on selectivity motivates them to apply. Participants in the second tercile have CATEs near zero. I follow Golub Capital Social Impact Lab (2023) to test for the presence of treatment effect heterogeneity and reject the null hypothesis of no heterogeneity with a p-value of 0.015.¹⁴



Note: Terciles according to participants’ CATE. A “high” grade refers to a 5.5 or 6 on a scale from 1 to 6. The qualitative part of the transcript is the teacher’s assessment of the students’ study habits, work ethic, and social behavior.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure 4: CATE Terciles Characteristics

Figure 4 displays the analysis for those observations with a complete baseline survey.¹⁵ Beyond gender differences, participants in the first tercile have higher grades compared to those in the third tercile. Grades are a useful proxy for applicant quality because they are highly correlated with the applicants’ success in the hiring process (see Table A5 in the

¹⁴This test evaluates how the CATE predictions on held-out data correlate with the true CATE.

¹⁵This is the same sample as, e.g., in columns (2),(4),(5), and (6) of Table 2. There are not enough dependent variables for a meaningful analysis of the full sample.

appendix). This is not surprising because ZKB hires directly from secondary school and, therefore, the applicants CVs are very similar, with the primary quantitative measures of participant quality being their transcripts.¹⁶

Participants in the first tercile are also less likely to be competitive and overconfident in communication, and more likely to be shy and under-confident in both math and communication than those in the third tercile. I define under- and overconfidence based on the participants' math and German grade relative to their own assessment how many out of 10 students are better than them. If their assessment is wrong by at least 3 students, I classify them as under- or overconfident respectively.¹⁷ Additionally, participants in the first decile are considerably more likely to have parents with degrees and less likely to have parents who completed an apprenticeship. These measures can serve as a proxy for household wealth, suggesting that participants from wealthier households are more likely to be deterred by the emphasis on selectivity.

Analyzing the terciles separately by gender yields results similar to those in Figure 4 for both boys and girls (see Figures A1 and A2 in the appendix). However, the deterrent effect of the treatment on high-performing participants is particularly pronounced for girls. Girls also drive the observed increase in competitiveness among those attracted by the treatment. Firms are unlikely to benefit from such a selection, as Lüthi and Wolter (2023) show that competitiveness does not necessarily enhance apprenticeship quality or stability, especially for girls. The vast majority of those attracted to apply by the treatment are boys with relatively low grades.

Complementing the causal forest analysis, I also directly compare the grades of applicants between the treated and control groups. Once again, there is no evidence supporting a signaling model in which stressing selectivity attracts higher-quality applicants. Nor do I find the pattern observed in Abraham et al. (2024), where their treatment—listing

¹⁶In principle, the school type would be another indicator of applicant quality. However, only 4 participants who have applied are from track B.

¹⁷I compare their grades to the grade distribution obtained by Brenøe and Wasserman (2024).

more qualifications in the job ad—primarily deterred low-quality women from applying. In fact, as shown in Table A6 in the appendix, the opposite is true in my setting, consistent with the result in Figure 4.

5.5 The Female Dummy as Proxy Variable

Interacting the female dummy with the treatment reveals considerable heterogeneity. However, is this dummy merely a proxy for gender-typical characteristics, and if so, which ones? To address this question, I interact non-cognitive variables with the treatment and examine when the treatment’s interaction with the female dummy ceases to be significant. I limit the sample to participants who not only completed the baseline survey but also spent at least 30 seconds answering the questions about being competitive, social, shy, and risk averse. Additionally, I include the measures of under- and overconfidence defined in Section 5.4.

Adding interactions between treatment and these measures, both separately and in pairs (resulting in triple interactions), reveals that the coefficient on the treatment-female interaction remains remarkably stable. Drawing from the literature on gender differences in competitiveness (e.g., Niederle and Vesterlund, 2007) and overconfidence (e.g., Buser et al., 2020), I report the regressions with these covariates both separately (columns 3 and 4) and together (column 5) in Table 7. The coefficient on the treatment-female interaction remains stable and closely resembles the estimate from my main specification in Section 4 (reported in column 1 for the current subsample), as well as from a specification that includes additional control variables for non-cognitive factors (column 2).

The point estimate becomes insignificant and nearly halves in size only when I use both being social and under-confident in communication (column 6). Being both social and under-confident in communication is also indeed correlated with being female with a correlation coefficient of 0.22. This indicates that at least part of the female-treatment interaction is due to this combination of non-cognitive characteristics. In terms of inter-

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.137 (0.093)	0.134 (0.094)	0.244* (0.145)	0.116 (0.096)	0.244 (0.155)	0.260 (0.167)
Treated*Female	-0.191* (0.130)	-0.189* (0.131)	-0.212* (0.137)	-0.191* (0.131)	-0.229** (0.136)	-0.113 (0.134)
Treated*Competitive			-0.132 (0.128)		-0.158 (0.137)	
Tr.*Overconf. Math				0.240 (0.224)	-0.305 (0.240)	
Tr.*Comp.*Overconf. Math					0.669** (0.314)	
Treated*Social						-0.133 (0.165)
Treated*Underconf. Comm.						-0.065 (0.320)
Tr.*Social*Underconf. Comm.						-0.324 (0.372)
Prev. Controls + FE	✓	✓	✓	✓	✓	✓
Levels 4 Non-Cognitive Var.		✓	✓	✓	✓	✓
Lower Interactions					✓	✓
R-squared	0.135	0.144	0.147	0.147	0.160	0.167
N	280	280	280	280	280	280
Control Mean of Dep. Var.	0.509	0.509	0.509	0.509	0.509	0.509

Note: The dependent variable is if a participant has applied. “Levels 4 Non-Cognitive Var.” indicates the inclusion of the four levels of the non-cognitive variables—competitive, overconfident in math, social, and under-confident in communication—as control variables. “Lower interactions” refers to controlling for all pairwise combinations of the variables involved in the triple-interaction. Standard errors are clustered at the level of randomization. One-sided p-values are used for the treatment-gender interaction due to one-sided pre-registration. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Explorative Treatment Interactions

pretation, this is in line with the result that the changed peer behavior is an important driver of the effect (see Section 5.3). Girls who are social but under-confident in their communication skills show a pattern of being deterred by events where the social interactions are increasingly dominated by peers signaling their interest and competence. The result also aligns with the finding in Section 5.4 that under-confidence is a significant driver of the treatment effect.

5.6 Endline Results

The endline survey took place when nearly 80% of respondents had already signed an apprenticeship contract and was completed by 49% of participants, with no significant difference in attrition between the treatment and control groups. Most participants had applied to other banks, and treatment had no significant effect on this behavior or on their applications to the occupation encountered at the information event (see Table A7 in the appendix). Moreover, treatment had no effect on how participants perceived ZKB’s work environment relative to other banks (see column 1 in Table A8 in the appendix), or on whether they believed it was easier to secure an apprenticeship at another bank compared to ZKB (see column 2 in Table A8). These findings confirm the earlier results reported in Section 5.2 on treatment not affecting the perceived work environment or subjective probability of succeeding in an application.

6 Conclusion

Stressing selectivity does not significantly affect the total number of applications but triggers crucial heterogeneity in participant responses. In terms of diversity, stressing selectivity discourages female participants from applying and reduces the share of both female participants and children of migrants from lower-GDP countries in the applicant pool. In contrast, it motivates local male participants to apply (in line with Delfino, 2024) and to engage more actively during the information events. In terms of applicant quality, it deters—primarily female—high-quality applicants while attracting—primarily male—lower-quality applicants. These findings show that there is no quality-diversity tradeoff.

Regarding mechanisms, the social interaction effect of treated participants being exposed to other treated participants—who behave more competitively than control participants—is key. A back-of-the-envelope calculation reveals that this accounts for approx-

imately half of the treatment effect on the gender gap in application rates. By contrast, the treatment does not alter participants' perceptions of workplace characteristics or their subjective probability of receiving a job offer.

Absent of treatment, the partner firm in this study differs from privately owned firms by actively de-emphasizing selectivity. The treatment was designed to reverse this approach, simulating the more common, selectivity-focused status quo. A key recommendation based on my findings is for firms to consider de-emphasizing selectivity. If this is not feasible, firms should at least delay highlighting selectivity until after peer interactions have concluded. When neither option is possible, my results also support an alternative approach: female-only information events, as practiced by several firms in fields like IT. This recommendation reflects the results that it is male participants who behave differently at the events when selectivity is stressed but they are not discouraged by this competitive peer behavior.

References

- Abraham, L., Hallermeier, J., and Stein, A. (2024). Words matter: Experimental evidence from job applications. *Journal of Economic Behavior Organization*, 225:348–391.
- Athey, S., Tibshirani, J., and Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47:1148–1178.
- Brenøe, A. and Wasserman, M. (2024). Math skills, perceptions of fit, and occupational choice. *AEA RCT Registry*, May 31. <https://doi.org/10.1257/rct.10215-2.2>.
- Buser, T., Grimalda, G., Putterman, L., and van der Weele, J. (2020). Overconfidence and gender gaps in redistributive preferences: Cross-country experimental evidence. *Journal of Economic Behavior Organization*, 178:267–286.
- Coffman, K. B., Collis, M. R., and Kulkarni, L. (2024). Whether to apply. *Management Science*, 70:4649–4669.
- Connelly, B. L., Certo, S. T., Reutzell, C. R., DesJardine, M. R., and Zhou, Y. S. (2024). Signaling theory: State of the theory and its future. *Journal of Management*.
- Del Carpio, L. and Guadalupe, M. (2022). More women in tech? evidence from a field experiment addressing social identity. *Management Science*, 68:3196.
- Delfino, A. (2024). Breaking gender barriers: Experimental evidence on men in pink-collar jobs. *American Economic Review*, 114:1816–53.
- Dobbin, F. and Kalev, A. (2016). Why diversity programs fail. *Harvard Business Review*, Aug. 21.
- Duchene, P. (2024). What companies get wrong in their dei efforts. <https://www.forbes.com/sites/patriciaduchene/2024/03/08/what-companies-get-wrong-in-their-dei-efforts/>. Accessed: 2025-03-17.

- Flory, J., Leibbrandt, A., Rott, C., and Stoddard, O. (2021). Increasing workplace diversity. *Journal of Human Resources*, 56:73–92.
- Flory, J., Leibbrandt, A., Rott, C., and Stoddard, O. (2023). Leader signals and “growth mindset”: A natural field experiment in attracting minorities to high-profile positions. *Management Science*, 0.
- Flory, J. A., Leibbrandt, A., and List, J. A. (2015). Do competitive workplaces deter female workers? a large-scale natural field experiment on job entry decisions. *The Review of Economic Studies*, 82:122–155.
- Gee, L. K. (2019). The more you know: Information effects on job application rates in a large field experiment. *Management Science*, 65:1949–2443.
- Global Finance (2024). The world’s largest banks by assets, 2024. <https://gfmag.com/award/winner-announcements/press-releases-global-finances-worlds-safest-banks-2024/>. Press release, 2024-10-07.
- Golub Capital Social Impact Lab (2023). Machine Learning-based Causal Inference Tutorial. <https://bookdown.org/stanfordgsbsilab/ml-ci-tutorial/>. Stanford University, Accessed 2025-03-17.
- Jimenea, A., Wu, J., and Terris, H. (2024). The world’s largest banks by assets, 2024. <https://www.spglobal.com/marketintelligence/en/news-insights/research/the-worlds-largest-banks-by-assets-2024>. Accessed: 2025-03-17.
- Lazear, E. P., Shaw, K. L., and Stanton, C. T. (2018). Who gets hired? the importance of competition among applicants. *Journal of Labor Economics*, 36:S133–S181.
- Lüthi, S. and Wolter, S. C. (2023). Is being competitive always an advantage? competitiveness, gender, and labour market success. *Labour Economics*, 85:102457.

- McKinsey (2024). Find your future at mckinsey. <https://www.mckinsey.com/careers/students>. Accessed: 2024-05-29.
- Microsoft (2025). Aspire programme. <https://www.microsoft.com/en-ie/earlycareers/aspire-program>. Accessed: 2025-02-04.
- Miller, A. R., Petrie, R., and Segal, C. (2024). Effects of workplace competition on work time and gender inequality. *ILR Review*, 77:251–272.
- Mocanu, T. (2023). Designing gender equity: Evidence from hiring practices and committees. *mimeo*.
- Niederle, M., Segal, C., and Vesterlund, L. (2013). How costly is diversity? affirmative action in light of gender differences in competitiveness. *Management Science*, 59:1–16.
- Niederle, M. and Vesterlund, L. (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics*, 122:1067–1101.
- Samek, A. (2019). Gender differences in job entry decisions: A university-wide field experiment. *Management Science*, 65:3272–3281.
- Wager, S. and Athey, S. (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113:1228–1242.

Appendix

Table A1: Application Behavior: Probit Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.008 (0.134)	0.146 (0.168)	0.229 (0.175)	0.375* (0.211)	0.285 (0.197)	0.499** (0.229)
Treated*Female			-0.475** (0.245)	-0.505** (0.293)		-0.488** (0.293)
Tr*Par. born low-GDP C.					-0.441* (0.322)	-0.418* (0.325)
Female	-0.154 (0.123)	-0.126 (0.149)	0.120 (0.177)	0.176 (0.222)	-0.130 (0.149)	0.162 (0.223)
Parents born low-GDP C.					0.298 (0.246)	0.277 (0.245)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
N	511	352	511	352	352	352
Sum Tr. + Interaction(s)			-0.246* (.188)	-0.13 (.235)	-0.156 (.276)	-0.407* (.314)

Note: The dependent variable is if a participant has applied. Columns (2) and (4) replicate the specification in columns (1) and (3) respectively, but with the same sample that also contains information on migration background (as in columns 5 and 6). Standard errors are clustered at the level of randomization. One-sided p-values are used for the interactions in columns (3), (4), (5), and (6) due to one-sided pre-registration (including the sums of the interactions with the treatment coefficients). Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: “Spillover-Proof” Application Behavior

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.119 (0.086)	0.124 (0.088)	0.340*** (0.102)	0.298*** (0.110)	0.171* (0.096)	0.335*** (0.113)
Treated*Female			-0.515*** (0.143)	-0.422*** (0.158)		-0.407*** (0.159)
Tr*Par. born low-GDP C.					-0.246* (0.181)	-0.220 (0.179)
Female	-0.107 (0.078)	-0.075 (0.080)	0.234* (0.119)	0.201 (0.124)	-0.082 (0.082)	0.186 (0.127)
Parents born low-GDP C.					0.234 (0.167)	0.207 (0.166)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared	0.191	0.205	0.239	0.236	0.215	0.244
N	255	244	255	244	244	244
Control Mean of Dep. Var.	0.515	0.506	0.515	0.506	0.506	0.506
Sum Tr. + Interaction(s)			-0.174* (.116)	-0.124 (.124)	-0.075 (.168)	-0.292** (.174)

Note: The sample is reduced to those from school classes that either have only control or only treated participants, and p-weights are used to reflect the probability of remaining in the sample. The dependent variable is if a participant has applied. Columns (2) and (4) replicate the specification in columns (1) and (3) respectively, but with the same sample that also contains information on migration background (as in columns 5 and 6). Standard errors are clustered at the level of randomization. One-sided p-values are used for the interactions in (3), (4), (5), and (6) due to one-sided pre-registration (including the sums of the interactions with the treatment coefficients). Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A3: Offered Positions Depending on Treatment: Probit Estimation

	Everybody		Spillover-Proof		Cond. on Application	
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.201 (0.175)	0.425* (0.225)	0.422 (0.324)	1.267*** (0.427)	0.234 (0.211)	0.408 (0.273)
Treated*Female		-0.508* (0.322)		-1.974*** (0.579)		-0.395 (0.407)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared						
N	466	466	204	204	255	255
C. Mean of Dep. Var.	0.11	0.11	0.12	0.12	0.22	0.22

Note: The dependent variable is if a participant has received an job offer. For columns (1) to (4), this is the combination of participants applying and succeeding in their application. Columns (5) and (6) focus on the subsample of those participants that apply. Columns (3) and (4) use the reduced sample from school classes that either have only control or only treated participants, and p-weights are used to reflect the probability of remaining in the sample. Standard errors are clustered at the level of randomization. One-sided p-values are used for the interactions in Columns (2), (4), and (6) in line with Table 2. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* p < 0.10, ** p < 0.05, *** p < 0.01

Table A4: Noted as Desirable Applicant: Probit Estimation

	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.374** (0.181)	0.458** (0.218)	0.726*** (0.280)	0.891*** (0.313)	0.664** (0.276)	1.063*** (0.348)
Treated*Female			-0.692* (0.355)	-0.809** (0.404)		-0.775* (0.403)
Tr*Par. born low-GDP C.					-0.544 (0.430)	-0.515 (0.433)
Female	0.058 (0.164)	0.125 (0.178)	0.497* (0.287)	0.660* (0.338)	0.108 (0.179)	0.624* (0.338)
Parents born low-GDP C.					0.482 (0.345)	0.431 (0.348)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
N	485	331	485	331	331	331
Sum Tr. + Interaction(s)			.033 (.231)	.082 (.285)	.12 (.342)	-.227 (.38)

Note: The dependent variable is if participants have been noted by ZKB's staff to have behaved sufficiently interested and competent such that they encourage their application in a phone call after the event. Columns (2) and (4) replicate the specification in columns (1) and (3) respectively, but with the same sample that also contains information on migration background (as in columns 5 and 6). Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A5: Quality Measures of Successful vs. Unsuccessful Applicants

	Unsuccessful Appl.			Successful Appl.			Diff
	n	mean	sd	n	mean	sd	
High Math Grade	177	0.21	0.41	65	0.31	0.47	0.093
High German Grade	177	0.14	0.34	65	0.35	0.48	0.218***
High English Grade	177	0.18	0.38	65	0.42	0.50	0.240***
High French Grade	177	0.14	0.35	65	0.42	0.50	0.274***
Perfect Qualitat. Transscript	177	0.20	0.40	65	0.18	0.39	-0.019
Low School Type	232	0.02	0.13	78	0.00	0.00	-0.017

Note: A "high" grade refers to a 5.5 or 6 on a scale from 1 to 6. The qualitative part of the transcript is the teacher's assessment of the students' study habits, work ethic, and social behavior. Successful applicants are those that have received a job offer.

Table A6: Quality Comparison in the Pool of Applicants

	Control			Treatment			Diff
	n	mean	sd	n	mean	sd	
High Math Grade	101	0.30	0.46	143	0.20	0.40	-0.094*
High German Grade	101	0.20	0.40	143	0.19	0.39	-0.009
High English Grade	101	0.21	0.41	143	0.27	0.44	0.058
High French Grade	101	0.27	0.44	143	0.17	0.38	-0.093*
Perfect Qualitat. Transscript	101	0.21	0.41	143	0.20	0.40	-0.012

Note: A “high” grade refers to a 5.5 or 6 on a scale from 1 to 6. The qualitative part of the transcript is the teacher’s assessment of the students’ study habits, work ethic, and social behavior.

Table A7: Application Behavior beyond the Partner Firm

	Applied to Occ. Visited			Applied to other Banks		
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	-0.045 (0.053)	-0.065 (0.068)	-0.063 (0.073)	-0.011 (0.080)	-0.075 (0.105)	-0.137 (0.120)
Treated*Female		0.042 (0.103)	0.131 (0.108)		0.134 (0.133)	0.116 (0.134)
Tr*Par. born low-GDP C.			-0.175* (0.099)			0.133 (0.163)
Female	-0.006 (0.048)	-0.032 (0.084)	-0.062 (0.080)	0.080 (0.064)	-0.004 (0.110)	0.030 (0.102)
Parents born low-GDP C.			0.100 (0.068)			-0.083 (0.135)
Occupation FE	✓	✓	✓	✓	✓	✓
Location FE	✓	✓	✓	✓	✓	✓
Date FE	✓	✓	✓	✓	✓	✓
Individ. Controls	✓	✓	✓	✓	✓	✓
R-squared	0.226	0.226	0.319	0.180	0.184	0.223
N	243	243	188	235	235	184
Mean of Dependent Var.	0.774	0.774	0.787	0.613	0.613	0.652

Note: The dependent variable is if participants have applied (or are planning to apply) to an apprenticeship in the occupation they have seen at the information events (first three columns) or if they have applied (or are planning to apply) to other banks (last three columns). Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Ex Post Comparison to Other Banks

	Oth Banks less comp WE	Get Job at oth Banks Easier
	(1)	(2)
Treated	-0.052 (0.079)	-0.000 (0.082)
Female	-0.104 (0.065)	0.014 (0.077)
Occupation FE	✓	✓
Location FE	✓	✓
Date FE	✓	✓
Individ. Controls	✓	✓
R-squared	0.148	0.075
N	204	206
Control Mean of Dep. Var.	0.247	0.235

Note: The dependent variable is if participants perceive other banks to have a less competitive work environment than ZKB (column 1), or perceive it to be easier to receive a job offer from them (column 2). The sample is reduced to those who have informed themselves about other banks (nearly 90% of participants). Standard errors are clustered at the level of randomization. Individual controls include a dummy for filling in the baseline survey, age, and school type.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A1: CATE Terciles Characteristics (girls)

	N	58	74	18	
Age at Last Event		14.2 (0.0539)	14.3 (0.0511)	14.2 (0.0904)	
Both Parents Apprenticeship		0.276 (0.0592)	0.5 (0.0585)	0.389 (0.118)	***
Both Parents Degree		0.724 (0.0592)	0.0676 (0.0294)	0 (0)	*** *** **
Competitive		0.517 (0.0662)	0.595 (0.0575)	0.778 (0.101)	**
High English Grade		0.534 (0.0661)	0.392 (0.0571)	0.222 (0.101)	***
High French Grade		0.414 (0.0652)	0.27 (0.052)	0 (0)	* *** ***
High German Grade		0.379 (0.0643)	0.23 (0.0492)	0.111 (0.0762)	* ***
High Math Grade		0.397 (0.0648)	0.216 (0.0482)	0 (0)	** *** ***
Low School Track		0.0172 (0.0172)	0.108 (0.0363)	0 (0)	** ***
No Parent Apprenticeship		0.414 (0.0652)	0.257 (0.0511)	0.167 (0.0904)	* **
No Parent Degree		0.19 (0.0519)	0.581 (0.0577)	0.111 (0.0762)	*** ***
Occupation Commercial		0.845 (0.048)	0.743 (0.0511)	0.833 (0.0904)	
Occupation Digital Business		0.0172 (0.0172)	0.0541 (0.0265)	0 (0)	**
Occupation IT		0.0517 (0.0293)	0.149 (0.0416)	0 (0)	* * ***
Occupation Mediamatics		0.0862 (0.0372)	0.0541 (0.0265)	0.167 (0.0904)	
Overconfident Comm.		0.0517 (0.0293)	0.149 (0.0416)	0.111 (0.0762)	*
Overconfident Math		0.0345 (0.0242)	0.0676 (0.0294)	0.167 (0.0904)	
Parents Born Low-GDP C.		0.31 (0.0613)	0.378 (0.0568)	0.0556 (0.0556)	*** ***
Perfect Qualitat. Transcript		0.155 (0.048)	0.27 (0.052)	0.278 (0.109)	
Risk-Loving		0.534 (0.0661)	0.554 (0.0582)	0.444 (0.121)	
Shy		0.379 (0.0643)	0.392 (0.0571)	0.278 (0.109)	
Social		0.81 (0.0519)	0.757 (0.0502)	1 (0)	*** ***
Underconfident Comm.		0.345 (0.063)	0.135 (0.04)	0.111 (0.0762)	*** **
Underconfident Math		0.207 (0.0537)	0.162 (0.0431)	0 (0)	*** ***
		T1	T2	T3	1vs2 1vs3 2vs3

Note: Terciles according to participants' CATE (over participants of all genders). A "high" grade refers to a 5.5 or 6 on a scale from 1 to 6. The qualitative part of the transcript is the teacher's assessment of the students' study habits, work ethic, and social behavior.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Figure A2: CATE Terciles Characteristics (boys)

	N	62	42	102				
	Age at Last Event	14.1 (0.0602)	14.3 (0.0687)	14.2 (0.0454)				
	Both Parents Apprenticeship	0.323 (0.0599)	0.381 (0.0758)	0.637 (0.0478)		***	***	
	Both Parents Degree	0.952 (0.0275)	0.0714 (0.0402)	0 (0)	***	***	*	
	Competitive	0.806 (0.0506)	0.738 (0.0687)	0.804 (0.0395)				
	High English Grade	0.29 (0.0581)	0.238 (0.0665)	0.225 (0.0416)				
	High French Grade	0.21 (0.0521)	0.381 (0.0758)	0.108 (0.0309)	*	*	***	
	High German Grade	0.177 (0.0489)	0.238 (0.0665)	0.0882 (0.0282)			**	
	High Math Grade	0.355 (0.0613)	0.262 (0.0687)	0.147 (0.0352)		***		
	Low School Track	0.0161 (0.0161)	0.0476 (0.0333)	0.0196 (0.0138)				
	No Parent Apprenticeship	0.435 (0.0635)	0.19 (0.0613)	0.049 (0.0215)	***	***	**	
	No Parent Degree	0 (2.73e-17)	0.643 (0.0748)	0.441 (0.0494)	***	***	**	
	Occupation Commercial	0.613 (0.0624)	0.619 (0.0758)	0.696 (0.0458)				
	Occupation Digital Business	0.0968 (0.0379)	0.0476 (0.0333)	0.0392 (0.0193)				
	Occupation IT	0.29 (0.0581)	0.262 (0.0687)	0.216 (0.0409)				
	Occupation Mediamatics	0 (0)	0.0714 (0.0402)	0.049 (0.0215)	*	**		
	Overconfident Comm.	0.0968 (0.0379)	0.0952 (0.0458)	0.167 (0.0371)				
	Overconfident Math	0.0806 (0.0349)	0.119 (0.0506)	0.0392 (0.0193)				
	Parents Born Low-GDP C.	0.161 (0.0471)	0.643 (0.0748)	0.186 (0.0387)	***		***	
	Perfect Qualitat. Transcript	0.274 (0.0571)	0.286 (0.0706)	0.167 (0.0371)				
	Risk-Loving	0.565 (0.0635)	0.619 (0.0758)	0.52 (0.0497)				
	Shy	0.323 (0.0599)	0.357 (0.0748)	0.225 (0.0416)				
	Social	0.742 (0.056)	0.714 (0.0706)	0.775 (0.0416)				
	Underconfident Comm.	0.129 (0.0429)	0.19 (0.0613)	0.0294 (0.0168)		**	**	
	Underconfident Math	0.129 (0.0429)	0.0714 (0.0402)	0.0784 (0.0268)				
		T1	T2	T3		1vs2	1vs3	2vs3

Note: Terciles according to participants' CATE (over participants of all genders). A "high" grade refers to a 5.5 or 6 on a scale from 1 to 6. The qualitative part of the transcript is the teacher's assessment of the students' study habits, work ethic, and social behavior.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$