Self-selection of peers and performance*

Lukas Kiessling Jonas Radbruch Sebastian Schaube

June 22, 2021

Abstract

This paper studies how the presence of peers and different peer assignment rules—self-selection versus random assignment—affect individual performance. Using a framed field experiment, we find that the presence of a randomly assigned peer improves performance by 28% of a standard deviation (SD), while self-selecting peers induces an additional 15-18% SD improvement in performance. Our results document peer effects in multiple characteristics and show that self-selection changes these characteristics. However, a decomposition reveals that variations in the peer composition contribute only little to the performance differences across peer assignment rules. Rather, we find that self-selection has a direct effect on performance.

Keywords: Peer effects, Self-selection, Autonomy, Framed field experiment
JEL-Codes: C93, D01, I20, J24, L23

*Lukas Kiessling: Max Planck Institute for Research on Collective Goods, lukas.kiessling@gmail.com; Jonas Radbruch: Institute of Labor Economics (IZA), radbruch@iza.org; Sebastian Schaube: University of Bonn, sebastian.schaube@uni-bonn.de. We thank Viola Ackfeld, Philipp Albert, Thomas Dohmen, Lorenz Goette, Ingo Isphording, Sebastian Kube, Pia Pinger, Matthias Sutter, Ulf Zölitz, and numerous seminar and conference audiences for helpful feedback and comments. We also thank the schools and students that participated in the experiments. This research was undertaken while all authors were at the University of Bonn. We did not obtain an IRB approval for this project because at the time of the experiment there was no IRB at the University of Bonn's Department of Economics. However, we would like to stress that the schools' headmasters approved the study, written parental consent was required for students to take part in the study, and participation was voluntary. Moreover, the experiment is in line with the requirements of the BonnEconLab. Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (projects A01 and A02) is gratefully acknowledged. This study is registered in the AEA RCT Registry as AEARCTR-0005562.

1 Introduction

In various dimensions of life—ranging from cashiers in supermarkets (Mas and Moretti, 2009) and fruit pickers on strawberry fields (Bandiera, Barankay, and Rasul, 2009, 2010) to fighter pilots during World War II (Ager et al., 2019) and students in educational settings (Sacerdote, 2001)—people affect each other through their presence, performance, and choices. Yet, these social influences often stem from specific persons—frequently interacting coworkers, friends, or (former) colleagues—whom individuals select themselves. This is in stark contrast to settings in which peers are exogenously assigned by, e.g., supervisors. But what actually changes once we allow peers to be self-selected? In general, these settings differ in two aspects: first, self-selection changes with whom one interacts; and second, the opportunity to self-select peers fundamentally changes the mode of peer assignment. The change from exogenous (or random) assignment to self-selection can, thus, change how the situation is perceived by individuals. Both of these channels potentially alter an individual's motivation and performance. Thus, leveraging any positive effect of peer self-selection might complement a firm's tools to promote performance on the job similar to other non-monetary incentives (e.g., Cassar and Meier, 2018).

In this paper, we study how the presence of peers and different peer assignment rules—self-selection versus random assignment—affect individual performance. Studying this question requires objective performance measures and a setting in which we can precisely isolate the role of specific peers. In addition, it requires an environment in which we can vary the presence of peers and implement exogenous as well as endogenous selection of peers based on, e.g., pre-existing social ties. Meeting all of these requirements jointly is difficult to achieve within firms or organizations. We therefore implemented a controlled framed field experiment in a setting with existing social ties and personal knowledge amongst subjects.

We conducted a framed field experiment with over 700 subjects in physical education classes at secondary schools. Subjects participated in a real-effort task (running task) twice and filled out a survey in between that elicited preferences for peers, personal characteristics, and the social network within each class. In a control treatment, subjects ran alone in both runs (NoPeer). Three additional treatments introduced peers in the second run according to different peer assignment rules. We implement a random matching of pairs (Random), as well as two matching rules that apply two notions of self-selection: first, the classroom environment enabled subjects to select known peers based on their names (Name); and second, using a running task yielded direct measures of performance and thus could be used to select peers based on their relative performance in the first run (Performance). This setup allows us to document differences in performance between treatments which vary the presence of peers and the peer assignment rule. Hence, we can cleanly identify the overall effect of the presence of peers as well as the effect of self-selection on performance. Subsequently, we analyze the underlying mechanisms of performance differences across different peer assignment rules and decompose them into their two possible sources: an indirect effect stemming from interacting with different peers and a direct effect from being able to self-select rather than being assigned to a specific peer.

We find that the presence of an exogenously and randomly assigned peer improves performance by about 28% of a standard deviation (SD). Comparing the three different peer assignment rules shows that self-selection of peers yields additional performance improvements relative to random assignment. In treatments with self-selection, performance improves by an additional 15% SD in Name and 18% SD in Performance. In a second step, we decompose the treatment effects into the indirect effects and the direct effects of peer self-selection. Although self-selection changes with whom a subject interacts and peer effects exist in multiple dimensions (e.g., the relative performance in the first run matters for performance in the second run), these changes in the peer composition do not help to explain the differences in performance across treatments. Correspondingly, we estimate the indirect effects to be close to zero. Instead, our estimates provide evidence for a sizable direct effect of peer self-selection on performance. Borrowing from self-determination theory (Deci and Ryan, 1985, 2000), we interpret this direct effect as a positive effect of having autonomy: the opportunity to self-select peers has a psychological effect that enhances intrinsic motivation and improves performance. In a final step, we ask whether a principal can use other exogenous peer assignment rules to achieve similar performance improvements. To do so, we simulate the results of additional peer assignment rules that aim at increasing aggregate performance. We document that these alternative policies yield performance improvements close to those observed with randomly assigned peers and consistently lower than those allowing for peer self-selection. These findings, thus, support our interpretation that self-selection of peers may have an intrinsic value beyond changing the peer composition.

This paper contributes to the literature on peer effects, social interactions, and autonomy by showing that self-selection of peers can directly affect behavioral outcomes and performance. We thereby provide first experimental evidence on autonomy in a field setting, and highlight a novel channel through which the selection of peers can affect behavior.

The mere existence of peer effects in general already raises the question whether firms are able to exploit these effects strategically. Theoretical considerations suggest that the (exogenous) assignment of peers can be leveraged to improve aggregate performance (e.g., Kräkel, 2016; Roels and Su, 2014). However, the empirical evidence on interventions that change group compositions based on the ability distribution remains mixed (e.g., Carrell, Sacerdote, and West, 2013; Duflo, Dupas, and Kremer, 2011; Garlick, 2018). The results in this paper suggest that the presence of peers in general and, in particular, changing the mode of peer assignment to self-selection might be leveraged to increase overall performance.

In particular, our results show that performance increases if individuals can self-select with whom they interact. Therefore, our findings add to the existing research that analyzes the effects of autonomy and decision rights on behavioral outcomes by providing novel field evidence that self-selection can have a direct effect that increases performance beyond changing peer characteristics. Thus, we complement laboratory studies by Bartling, Fehr, and Herz (2014) and Owens, Grossman, and Fackler (2014), which demonstrate that people are willing to pay for autonomy, i.e., the opportunity to select relevant aspects of their decision environment actively (Deci and Ryan, 1985). Similarly, autonomy in the workplace is associated with higher wages and employee happiness (Bartling, Fehr, and Schmidt, 2013) and leads to increased labor supply (Chevalier et al., 2019), while removing autonomy has been found to have negative consequences on employee effort (Falk and Kosfeld, 2006). Our results highlight an additional channel through which autonomy might provide value to employers or policy-makers: the freedom to choose one's own peers or teammates can boost performance, similar to other non-monetary incentives such as recognitions and awards (Bradler et al., 2016; Kosfeld and Neckermann, 2011), framing of rewards (Levitt et al., 2016), or personal goals (Corgnet, Gómez-Miñambres, and Hernán-González, 2015; Koch and Nafziger, 2011).

Our results also inform the design, usage, and evaluation of peer assignment rules. The data suggest that if managers assign co-workers or groups based on peer effects in a single dimension only (e.g., past performance), they neglect the fact that such assignments simultaneously change other peer characteristics. This may lead to peer effects in dimensions apart from the targeted one, potentially counteracting a manager's objective. In fact, we allow a comprehensive set of peer characteristics such as productivity, friendship ties, and personality measures to exert peer effects.[1] In our setting, we observe that these peer effects counterbalance each other, leading to

[1]We therefore also join a small set of studies explicitly considering the impact of personality traits on performance or educational outcomes (e.g., Chan and Lam, 2015; Golsteyn, Non, and Zölitz, 2021; Záraté, 2020) or analyzing the role of friendship ties for peer effects (e.g., Park, 2019). Yet, these

a net effect that is close to zero and is in general ambiguous. Hence, when designing policies aimed at exploiting peer effects in organizations, managers need to take into account this multidimensionality of peer effects as well as a potential direct effect of self-selection.

More generally, the experimental design in this paper allows us to document a strong causal difference in performance between widely-used randomly assigned peer groups and self-selected peers.[2] The literature's focus on random peer assignment is understandable given that researchers aim to identify a clean causal effect of being exposed to certain peers or certain characteristics of peers. However, similar to what has been found in previous studies exploring the selection of students into peer groups (e.g., Cicala, Fryer, and Spenkuch, 2018; Tincani, 2017), our results indicate that the relevant *and* self-selected peer within a group does not correspond to a random peer. Related to our paper, Chen and Gong (2018) study self-selection of team members as peers in a setting where skill complementarities as well as peer pressure are relevant for the production; and they document, in line with our findings, that endogenously formed teams outperform randomly assigned ones. They focus on teams consisting of four, which interact over several weeks and prepare a presentation for an undergraduate business class. Our approach differs from and advances their approach in at least two aspects. First, we focus on a setting with a single peer, a single interaction, and individual incentives, rather than teams interacting over a longer time span. Thus, we can isolate the source of peer effects. Second, we lever a rich set of characteristics (past performance, friendship ties, personality traits) and, in particular, impose structure on the self-selection process, which allows us to measure preferences for peers—a normally unobserved dimension. Specifically, this allows us to gain experimental control and to condition on the quality of the match even under random assignment of peers. Hence, we can move beyond treatment comparisons and can perform a detailed decomposition of treatment effects.

Methodologically, we build on two well-established experimental paradigms. First, we study self-selection of peers using an established real-effort task (see, e.g., Belot and van de Ven, 2011; Fershtman and Gneezy, 2011; Gneezy and Rustichini, 2004; Rao, 2019; Sutter and Glätzle-Rützler, 2015, for other studies employing similar running tasks to study a variety of phenomena). Second, we study the consequences of imposed or self-selected environments, in which subjects who self-selected an envi-

studies do not consider the implications of multidimensional peer effects and focus on a smaller set of peer characteristics.

[2]The literature on peer effects builds on (conditional) random assignment to identify peer effects and circumvent statistical issues outlined in Manski (1993). See also Villeval (2020) and Herbst and Mas (2015) for literature reviews on peer effects in the workplace and laboratory, and a comparison of peer effects from field and lab settings, respectively.

ronment are compared to others who were exogenously assigned, but who would have chosen it anyway. Previous laboratory studies show that having the opportunity to decide on leaders or vote for institutions positively affects the quality of leadership (e.g., Brandts, Cooper, and Weber, 2014), as well as the effectiveness of institutions in the presence of social dilemmas (e.g., Bó, Foster, and Putterman, 2010; Sutter, Haigner, and Kocher, 2010). We translate this idea to our field setting and elicit revealed preferences for peers. This allows us to compare the impact of a self-selected peer to a randomly assigned one, while controlling both for the peer's characteristics and—crucially—for an individual's preference for a particular peer, an usually unobserved dimension.

While our results document a direct effect of self-selection on performance in a particular stylized setting that mimicks important features of many workplace interactions, we do not claim that the effect will quantitatively or qualitatively carry over to all settings. Rather, we view our results as a proof-of-concept that the opportunity to self-select peers can affect performance. This process of self-selecting peers, however, should also be important for settings in which peer effects do not arise due to social comparisons or peer pressure, but from effort or skill complementarities (e.g., Bandiera, Barankay, and Rasul, 2010; Mas and Moretti, 2009), or settings in which peers learn from each other (e.g., Bursztyn et al., 2014; Jackson and Bruegmann, 2009). The settings across these studies differ enormously, as does the underlying mechanism which leads to the observed peer effects. Nonetheless, all of these share the notion that the behavior or action of peers imposes an externality on the action or behavior of others and that peers can in principle also be self-selected, affecting subsequent peer interactions.

The remainder of the paper is structured as follows. The next section presents our experimental design. Section 3 describes the data and we outline our empirical framework in Section 4. In Section 5, we present the empirical results and discuss the limits and consequences of peer assignment rules. Finally, Section 6 concludes.

2 Experimental design

The aim of this paper is to study how the presence of peers and the mode of peer assignment—most importantly self-selection of peers—impacts performance. This requires an environment in which subjects can choose peers themselves, but where we can also assign peers exogenously. Subjects must be able to compare their own performance with that of a peer in a task that lends itself to natural up- and downward comparisons. One complication in many settings is that it is difficult to isolate the person who serves as the relevant point of comparison. This is especially true if several potential peers are present at all times, while only some constitute relevant peers. As subjects might select those peers for reasons besides their (relative) performance, it is essential to measure additional characteristics of subjects and to collect data on existing social groups. In such groups, subjects have a clear impression of other group members and are able to select peers based on characteristics such as their social ties.

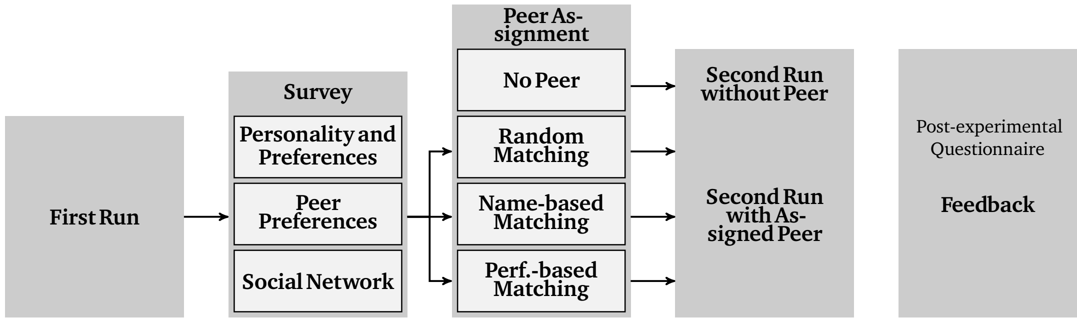
In this study, we used the controlled environment of a framed field experiment to overcome these challenges. We embedded our experiment in physical education classes of German secondary schools. We built on a well-established real-effort task from the economic and management literature—running exercises in schools (e.g., Belot and van de Ven, 2011; Fershtman and Gneezy, 2011; Gneezy and Rustichini, 2004; Rao, 2019; Sutter and Glätzle-Rützler, 2015). Specifically, students from grades 7 to 10 participated in two running tasks, first alone and then, depending on the treatment, either alone or simultaneously with one peer. By focusing on pairs of subjects, we circumvent the issue of multiple potential peers being present, which allows us to identify the impact of this particular peer and his or her characteristics.[3]

We deliberately chose this particular task to narrow the channels through which subjects can affect each other. In particular, subjects can easily compare their performance with faster and slower subjects, inducing social comparisons in performance and peer encouragement. The mere presence of a peer and his or her simultaneous performance can therefore increase the performance of an individual as, for example, in Falk and Ichino (2006); a phenomenon the social psychology literature commonly refers to as a social facilitation effect.[4] Our task reduces the role of other mechanisms such as learning or skill complementarities between subjects to a minimum. The task is sufficiently short to avoid learning a better technique from peers during the second run. Moreover, we designed the experiment with a focus on individual and low-stake incentives by reporting the individual times to the teachers for their class evaluation. Thus, we minimize the role of skill complementarities in the experiment.

[3]The task was chosen for several reasons: (1) the task is not a typical part of the German physical education curriculum, yet it is easily understandable for the students; (2) in contrast to a pure and very familiar sprint exercise, students should only have a vague idea of their classmates' performance and cannot precisely target specific individuals based on their performance; and (3) due to the different aspects of the task (general speed, quickness in turning, as well as some level of endurance or perseverance), the performance across age groups was not expected to (and did not) change dramatically.

[4]Several studies in psychology suggest that the mere presence of others can increase performance in simple tasks (e.g., Allport, 1920; Hunt and Hillery, 1973; Towler, 1986; Zajonc, 1965).

Figure 1: Experimental design



Notes: This figure illustrates the experimental design. Treatments or peer assignment rules (NoPeer, Random, Name, Performance) are randomly assigned at the classroom level.

2.1 Experimental design

Figure 1 illustrates the experimental design. Subjects participated in running tasks commonly known as "suicide runs", a series of short sprints to different lines of a volleyball court (cf. Appendix A). The first run, in which subjects ran alone, served two purposes: first, recorded times can be used as a measure of productivity and to evaluate the time improvement between the two runs; and second, they allow subjects in one of the treatments to select their peers based on relative times from the first run. The second run mirrored the first one, but subjects were assigned to a peer in three of the four treatments, in which two subjects performed the task simultaneously on neighboring tracks, while their times were recorded individually. We provided feedback about performance in both runs only at the end of the experiment.

Between the two runs, subjects completed a survey comprising three parts, eliciting preferences for peers, non-cognitive skills, and the social network within each class. First, we elicited two kinds of preferences for peers by initially asking subjects to state the names of those classmates with whom they would like to perform the second run; we then asked them to state the relative performance level of their most-preferred peers. Second, and in addition to these preferences, the survey included sociodemographic questions and measures of personality and economic preferences: the Big Five inventory as used in the youth questionnaire of the German socio-economic panel (M. Weinhardt and Schupp, 2011), a measure of locus of control (Rotter, 1966), competitiveness[5], general risk attitude (Dohmen et al., 2011), and a short version of

[5]We implemented a continuous survey measure of competitiveness using a four-item scale. For this, we asked subjects about their agreement to the following four statements on a seven-point Likert scale: (i) "I am a person who likes to compete with others", (ii) "I am a person who gets motivated through competition", (iii) "I am a person who performs better when competing with somebody", and (iv) "I

the INCOM scale for social comparison (Gibbons and Buunk, 1999; Schneider and Schupp, 2011). Finally, the survey concluded by eliciting the social network within every class. Subjects were asked to state up to six of their closest friends within the class.

Before and after the second run, we asked subjects a short set of questions about their peer and their experience during the task. Before the run, we elicited their belief about the relative performance of their peer in the first run, namely who they thought was faster. Following the second run, we asked them whether they would rather run alone or in pairs the next time, how much fun they had had, as well as how pressured they felt in the second run, on a five-point Likert scale, due to their peer running with them.

2.2 Preference elicitation

We elicited the two sets of peer preferences using the strategy method, i.e., independently of the treatment to which a subject is eventually assigned. The first set elicited preferences for situations in which social information is available (*name-based preferences*). Accordingly, we asked each student to state his or her six most-preferred peers from the same gender within their class, i.e., those people with whom they would like to be paired in the second run. Appendix Figure A.2 presents a screenshot of the elicitation screen. Subjects could select any person of the same gender, irrespectively of this person's actual participation in the study or their attendance in class.[6] These classmates had to be ranked, creating a partial ranking of their potential peers.

This first set of preferences allows us to study a natural form of peer selection based on names. Yet, in other situations, we often have only limited information about our potential peers, e.g., on leaderboards of sales teams and selecting into certain schools or workplaces. To mimic these settings, we elicited a second set of preferences solely based on the relative performance in the first run, ignoring the identities of the potential running partners (*performance-based preferences*). For this purpose, we presented subjects with ten categories comprising one-second intervals starting from $(4, 5]$ seconds slower than their own performance in the first run, to $(0, 1]$ seconds slower and $(0, 1]$ seconds faster up to $(4, 5]$ seconds faster (corresponding to a range of approx. ± 2 SD from their own performance). Appendix Figure A.2 presents a screenshot of the elicitation screen. Subjects had to indicate from which time interval

am a person who feels uncomfortable in competitive situations". We then extracted a single principal component factor from those four items, of which the fourth item was scaled reversely.

[6]All subjects were informed that peers in the second run would always have the same gender as themselves and would also need to participate in the study.

they would prefer a peer for the second run, irrespectively of the potential peer's identity. Similar to the name-based preferences, we elicited a partial ranking for those performance-based preferences. Accordingly, subjects had to indicate their most-preferred relative time interval, second-most-preferred relative time interval, and so on.[7]

2.3 Treatments

Across treatments, we varied if and how pairs in the second run were formed. Specifically, we implemented either one of three peer assignment rules or a NoPeer condition in which subjects ran alone twice. The first peer assignment rule matched subjects randomly—i.e., we employed a random matching (Random)—and serves as a natural baseline treatment. The second rule used the elicited name-based preferences (Name) to form pairs, whereas the third rule employed performance-based preferences (Performance). Note that the problem of matching pairs based on their preferences constitutes a typical roommate problem. We thus implemented a "stable roommate" algorithm proposed by Irving (1985) to form stable pairs using the elicited preferences.[8]

In all treatments with self-selected peers, subjects did not know the specific matching algorithm, but were told that their preferences would be taken into account when forming pairs. Furthermore, we highlighted that the mechanism was incentive-compatible by telling subjects that it was in their best interest to reveal their true preferences. We informed subjects about the existence of all three matching rules in the survey to elicit both sets of preferences, irrespectively of the implemented treatment. Just before the second run took place, they were informed about the specific matching rule employed in their class and the resulting pairs.

In the additional NoPeer treatment, subjects ran alone twice. Subjects were told this in advance to avoid deception. Moreover, subjects in this treatment only participated in a shortened survey. The survey only asked subjects for their preferences

[7]Naturally, each time interval could only be chosen once in the preference elicitation, although each interval could potentially include several peers (e.g., if several subjects had similar times and thus belonged to the same interval). Similarly, some intervals may not contain any peers if no subject in the class had a corresponding time.

[8]Given the mechanism proposed by Irving (1985), it is a (weakly) dominant strategy for all participants to reveal their true preferences. The matching algorithm requires a full ranking of all potential peers to implement a matching. Since we only elicited a partial ranking, we randomly filled the preferences for each student to generate a full ranking. However, in most cases, subjects were assigned a peer according to one of their first three preferences. Nonetheless, if groups were small, it could be the case that subjects were not assigned one of their most-preferred peers. This is especially the case for performance-based preferences. See also the discussion of our manipulation checks in Section 3 below.

for peers, their sociodemographics and their social network.[9] This treatment allows to identify the impact of the presence of peers in the second run.

2.4 Procedures

We conducted the experiment in physical education lessons at three secondary schools in Germany. All students from grades 7 to 10 (corresponding to the ages 12 to 16) of those schools were invited to participate in the experiment. Approximately two weeks prior to the experiment, teachers distributed parental consent forms. These forms contained a brief, very general description of the experiment. Only those students who had handed in the parental consent form before the study took place participated in the study.

The experiment started with a short explanation of the following lesson and a demonstration of the experimental task. A translation of this explanation as well as screenshots detailing the preference elicitation are presented in Appendix A.

The students themselves did not receive any information on their performance until the completion of the experiment. We stressed that both running times would be graded by their teacher—thus incentivizing both runs—and that the objective was to run as fast as possible in both runs.[10] Teachers received students' times from both runs after the experiment for grading, but no information about the pairings during the second run. In addition to these formal incentives, most students themselves were very interested in their own times. The introduction concluded with a short warm-up period. After this, the subjects were led to a location outside the gym.

Students entered the gym individually, which ruled out any potential audience effects from classmates being present. They completed the first suicide run and subsequently were handed a laptop to answer the survey. Answering the survey took place in a separate room and in the presence of an experimenter. After the completion of the survey, subjects waited outside the gym. Upon completion of the survey by all students, they returned to the gym to receive further instructions for the second run. In particular, for treatments featuring peers in the second run, we reminded them of the existence of the three matching rules. We then announced which randomly assigned rule was implemented in their class, as well as informing them of the resulting pairs from the matching process. Following these instructions, the entire group waited

[9]This treatment was only conducted at one school, where some (random) classes did not have enough time for the full survey due to scheduling issues.

[10]In order for the teacher to grade the entire set of students, the students who did not participate in the study also had to run twice. Their times were recorded for the teacher only and were never stored by us.

outside the gym again. Pairs were called into the gym and both students participated in the second run simultaneously on neighboring tracks.

After all pairs had finished their second suicide run, the experiment concluded with a short statement by the experimenters, thanking the students for their participation. The teacher received a list of the students' times in both runs and students were informed about their performance. We then asked the teacher to evaluate the general atmosphere within the class.[11]

3 Data description and manipulation check

In total, 48 classes from three schools with an average class size of about 25 students participated in the experiment. On average, 68.2% of the students within each class consented to participate and 62.8% of the students subsequently took part in the experiment.[12] This amounts to 754 participating students. Due to odd numbers of students within some matching groups in treatments with peers, we randomly dropped one student in those groups to match subjects into pairs. Therefore, some subjects participated in the experiment, but were only recorded once and are dropped for estimating the treatment effects in the next section. Our resulting estimation sample comprises 715 subjects (588 for treatments with peers).

Summary statistics. We present summary statistics of our sample in Table 1. On average, subjects are 14.4 years old and 65% are female, since one school in our sample—the smallest one—was a female-only school. On average, female subjects took 27.51 seconds in the first run. Their performance is quite stable across grades, with subjects from the seventh grade being somewhat slower. Male subjects' times improved with age: while male subjects took on average 25.29 seconds in the first run in grade 7, they ran about two seconds faster in grade 10.

Preferences for peers. Most subjects nominated friends as their most-preferred peers (89%), and subjects preferred on average to run with a slightly faster peer

[11]Teachers indicated their agreement with three statements on a seven-point Likert scale: (1) "The class atmosphere is very good", (2) "Some students get excluded from the group", and (3) "Students stick together when it really matters".

[12]We aimed to recruit all students from a class. However, due to numerous reasons (e.g., absences, sickness, injuries, or missing consent) this was not possible in every class. We do not have concerns of non-random selection into the study, since students did not know in advance the exact day when the experiment was scheduled and most reasons for non-participation were rather exogenous (like injuries or sickness). Moreover, treatment randomization was at the class level within schools and therefore selection into treatments was ruled out by design.

Table 1: Summary statistics

	7th grade	8th grade	9th grade	10th grade	Total
Sociodemographic Variables					
Age	12.76	13.76	14.77	15.84	14.40
	(0.44)	(0.45)	(0.39)	(0.53)	(1.24)
Female	0.59	0.62	0.67	0.70	0.65
	(0.49)	(0.49)	(0.47)	(0.46)	(0.48)
Times (in sec)					
Time 1 (Females)	27.71	27.13	27.34	27.80	27.51
	(2.65)	(1.98)	(2.25)	(2.72)	(2.44)
Time 2 (Females)	26.97	26.83	26.58	26.96	26.83
	(1.90)	(1.85)	(2.42)	(2.40)	(2.20)
Time 1 (Males)	25.29	24.59	23.71	23.12	24.19
	(2.02)	(2.46)	(2.00)	(2.08)	(2.30)
Time 2 (Males)	24.94	24.13	22.93	22.33	23.63
	(2.23)	(2.39)	(1.76)	(1.42)	(2.23)
Observations	165	180	192	217	754

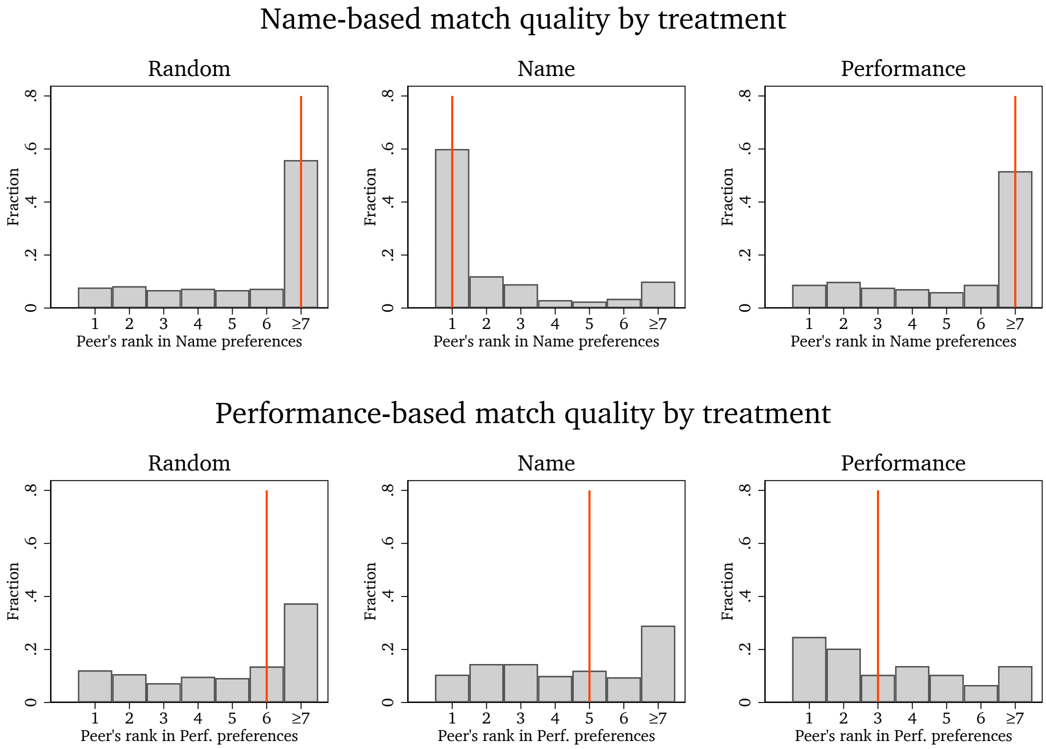
Notes: Standard deviations are presented in parentheses. Note that some subjects only participated in the survey in cases in which they were allowed to participate in the study, but were unable to take part in the regular physical education lesson, while some others only took part in the first run if there was an odd number of subjects in the matching group. See the text for details.

(about 30%), but there is a strong heterogeneity in this preference. We summarize the preferences for peers in Appendix B and analyze these preferences in further detail in Kiessling, Radbruch, and Schaube (2020).

Randomization checks. We randomized classes into treatments within schools and grades. In Appendix Tables B.1 and B.2, we check whether observable characteristics and preferences differ between our treatments. Overall, randomization was successful. Small (but always insignificant) pre-treatment difference can be entirely explained by our block randomization. Once we condition on the randomization strata (grade-by-school fixed effects), the remaining differences disappear. Additionally, Panel A of Appendix Table B.3 provides evidence that there is no correlation between a peer's time in the first run and a subject's time in RANDOM, as it should be the case under random assignment. Panel B summarizes a series of further randomization checks, examining correlations between a subject's and her peer's characteristics, and providing further evidence that the random assignment in RANDOM was successful.

Manipulation check. Figure 2 provides evidence that our peer assignment indeed changed the actual match quality, which we define as the rank of the assigned peer in the elicited preference rankings. The upper panel shows the realized match qual-

Figure 2: Match quality across treatments



Notes: The figure presents a histogram of match qualities for each treatment measured by the rank of the realized peer in an individual's name-based (upper panel) or performance-based preferences (lower panels). Vertical red lines denote median ranks.

ity according to name-based preferences for each of the three treatments featuring peers. While the median peer in Name corresponds to the most-preferred peer according to the elicited name-based preferences, the median peer is not part of the elicited preferences (i.e., not among the six most-preferred peers) for Random and Performance. A similar, albeit less pronounced, picture arises when analyzing the match quality according to the preferences over relative performance in the lower panel of Figure 2. We observe that subjects in Performance were paired with more preferred peers according to their preferences relative to the other two treatments.[13] Appendix Table E.3 provides evidence that those changes in the match quality across treatments are significant. Hence, we observe that subjects in treatments with self-selected peers had a higher probability of being matched with someone whom they

[13] Subjects might have preferred other subjects or relative times that were not available to them, which mechanically affects the match quality. In Appendix B, we check that once we take this mechanical effect into account, the median match quality in Performance corresponds to the second-most-preferred peer, i.e., we obtain a similarly pronounced pattern as in Name.

preferred more, i.e., who ranked higher in their name-based or performance-based preferences, providing evidence that our manipulation was indeed successful.

4 Empirical strategy

In a first step, we lever the random assignment of classes to treatments to study how the presence of peers and how changing the mode from random assignment to self-selection of peers affects performance. In what follows, we take random assignment of peers (Random) as the baseline. Let $D^d = 1$ with $d \in \{Name, Perf, NoPeer\}$ denote treatment assignment to Name, Performance, and NoPeer, respectively, and zero otherwise. In our main specification, we focus on percentage-point improvements from the first to the second run as the outcome of interest, but also consider other specifications with time in the second run. Let the outcome of individual i in class c, grade g of school s be denoted as y_{icgs}. We estimate the following specification:

$$(1) \qquad y_{icgs} = \tau^{Name} D_{ic}^{Name} + \tau^{Perf} D_{ic}^{Perf} + \tau^{NoPeer} D_{ic}^{NoPeer} + \gamma X_i + \rho_{sg} + u_{icgs}$$

The main parameters of interest are τ^{Name}, τ^{Perf}, and τ^{NoPeer}, the effect of being assigned to one of our treatments relative to Random. Grade-by-school fixed effects, ρ_{sg}, control for variation due to different schools and grades (i.e., as a result of different locations and timing of the experiment) and correspond to our randomization strata. The vector X_i captures predetermined characteristics including gender-specific age trends, personality characteristics, and—in some specifications—class-level control variables, and u_{icgs} is a mean zero error term clustered at the class-level. Note that we do not observe personality characteristics in the treatment NoPeer, as this treatment only featured a shorter questionnaire (see Section 2 for details). In specifications using times in the second run as an outcome, we additionally control for performance in the first run.

In a second step, we decompose this total treatment effect, i.e., differences in performance across those treatments featuring peers, into their two potential sources: first, different peer assignment mechanisms may affect peer interactions directly (*direct effect of self-selection*); and second, self-selection may change the peer composition and therefore the difference between the student's and his or her peer's characteristics, potentially affecting performance through peer effects (*indirect effect*).[14] We

[14]The direct effect mainly captures changes in performance due to being able to self-select a peer, which we interpret as an increase in autonomy (see Section 5.2.3 for a discussion of the psychological underpinnings). We acknowledge that our definition of a direct effect in principle also captures inputs that (i) differ across treatments, and (ii) but are not measured in our rich set of potential mediators

implement this decomposition using the following specification, which we derive in more detail in Appendix C:

$$(2) \qquad y_{icgs} = \underbrace{\bar{\tau}^{Name} D_{ic}^{Name} + \bar{\tau}^{Perf} D_{ic}^{Perf}}_{\substack{\text{Treatments} \\ \text{(Direct effects)}}} + \underbrace{\beta\theta_i\left(D_{ic}^{Name}, D_{ic}^{Perf}\right)}_{\substack{\text{Peer} \\ \text{characteristics}}} + \underbrace{\delta X_i + \lambda_{sg}}_{\substack{\text{Ind. characteristics} \\ \text{and fixed effects}}} + u_{icgs}$$

We are interested in $\bar{\tau}^{Name}$ and $\bar{\tau}^{Perf}$, the direct effects of our treatments relative to Random. While the treatment indicators capture the direct effects, the indirect effects can be recovered by multiplying changes in the peer composition with their corresponding coefficients. Changes in peer characteristics through our treatments are captured by changes in the vector $\theta_i\left(D^{Name}, D^{Perf}\right)$, and β denotes the influence of these peer characteristics on the outcome. We allow our effects to be mediated in various ways: a first set of mediators captures the quality of the match measured by the rank of the peer in an individual's preferences[15], productivity differences measured by absolute differences of times in the first run, and (directed) friendship ties.

Second, we allow within-pair ranks to mediate the effects based on subjects' times in the first run. Importantly, we interact the previous channels with the rank indicators to allow them to differ depending on this within-pair ranking (e.g., to allow initially faster subjects to slow down and slower subjects to improve their performance with increasing absolute differences in times of the first run).[16] As a third set of mediators, we introduce variables capturing the peer's personality and preference measures (i.e., Big Five, locus of control, competitiveness, risk attitudes, social comparison). Additionally, we also include the absolute differences in these personality measures to capture potential non-linear effects.

(match quality, friendship ties, productivity differences, ranks and personality differences). However, we show in robustness checks that in our setting this is of minor concern only (cf. Appendix E).

[15]In our main specification, we define two indicators to measure whether the assigned peer is nominated among the first three peers for name-based preferences or falls into the three highest ranked categories for performance-based preferences. We relax this definition in robustness checks.

[16]Previous research highlighted the importance of ranks among peers for a range of economic outcomes (e.g., Cicala, Fryer, and Spenkuch, 2018; Elsner and Isphording, 2017; Gill et al., 2019; Kiessling and Norris, 2020; Murphy and F. Weinhardt, 2020). In contrast to the previous literature, we allow peer characteristics to interact with the rank in the pair rather than focusing on the rank indicator only.

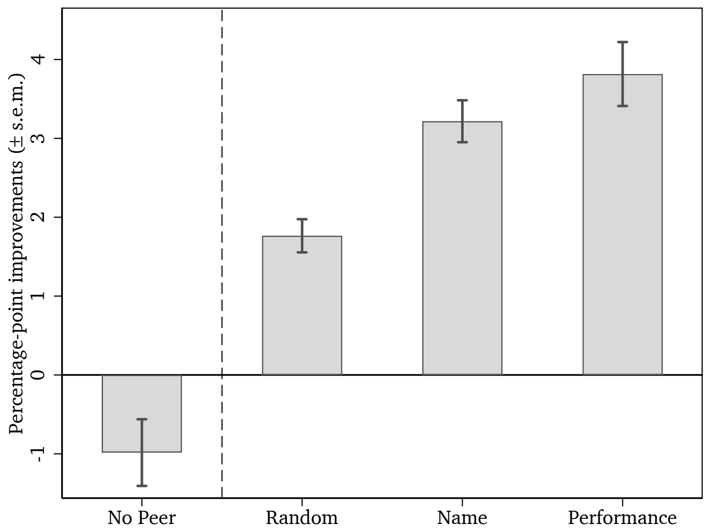
5 Results

We start by analyzing if the presence of peers matter for performance in our setting. To do so, we compare the NoPeer treatment to the three different treatments where peers are present. In a second step, which constitutes the main part of our investigation, we focus on the consequences of different peer assignment rules, with and without self-selection, and decompose the average effect into direct and indirect effects.

5.1 Average treatment effects on performance

We analyze how average performance differs across treatments. For this purpose, we use percentage-point improvements as outcomes and therefore base our comparisons on the performance in the first run. Arguably, this specification takes into account the notion that slower subjects (i.e., those with a slower time in the first run) can improve more easily by the same absolute value compared with faster subjects, as it is physically more difficult for the latter. We corroborate this by analyzing the effect on absolute performance, i.e., the time in the second run.

Figure 3: Average performance improvements



Notes: The figure presents percentage-point improvements from the first to the second run with corresponding standard errors for the treatment without peers in the second run (NoPeer) as well as the three treatments with peers—Random, Name, and Performance. The corresponding regression coefficients are shown in column (1) of Table 2. We control for school-by-grade and gender-by-grade fixed effects, and cluster standard errors at the class level.

Table 2: Average treatment effects

	(a) Percentage-point improvements			(b) Time 2 (sec.)			(c) Time 2 (std.)
	(1) All	(2) All	(3) Peers only	(4) All	(5) All	(6) Peers only	(7) Peers only
NAME	1.73***	1.45***	1.80***	-1.21**	-0.40***	-0.46***	-0.15***
	(0.48)	(0.33)	(0.38)	(0.58)	(0.09)	(0.09)	(0.03)
PERFORMANCE	2.19***	2.05***	2.22***	-1.21**	-0.48***	-0.49***	-0.18***
	(0.68)	(0.41)	(0.42)	(0.57)	(0.08)	(0.07)	(0.03)
NOPEER	-2.68***	-2.75***		0.03	0.76***		0.28***
	(0.46)	(0.52)		(0.57)	(0.15)		(0.06)
Time (First Run)					0.73***	0.67***	0.77***
					(0.04)	(0.04)	(0.04)
Fixed effects	No	Yes	Yes	No	Yes	Yes	Yes
Controls	No	No	Yes	No	No	Yes	No
p-value: NAME = PERF.	0.50	0.20	0.37	0.99	0.52	0.84	0.52
N	715	715	588	715	715	588	715
R^2	0.13	0.18	0.12	0.05	0.82	0.82	0.82

Notes: This table presents least squares regressions according to equation (1) using percentage-point improvements (panel (a)), time in the second run (panel (b)), and standardized times in the second run (panel (c)) as the dependent variable. Fixed effects include school-by-grade and gender-by-grade fixed effects. Controls include the Big 5, locus of control, social comparison, competitiveness, and risk attitudes. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

The presence of peers. Figure 3 presents our first result. Subjects without a peer significantly decrease their performance in the second run compared to the first run. This is in stark contrast to all other treatments in which peers are present (RANDOM, NAME, and PERFORMANCE) for which we observe performance improvements of at least 1.9 percentage points. The corresponding estimates are depicted in Table 2. Columns (1) and (2) show that the estimated percentage-point improvements in the treatment NOPEER are 2.75 percentage points (0.76 seconds, cf. column (5)) lower compared to subjects with randomly assigned peers and statistically significant. This provides strong evidence of the beneficial effects of the presence of peers in our setting.

Self-selection vs. random assignment. In a second step, we compare the three different peer assignment rules to understand how self-selection of peers affects individual performance in contrast to randomly assigned peers. As shown in Figure 3 and Table 2, the performance improvement of subjects with self-selected peers is 1.80 (NAME) to 2.22 percentage points (PERFORMANCE) larger compared to those with randomly assigned peers (cf. column (3) of Table 2). These improvements correspond to nearly 0.50 seconds (cf. column (4)-(6)) or 15-18% of a standard deviation (cf. column (7)). These results show that subjects with self-selected peers outperform those who interact with randomly assigned peers or perform the task without the presence of a peer. We do not observe significant differences between the two treatments featuring self-selection (all p-values between 0.20 and 0.89).[17]

5.2 Decomposition into direct and indirect effect of self-selection

The previous section provides evidence that self-selection of peers improves performance relative to interacting with randomly assigned peers. We now aim at decomposing these total treatment effects into their two possible sources, as outlined in Section 4: first, self-selection changes with whom someone interacts (*indirect effect*), e.g., as subjects chose to interact with their friends; second, self-selection changes the selection procedure from exogenous assignment to self-selection of peers (*direct effect*).

The existence of an indirect effect relies on two conditions. First, peer characteristics need to be important for individual outcomes. Second, self-selection needs to change the characteristics of peers relative to random peer assignment. In the following, we briefly provide evidence that these conditions are indeed satisfied in our setting.

Peer characteristics matter for performance improvements. Intuitively, not all persons exert the same influence on someone's performance: friends might have a different influence than other persons, or differences in productivity might determine performance. To show the relevance of these peer characteristics, we focus on subjects with randomly assigned peers and calculate how much of the variation in outcomes is attributable to the subjects' own or their peers' characteristics. Table E.1 in the Appendix shows that peer characteristics explain 23-31% of the overall variation in performance improvements, highlighting that a peer's characteristics matter for performance in the second run.[18]

[17] Appendix D presents additional specifications and robustness checks. In particular, we report results from difference-in-differences specifications, estimations using biased-reduced linearization or group means to account for the limited number of clusters in our sample, robustness checks that control for outliers, and we report the average treatment effects for different subgroups (by gender, grade, school) in our sample. Our conclusions are robust to all of these checks.

[18] Results from an additional permutation exercise confirm the importance of these results (see Appendix Figure E.1). In addition, Appendix Table E.2 provides estimates of peer effects in several dimensions.

Self-selection changes the peer composition. The second necessary condition is that subjects interact with systematically different peers when self-selecting them. In Section 3, we already showed that the matching quality varies across treatments, indicating differences regarding the peers' fit with subjects' preferences. In addition, the composition changes along the two dimensions which can be easily targeted by individuals in the two preferences. Appendix Figure E.2 shows that subjects are predominantly paired with friends in Name (76% of all peers are friends), whereas the share of peers being friends in Random and Performance is 49% and 37%, respectively. Furthermore, the average absolute difference in times from the first run is only 1.53 seconds in Performance, while it is larger than two seconds in the other two treatments (2.24 and 2.16 seconds in Random and Name). Thus, self-selection inherently changes with whom somebody interacts.[19]

5.2.1 Decomposition of the total effect

The results of the decomposition of the total effect into a direct and indirect effect are presented in Table 3 and depicted for the main specification in Figure 4. Column (1) of Table 3 replicates the total average treatment effect of column (3) of Table 2 for means of comparison. We start by allowing only match quality to mediate the effect, which measures how well a given peer fits someone's preferences. This specification therefore allows us to show that the performance differs across treatments, even holding the fit of a peer constant across treatments. Yet, only controlling for match quality does not allow other additional peer characteristics to mediate the total effect of self-selection. Thus, we document in columns (2) through (4) that some other peer characteristics also influence the performance in the second run. In particular, we find that productivity differences in pairs affect the initially faster and slower students differentially. Column (4) shows that slower subjects within a pair benefit by a 1.00 percentage point improvement from running with a one-second faster student, while the relatively faster student's performance suffers from this productivity difference by 0.44 percentage points. In combination, the average performance of a pair thus improves with increasing differences in productivity. As we observe lower differences in productivity in Performance (cf. Appendix Figure E.3b), this also implies that the indirect effect, with respect to that characteristic, should be negative in Performance.

[19]We also present how our treatments affect the peer composition along various other characteristics in Appendix Table E.3. We find that targeting specific peers also results in systematically different peers in terms of their personality.

Table 3: Decomposition of treatment effects

Panel A: Decomposition	Percentage-Point Improvements				
	(1)	(2)	(3)	(4)	(5)
	Baseline	Match Quality	Friend-ship ties	Time Difference	All
Direct Effects					
Name	1.80***	1.64***	1.80***	1.72***	1.64***
	(0.38)	(0.39)	(0.40)	(0.37)	(0.40)
Performance	2.22***	2.29***	2.13***	2.32***	2.57***
	(0.42)	(0.37)	(0.40)	(0.38)	(0.39)
Peer Characteristics					
Faster Student × High match quality (Name)		-0.08			0.32
		(0.40)			(0.45)
Slower Student × High match quality (Name)		0.39			0.30
		(0.60)			(0.69)
Faster Student × High match quality (Perf.)		1.29**			0.55
		(0.50)			(0.50)
Slower Student × High match quality (Perf.)		-2.11***			-0.76
		(0.65)			(0.69)
Faster Student × Peer is Friend			-0.67		-1.05**
			(0.46)		(0.51)
Slower Student × Peer is Friend			0.08		0.22
			(0.48)		(0.66)
Faster Student × $\lvert \Delta Time\ 1 \rvert$				-0.44***	-0.42***
				(0.13)	(0.14)
Slower Student × $\lvert \Delta Time\ 1 \rvert$				1.00***	0.98***
				(0.21)	(0.19)
Slower Student in Pair		3.76***	2.09***	-0.28	-0.16
		(0.44)	(0.47)	(0.44)	(0.66)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes	Yes
Peer Characteristics	No	No	No	No	Yes
Abs. Diff. in Characteristics	No	No	No	No	Yes
p-value: Name = Perf.	0.37	0.17	0.51	0.18	0.06
N	588	588	588	588	588
R^2	0.12	0.22	0.19	0.28	0.33

Panel B: Role of unobservables	Oster's δ		
	$R^2_{max} = 0.50$	$R^2_{max} = 0.75$	$R^2_{max} = 1.00$
Name	3.39	1.42	0.90
Performance	-27.89	-12.37	-7.95

Notes: Panel A of this table presents least squares regressions of equation (2) using percentage-point improvements as the dependent variable. High match quality is an indicator that equals one if the peer was ranked within a subject's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and absolute differences in personality include the corresponding characteristics of the peer and their absolute differences. Appendix Table E.11 presents the omitted coefficients of column (5). Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors are clustered at the class level. Panel B quantifies the importance of unobservables relative to observable characteristics for column (5) that would imply zero direct effects under different assumption on the theoretical maximum R^2_{max} based on a measure proposed by Oster (2019).

While the parsimonious specifications in columns (2) through (4) support our overall conclusion of a significant direct effect of self-selection, they share the caveat that there could be other characteristics that explain the direct effect. Therefore, we control for all mediating factors jointly in column (5), where we additionally add a rich set of peer personality characteristics as additional mediators. Figure 4 depicts the resulting direct and indirect effects of this specification visually.[20] The estimates indicate the presence of robust direct effects of self-selection on performance for both Name (1.64 pp., SE: 0.40) and Performance (2.57 pp., SE: 0.39). The indirect effects amount to only 0.18 pp. (SE: 0.25) in Name and -0.39 pp. (SE: 0.20) in Performance, contributing only little to the total effect. We present an additional specification in Appendix Table E.7, where we use time in the second run as an alternative outcome variable, and find that the direct effects correspond to time improvements of 0.35 (Name) and 0.49 sec. (Performance). Hence, even the rich set of peer characteristics do not explain the total effect of self-selection, providing evidence that self-selection per se has a direct effect on performance.

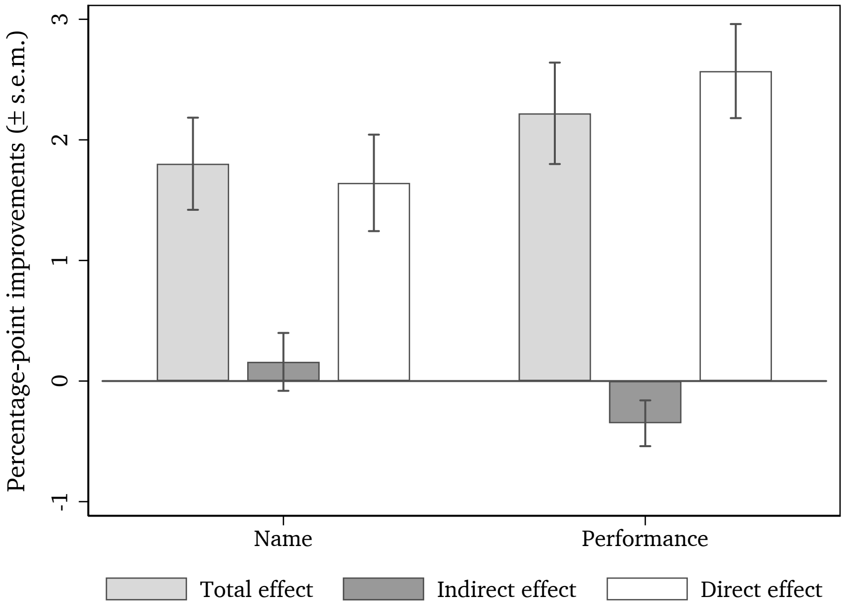
5.2.2 Robustness of the decomposition

In the following, we provide several robustness checks for the decomposition analysis.

Omitted and unobserved variables do not seem to drive our results. In Panel B of Table 3, we address the possible concern that other characteristics which we cannot account for are driving the direct effect. Reassuringly, and as mentioned above, our results above remain relatively stable when adding different sets of peer controls. A more formal approach to tackle this concern is to ask how important unobserved characteristics would have to be to explain our direct treatment effects (Altonji, Elder, and Taber, 2005; Oster, 2019). We follow Oster (2019) and calculate δ, a measure for the relative importance of unobserved characteristics compared to observed ones to explain the direct effects, i.e., to drive down the estimated direct effects to zero. Absolute values of δ larger than one indicate that these omitted variables would have to be more important than observed variables. We calculate these measures for three scenarios that differ in the maximum amount of variance that would theoretically be explained if all factors that affect the outcomes were observed. In all but one extreme

[20] In Figure 4, the indirect effects are defined as the changes in performance improvements that are due to changes in peer characteristics relative to Random based on the specification in column (5) of Table 3. Thus, the indirect effect corresponds to the difference between the total effect and the direct effect. More precisely, it can be calculated by $\beta \times (\overline{\theta}(D^k) - \overline{\theta}(D^{Random}))$ with $k = \{Name, Perf.\}$, where β denotes the influence of peer characteristics and $\overline{\theta}^j$ denotes the average peer characteristics in treatment $j = \{R, N, P\}$. See Appendix C for more details.

Figure 4: Decomposition of average treatment effects into direct and indirect effects



Notes: This figure summarizes the results of the decomposition presented in specification (5) of Table 3. It decomposes the total effect (average treatment effect as shown in column (1) of Table 3) into an indirect effect due to changes in the peer composition and a direct effect of self-selection. The indirect effect comprises the changes in percentage-point improvements that are due to changes in the peer characteristics in NAME and PERFORMANCE relative to RANDOM. Standard errors are clustered at the class level.

scenario, the omitted peer characteristics are required to be more important than the observed peer characteristics. This suggests that such unobserved variables need to have a larger effect than productivity differences, friendship ties, match quality, and all other controls—including personality traits—combined. Compared to other studies, our analysis already allows for many peer characteristics to influence subjects' behavior. Therefore, we can conclude that unobserved characteristics are highly unlikely to drive the estimated direct effects.

Control group comprising only high-match peers does not alter our conclusions. A different approach to test if performance improvements are due to a direct effect or solely due to changing the match quality, would be to focus on subjects with the same (high) match quality across treatments. For this purpose, we restrict the control group to those subjects that received one of their preferred peers by pure chance and estimate the direct effects using this control group in Table E.4. These matches occurred by pure chance and are not due to self-selection. The table, thus, compares only subjects who interact with someone they would self-select or did self-select and therefore only differ in the assignment procedure, holding the match quality constant.

We find that the direct effects persist when restricting to these subgroups as a control group.

Further robustness checks. We perform a series of further robustness checks, discussed in more detail in Appendix E. First, we provide evidence that our results do not depend on the exact definition of key variables such as the definition of match quality, friendship ties, or the exact functional form (e.g., allowing for non-linear effects in productivity differences). Second, we show that an alternative estimation procedure, where we estimate peer effects in Random only and impose those effects on the treatments with self-selection, neither changes our conclusions qualitatively nor in terms of statistical significance. Third, we estimate additional specifications controlling for further aspects of the classroom environment. Fourth, we address a potential concern that our results may be an artifact of over-fitting control variables and thus adopt a post-double LASSO estimator proposed by Belloni, Chernozhukov, and Hansen (2014) to alleviate these concerns. Our results and conclusions remain unaffected by all of these robustness checks.

Taken together, the analysis shows that self-selection improves individual performance directly and not due to a change in the peer composition. Characteristics of peers are important in determining outcomes, but they do not explain the average treatment effects of self-selection. Instead, our results suggest that the treatment effects are driven by a direct effect of self-selection.

5.2.3 Interpretation of the direct effect of self-selection

We interpret the direct effect as a positive effect of self-selection due to increased control or autonomy over the peer assignment mechanism. However, one might worry that knowledge of all three treatment conditions—we elicited preferences for peers irrespective of the treatment—could lead subjects in Random to react negatively due to disappointment that their preferences have not been taken into account. If these disappointed subjects drove our findings, we would falsely attribute effects to self-selection, even if subjects in Name and Performance do not react positively.[21] If the direct effect originated from disappointment, we would expect subjects in Random

[21]At the same time, this also describes a feature of many real-world settings. Imagine that a person is randomly assigned a partner from a group of available people. Even if this person has not been asked explicitly with whom she would like to interact, she still has preferences about interacting with certain people. Therefore, disappointment could also play a role in these settings. This might be true for all settings that feature exogenous assignment and overrule the underlying preferences of the involved persons.

to have enjoyed the experimental task less. Therefore, in column (1) of Appendix Table E.12, we analyze the extent to which subjects across treatments had different perceptions regarding their fun in the second run. We find zero effects. The absence of direct effects in the fun dimension therefore alleviates the concern that knowledge of all three treatments leads to disappointment when subjects are assigned to Random.

We therefore conclude that the direct effects in our experiment are due to positive effects of self-selection. More specifically, we argue that the opportunity to self-select key aspects of one's environment—having autonomy over the peer selection in our experiment—has a direct effect beyond the instrumental value of changing peer characteristics. Self-determination theory (Deci and Ryan, 1985, 2000) provides a credible explanation through which self-selection can impact performance directly. The theory identifies autonomy as a crucial determinant of motivation: individuals who can actively select parts of their environment—most importantly their tasks in work environments—display higher intrinsic motivation.[22] Applying this explanation to our setting suggests that not the selected peer herself increases motivation, but the mere act of selecting her. However, we do not argue that this behavioral effect stems from self-selecting any aspect, but selecting a relevant aspect of one's environment.

Self-determination theory and autonomy in particular have recently gained increasing attention from economists. Cassar and Meier (2018) review the economic literature on non-monetary aspects of work environments in light of self-determination theory and highlight the importance of autonomy for various behavioral outcomes. A related argument to ours also underlies the findings of Bartling, Fehr, and Herz (2014) and Owens, Grossman, and Fackler (2014). Although they do not focus on the effect of autonomy on subsequent outcomes, their studies demonstrate that people have a willingness to pay for making decisions by themselves and maintaining autonomy. Similarly, a growing body of literature demonstrates that restricting subjects' choice sets and therefore restricting their autonomy and freedom can negatively influence outcomes (e.g., Falk and Kosfeld, 2006). Therefore, our results add to this literature by highlighting the motivational benefits of autonomy and self-determination, and provide novel field evidence that having control positively affects outcomes.

[22]Two other components of self-determination theory are relatedness and competence, referring to the need to care about something and the need to feel challenged, respectively. In our experiment, we hold these other components constant across treatments.

5.3 Self-selection vs. exogenous assignment rules: Limits and consequences

Our results show that self-selected peers lead to substantially larger performance improvements than randomly assigned peers. Independently of the exact underlying mechanism, managers and policy-makers therefore face a choice if they want to let workers or students self-select their peers, assign peers at random, or employ another exogenous peer assignment rule. Examples for other assignment rules include tracking in schools (e.g., Betts, 2011; Duflo, Dupas, and Kremer, 2011; Fu and Mehta, 2018; Garlick, 2018) or pairing high-performing students with low-performing ones (e.g., Carrell, Sacerdote, and West, 2013). While we have not implemented these assignment rules in our context, we can use our estimates to simulate the effect of such exogenous peer assignment rules and compare their effect to outcomes under self-selection.

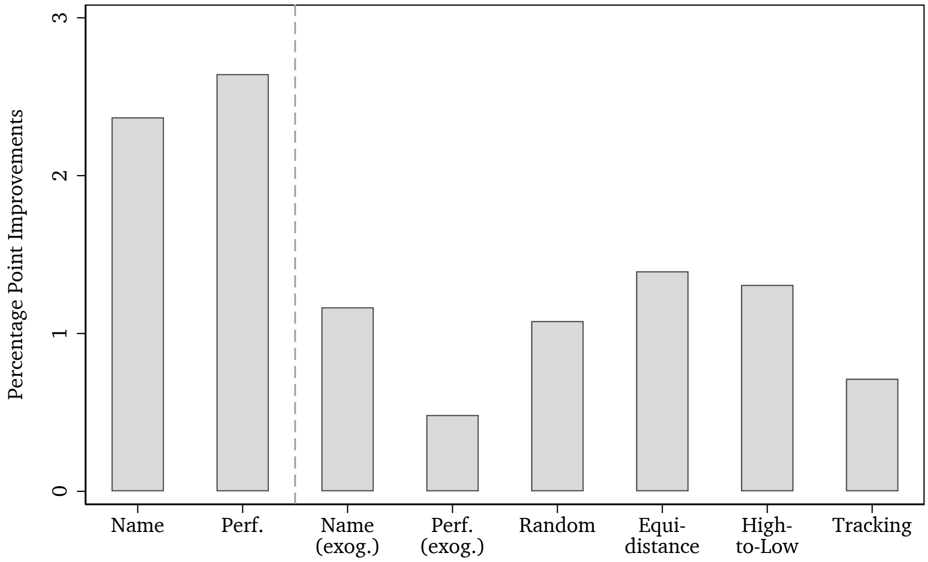
We base our simulations on the peer effects estimated in Section 5.2, using the whole set of peer characteristics (column (5) of Table 3). Given these coefficients, we examine different (exogenous) assignment rules by hypothetical matching students, calculating the resulting effects on performance, and comparing them to the observed performance improvements from our experiment.[23] We start by providing a counterfactual when assigning the same peers as in Name and Performance, but without the direct effect of self-selection, i.e., calculating the performance improvement without the direct effect. Second, we simulate the expected performance improvements under a random matching for all subjects in our experiment. Third, we use several assignment rules that base the assignment on one single and commonly employed peer characteristic, namely past performance. The results in Section 5.2 suggest that pairs with a higher difference in initial performances will improve their performance on average more strongly than pairs with lower differences in initial performance.[24] We consider two matching rules that maximize these productivity differences within pairs and thus should increase performance (Equidistance and High-to-Low), but keep the distance in ranks within the matching group constant.[25] Additionally, we look at the effect of tracking (i.e., pairing the best student with the

[23]We provide details on the prediction of performance improvements and the peer assignment rules in Appendix F.

[24]If this is the only characteristic of a peer that affects performance, aggregate performance would be maximized as long as the sum of productivity differences within a pair is maximized. Given our specification, this is true for all peer assignment rules that match each student from the bottom half of the productivity distribution with a student from the top half.

[25]In particular, we pair students from the top half with the similarly ranked student in the bottom half of performance within their class in Equidistance, or pair the best-performing student with the slowest student and the second-best with the second-slowest student in High-to-Low.

Figure 5: Simulation of different peer assignment rules



Notes: The figure presents predicted percentage-point improvements for the two treatments (NAME, PERFORMANCE) with and without the effect of self-selection, the RANDOM treatment, as well as three simulated peer assignment rules (EQUIDISTANCE, HIGH-TO-LOW and TRACKING). We fix the personal characteristics and other covariates not at the pair level to 0; effect sizes are therefore not directly comparable to treatment effects above. More details are provided in Appendix F.

second-best, third with the fourth, etc.; TRACKING). Importantly, while all of these exogenous assignment rules are based on past performance, we take the effects of all peer characteristics into account.

Figure 5 presents the results from these simulations. The figure reveals that no other peer assignment rule is able to induce similar performance improvements as those featuring self-selection. This is due to the fact that individuals under exogenous peer assignment rules do not benefit from the additional intrinsic value of self-selection. Interestingly, we also observe that reassignment rules that maximize productivity differences in pairs—EQUIDISTANCE and HIGH-TO-LOW—do not substantially improve average performance compared to the random assignment of peers. This suggests that other changes in peer characteristics offset the positive effect of increased productivity differences.

Our results therefore suggest a limited effectiveness of exogenous reassignment due to the fact that rearranging peers based on only one characteristic simultaneously changes other characteristics of peers, which may counteract the intended effect. Thus, the simulations imply that the consequences of peer assignment rules are difficult to predict or even ambiguous if peer effects exist in multiple dimensions. This insight further helps us to understand why we observe a very small indirect effect in the decomposition of the treatment effects despite the fact that peer characteristics help to explain much of the variation in individual outcomes (cf. Table 3 and Appendix Table E.1).

The simulations suggest that self-selection of peers can be an attractive alternative compared to traditional peer assignment rules to increase individual performance. While managers and policy-makers certainly care about overall performance and efficiency, they might also be interested in effects on inequality and possible unintended psychological side effects of such peer assignment mechanisms. These may affect workers' well-being and, in turn, further decisions and outcomes such as retention. In this experiment, both treatments do not lead to larger overall inequality in performance. However, we find some evidence that individual ranks are more perturbed between the two runs in NAME and RANDOM relative to PERFORMANCE and the share of subjects performing at insufficient levels increases slightly in PERFORMANCE (cf. Appendix Table F.2). Turning to psychological and social side effects, Appendix Table F.2 provides evidence that subjects in PERFORMANCE experience significantly more pressure compared to the other two treatments with peers. A second concern and potential challenge for policy-makers is to prevent (social) exclusion of individuals if some of them are not selected. Yet, the likelihood of subjects not being nominated as part of the name-based preferences is only around 1.3%, suggesting only a minor role of (social) exclusion in our setting.

6 Discussion and Conclusion

Across many workplace environments, employees actively select peers who affect their behavior and performance. Thus, identifying and understanding the effects of actively chosen peers is important to appropriately leverage social interactions within organizations. Our framed field experiment introduces a new way to study self-selection of peers in a controlled manner and is able to separate the effect of a specific peer on a subject's performance from the overall effect of self-selection. We find that the presence of peers in general improves performance, but that self-selecting peers yields additional performance improvements of 15-18% of a standard deviation relative to random assignment of peers. While peer characteristics are important to explain the variation in performance improvements, they do not drive the estimated treatment effects. Rather, our results suggest that these improvements stem from a direct effect of self-selection. Based on self-determination theory (Deci and Ryan, 1985), we interpret this direct effect such that the ability to select one's own peer enhances a student's intrinsic motivation and subsequently increases an individual's performance.

One might be eager to infer that our results give rise to a trade-off between performance improvements as a result of self-selection per se and the exogenous assignment of performance-maximizing peers. However, our simulations show that exogenous peer assignment rules, which try to lever peer effects in ability, have an impact close to zero compared to random assignment in our setting and are in general ambiguous in size and sign. This result relies on the existence of peer effects in multiple dimensions, which partially offset each other, limiting the effectiveness of exogenous reassignment rules. Hence, positive effects of peer self-selection might be performance-maximizing even in absence of subjects choosing "optimal" peers.

More generally, our data show that peer assignment rules might not only affect performance. They also may lead to potentially negative and unintended consequences. Managers who do not care only about the performance levels, but also about other aspects of a workers' utility, have to take these potential side effects into account, eventually affecting employee retention. This can be of particular importance when changing from established rules to different and new peer assignment rules in firms or organizations.

The results in this paper constitute a first proof-of-concept that self-selection of peers can directly affect performance. It is therefore crucial to investigate whether and how this effect transfers to other situations and other mechanisms of peer effects. In particular, it is not clear ex-ante whether such effects persist over an extended period of time or under different incentive structures such as in cooperative environments. We note, however, that our experiment mimics important features of other settings. For example, many effort decisions of workers have only low stakes or are only implicitly incentivized. This includes, for example, cashiers working faster when a productive peer enters their shift (Mas and Moretti, 2009), and employees or volunteers working longer when their peers stay longer (Linardi and McConnell, 2011; Rosaz, Slonim, and Villeval, 2016). Similar, in this study, individuals are incentivized to perform well in a given task, but their behavior does not not necessarily have immediate consequences.

More generally, self-selection can also be interpreted as a non-monetary aspect of (work) environments. Workers in firms increasingly form self-managed work teams (Lazear and Shaw, 2007), study groups at universities often form endogenously (Chen and Gong, 2018), researchers select their co-authors, employees self-select with whom they work by referring others to their employer (Friebel et al., 2020; Lazear and Oyer, 2012), and workers decide when to work and where to work (Bloom et al., 2015). Hence, managers, teachers, or supervisors might be interested in adopting different forms of self-selection and autonomy as another form of motivational tool complementing non-monetary incentives used in schools (Levitt et al., 2016) or workplaces (Cassar and Meier, 2018).

Based on the fact that decision rights have an intrinsic value to humans (e.g., Bartling, Fehr, and Schmidt, 2013; Owens, Grossman, and Fackler, 2014) and the fact that workers or students have a willingness to pay for more flexible working arrangements (e.g., Mas and Pallais, 2017; Wiswall and Zafar, 2017), one might conclude that increasing autonomy decreases earnings. Yet, our findings point to a second potential effect: workers' performance increases. While the net effect is ambiguous in general, previous evidence documents a positive correlation between work time flexibility and earnings (e.g., Beckmann, Cornelissen, and Kräkel, 2017; Mas and Pallais, 2017). Taken together, these patterns suggest that the consequences of autonomy for wages depend on the interplay of job attributes and the selection of workers into such workplaces, their compensating wage differentials, and the corresponding effort change.

In this paper, we highlight that self-selecting peers can serve as a complement to other established methods such as incentives and exogenous peer assignment policies aimed at increasing individual performance. However, further research on the interplay between endogenous group formation, social interactions, and production environments remains imperative to understand how peer effects work.

References

Ager, Philipp, Leonardo Bursztyn, Lukas Leucht, and Hans-Joachim Voth (2019). "Killer Incentives: Relative Position, Performance and Risk-Taking among German Fighter Pilots, 1939-45".

Allport, Floyd H (1920). "The Influence of the group upon association and thought". In: *Journal of Experimental Psychology* 3.3, p. 159.

Altonji, Joseph G., Todd E. Elder, and Christopher R. Taber (2005). "Selection on observed and unobserved variables: Assessing the effectiveness of catholic schools". In: *Journal of Political Economy* 113.1, pp. 151–184.

Bandiera, Oriana, Iwan Barankay, and Imran Rasul (2009). "Social Connections and Incentives in the Workplace: Evidence From Personnel Data". In: *Econometrica* 77.4, pp. 1047–1094.

— (2010). "Social Incentives in the Workplace". In: *Review of Economic Studies* 77.2, pp. 417–458.

Bartling, Björn, Ernst Fehr, and Holger Herz (2014). "The intrinsic value of decision rights". In: *Econometrica* 82.6, pp. 2005–2039.

Bartling, Björn, Ernst Fehr, and Klaus M. Schmidt (2013). "Discretion, productivity, and work satisfaction". In: *Journal of Institutional and Theoretical Economics* 169.1, pp. 4–22.

Beckmann, Michael, Thomas Cornelissen, and Matthias Kräkel (2017). "Self-managed Working Time and Employee Effort: Theory and Evidence". In: *Journal of Economic Behavior and Organization* 133, pp. 285–302.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen (2014). "Inference on Treatment Effects after Selection among High-Dimensional Controls". In: *Review of Economic Studies* 81.2, pp. 608–650.

Belot, Michèle and Jeroen van de Ven (2011). "Friendships and Favouritism on the Schoolground – A Framed Field Experiment". In: *Economic Journal* 121.557, pp. 1228–1251.

Betts, Julian R. (2011). "The Economics of Tracking in Education". In: ed. by Eric A. Hanushek, Stephen Machin, and Ludger Woessmann. Vol. 3. Handbook of the Economics of Education. Elsevier, pp. 341–381.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying (2015). "Does Working from Home Work? Evidence from a Chinese Experiment". In: *Quarterly Journal of Economics* 130.1, pp. 165–218.

Bó, Pedro Dal, Andrew Foster, and Louis Putterman (2010). "Institutions and Behavior: Experimental Evidence on the Effects of Democracy". In: *American Economic Review* 100.5, pp. 2205–2229.

Bradler, Christiane, Robert Dur, Susanne Neckermann, and Arjan Non (2016). “Employee Recognition and Performance: A Field Experiment”. In: *Management Science* 62.11, pp. 3085–3099.

Brandts, Jordi, David Cooper, and Roberto Weber (2014). “Legitimacy, Communication, and Leadership in the Turnaround Game”. In: *Management Science* 61.11, pp. 2627–2645.

Bursztyn, Leonardo, Florian Ederer, Bruno Ferman, and Noam Yuchtman (2014). “Understanding Mechanisms Underlying Peer Effects: Evidence From a Field Experiment on Financial Decisions”. In: *Econometrica* 82.4, pp. 1273–1301.

Carrell, Scott, Bruce Sacerdote, and James West (2013). “From Natural Variation to Optimal Policy? The Importance of Endogenous Peer Group Formation”. In: *Econometrica* 81.3, pp. 855–882.

Cassar, Lea and Stephan Meier (2018). “Nonmonetary Incentives and the Implications of Work as a Source of Meaning”. In: *Journal of Economic Perspectives* 32.3, pp. 215–38.

Chan, Tszkin Julian and Chungsang Tom Lam (2015). “Type of Peers Matters: A Study of Peer Effects of Friends Studymates and Seatmates on Academic Performance”.

Chen, Roy and Jie Gong (2018). “Can self selection create high-performing teams?” In: *Journal of Economic Behavior and Organization* 148, pp. 20–33.

Chevalier, Judith A., M. Keith Chen, Peter E. Rossi, and Emily Oehlsen (2019). “The Value of Flexible Work: Evidence from Uber Drivers”. In: *Journal of Political Economy* 127.6, pp. 2735–2794.

Cicala, Steve, Roland Fryer, and Jörg Spenkuch (2018). “Self-Selection and Comparative Advantage in Social Interactions”. In: *Journal of the European Economic Association* 16.4, pp. 983–1020.

Corgnet, Brice, Joaquín Gómez-Miñambres, and Roberto Hernán-González (2015). “Goal Setting and Monetary Incentives: When Large Stakes Are Not Enough”. In: *Management Science* 61.12, pp. 2926–2944.

Deci, Edward and Richard Ryan (1985). *Intrinsic motivation and self-determination in human behavior*. Springer Science & Business Media.

— (2000). “The “what” and “why” of goal pursuits: Human needs and the self-determination of behavior”. In: *Psychological Inquiry* 11.4, pp. 227–268.

Dohmen, Thomas, Armin Falk, David Huffman, Uwe Sunde, Jürgen Schupp, and Gert G. Wagner (2011). “Individual Risk Attitudes: Measurement, Determinants, and Behavioral Consequences”. In: *Journal of the European Economic Association* 9.3, pp. 522–550.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2011). "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya". In: *American Economic Review* 101.5, pp. 1739–1774.

Elsner, Benjamin and Ingo Isphording (2017). "A Big Fish in a Small Pond: Ability Rank and Human Capital Investment". In: *Journal of Labor Economics* 35.3, pp. 787–828.

Falk, Armin and Andrea Ichino (2006). "Clean Evidence on Peer Effects". In: *Journal of Labor Economics* 24.1, pp. 39–57.

Falk, Armin and Michael Kosfeld (2006). "The Hidden Costs of Control". In: *American Economic Review* 96.5, pp. 1611–1630.

Fershtman, Chaim and Uri Gneezy (Apr. 2011). "The Tradeoff between Performance and Quitting in High Power Tournaments". In: *Journal of the European Economic Association* 9.2, pp. 318–336. ISSN: 1542-4766.

Friebel, Guido, Matthias Heinz, Mitchell Hoffman, and Nick Zubanov (2020). "What Do Employee Referral Programs Do? Measuring the Direct and Overall Effects of a Management Practice".

Fu, Chao and Nirav Mehta (2018). "Ability Tracking, School and Parental Effort, and Student Achievement: A Structural Model and Estimation". In: *Journal of Labor Economics* 36.4, pp. 923–979.

Garlick, Robert (2018). "Academic Peer Effects with Different Group Assignment Policies: Residential Tracking versus Random Assignment". In: *American Economic Journal: Applied Economics* 10.3, pp. 345–369.

Gibbons, Frederick and Bram Buunk (1999). "Individual Differences in Social Comparison: Development of a Scale of Social Comparison Orientation." In: *Journal of Personality and Social Psychology* 76.1, pp. 129–147.

Gill, David, Zdenka Kissová, Jaesun Lee, and Victoria Prowse (2019). "First-Place Loving and Last-Place Loathing: How Rank in the Distribution of Performance Affects Effort Provision". In: *Management Science* 65.2, pp. 494–507.

Gneezy, Uri and Aldo Rustichini (2004). "Gender and Competition at a Young Age". In: *American Economic Review* 94.2, pp. 377–381.

Golsteyn, Bart, Arjan Non, and Ulf Zölitz (2021). "The Impact of Peer Personality on Academic Achievement". In: *Journal of Political Economy* 129.4, pp. 1052–1099.

Guryan, Jonathan, Kory Kroft, and Matthew Notowidigdo (2009). "Peer Effects in the Workplace: Evidence from Random Groupings in Professional Golf Tournaments". In: *American Economic Journal: Applied Economics* 1.4, pp. 34–68.

Heckman, James and Rodrigo Pinto (2015). "Econometric Mediation Analyses: Identifying the Sources of Treatment Effects from Experimentally Estimated Production

Technologies with Unmeasured and Mismeasured Inputs". In: *Econometric Reviews* 34.1-2, pp. 6–31.

Herbst, Daniel and Alexandre Mas (2015). "Peer Effects on Worker Output in the Laboratory Generalize to the Field". In: *Science* 350.6260, pp. 545–549.

Huettner, Frank and Marco Sunder (2012). "Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values". In: *Electronic Journal of Statistics* 6, pp. 1239–1250.

Hunt, Peter J. and Joseph M. Hillery (1973). "Social Facilitation in a Coaction Setting: An Examination of the Effects over Learning Trials". In: *Journal of Experimental Social Psychology* 9.6, pp. 563–571.

Irving, Robert (1985). "An Efficient Algorithm for the "Stable Roommates" Problem". In: *Journal of Algorithms* 6.4, pp. 577–595.

Jackson, C. Kirabo and Elias Bruegmann (2009). "Teaching Students and Teaching Each Other: The Importance of Peer Learning for Teachers". In: *American Economic Journal: Applied Economics* 1.4, pp. 85–108.

Kiessling, Lukas and Jonathan Norris (2020). "The Long-run Effects of Peers on Mental Health". Working Paper.

Kiessling, Lukas, Jonas Radbruch, and Sebastian Schaube (2020). "Determinants of Peer Selection". Working Paper.

Koch, Alexander K. and Julia Nafziger (2011). "Self-regulation through Goal Setting". In: *Scandinavian Journal of Economics* 113.1, pp. 212–227.

Kosfeld, Michael and Susanne Neckermann (2011). "Getting More Work for Nothing? Symbolic Awards and Worker Performance". In: *American Economic Journal: Microeconomics* 3.3, pp. 86–99.

Kräkel, Matthias (2016). "Peer effects and incentives". In: *Games and Economic Behavior* 97, pp. 120–127.

Lazear, Edward and Paul Oyer (2012). "Personnel Economics". In: *The Handbook of Organizational Economics*. Princeton University Press, pp. 479–519.

Lazear, Edward and Kathryn Shaw (2007). "Personnel Economics: The Economist's View of Human Resources". In: *Journal of Economic Perspectives* 21.4, pp. 91–114.

Levitt, Steven, John List, Susanne Neckermann, and Sally Sadoff (2016). "The Behavioralist Goes to School: Leveraging Behavioral Economics to Improve Educational Performance". In: *American Economic Journal: Economic Policy* 8.4, pp. 183–219.

Linardi, Sera and Margaret A. McConnell (2011). "No excuses for good behavior: Volunteering and the social environment". In: *Journal of Public Economics* 95, pp. 445–454.

Manski, Charles (1993). “Identification of Endogenous Social Effects: The Reflection Problem”. In: *Review of Economic Studies* 60.3, pp. 531–542.

Mas, Alexandre and Enrico Moretti (2009). “Peers at Work”. In: *American Economic Review* 99.1, pp. 112–145.

Mas, Alexandre and Amanda Pallais (Dec. 2017). “Valuing Alternative Work Arrangements”. In: *American Economic Review* 107.12, pp. 3722–59.

Murphy, Richard and Felix Weinhardt (2020). “Top of the Class: The Importance of Ordinal Rank”. In: *Review of Economic Studies* 87.6, pp. 2777–2826.

Oster, Emily (2019). “Unobservable Selection and Coefficient Stability: Theory and Evidence”. In: *Journal of Business & Economic Statistics* 37.2, pp. 187–204.

Owens, David, Zachary Grossman, and Ryan Fackler (2014). “The Control Premium: A Preference for Payoff Autonomy”. In: *American Economic Journal: Microeconomics* 6.4, pp. 138–161.

Park, Sangyoon (2019). “Socializing at Work: Evidence from a Field Experiment with Manufacturing Workers”. In: *American Economic Journal: Applied Economics* 11.3, pp. 424–455.

Rao, Gautam (2019). “Familiarity Does Not Breed Contempt: Generosity, Discrimination, and Diversity in Delhi Schools”. In: *American Economic Review* 109.3, pp. 774–809.

Roels, Guillaume and Xuanming Su (2014). “Optimal Design of Social Comparison Effects: Setting Reference Groups and Reference Points”. In: *Management Science* 60.3, pp. 606–627.

Rosaz, Julie, Robert Slonim, and Marie Claire Villeval (2016). “Quitting and peer effects at work”. In: *Labour Economics* 39, pp. 55–67.

Rotter, Julian B. (1966). “Generalized Expectancies for Internal Versus External Control of Reinforcement”. In: *Psychological Monographs: General and Applied* 80.1, pp. 1–28.

Sacerdote, Bruce (2001). “Peer Effects with Random Assignment: Results for Dartmouth Roommates”. In: *Quarterly Journal of Economics* 116.2, pp. 681–704.

Schneider, Simone and Jürgen Schupp (2011). “The Social Comparison Scale: Testing the Validity, Reliability, and Applicability of the IOWA-Netherlands Comparison Orientation Measure (INCOM) on the German Population”. In: *DIW Data Documentation*.

Sutter, Matthias and Daniela Glätzle-Rützler (2015). “Gender Differences in the Willingness to Compete Emerge Early in Life and Persist”. In: *Management Science* 61.10, pp. 2339–2354.

Sutter, Matthias, Stefan Haigner, and Martin G. Kocher (2010). “Choosing the Carrot or the Stick? Endogenous Institutional Choice in Social Dilemma Situations”. In: *Review of Economic Studies* 77.4, pp. 1540–1566.

Tincani, Michela (2017). “Heterogeneous Peer Effects and Rank Concerns: Theory and Evidence”.

Towler, Gary (1986). “From Zero to One Hundred: Coaction in a Natural Setting”. In: *Perceptual and Motor Skills* 62.2, pp. 377–378.

Villeval, Marie Claire (2020). “Performance Feedback and Peer Effects: A Review”.

Weinhardt, Michael and Jürgen Schupp (2011). “Multi-Itemskalen im SOEP Jugendfragebogen”. In: *DIW Data Documentation*.

Wiswall, Matthew and Basit Zafar (Aug. 2017). “Preference for the Workplace, Investment in Human Capital, and Gender”. In: *Quarterly Journal of Economics* 133.1, pp. 457–507.

Zajonc, Robert B. (1965). “Social Facilitation”. In: *Science* 149.3681, pp. 269–274.

Záraté, Roman Andrés (2020). “An Experimental Approach to Peer Effects: Evidence on Social and Academic Skills”.

Appendix – For Online Publication

A Experimental instructions and protocol

The instructions below are translations of the German instructions for the experiment.

Introduction to the experiment

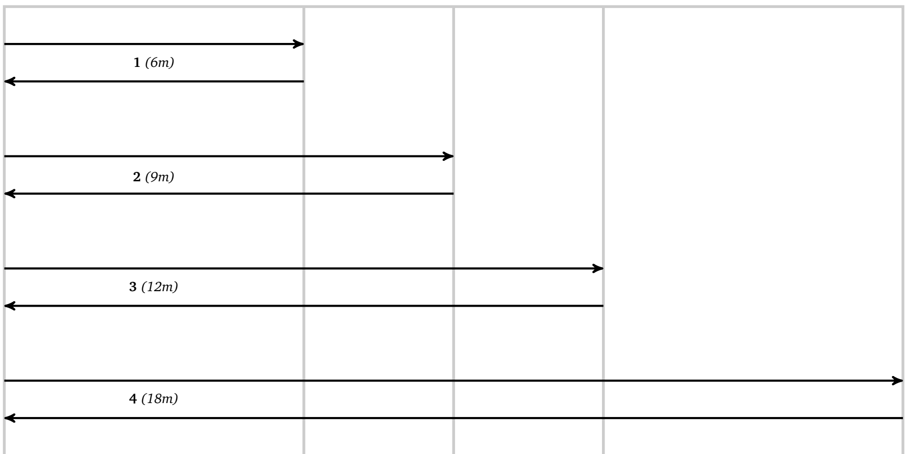
Welcome everyone to today's physical education session. As you might have already noticed, today's session is going to be different. As you already know, you will take part in a scientific study. For that purpose, you received a parental consent form and handed it back to your teacher. If you have not handed it back to your teacher, you will not take part in the study.
If you have any questions throughout the study, you can address us at any point in time.
The study comprises several parts. For the first part, we would like you to do a running task called suicide runs. My colleague will shortly demonstrate this exercise.
(The following verbal explanation was accompanied with physical demonstration of the exercise; below we present a further description of the task)
You start at the baseline of the volleyball court and run to this first line. You touch it with your hand and run back to the baseline. You touch the baseline with your hand and run to the next line. Touch it again, back to the baseline; touch it, and then to the third line, back to the baseline, to the fourth line, and then you return to the baseline. Everyone of you will run alone and the goal is to be as fast as possible. After this run, we will hand you a computer to fill out a survey.
After all of you have run and filled out the survey, you will run for a second time. This time at the same time as another student. During the survey we will ask you—among other questions—with whom you would like to run. You will receive detailed information about this later on.
The goal during both runs is to be as fast as possible. We will record your running times and hand it to your teacher. Your teacher will grade your performance during both runs.
Before we start with the study, we would like to remind you again that your participation is voluntary. If anyone does not want to take part in the study, then please inform us now.
Do you have any further questions? If this is not the case, please start with the warm-up, before we start with the experiment.
(This introduction was followed by short warm-up exercise by students. After a short warm-up all students were asked to leave the gym and wait in an accompanying the hallway until they were called in the gym to take part in the first run. We asked students

whether they understood the task and, if necessary, explained the task again. Directly afterwards, they were asked to leave the gym and were led to a different room. There we asked them to complete the survey on a computer we handed them.)

Detailed Description of Task

The exact task—often called suicide runs—is to sprint and turn at every line of the volleyball court. Subjects had to line up at the baseline. From there, they started running to the first attack line of the court (6 meters). After touching this line, they returned to the baseline again, touching the line on arrival. The next sprint took the subjects to the middle of the court (9 meters), the third to the second attack line (12 meters) and the last to the opposite baseline (18 meters), each time returning to the baseline. They finished by returning to the starting point. The total distance of this task was 90 meters. Figure A.1 illustrates the task.

Figure A.1: Illustration of the running task



Notes: This figure illustrates the running task (suicide run) subjects performed as part of the experiment. See text for details.

Screenshots of the preferences-elicitation during the survey

The following two screenshots, Figures A.2 and A.3, display translated elicitation screens for performance- and name-based preferences for peers.

Introduction to the second run for the whole class

(Class was gathered for announcement)

Figure A.2: Performance-based preferences

	4-5 seconds slower	3-4 seconds slower	2-3 seconds slower	1-2 seconds slower	0-1 seconds slower	Own time	0-1 seconds faster	1-2 seconds faster	2-3 seconds faster	3-4 seconds faster	4-5 seconds faster
1st Preference	○	○	○	○	○	x	○	○	○	○	○
2nd Preference	○	○	○	○	○	x	○	○	○	○	○
3rd Preference	○	○	○	○	○	x	○	○	○	○	○
4th Preference	○	○	○	○	○	x	○	○	○	○	○
5th Preference	○	○	○	○	○	x	○	○	○	○	○
6th Preference	○	○	○	○	○	x	○	○	○	○	○
7th Preference	○	○	○	○	○	x	○	○	○	○	○

Figure A.3: Name-based preferences

	ID of running mate	4-5 seconds faster	3-4 seconds faster	2-3 seconds faster	1-2 seconds faster	0-1 seconds faster	Own time	0-1 seconds slower	1-2 seconds slower	2-3 seconds slower	3-4 seconds slower	4-5 seconds slower
1st Preference	---------	○	○	○	○	○	x	○	○	○	○	○
2th Preference	---------	○	○	○	○	○	x	○	○	○	○	○
3rd Preference	---------	○	○	○	○	○	x	○	○	○	○	○
4th Preference	---------	○	○	○	○	○	x	○	○	○	○	○
5th Preference	---------	○	○	○	○	○	x	○	○	○	○	○
6th Preference	---------	○	○	○	○	○	x	○	○	○	○	○

We will shortly start with the second run. For this purpose a partner for you has been selected. In your class, the partner has been selected *randomly [based on your indication how fast you want your partner to be] [based on the classmates you nominated]*. We would like to remind you that the objective is to be as fast as possible and it is only about your own time. Your teacher will receive a list with your performance, but no information about the pairs.

(The list with pairs was read out aloud to the students and students were accompanied to the waiting zone. Students were called into the gym one pair after the other. In the gym they were led to separate, but adjacent tracks. Each student was accompanied by one experimenter, who recorded their time as well their responses to four additional questions.)

Individual introduction directly before the second run

The two of you will now run simultaneously. Your partner has been selected randomly [based on your indication how fast you want your partner to be] [based on the classmates you nominated].
(We then asked each subject to assess their relative performance in the first run) Please guess, who of you two was faster during the first run?

Post-run questionnaire after the second run

(Directly after a pair participated in the second run, we asked each of the two subjects the following three questions in private)

(1) How much fun did you have during the second run? Please rate this on a scale from 1—no fun at all—to 5—a lot of fun

(2) If you were to run again, would you prefer to run alone or with a partner)

(3) How much pressure did you feel form your partner during the second run? Please rate this on a scale from 1—no pressure at all—to 5—a lot of pressure.

B Randomization and manipulation check

B.1 Randomization Checks

Table B.1 and B.2 present the randomization checks for our experiment. For each variable we check if there are differences across treatments without controls (always first column) or after taking into account the randomization procedure, i.e., with grade-by-school fixed effects (always 2nd column). Any small differences in the first run can be explained by grade-by-school fixed effects and hence are an artifact of the block randomization, as classrooms rather than individuals were randomly assigned to treatments. Note that for these randomization checks, we use the sample of all students independently of their participation in the second run.

Table B.3 additionally checks if the random assignment in RANDOM was successful. For this purpose, column (1) in Panel A tests whether the peer's performance in the first run predicts a subjects own performance, taking into account a correction for exclusions bias (Guryan, Kroft, and Notowidigdo, 2009). In line with random assignment, this is not the case. In Panel B, we summarize the results of a series of additional randomization checks. Specifically, if peers are indeed randomly assigned, we should not observe an unexpected high number of significant correlations when examining if a peer's characteristic predicts a subject's characteristic. Similar to Panel A, we run those randomization checks with 11 different own and peer characteristics (times in the first run, age, agreeableness, conscientiousness, extraversion, neuroticism, openness to experience, locus of control, risk attitudes, attitudes towards social comparisons, and competitiveness), resulting in $11 \times 11 = 121$ separate randomization checks. In line with random assignment, we observe that less than 10% of those regressions are significant at the 10% level, less then 5% at the 5% level, and less than 1% at the 1% level. Thus, these randomization checks support that the randomization of peers in RANDOM was indeed successful.

Table B.1: Randomization check: Subject characteristics

	Time in the first run (in sec.)		Age		Female		Number of friends		Perf.-based preference 1		Perf.-based preference 2		Perf.-based preference 3	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
NAME	-0.73	0.35	0.13	-0.03	-0.11	0.10*	0.05	0.10	0.12	0.03	0.16	0.13	-0.08	-0.05
	(0.57)	(0.32)	(0.46)	(0.04)	(0.10)	(0.05)	(0.15)	(0.15)	(0.23)	(0.17)	(0.22)	(0.17)	(0.23)	(0.21)
PERFORMANCE	-0.62	0.43	0.15	-0.03	-0.12	0.06	-0.10	-0.02	0.21	0.23	0.21	0.27	0.08	0.29
	(0.54)	(0.32)	(0.46)	(0.05)	(0.10)	(0.05)	(0.20)	(0.15)	(0.23)	(0.20)	(0.22)	(0.21)	(0.28)	(0.22)
NOPEER	-0.68	0.25	-0.62	-0.06	-0.13	0.09	0.03	-0.13	0.05	-0.02	-0.02	0.19	0.03	0.27
	(0.55)	(0.44)	(0.51)	(0.06)	(0.09)	(0.08)	(0.18)	(0.14)	(0.28)	(0.22)	(0.23)	(0.34)	(0.28)	(0.23)
Grade-by-school FEs	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
N	754	754	754	754	754	754	754	754	754	754	754	754	754	754
R^2	.012	.14	.049	.87	.013	.15	.0035	.088	.0016	.028	.0027	.03	.00095	.034
p-value: NAME vs. PERF.	.82	.82	.95	.88	.92	.57	.4	.4	.66	.32	.83	.42	.5	.13
p-value: NAME vs. NO PEER	.91	.81	.15	.57	.77	.9	.94	.05	.77	.81	.43	.83	.66	.13
p-value: PERF. vs. NO PEER	.91	.72	.14	.57	.87	.79	.49	.53	.52	.32	.33	.78	.84	.95

Notes: This table presents least squares regressions using the variable in the table header as the dependent variable. Note that we use unincentivized measures of preferences for peers in the NOPEER treatment in columns (9)-(14). *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors are clustered on the class level.

Table B.2: Randomization check: Class characteristics

	Number of students in class		Share of participating students		Grade
	(1)	(2)	(3)	(4)	(5)
NAME	0.03	0.52	0.08	0.07	-0.02
	(1.02)	(1.06)	(0.07)	(0.07)	(0.44)
PERFORMANCE	-0.43	-0.92	0.06	0.05	0.20
	(1.28)	(1.29)	(0.07)	(0.06)	(0.44)
NOPEER	-1.43	0.15	-0.02	-0.05	-0.71
	(1.11)	(1.40)	(0.07)	(0.11)	(0.44)
Grade-by-school FEs	No	Yes	No	Yes	No
N	48	48	48	48	48
R^2	.039	.43	.065	.29	.079
p-value: NAME vs. PERF.	.69	.2	.81	.79	.64
p-value: NAME vs. NO PEER	.14	.75	.081	.24	.14
p-value: PERF. vs. NO PEER	.42	.42	.13	.37	.057

Notes: This table presents least squares regressions using the variable in the table header as the dependent variable. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors are clustered on the class level.

Table B.3: Randomization checks: Randomization of peers

Panel A: Randomization check for time in the first run in RANDOM

	Time in the first run (in sec)
Peer's time in the first run (in sec)	0.00
	(0.16)
N	206
R^2	.44

Panel B: Randomization checks using all observable characteristics

	Share (number) of randomization checks significant at		
	10%	5%	1%
Expected under random assign.	10.0% (12)	5.0% (6)	1.0% (1)
Observed in RANDOM	6.6% (8)	3.3% (4)	0.8% (1)
Number of tests performed	121		

Notes: In column (1) of Panel A, we present a randomization check for times in the first run. We adopt the Guryan, Kroft, and Notowidigdo (2009) correction to account for exclusion bias, i.e. control for leave-one-out mean on matching group level. Panel B of this table extends the randomization checks for the RANDOM treatment by calculating randomization checks for each own and peer characteristic (times in the first run, age, agreeableness, conscientiousness, extraversion, neuroticism, openness to experience, locus of control, risk attitudes, attitudes towards social comparisons and competitiveness) resulting in $11x11$ = 121 separate randomization checks. All regressions include class fixed effects. We include the correction terms for the exclusion bias as suggested by Guryan, Kroft, and Notowidigdo (2009) by controlling for the leave-one-out mean of the corresponding characteristic on the matching-group level.

B.2 Preferences for Peers

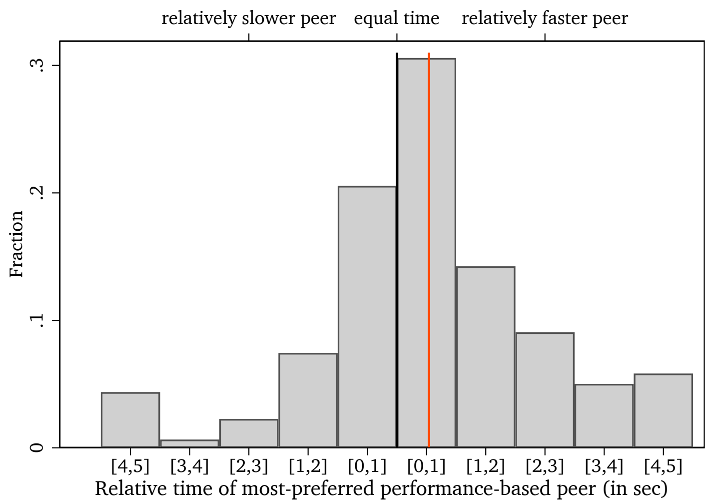
We summarize the preferences for peers according to name- and performance-based preferences in Table B.4 and Figure B.1, respectively. Two findings emerge: first, most subjects nominated friends as their most-preferred peer; and second, while subjects on average preferred to run with a slightly faster peer, there is a strong heterogeneity in this preference. We analyze these preferences in further detail in Kiessling, Radbruch, and Schaube (2020).

Table B.4: Share of name-based preferences being friends

Name-based preference	1st	2nd	3rd	4th	5th	6th	Average
Share of peers being friends	0.89	0.79	0.73	0.60	0.49	0.41	0.65

Notes: This table presents the share of friends for each name-based preference (most-preferred peer to sixth most-preferred peer as well as pooled over all six preferences) as elicited in the survey.

Figure B.1: Most-preferred performance-based peer



Notes: The figure presents a histogram of the peer preferences over relative performance as elicited in the survey. Vertical lines indicate own time (black line; equals zero by definition) and the mean preference of all individuals (red line; 0.56 sec faster on average, where we used the midpoint of each interval to calculate the mean).

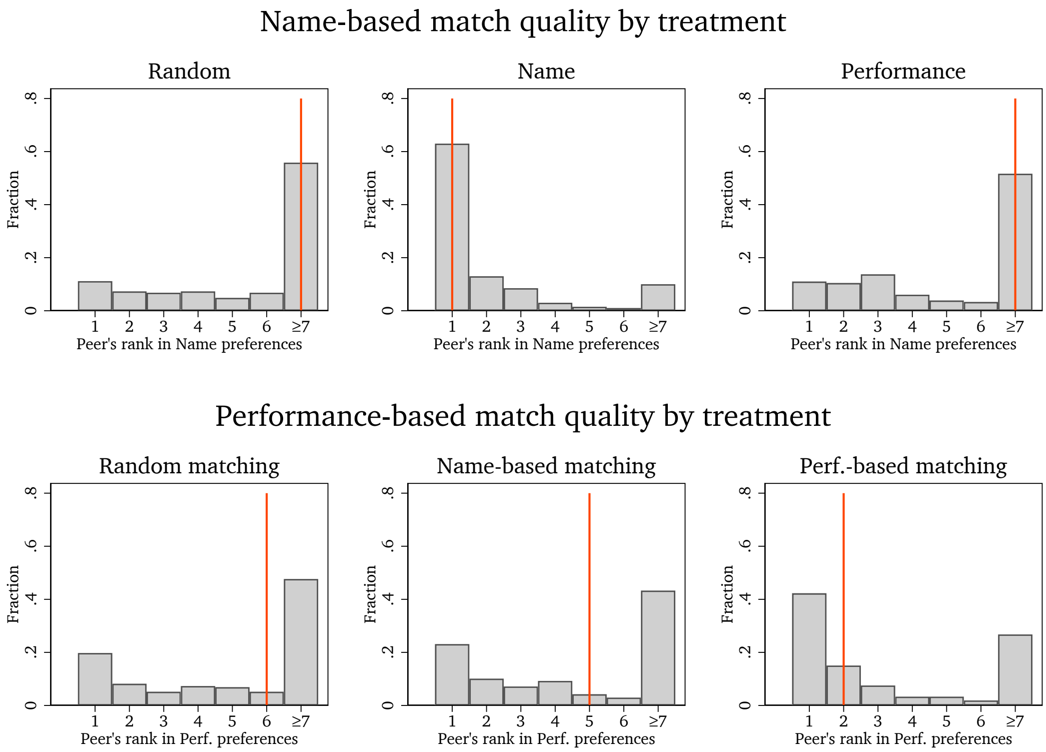
B.3 Manipulation Check

In Section 3, we presented the resulting match qualities using the preferences as elicited in the survey. However, some subjects may prefer relative times, which are not available to them. For example, the fastest subject in the class might want to run with someone who is even faster, or a student wants to run with somebody else who is 1-2 seconds faster but by chance there is no one in the class with such a time. Similarly, subjects in NAME may rank other subjects which were not present during the experiment or did not participate. We therefore present an alternative approach to evaluate the match quality by taking the availability of peers into account. This implies that the quality of a match does not correspond directly to the elicited preferences; rather, based on these preferences all available subjects (i.e., the students participating in the study) are ranked. The quality of the match is then calculated based on this new ranking and results in a realized feasible match quality.

Consequently, we determine the feasible match quality by calculating how high a classmate is ranked in a list of available classmates. [1] In NAME, this can only increase the match quality. If someone nominates another student who is not available as her most-preferred peer and she received her second highest ranked choice, this means that she is matched with her most-preferred feasible peer. Similar arguments can increase the match quality for preferences over relative performance. However, the match quality in performance can also be lower. Suppose that a student ranks the category "1-2 seconds faster" highest and there are three students in that category. However, she is only matched with her second highest ranked category. There would have been three subjects whom she would have preferred more, generating a feasible match quality of 4. We present the corresponding histograms in Figure B.2 and observe that the median of the feasible match quality is actually higher for both treatments relatively to the match qualities depicted in Figure 2.

[1]We code peers who are not ranked among the first six preferences with a match quality of 7.

Figure B.2: Feasible match quality across treatments



Notes: The figure presents a histogram of match qualities for each treatment evaluated according to either the subjects' name-based preferences (upper panel) or performance-based preferences (lower panel). Vertical lines denote median match qualities.

C Econometric Framework

In this appendix, we outline how to interpret our estimates in light of a mediation analysis similar to Heckman and Pinto (2015). A key difference between their framework and ours is that we are interested in the direct effect of our treatments as well as indirect effects of a change in the production inputs, rather than only the latter.

In general, any observed change in outcomes of our experiment can be attributed to one of two main sources: first, different peer assignment mechanisms may affect peer interactions directly; and second, self-selection changes the peers and therefore the difference between the subject's and his or her peer's characteristics. We therefore decompose the total effect into a direct effect of self-selection as well as a pure peer composition effect. This takes into account the change in relative peer characteristics across treatments.[1]

Consider the following potential outcomes framework. Let Y^P and Y^N and Y^R denote the counterfactual outcomes in the three treatments. Naturally, we only observe the outcome in one of the treatments:

$$Y = D^N Y^N + D^P Y^P + (1 - D^P)(1 - D^N) Y^R \tag{C.1}$$

Let θ_d be a vector characterizing a peer's relative characteristics in treatment $d \in \{R, N, P\}$.[2] Similar to the potential outcomes above, we can only observe the peer composition vector θ in one of the treatments and thus $\theta = D_P\theta_P + D_N\theta_N + (1 - D_P)(1 - D_N)\theta_R$ and define an intercept α analogously. The outcome in each of the treatments is therefore given by

$$Y_d = \alpha_d + \beta_d\theta + \delta X + \epsilon_d \tag{C.2}$$

where we implicitly assume that we have a linear production function, which can be interpreted as a first-order approximation of a more complex non-linear function. The outcome depends on own characteristics X as well as treatment-specific effects of relative characteristics of the peer θ and a zero-mean error term ϵ_d, independent of X and θ.

[1]Our treatments do not change the distribution of characteristics or skills within the class or of a particular subject; rather, the treatments change with whom from the distribution a subject interacts. Due to the random assignment, we assume independence of own characteristics and the treatment.

[2]In our estimations, we include the following characteristics in θ_d: indicators whether the peer ranked high in the individual preference rankings, effects of absolute time differences for slower and faster subjects within pairs, the rank and presence of friendship ties within pairs, and absolute differences in personal characteristics (Big 5, locus of control, competitiveness, social comparison and risk attitudes).

Potentially, there are unobserved factors in θ. We therefore split θ in a vector with the observed inputs ($\bar{\theta}$) and unobserved inputs ($\tilde{\theta}$)[3] with corresponding effects $\bar{\beta}_d$ and $\tilde{\beta}_d$ and can rewrite equation (C.2) as follows:

$$Y_d = \alpha_d + \bar{\beta}_d\bar{\theta} + \tilde{\beta}_d\tilde{\theta} + \delta X + \epsilon_d \tag{C.3}$$

$$= \tau_d + \bar{\beta}_d\bar{\theta} + \delta X + \tilde{\epsilon}_d \tag{C.4}$$

where $\tau_d = \alpha_d + \tilde{\beta}_d\mathbb{E}[\tilde{\theta}]$ and $\tilde{\epsilon}_d = \epsilon_d + \tilde{\beta}_d(\tilde{\theta} - \mathbb{E}[\tilde{\theta}])$. We assume $\tilde{\epsilon}_d \overset{d}{=} \epsilon$, i.e., are equal in their distribution with a zero-mean. We can express the effect of $\bar{\theta}$ in Name and Performance relative to the effect in Random by rewriting $\beta_d = \beta + \Delta_{R,d}$. Accordingly, we rewrite the coefficients $\bar{\beta}_d$ of θ_i as the sum of the coefficients in Random denoted by β and the distance of the coefficients between treatment d and Random (denoted by $\Delta_{R,d}$).

$$Y_d = \tau_d + \bar{\beta}\bar{\theta} + \bar{\Delta}_{R,d}\bar{\theta} + \delta X + \tilde{\epsilon}_d \tag{C.5}$$

$$= \hat{\tau}_d + \bar{\beta}\bar{\theta} + \delta X + \tilde{\epsilon}_d \tag{C.6}$$

In what follows, we are interested in $\bar{\tau}_d = \mathbb{E}[\hat{\tau}_d - \hat{\tau}_R]$ $(d \in \{N, P\};\ \hat{\tau}_d = \tau_d + \bar{\Delta}_{R,d}\bar{\theta})$, i.e., the direct treatment effect of Name and Performance conditional on indirect effects from changes in the peer composition captured in $\bar{\theta}$. In general, this direct effect subsumes the effect of the treatment itself ($\alpha_d - \alpha_R$), the changed impact of the same peer's observables ($\bar{\Delta}_{R,d}\bar{\theta}$), and changes in unmeasured inputs as well as their effect $((\tilde{\beta} + \tilde{\Delta}_{R,d})\tilde{\theta})$. Yet, we show in a series of robustness checks that the latter two effects play only a minor role for the estimated direct effect. Hence, we can interpret this direct effect in light of self-determination theory (Deci and Ryan, 1985) as an additional motivation due to being able to self-select a peer. This focus on the direct effect is a key difference compared with Heckman and Pinto (2015), who are mainly interested in the indirect effects of the mediating variables. The empirical specification of C.6 is given by

$$y_{icgs} = \bar{\tau}^{Name} D_{ic}^{Name} + \bar{\tau}^{Perf} D_{ic}^{Perf} + \beta\theta_i\big(D_{ic}^{Name}, D_{ic}^{Perf}\big) + \delta X_i + \lambda_{sg} + u_{icgs} \tag{C.7}$$

where we are interested in $\bar{\tau}_N$ and $\bar{\tau}_P$, the direct effects of our treatments relative to Random. Indirect effects are captured by $\beta\theta_i$, the effect of changed peer characteristics on the outcome y_{icgs}.

[3]Furthermore, we assume that unobserved and observed inputs are independent conditional on X and D.

D Robustness checks for average treatment effects

Difference-in-difference specifications. In Table D.1, we present difference-in-differences specifications of the treatment effects. Reiterating our findings from Table 2, we find significant performance improvements for subjects matched to a peer, and in particular among those with self-selected peers.

Table D.1: Robustness checks: Difference-in-difference specification

	Time				
	(1) sec.	(2) sec.	(3) sec.	(4) std.	(5) log.
Name	-0.73	0.03			
	(0.57)	(0.20)			
Performance	-0.62	0.25			
	(0.54)	(0.23)			
NoPeer	-0.68	0.08			
	(0.55)	(0.26)			
Second Run	-0.46***	-0.51***	-0.49***	-0.17***	-0.02***
	(0.12)	(0.11)	(0.13)	(0.05)	(0.00)
Name × Second Run	-0.47***	-0.46***	-0.45**	-0.16**	-0.02**
	(0.16)	(0.15)	(0.19)	(0.06)	(0.01)
Performance × Second Run	-0.59***	-0.57**	-0.55**	-0.19**	-0.02**
	(0.22)	(0.22)	(0.27)	(0.09)	(0.01)
NoPeer × Second Run	0.70***	0.76***	0.73***	0.25***	0.03***
	(0.14)	(0.14)	(0.17)	(0.06)	(0.01)
Fixed effects	No	Yes	No	No	No
Individual FEs	No	No	Yes	Yes	Yes
p-value: Name = Perf.	0.59	0.59	0.71	0.71	0.63
N	1469	1469	1469	1469	1469
R^2	0.04	0.41	0.94	0.94	0.95

Notes: This table presents results from difference-in-difference specifications using running times in seconds in columns (1)-(3), standardized times in column (4), and logarithms of running times in column (5) as the dependent variable. The sample includes all subjects participating either in the first or second run. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Limited number of clusters and outliers. In Tables D.2 and D.3, we compare the clustered standard errors with clustered standard errors using a biased-reduced linearization to account for the limited number of clusters. Comparing the first two columns, we observe that the results are robust to this alternative specification of the standard errors. In column (3), we additionally check whether looking at matching group-specific group means—i.e., the average percentage point improvement for males and females in each class—affects the estimates. While the power is reduced due to the small number of observations, the treatment effects persist and the coefficients on the treatment effects do not change. Finally, columns (4) and (5) analyze the sensitivity of our estimates with respect to outliers. We use two different strategies. First, we apply a 90% winsorization, which replaces all observations with either a time or a percentage point improvement below or above the threshold with the value at the threshold. We replace a time of improvement below the 5th percentile with the corresponding value of the 5th percentile and all observations above the 95th percentile with the 95th percentile. Second, we truncate the data and keep only those pairs where no time or no improvement falls into the bottom 5% or top 5%. Neither winsorization nor truncation changes our conclusions.

Table D.2: Robustness checks: Limited number of clusters

	Percentage-point improvements				
	(1)	(2)	(3)	(4)	(5)
	Baseline	BRL	Group means	Winsorization	Truncation
NAME	1.26***	1.26***	1.24**	1.22***	1.15***
	(0.44)	(0.48)	(0.57)	(0.33)	(0.32)
PERFORMANCE	2.10***	2.10***	2.24***	1.76***	1.68***
	(0.49)	(0.54)	(0.58)	(0.39)	(0.36)
NOPEER	-2.98***	-2.98***	-3.20***	-2.27***	-1.54***
	(0.48)	(0.52)	(0.64)	(0.38)	(0.38)
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.14	0.17	0.08	0.21	0.19
N	715	715	88	715	608
R^2	0.16	0.16	0.60	0.22	0.22

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) presents the baseline specifications as used in Table 2. Columns (2) uses biased-reduced linearization (BRL) to account for the limited number of clusters. Column (3) uses matching group-specific means as the unit of observation. Finally, columns (4) and (5) apply a 90% winsorization and truncation, respectively. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table D.3: Robustness checks: Limited number of clusters using time in the second run

	Time in second run (in sec.)				
	(1)	(2)	(3)	(4)	(5)
	Baseline	BRL	Group means	Winsorization	Truncation
NAME	-0.43***	-0.43***	-0.36**	-0.29***	-0.19
	(0.12)	(0.13)	(0.14)	(0.10)	(0.12)
PERFORMANCE	-0.59***	-0.59***	-0.57***	-0.35***	-0.31***
	(0.12)	(0.13)	(0.15)	(0.10)	(0.11)
NOPEER	0.73***	0.73***	0.87***	0.73***	0.82***
	(0.14)	(0.15)	(0.16)	(0.17)	(0.17)
Fixed effects and Time 1	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.28	0.31	0.14	0.61	0.34
N	715	715	88	715	612
R^2	0.81	0.81	0.95	0.84	0.78

Notes: This table presents least squares regressions using times in the second run as the dependent variable. Column (1) presents the baseline specifications as used in Table 2. Columns (2) uses biased-reduced linearization (BRL) to account for the limited number of clusters. Column (3) uses matching group-specific means as the unit of observation. Finally, columns (4) and (5) apply a 90% winsorization and truncation, respectively. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Subsample analyses. We further analyze the robustness of our results by looking at different subsamples. We therefore split our sample first by grades in the upper panels of Tables D.4 and D.5 and by schools as well as gender in the lower panels, and estimate the treatment effects separately for those samples. The table shows the robustness of the estimated treatment effects as these effects persists for all subsamples with similar magnitude.

Table D.4: Robustness checks: Subsample analyses using performance improvements

	Percentage-Point Improvements				
	(1)	(2)	(3)	(4)	(5)
	Baseline	7th grade	8th grade	9th grade	10th grade
NAME	1.45***	1.92***	2.57***	1.47**	1.12*
	(0.33)	(0.06)	(0.28)	(0.61)	(0.62)
PERFORMANCE	2.05***	2.75***	2.45***	2.43***	1.31
	(0.41)	(0.51)	(0.12)	(0.67)	(0.90)
NOPEER	-2.75***	-1.24*	-3.16***	-4.62***	-2.92***
	(0.52)	(0.57)	(0.44)	(0.53)	(0.82)
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.20	0.12	0.71	0.21	0.85
N	715	158	172	184	201
R^2	0.18	0.23	0.20	0.20	0.07
	(6)	(7)	(8)	(9)	(10)
	Female	Male	School 1	School 2	School 3
NAME	1.45***	1.38***	1.73***	1.43**	1.49***
	(0.49)	(0.45)	(0.16)	(0.65)	(0.45)
PERFORMANCE	2.36***	1.49*	1.33***	2.29***	1.80
	(0.47)	(0.78)	(0.00)	(0.54)	(1.23)
NOPEER	-3.08***	-2.41**			-2.78***
	(0.78)	(0.98)			(0.47)
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.03	0.88	0.04	0.14	0.78
N	466	249	148	274	293
R^2	0.16	0.22	0.04	0.10	0.25

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) presents the estimates using the whole sample as in Table 2. Columns (2)-(5) restrict the sample to one grade, columns (6) and (7) to each gender and columns (8)-(10) to one school. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table D.5: Robustness checks: Subsample analyses using time in second run

	Time in second run (in sec.)				
	(1) Baseline	(2) 7th grade	(3) 8th grade	(4) 9th grade	(5) 10th grade
NAME	-0.40***	-0.44***	-0.67***	-0.39**	-0.42**
	(0.09)	(0.04)	(0.14)	(0.14)	(0.17)
PERFORMANCE	-0.48***	-0.57**	-0.52***	-0.58***	-0.44**
	(0.08)	(0.18)	(0.06)	(0.14)	(0.17)
NOPEER	0.76***	0.34**	0.92***	1.46***	0.62***
	(0.15)	(0.14)	(0.12)	(0.14)	(0.17)
Time (First Run)	0.73***	0.74***	0.80***	0.72***	0.67***
	(0.04)	(0.06)	(0.03)	(0.09)	(0.08)
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.52	0.53	0.36	0.35	0.94
N	715	158	172	184	201
R^2	0.82	0.78	0.87	0.76	0.86
	(6) Female	(7) Male	(8) School 1	(9) School 2	(10) School 3
NAME	-0.41***	-0.36***	-0.63***	-0.41**	-0.32**
	(0.11)	(0.11)	(0.05)	(0.16)	(0.12)
PERFORMANCE	-0.64***	-0.26	-0.43***	-0.54***	-0.34
	(0.09)	(0.18)	(0.05)	(0.11)	(0.23)
NOPEER	0.90***	0.62***			0.83***
	(0.21)	(0.21)			(0.14)
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.	0.04	0.60	0.00	0.43	0.91
N	466	249	148	274	293
R^2	0.67	0.85	0.51	0.85	0.85

Notes: This table presents least squares regressions using times in the second run as the dependent variable. Column (1) presents the estimates using the whole sample as in Table 2. Columns (2)-(5) restrict the sample to one grade, columns (6) and (7) to each gender and columns (8)-(10) to one school. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Adopting post-double selection Lasso estimators for ATEs. Table D.6 presents results using a post-double selection (PDS) Lasso method described by Belloni, Chernozhukov, and Hansen (2014) that penalizes control variables (in our case personality traits), but allows for inference on treatment indicators. We restrict the robustness checks to the treatments with peers, as we only elicited the large set of control variables for these treatments. Our results are robust to this data-driven selection of control variables.

Table D.6: Post-double selection Lasso estimates of average treatment effects

	(a) Percentage-point improvements		(b) Time 2 (sec.)	
	(1) OLS	(2) PDS Lasso	(3) OLS	(4) PDS Lasso
Name	1.80***	1.64***	-0.46***	-0.44***
	(0.38)	(0.36)	(0.09)	(0.09)
Performance	2.22***	2.20***	-0.49***	-0.48***
	(0.42)	(0.40)	(0.07)	(0.07)
Time (First Run)			0.67***	0.67***
			(0.04)	(0.04)
Fixed effects	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes
p-value: Name = Perf.	0.37	0.20	0.84	0.72
N	588	588	588	588
R^2	0.12		0.82	

Notes: This table presents least squares regressions using percentage-point improvements or time in the second run as the dependent variables. Odd columns present OLS estimates, whereas even columns adopt the post-double selection (PDS) Lasso method proposed by Belloni, Chernozhukov, and Hansen (2014). The PDS Lasso always includes the treatment indicators and grade-by-school and gender-by-grade fixed effects and penalizes the remaining control variables to avoid over-fitting. Standard errors in this specification are only valid for the treatment indicators and fixed effects. See Belloni, Chernozhukov, and Hansen (2014) for further details. Fixed effects include school-by-grade and gender-by-grade fixed effects. Controls include the Big Five, locus of control, social comparison, competitiveness and risk attitudes, as well as time in the first run (only when using times in the second run as an outcome). *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

E Additional Material: Decomposition

E.1 Necessary conditions for the decomposition: Relevance of peers and changes in peer composition

Table E.1 presents the results of a variance decomposition using Shapley values based on equation (2) (Huettner and Sunder, 2012). More specifically, we estimate equation (2) and decompose the corresponding coefficient of determination, R^2, into variation that is attributable to individual characteristics and peer characteristics. For the latter, we include the rank within a pair itself as well as the rank interacted with match quality with respect to both sets of preferences, friendship indicators and productivity differences. We also include personality traits of a peer and absolute differences in personality traits between peers. This corresponds to the full specification that we also use in our decomposition (col. 5 of Table 3). We then account for correlations among different explanatory variables by employing a variance decomposition based on Shapley values to calculate the marginal contribution of each group of variables (see, e.g., Huettner and Sunder, 2012). In addition to our main specification using percentage-point improvements as the outcome, we also report two specifications using time in the second run as an outcome (with and without controlling for times in the first run). Across all specifications, we see that peer characteristics are important to understand the variation in subjects' performance—a necessary condition to be satisfied for the relevance of indirect effects in our decomposition in Section 5.2. Additional evidence is provided by the placebo check in Figure E.1. We randomly form artificial new pairs within each matching group. For each random draw of pairs, we then calculate the share of the variance explained by peer characteristics (i.e., partial R-squared). We repeat this exercise 1,000 times. We find that our observed partial R-squared exceeds over 99% of all partial R-squared we find for simulated pairs.

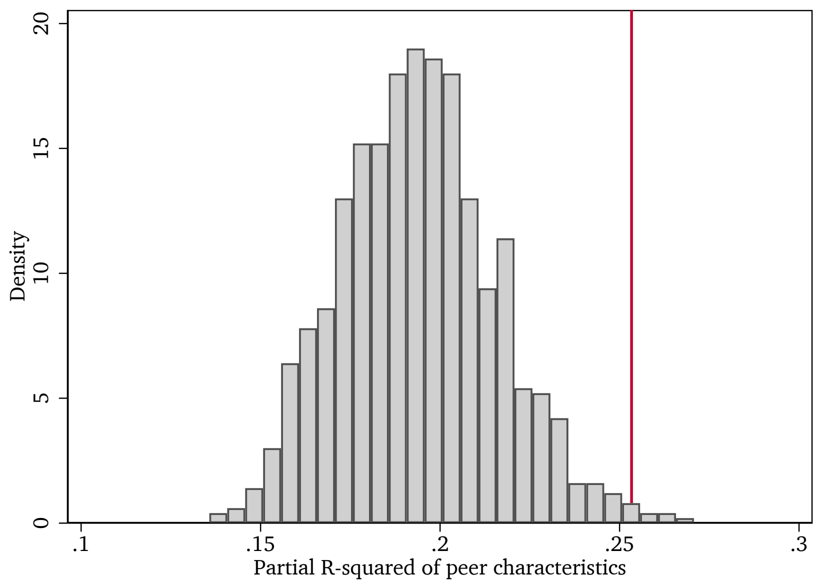
In addition to this variance decomposition, Table E.2 provides complementary evidence on peer effects in several dimensions using data from Random only. In particular, we observe that there exist significant and non-linear peer effects in performance differences as well as several personality traits. A second necessary condition described in Section 5.2 pertains changing peer characteristics when allowing for self-selection. In Figure E.2, we show that this is indeed the case. Specifically, in Figure E.3a we observe more friendship ties among running pairs in Name (76%)than in the other two treatments, Random (49%) and Performance (37%). Similarly, Figure E.3b shows that subjects have lower absolute differences in times in the first run in Performance (1.53 seconds) compared to Random (2.24 sec.) and Name (2.16 sec.). We quantify these and other differences in a regression setup in Table E.3.

Table E.1: Variance decomposition of performance improvements

Outcome	Explained variation (R^2)		Variation attributable to: Treatments		Peer characteristics		Individual characteristics	
Panel A: Random only								
Percentage-point improvement	.27	*(100%)*	—		.23	*(85%)*	.04	*(15%)*
Time in second run	.79	*(100%)*	—		.25	*(32%)*	.54	*(68%)*
Time in second run (w/o time 1)	.7	*(100%)*	—		.31	*(45%)*	.39	*(55%)*
Panel B: All treatments combined								
Percentage-point improvement	.29	*(100%)*	.03	*(11%)*	.21	*(73%)*	.04	*(15%)*
Time in second run	.82	*(100%)*	.02	*(2%)*	.18	*(22%)*	.62	*(75%)*
Time in second run (w/o time 1)	.67	*(100%)*	.02	*(3%)*	.24	*(36%)*	.41	*(62%)*

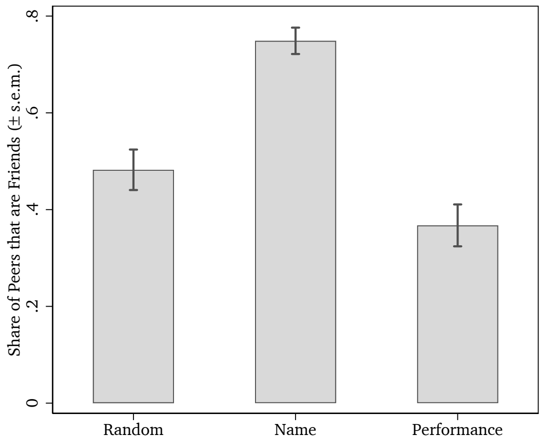
Notes: This table presents decompositions of the coefficient of determination, R^2, using Shapley values and is based on equation (2) estimated on Random only (Panel A) and all three treatments (Random, Name, Performance; Panel B). The specification using percentage-point improvements corresponds to our main specification. For reference, we also report a similar decomposition using times in the second run (with and without controlling for times in the first run) as the outcome.

Figure E.1: Placebo check: Variance decomposition with simulated random peers

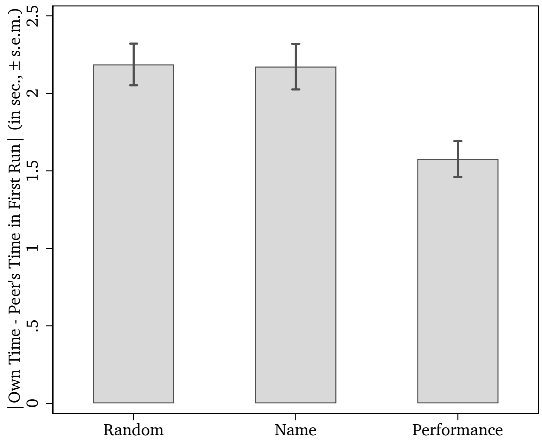


This figure presents the results from a placebo variance decomposition in the Random treatment. Specifically, we randomly form pairs within matching groups, perform a variance decomposition of actual times in the second run in (i) own characteristics (including time in the first run) and fixed effects and (ii) peer characteristics, similar to Table E.1. We then plot the resulting partial R^2 of 1,000 of these placebo decompositions. The red line indicates the actual partial R^2 observed in our data, which exceeds 99.5% of all placebo estimates. Excluding times in the first run from the own characteristics shifts the distribution by approx. 0.05 to the right, but leaves the qualitative conclusions unaffected. In particular, the observed partial R^2 is larger than 99.6% of all placebo estimates.

Figure E.2: Changes in peer composition



(a) Share of peers being friends



(b) Absolute differences in productivity

Notes: Figure E.3a presents the share of all subjects who nominated their assigned peer as a friend for each of the three treatments including standard errors. Figure E.3b shows the average absolute within-pair difference in productivity (measured in times from the first run) and including standard errors for each treatment. We include school-by-grade and gender-by-grade fixed effects. The corresponding regressions are depicted in Appendix Table E.3.

Table E.2: Peer effects in RANDOM

	(1) Match Qual.	(2) Friend	(3) Time Diff.	(4) Personality	(5) All
Faster Student × High match quality (NAME)	0.71				0.35
	(0.95)				(0.92)
Slower Student × High match quality (NAME)	-0.57				-0.65
	(1.52)				(1.39)
Faster Student × High match quality (PERF.)	1.00				0.23
	(1.23)				(0.95)
Slower Student × High match quality (PERF.)	-2.06				-0.79
	(1.54)				(1.26)
Faster Student × Peer is Friend		-0.20			-0.06
		(0.90)			(0.89)
Slower Student × Peer is Friend		-0.25			0.09
		(1.03)			(1.23)
Faster Student × $\lvert\Delta Time\ 1\rvert$			-0.47**		-0.65**
			(0.18)		(0.24)
Slower Student × $\lvert\Delta Time\ 1\rvert$			0.83**		0.61
			(0.36)		(0.37)
$\lvert\Delta$ Agreeableness$\rvert$				-0.79	-0.73
				(0.95)	(0.92)
$\lvert\Delta$ Conscientiousness$\rvert$				-0.05	-0.04
				(0.74)	(0.77)
$\lvert\Delta$ Extraversion$\rvert$				-0.27	-0.28
				(0.41)	(0.51)
$\lvert\Delta$ Openness$\rvert$				-1.36*	-1.32*
				(0.66)	(0.73)
$\lvert\Delta$ Neuroticism$\rvert$				1.32	1.38
				(0.99)	(1.05)
$\lvert\Delta$ Locus of Control$\rvert$				-0.33	-0.35
				(0.90)	(0.85)
$\lvert\Delta$ Social Comp.$\rvert$				-0.45	-0.43
				(0.66)	(0.72)
$\lvert\Delta$ Competitiveness$\rvert$				-0.53	-0.54
				(0.54)	(0.53)
$\lvert\Delta$ Risk attitudes$\rvert$				0.92**	0.85*
				(0.37)	(0.41)
Slower Student in Pair	3.92***	2.65***	-0.40	2.76***	0.41
	(0.64)	(0.61)	(0.89)	(0.58)	(1.07)
Peer Characteristics	No	No	No	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes
p-value: NAME = PERF.					
N	206	206	206	206	206
R^2	0.14	0.12	0.19	0.22	0.29

Notes: This table presents least squares regressions according to equation (2) using percentage-point improvements as the dependent variable on RANDOM only. High match quality is an indicator that equals one if the partner was ranked within an individual's first three preferences. Personality characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table E.3: Effects of treatments on peer composition

	Match Qual. (name)	Match Qual. (time)	Friendship Ties	Time 1	
Name	0.47***	0.04	0.31***	-0.08	
	(0.05)	(0.04)	(0.06)	(0.19)	
Performance	-0.08	0.23***	-0.09	-0.70***	
	(0.07)	(0.04)	(0.08)	(0.21)	
N	588	588	294	294	
R^2	.35	.082	.21	.09	
p-value: Name vs. Perf.	4.8e-13	.000056	2.6e-08	.0037	
Mean in Random	.23	.3	.4	2.4	
	Extra-version	Agree-ableness	Conscien-tiousness	Neuroticism	Openness to Experience
Name	-0.15	0.07	-0.15	0.13	-0.16
	(0.15)	(0.10)	(0.11)	(0.13)	(0.11)
Performance	0.01	0.13	-0.20	0.30**	0.12
	(0.17)	(0.09)	(0.12)	(0.13)	(0.11)
N	294	294	294	294	294
R^2	.052	.057	.049	.041	.032
p-value: Name vs. Perf.	.18	.5	.66	.2	.029
Mean in Random	1.3	1.1	1.1	.99	1.2
	Locus of Control	Social Comparison	Compe-titiveness	Risk	
Name	0.12	0.02	0.04	0.08	
	(0.11)	(0.11)	(0.13)	(0.12)	
Performance	0.48***	-0.19**	0.12	0.06	
	(0.12)	(0.09)	(0.11)	(0.11)	
N	294	294	294	294	
R^2	.066	.033	.033	.018	
p-value: Name vs. Perf.	.0028	.076	.41	.78	
Mean in Random	.99	1.1	1.1	1.1	

Notes: This table presents least squares regressions using absolute differences in pairs' characteristics except for match quality and friendship as the dependent variable. All regressions include school-by-grade and gender-by-grade fixed effects in regressions with individual outcomes. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

E.2 Robustness checks and additional material for decomposition

Control group comprising only high-match peers does not alter our conclusions. Table E.4 restricts the control group sample to subjects with a high match quality within RANDOM to show that the treatment effects persist for these subjects and the coefficients on peer compositional effects do not substantially change.

Table E.4: Only high match quality sample as comparison group

	Percentage-Point Improvements				
	(1) All	(2) RANDOM & NAME	(3) with Controls	(4) RANDOM & PERF.	(5) with Controls
Direct Effects					
NAME	1.64***	2.63***	2.52***		
	(0.40)	(0.34)	(0.43)		
PERFORMANCE	2.57***			3.46***	2.38***
	(0.39)			(0.78)	(0.63)
Peer Characteristics					
Faster Student × Match Quality (name-based)	0.32				0.29
	(0.45)				(1.20)
Slower Student × Match Quality (name-based)	0.30				-0.66
	(0.69)				(1.16)
Faster Student × Match Quality (perf.-based)	0.55		0.22		
	(0.50)		(0.65)		
Slower Student × Match Quality (perf.-based)	-0.76		-1.11		
	(0.69)		(0.89)		
Faster Student × Peer is friend	-1.05**		-1.18		-1.17
	(0.51)		(0.96)		(1.83)
Slower Student × Peer is friend	0.22		-0.96		-1.23
	(0.66)		(0.83)		(1.12)
Faster Student × $\lvert\Delta Time\ 1\rvert$	-0.42***		-0.72**		-0.26
	(0.14)		(0.31)		(0.49)
Slower Student × $\lvert\Delta Time\ 1\rvert$	0.98***		1.18***		0.89**
	(0.19)		(0.34)		(0.42)
Slower Student in Pair	-0.16		0.05		-0.99
	(0.66)		(1.66)		(1.36)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes	Yes
Peer Characteristics	Yes	No	Yes	No	Yes
Abs. Diff. in Characteristics	Yes	No	Yes	No	Yes
N	588	209	209	163	163
R^2	.33	.21	.56	.26	.44

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) presents the last specification of Table 3 for reference. Columns (2) to (5) restrict the comparison group to the sample of individuals in RANDOM that received a peer with high match quality according to their name- (columns (2) and (3)) or performance-based preferences (columns (4) and (5)), respectively. The direct effects persist and the coefficients on peer compositional effects do not change much. High match quality is an indicator that equals one if the partner was ranked within an individual's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and the abs. diff. in personality includes the corresponding characteristics of th peer and the respective absolute difference. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Results are robust to different definition of key variables and the exact functional form. We conduct several robustness checks relaxing the definition of key variables and addressing functional form assumptions used in our decomposition. Recall that our empirical specification covers a series of potential channels for peers to affect individual behavior (e.g., the quality of the match, presence of friendship ties, productivity and personality differences) and allows for non-linearities through interacting them with a student's rank in the pair. We now relax several of the assumptions underlying our main specification.

First, in Table E.5 we use different specifications for match quality. We consider the partner's match quality, an interaction between one's own and the partner's match quality, and feasible match quality as defined in Appendix B, and find that the estimates of our direct effects are qualitatively and quantitatively the same.

Second, in Table E.6, we show that our results do not depend on the precise definition of friendship ties. We check whether our results change when we define friendship ties as undirected or reciprocal rather than directed. As can be seen from the table, the coefficients on the direct effects as well as on other peer characteristics remain the same.

Third, Appendix Table E.7 considers the time in the second run as an outcome instead of percentage-point improvements. The results for these robustness checks remain qualitatively and quantitatively similar.

Fourth, we control for differences in productivity in a more flexible way in Table E.8 by allowing for quartic rather than linear effects of productivity differences in column (2) (see also Figure E.3 comparing linear and quartic terms graphically). In addition, we allow for a second flexible specification using fixed effects for productivity differences. More specifically, we include an indicator for each one-second interval of productivity differences between subjects within a pair. This allows for a potential non-linear influence of productivity differences on our estimates. Comparing the estimates shows that neither the quartic functional form nor the fixed effect specification is restrictive. The results from these specification checks thus alleviate potential concerns about our results being driven by specific functional form assumptions or the definition of key variables.

Table E.5: Robustness Checks for match quality

	Percentage-Point Improvements				
	(1) Partner's MQ	(2) Interaction	(3) Feasible	(4) MQ FEs	(5) MQ FEs
Direct Effects					
NAME	1.60***	1.57***	1.62***	1.59***	1.23**
	(0.43)	(0.47)	(0.39)	(0.46)	(0.46)
PERFORMANCE	2.58***	2.56***	2.46***	2.43***	2.57***
	(0.41)	(0.40)	(0.40)	(0.45)	(0.42)
Peer Characteristics					
High match quality (partner; NAME)	0.12	-0.04			
	(0.41)	(0.47)			
High match quality (partner; PERF.)	-0.03	0.24			
	(0.41)	(0.45)			
High match quality (own and partner; NAME)		0.32			
		(0.74)			
High match quality (own and partner; PERF.)		-0.55			
		(0.87)			
Faster Student × High match quality (feasible; NAME)			-0.20		
			(0.41)		
Slower Student × High match quality (feasible; NAME)			1.13		
			(0.78)		
Faster Student × High match quality (feasible; PERF.)			0.91**		
			(0.43)		
Slower Student × High match quality (feasible; PERF.)			0.02		
			(0.95)		
Faster Student × Match Quality (name-based)	0.29	0.15			
	(0.41)	(0.42)			
Slower Student × Match Quality (name-based)	0.28	0.10			
	(0.69)	(0.76)			
Faster Student × Match Quality (perf.-based)	0.54	0.86			
	(0.49)	(0.61)			
Slower Student × Match Quality (perf.-based)	-0.76	-0.54			
	(0.68)	(0.63)			
Fixed effects	Yes	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes	Yes
Peer Characteristics	Yes	Yes	Yes	No	Yes
Abs. Diff. in Characteristics	Yes	Yes	Yes	No	Yes
Match Qual. FEs	No	No	No	Yes	No
Match Qual. FEs × Rank in pair	No	No	No	No	Yes
Friendship Ties and Performance Differences	Yes	Yes	Yes	No	Yes
p-value: NAME = PERF.	0.09	0.09	0.10	0.12	0.02
N	588	588	588	588	588
R^2	0.33	0.33	0.33	0.15	0.38

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) adds the partner's match quality in addition to own match quality as in Table 3, while column (2) additionally controls for the interaction of own and partner's match quality. Finally, column (3) uses a different measure of match quality, (feasible match quality—see also Appendix B), which acknowledges the fact that certain preferred peers may not be available. High match quality is an indicator that equals one if the partner was ranked within an individual's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and the abs. diff. in personality includes the corresponding characteristics of th peer and the respective absolute difference. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table E.6: Different definitions of friendship ties

	Percentage-Point Improvements			
	(1) directed	(2) undirected	(3) reciprocal	(4) dir. & rec.
Direct Effects				
Name	1.64***	1.60***	1.18**	1.11**
	(0.40)	(0.38)	(0.50)	(0.49)
Performance	2.57***	2.52***	2.19***	2.18***
	(0.39)	(0.39)	(0.65)	(0.65)
Faster Student × Peer is friend	-1.05**			-1.59*
	(0.51)			(0.82)
Slower Student × Peer is friend	0.22			-0.52
	(0.66)			(0.82)
Faster Student × Peer is friend (undirected)		-1.47**		
		(0.61)		
Slower Student × Peer is friend (undirected)		0.23		
		(0.84)		
Faster Student × Peer is friend (reciprocal)			-0.55	0.73
			(0.60)	(0.93)
Slower Student × Peer is friend (reciprocal)			0.43	0.78
			(0.52)	(0.66)
Faster Student × $\|\Delta Time\ 1\|$	-0.42***	-0.40***	-0.35**	-0.35**
	(0.14)	(0.14)	(0.16)	(0.15)
Slower Student × $\|\Delta Time\ 1\|$	0.98***	0.98***	1.06***	1.06***
	(0.19)	(0.19)	(0.19)	(0.20)
Slower Student in Pair	-0.16	-0.43	-0.03	-0.23
	(0.66)	(0.72)	(0.67)	(0.67)
Fixed effects	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes
Peer Characteristics	Yes	Yes	Yes	Yes
Abs. Diff. in Characteristics	Yes	Yes	Yes	Yes
Match quality and performance differences	Yes	Yes	Yes	Yes
N	588	588	588	588
R^2	.33	.33	.28	.29

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) presents the last specification of Table 3 for reference using directed friendship ties. Column (2) uses undirected friendship ties, column (3) reciprocal directed friendship ties, while column (4) allows for a differential effect of directed and reciprocal friendship ties. High match quality is an indicator that equals one if the partner was ranked within an individual's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and the abs. diff. in personality includes the corresponding characteristics of th peer and the respective absolute difference. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table E.7: Decomposition of treatment effects using time in second run

	Time (Second Run; in sec.)				
	(1) Baseline	(2) Match Quality	(3) Friend-ship ties	(4) Time Difference	(5) All
Direct Effects					
NAME	-0.46***	-0.36**	-0.41***	-0.37***	-0.35**
	(0.09)	(0.15)	(0.13)	(0.12)	(0.14)
PERFORMANCE	-0.49***	-0.39***	-0.38**	-0.42***	-0.49***
	(0.07)	(0.14)	(0.15)	(0.14)	(0.15)
Peer Characteristics					
Faster Student × High match quality (NAME)		0.08			-0.05
		(0.11)			(0.13)
Slower Student × High match quality (NAME)		-0.09			-0.13
		(0.18)			(0.18)
Faster Student × High match quality (PERF.)		-0.28*			-0.14
		(0.14)			(0.14)
Slower Student × High match quality (PERF.)		0.29			0.17
		(0.18)			(0.19)
Faster Student × Peer is Friend			0.23*		0.29*
			(0.12)		(0.15)
Slower Student × Peer is Friend			0.01		0.02
			(0.17)		(0.18)
Faster Student × $\lvert\Delta Time\ 1\rvert$				0.08**	0.06
				(0.04)	(0.04)
Slower Student × $\lvert\Delta Time\ 1\rvert$				-0.16**	-0.15*
				(0.07)	(0.08)
Slower Student in Pair		-0.39**	-0.08	0.15	0.14
		(0.16)	(0.13)	(0.13)	(0.21)
Own Characteristics					
Time (First Run)	0.67***	0.70***	0.68***	0.75***	0.75***
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)
Age (standardized)	-0.09	-0.11	-0.12	-0.12	-0.11
	(0.12)	(0.11)	(0.11)	(0.11)	(0.11)
Fixed effects	Yes	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes	Yes
Peer Characteristics	No	No	No	No	Yes
Abs. Diff. in Characteristics	No	No	No	No	Yes
p-value: NAME = PERF.	0.83	0.86	0.86	0.75	0.37
N	588	588	588	588	588
R^2	0.82	0.81	0.81	0.81	0.82

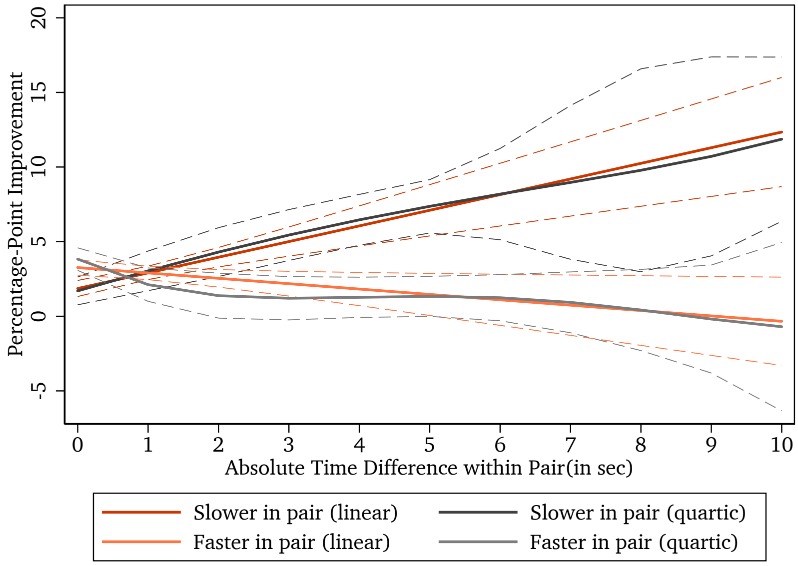
Notes: This table presents least squares regressions according to equation (2) using time in the second run as the dependent variable. High match quality is an indicator that equals one if the partner was ranked within an individual's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and the abs. diff. in personality includes the corresponding characteristics of th peer and the respective absolute difference. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table E.8: Robustness checks for absolute time differences

	Percentage-Point Improvements		
	(1) Linear	(2) Quartic	(3) FEs
Direct Effects			
Name	1.64***	1.67***	1.57***
	(0.40)	(0.39)	(0.42)
Performance	2.57***	2.61***	2.72***
	(0.39)	(0.39)	(0.43)
Faster Student × $\lvert\Delta Time\ 1\rvert$	-0.42***	-2.34*	
	(0.14)	(1.23)	
Slower Student × $\lvert\Delta Time\ 1\rvert$	0.98***	1.47	
	(0.19)	(1.74)	
Slower Student in Pair	-0.16	-1.79*	
	(0.66)	(0.90)	
Faster Student × $\lvert\Delta Time\ 1\rvert^2$		0.73	
		(0.53)	
Slower Student × $\lvert\Delta Time\ 1\rvert^2$		-0.06	
		(0.95)	
Faster Student × $\lvert\Delta Time\ 1\rvert^3$		-0.09	
		(0.08)	
Slower Student × $\lvert\Delta Time\ 1\rvert^3$		-0.00	
		(0.18)	
Faster Student × $\lvert\Delta Time\ 1\rvert^4$		0.00	
		(0.00)	
Slower Student × $\lvert\Delta Time\ 1\rvert^4$		0.00	
		(0.01)	
Fixed effects	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes
Peer Characteristics	Yes	Yes	Yes
Abs. Diff. in Characteristics	Yes	Yes	Yes
Time Diff. FEs	No	No	Yes
Match Quality and Friendship Ties	Yes	Yes	Yes
p-value: Name = Perf.	0.06	0.05	0.03
N	588	588	588
R^2	0.33	0.34	0.34

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Column (1) presents the last specification of Table 3 for reference. Column (2) includes polynomials of time differences in the first run and column (3) fixed effects for every one-second difference in productivity levels of the two subjects. Match quality controls for an indicator if the partner was ranked within an individual's first three preferences. Faster student is an indicator based on the performance in the first run. Own characteristics include the Big Five, locus of control, social comparison, competitiveness, and risk attitudes. Peer characteristics and the abs. diff. in personality includes the corresponding characteristics of th peer and the respective absolute difference. Fixed effects include school-by-grade and gender-by-grade fixed effects. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Figure E.3: Robustness of linear specification in time differences



Notes: The figure presents marginal effects (solid lines) from a least squares regression using percentage-point improvements as the dependent variable including 95% confidence intervals (dashed lines). It plots the linear specification (black lines) as used in the main text as well as a second specification using quartic polynomials (orange lines) of absolute time differences in the first run as regressors. We use the same set of controls as in column (5) of Table 3 and cluster standard errors at the class level. The corresponding regressions are presented in columns (1) and (2) of Appendix Table E.8.

Estimating peer effects on Random only does not affect our results. One might worry that estimating peer effects on all three treatments jointly may bias our estimates. We therefore perform the following alternative estimation strategy. We estimate the peer coefficients on the subsample of students in Random, resulting in unbiased estimates due to random assignment of peers. In a second step, we then impose these estimates on the two treatments featuring self-selected peers when estimating our main specification. Essentially, we are calculating the predicted performance improvements and compare them to the realized improvements to recover the direct effects of self-selection. Appendix Table E.9 shows that imposing peer effects from Random on the other two treatments does not change our conclusions. In particular, the direct effects remain significant, although slightly lower for the NAME treatment.

Table E.9: Restricting coefficients of peer characteristics

	Percentage-Point Improvements			
	Fixing only FEs		Fixing FEs & own char.	
	(1) only RANDOM	(2) all	(3) only RANDOM	(4) all
Direct Effects				
NAME		1.28*		1.34*
		(0.36)		(0.35)
PERFORMANCE		2.08**		2.10**
		(0.45)		(0.43)
Peer Characteristics				
Faster Student × High match quality (NAME)	0.55	0.55	0.54	0.54
	(0.79)		(0.69)	
Slower Student × High match quality (NAME)	-0.48	-0.48	-0.41	-0.41
	(1.16)		(1.10)	
Faster Student × High match quality (PERF.)	0.35	0.35	0.06	0.06
	(0.95)		(0.92)	
Slower Student × High match quality (PERF.)	-0.58	-0.58	-0.36	-0.36
	(1.14)		(1.24)	
Faster Student × Peer is friend	-0.12	-0.12	-0.16	-0.16
	(0.77)		(0.60)	
Slower Student × Peer is friend	0.16	0.16	0.03	0.03
	(1.28)		(1.13)	
Faster Student × $\lvert \Delta Time\ \ 1 \rvert$	-0.60*	-0.60	-0.57	-0.57
	(0.25)		(0.23)	
Slower Student × $\lvert \Delta Time\ \ 1 \rvert$	0.67**	0.67	0.72**	0.72
	(0.33)		(0.31)	
Slower Student in Pair	0.30	0.30	0.21	0.21
	(1.10)		(0.96)	
Own Characteristics	No	No	Yes	Yes
Peer Characteristics	Yes	Yes	Yes	Yes
Abs. Diff. in Characteristics	Yes	Yes	Yes	Yes
N	206	588	206	588
R^2	.25		.23	

Notes: This table presents least squares regressions using percentage-point improvements as the dependent variable. Own and peer characteristics include the Big Five, locus of control, social comparison, competitiveness and risk attitudes. Absolute differences in personality include the difference in those. We use residualized dependent and independent variables, where we take out the variation of individual-specific variables. The first two columns take out the variation of the set of fixed effects, while the last two columns additionally take out variation of own characteristics. Columns (1) and (3) present least squares regressions in RANDOM only, while columns (2) and (4) use all three treatments, but restrict the coefficients to equal the preceding columns. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Direct effects are robust to additional class-level controls and are not an artifact of over-fitting controls. In order to further probe the robustness of our findings, Table E.10 provides results from robustness checks that additional include a set of variables capturing the atmosphere within each class as reported by the teachers. While the estimates slightly differ in magnitude, the results are generally robust.

In addition, one might be worried that adding a series of own and peer characteristics results in over-fitting driving our results. In the second column of Table E.10 we therefore adopt a post-double selection (PDS) Lasso method described by Belloni, Chernozhukov, and Hansen (2014) that penalizes control variables, but allows for inference on treatment indicators. Our results are robust to this data-driven selection of control and peer variables.

Omitted Coefficients from Table 3 column (5). Table E.11 presents the omitted coefficients from our main specification in column (5) of Table 3.

Table E.10: Additional robustness checks using class-level controls and post-double selection Lasso

	Percentage-Point Improvements		Time in the second run (sec.)	
	(1) Class Con.	(2) PDS Lasso	(3) Class Con.	(4) PDS Lasso
Direct Effects				
Name	1.68***	1.36***	-0.43***	-0.39***
	(0.40)	(0.42)	(0.11)	(0.11)
Performance	2.18***	2.35***	-0.50***	-0.49***
	(0.40)	(0.37)	(0.09)	(0.07)
Peer Characteristics				
Faster Student × High match quality (Name)	0.55	0.68	-0.04	0.10
	(0.46)	(0.44)	(0.14)	(0.12)
Slower Student × High match quality (Name)	0.44	0.64	-0.11	-0.15
	(0.80)	(0.68)	(0.22)	(0.21)
Faster Student × High match quality (Perf.)	0.08		-0.02	
	(0.57)		(0.15)	
Slower Student × High match quality (Perf.)	-1.20	-0.65	0.30	0.18
	(0.77)	(0.61)	(0.21)	(0.14)
Faster Student × Peer is Friend	-0.85*	-1.08**	0.24*	
	(0.47)	(0.49)	(0.14)	
Slower Student × Peer is Friend	0.55	0.27	-0.08	-0.11
	(0.78)	(0.62)	(0.21)	(0.17)
Faster Student × $\lvert \Delta Time\ 1 \rvert$	-0.45***	-0.43***	0.09**	
	(0.15)	(0.13)	(0.04)	
Slower Student × $\lvert \Delta Time\ 1 \rvert$	0.78***	0.94***	-0.11	
	(0.19)	(0.21)	(0.08)	
Slower Student in Pair	0.10	-0.47	0.07	
	(0.71)	(0.61)	(0.22)	
Time (First Run)			0.79***	0.68***
			(0.06)	(0.05)
Fixed effects	Yes	Yes	Yes	Yes
Own Characteristics	Yes	Yes	Yes	Yes
Peer Characteristics	Yes	Yes	Yes	Yes
Abs. Diff. in Characteristics	Yes	No	Yes	No
Class-level Controls	Yes	No	Yes	No
p-value: Name = Perf.	0.27	0.05	0.53	0.48
N	518	588	518	588
R^2	0.34		0.85	

Notes: This table presents least squares regressions using percentage-point improvements (columns (1) and (2)) and times in the second run (columns (3) and (4)) as the dependent variable, and a set of class-level controls capturing the atmosphere within a class (missing for some classes) and results from the post-double selection (PDS) Lasso method by Belloni, Chernozhukov, and Hansen (2014). The PDS Lasso always includes the treatment indicators and grade-by-school as well as grade-by-gender fixed effects and penalizes the remaining control variables to avoid over-fitting. Standard errors in this specification are only valid for the treatment indicators and fixed effects. See Belloni, Chernozhukov, and Hansen (2014) for further details. Own and peer characteristics include the Big Five, locus of control, social comparison, competitiveness and risk attitudes. Absolute differences in personality include the difference in those (note that the PDS Lasso forces some coefficients such as all absolute differences in personality measures to zero). *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

Table E.11: Omitted Coefficients from Table 3 column (5)

	Own characteristics	Peer characteristics	Abs. Diff in characteristics
Agreeableness	0.13	-0.08	-0.42***
	(0.19)	(0.19)	(0.14)
Conscientiousness	0.00	0.20	0.98***
	(0.20)	(0.17)	(0.19)
Extraversion	0.01	0.03	-0.16
	(0.22)	(0.19)	(0.66)
Openness to Experience	-0.54***	-0.25	-1.05**
	(0.18)	(0.17)	(0.51)
Neuroticism	-0.05	-0.04	0.22
	(0.23)	(0.20)	(0.66)
Locus of Control	0.27	0.21	-0.38
	(0.21)	(0.21)	(0.29)
Social Comparison	0.36*	0.21	0.20
	(0.18)	(0.16)	(0.26)
Competitiveness	-0.14	-0.41**	-0.27
	(0.27)	(0.19)	(0.21)
Risk Attitudes	-0.02	-0.01	-0.51**
	(0.17)	(0.16)	(0.25)

Notes: This table presents omitted coefficients from Table 3 in the main text. Columns (1) and (2) show the coefficients on own and peer characteristics, respectively. Column (3) presents the coefficients on the absolute differences in personality measures. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

E.3 Additional Material: Discussion of direct effects

Table E.12 presents regressions to support section 5.2.3's discussion of the psychological effect underlying the direct effects. First, we show that subjects in RANDOM are not disappointed by having a partner assigned. If they were disappointed, they should have less fun during the second run. As column (1) show this is not the case. Second, we do not find evidence that subjects with self-selected perceive winning in the second run as more important as we do not see a differential effect on fun between being faster or slower in the second run.

Table E.12: Potential psychological mechanisms for the direct effect

	Fun (std.).	
	(1)	(2)
Direct Effects		
NAME	-0.04	-0.01
	(0.12)	(0.15)
PERFORMANCE	-0.11	-0.07
	(0.08)	(0.13)
NAME × Slower Student in Pair (2nd Run)		-0.07
		(0.19)
PERFORMANCE × Slower Student in Pair (2nd Run)		-0.08
		(0.17)
Peer Characteristics		
Faster Student (2nd Run) × $\|\Delta Time\ 2\|$	-0.01	-0.01
	(0.05)	(0.05)
Slower Student (2nd Run) × $\|\Delta Time\ 2\|$	-0.15***	-0.16***
	(0.05)	(0.05)
Slower Student in Pair (2nd Run)	0.04	0.08
	(0.18)	(0.20)
Fixed effects	Yes	Yes
Own Characteristics	Yes	Yes
Peer Characteristics	Yes	Yes
Abs. Diff. in Characteristics	Yes	Yes
Match quality	Yes	Yes
Friendship indicators	Yes	Yes
p-value: NAME = PERF.	0.62	0.75
N	588	588
R^2	0.33	0.33

Notes: This table presents least squares regressions using a standardized measure of fun in the second run as the dependent variable. Column (2) uses the full specification of Table 3 and additionally interacts the treatment indicators with one's own measure of agreeableness as a proxy of prosociality. Fun was elicited after the second run ("How much fun did you have during the second run? Please rate this on a scale from 1—no fun at all—to 5—a lot of fun.") and uses the full specification of Table 3 adapted using times and ranks from the second run. Column (2) additionally interacts treatment indicators with the final rank in the second run. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level.

F Additional Material: Self-selection vs. exogenous assignment rules: Limits and consequences

F.1 Simulation of matching rules

We simulate three matching rules and predict their impact on performance improvements using our estimates from Table 3. In a first step, we create artificial pairs, based on the employed matching rules described below. In a second step, we then calculate the vector θ of differences for the artificial pairs as well as the matching quality of artificial peers. Finally, we use the estimated coefficients from the column (5) of Table 3 to predict the performance improvements we would observe for the artificial pairs. As peer assignment rules only change θ, we are interested in the difference in the respective sums of the indirect effect and direct effect, that is between $\bar{\tau} + \beta\theta_i^{sim}$ and $\bar{\tau} + \beta\theta_i^{obs}$ from equation (2), where *sim* and *obs* denote simulated and observed pair characteristics, respectively. As we consider exogenous assignment rules, we assume that the direct effect of the simulated policies equals zero as in in RANDOM. We additionally fix the covariates X to 0 and leave out the fixed effects for the simulations and predictions. This means, we calculate the performance improvements for a particular baseline group for our treatments as well as the simulations. This enables us to compare our results of the simulations directly to the peer assignment rules using self-selection implemented in the experiment, as we compare the performance improvements for the same group.

In addition to our three treatments, we simulate four types of peer assignment rules. First, we simulate two settings in which we assign the self-selected peers exogenously (NAME (EXOG.) and PERFORMANCE (EXOG.)). Hence, the resulting pairs are the same as in the self-selection treatment, but we exclude the direct effect of self-selection. Second, we implement an ability tracking assignment rule, TRACKING, in the spirit of the matching also employed in Gneezy and Rustichini (2004). Students are matched in pairs, starting with the two fastest students in a matching group and moving down the ranking subsequently. This rule minimizes the absolute distance in pairs. Third, we employ a peer assignment rule that fixes the distance in ranks for all pairs (EQUIDISTANCE). We rank all students in a matching group and match the first student with the one in the middle and so forth. More specifically, if G denotes the group size, the distance in ranks is $G/2 - 1$ for all pairs. This rule is one way to maximize the sum of absolute differences in pairs, but keeps the distance across pairs similarly. Fourth, we match the highest ranked student with the lowest one, the second highest ranked with the second lowest one and so forth (HIGH-TO-LOW). This is similar to Carrell, Sacerdote, and West (2013), who match low-ability students with those students from whom they would benefit the most (i.e., the fastest students). Again, this assignment rule maximizes the sum of absolute differences in pairs. Table F.1 summarizes initial performance differences within pairs of the experimental treatments as well as the simulated assignment rules and the predicted performance improvements.

Table F.1: Overview of simulated peer assignment rules

Peer assignment rule	Mean absolut productivity differences (in sec)	Predicted improvement (in pp.)	Description
Self-selection of peers			
Name	2.09	2.37	Self-selected peers based on names
Performance	1.41	2.64	Self-selected peers based on relative performance
Exogeneous peer assignment			
Name (exog.)	2.09	1.17	Self-selected peers based on names without self-selection effect
Performance (exog.)	1.41	0.48	Self-selected peers based on relative performance without self-selection effect
Random	2.42	1.08	Randomly assigned peers
Equidistance	3.11	1.39	Same distance in ranks across pairs
High-to-Low	3.11	1.31	First to last, second to second to last etc.
Tracking	0.90	0.71	First to second, third to fourth etc.

F.2 Implications and consequences of self-selection and exogenous peer assignment

Our treatments also have implications for individual ranks of students within a class since slower subjects improve more than faster ones. As ranks are important in determining subsequent outcomes (Elsner and Isphording, 2017; Gill et al., 2019; Murphy and F. Weinhardt, 2020), a policy maker has to take the distributional effects of peer assignment mechanisms into account.[1] Since low-ability students improve relatively more than high-ability students in Name and Random, these treatments yield potentially large changes of a student's rank within the class between the two runs. By contrast, Performance will tend to preserve the ranking of the first run as improvements are distributed more equally relative to the two other treatments. We confirm this intuition in Table F.2 in which we regress the absolute change in percentile scores from the first to the second run on treatment indicators. The outcome

[1]Suppose that a policy maker wants to establish a rank distribution (ranks based on times in the second run) that mirrors the ability distribution (ranks based on times in the first run) due to some underlying fairness ideal (e.g., she wants to shift the distribution holding constant individual ranks). In other words, she might want to implement a peer assignment mechanism that preserves individual ranks rather than shuffle them.

variable measures the average perturbation of ranks within in a class across the two runs. The results show that PERFORMANCE shuffles the ranks of students less in comparison to RANDOM and NAME. While in RANDOM students change their position by about 15 out of 100 ranks, we find significantly less changes in the percentile score in PERFORMANCE relative to RANDOM. This change corresponds to a 27% reduction in reshuffling. However, in NAME we do not find any effect compared to RANDOM.

As another side effect we consider the pressure subjects experienced during the second run due to their peer. We find that in PERFORMANCE subjects experience significantly more pressure than subjects in the other two treatments. To check if self-selection does increase the share of subjects performing at insufficient levels, we compare if the performance of subjects in the second run falls below the 10th percentile (on grade-by-school level) of performance in the first run. Column (4) of Appendix Table F.2 shows that this share slightly increases with self-selection, but only significantly so in Performance. However, the lower part of Appendix Table F.2 shows that neither overall inequality (measured by the Gini coefficient of performance within each matching group), nor inequality in the lower or upper part of the distribution (using the ratios of the performances at the 50th and 10th percentile or the 90th to 50th percentile) differ substantially across the three peer assignment mechanisms.

Table F.2: Side effects of reassignment rules

	Absolute Change in Percentile Scores		Pressure (std.)	Prob. of slow perf.
	(1) within matching group	(2) within treatment	(3)	(4)
Name	-0.0062	-0.0199	0.1111	0.0595
	(0.0142)	(0.0122)	(0.2043)	(0.0408)
Performance	-0.0355**	-0.0358**	0.4533**	0.0902***
	(0.0158)	(0.0141)	(0.1534)	(0.0318)
Fixed effects	Yes	Yes	Yes	Yes
Other controls	No	No	Yes	Yes
p-value: Name vs. Performance	.052	.25	.22	.46
N	588	588	165	588
R^2	.061	.06	.31	.13
Mean in Random	.15	.14	-.16	.15

	Inequality measures		
	(1) Gini	(2) 50/10 ratio	(3) 90/10 ratio
Name	-0.0017	0.0079	-0.0019
	(0.0036)	(0.0092)	(0.0158)
Performance	0.0049	-0.0045	0.0209
	(0.0039)	(0.0082)	(0.0186)
Fixed effects	Yes	Yes	Yes
p-value: Name vs. Performance	.047	.23	.16
N	70	70	70
R^2	.3	.27	.18
Mean in Random	.038	.92	1.1

Notes: This table presents least squares regressions using absolute change in percentile scores, a standardized measure of pressure during the second run, or an indicator for low Performance as the dependent variable. *, **, and *** denote significance at the 10, 5, and 1 percent level. Standard errors in parentheses and clustered at the class level. Absolute changes in percentile scores within matching groups are calculated based on the change of individual ranks of subjects in the their class and gender from the first to the second. Percentile scores within treatment are calculated for all subjects within the same treatment and gender (i.e., across classrooms). Other controls include the same controls as the mediation model in Table 3, where we use times and ranks from the second rather than the first run as the pressure variable has been elicited after the second run. Note that information on pressure was only elicited at one of the three schools.