Do Requests to Sign a No-Cheating Declaration Affect Academic Cheating?

Tobias Cagala, Ulrich Glogowsky, Johannes Rincke

May 26, 2020

Abstract

Educators around the globe use academic integrity pledges to prevent students from cheating. This paper evaluates the effectiveness of this policy by testing how the request to sign a no-cheating declaration affects academic cheating. Exploiting a field experiment with undergraduate students, we identify cheating by comparing the similarity in multiple-choice answers of seat neighbors and counterfactual neighbors. We find that students plagiarize more after having signed a no-cheating declaration. This effect is driven by students of below-average ability. Regarding channels, complementary evidence suggests that the request to sign a no-cheating rule weakens the social norm of academic integrity.

Keywords: academic cheating; no-cheating rule; social norm; randomization inference; field experiment
1 Introduction

Academic cheating is a wasteful illicit activity. It distorts the incentives for students to invest in their human capital, threatens the usefulness of certificates as quality signals, and, hence, undermines the efficiency of the job-matching process (Spence, 1973). Despite being harmful, academic cheating is also widespread. In surveys, between 42% and 64% of participants stated that they had cheated in college at least once (Davis and Ludvigson, 1995). Moreover, the Center for Academic Integrity at Duke University reports that, between 2002 and 2005, 21% of undergraduates admitted to having cheated on exams at least once a year (McCabe, 2005).

Given that academic cheating is harmful and prevalent, educators around the world have developed various countermeasures. One widely used instrument involves academic integrity pledges of various forms. For example, when handing in assignments, term papers, or theses, students in many countries are commonly requested to sign a no-cheating declaration. Despite being commonly used, however, there is little evidence on the causal impact of such academic integrity pledges.

In this study, we provide insights on how signed no-cheating declarations affect plagiarism in exams. The evidence originates from a field experiment in written multiple-choice exams of undergraduate courses at a German university. All students in the experiment were subject to the same monitoring conditions; they also faced the same no-cheating rule that the supervisors publicly announced before the exam. As for the experimental variation, we randomly allocated students to a control group and a signature treatment group. In the signature treatment, before the beginning of the exam, students were required to sign a declaration of compliance with the no-cheating rule. By contrast, the control group featured the university’s standard exam conditions without any form of a pledge to academic integrity.

---

1 For further evidence on the prevalence of academic cheating, see, e.g., Schab (1991) and Davis et al. (1992). Because of social-desirability biases or because subjects may not understand the principles of academic integrity, self-reported data can be biased (Power, 2009; Dee and Jacob, 2012).

2 A well-known form of a pledge to academic integrity is implied by the honor code system. All of the top 10 U.S. universities according to the U.S. News & World Report 2019 have an honor code or code of conduct that explicitly refers to academic integrity, and four out of the ten require undergraduate students to sign or to pledge adherence to this code. While many practitioners question the usefulness of honor codes (Cheung, 2012), descriptive work comparing self-reported academic cheating across institutions found that cheating tends to be lower at honor-code institutions (Bowers, 1964; McCabe and Trevino, 1993; McCabe et al., 2001). To the best of our knowledge, the causal effect of honor codes on academic cheating has not been analyzed.

3 We also implemented a monitoring treatment that imposed close monitoring of students during the exam (but no signature of a no-cheating declaration). The purpose of the monitoring treatment was twofold: First, under the assumption that students could cheat less
Because academic cheating is a hidden activity, the first step of our analysis is to develop techniques that make cheating observable. We focus on plagiarism of multiple-choice answers, as copying solutions from seat neighbors leaves identifiable traces in the data. In particular, if students plagiarize, we expect that the similarities in seat neighbors’ answers are higher than in a counterfactual situation without any cheating and only randomly occurring similarities. In practice, the similarity in answers under the counterfactual scenario is, of course, not observable. However, we can approximate such a counterfactual scenario by considering the similarity in the answers of counterfactual neighbors (i.e., students who were not sitting side by side and, hence, could not copy from each other). Counterfactual neighbors serve as a viable control group because we randomly assigned students to seats. This feature eliminates other than cheating-related differences in the similarity of actual and counterfactual neighbors’ answers. Hence, a comparison of the similarity in the answers of actual and counterfactual neighbors solves the “reflection problem” (Manski, 1993) and identifies the amount of cheating. We exploit this basic idea in two types of tests that both simulate a large number of counterfactual neighbors: Treatment-specific non-parametric randomization tests and regression-based tests.

Our main results are as follows. First, using the treatment-specific randomization test, we document that in both the signature treatment and the control group, students plagiarize by copying answers from their neighbors: The similarity in neighbors’ answers is significantly higher than in the answers of counterfactual neighbors. Second, exploiting the regression-based test, we demonstrate that cheating among low-ability students (i.e., students with poor high-school GPAs) explains the above-counterfactual similarity. Third, we exploit the regression-based test to evaluate the effect of the request to sign a no-cheating declaration. Our results demonstrate that the above-normal similarity in answers among neighbors is significantly higher in the signature treatment than in the control group, implying that students plagiarize more in response to the request.

To gather suggestive evidence on channels, we repeated the experiment with later cohorts of freshmen and conducted a post-exam survey. The survey elicits the students’ perceptions of detection probabilities, sanctions, and the social norm of academic integrity. Our findings suggest that the signature treatment weakens the perceived social norm of academic integrity. By contrast, perceived detection probabilities and perceived sanctions are not affected by the treatment.

under close monitoring, the treatment allows us to substantiate that our methods identify cheating. Second, in a different paper, we used the monitoring treatment to identify intertemporal spillovers of monitoring (Cagala et al., 2014). In the main part of this paper, we focus on the control group and the signature treatment. We refer the reader to the Online Appendix for a full collection of results for the monitoring treatment.
Our work relates to two strands of literature. First, we extend the literature studying causal effects on academic cheating. This literature shows that social interactions amplify academic cheating (Lucifora and Tonello, 2015) and that higher levels of cheating among peers lead to a higher probability that an individual cheats (Carrell et al., 2008). These findings are in line with our finding that the perverse effect of letting students sign a no-cheating declaration likely works through the perceived social norm of academic integrity. Other causal studies demonstrate that classroom cheating responds to monetary incentives (Jacob and Levitt, 2003; Martinelli et al., 2018). Furthermore, anti-plagiarism tutorials reduce plagiarism in term papers (Dee and Jacob, 2012), and close monitoring eliminates plagiarism in exams (Levitt and Lin, 2015).

Second, our study is linked to literature, mostly from psychology, discussing the effects of oaths, moral reminders, or commitment requests on cheating behavior. These studies typically explore the effects of such interventions on misreporting in laboratory cheating games along the lines of Fischbacher and Föllmi-Heusi (2013). The literature suggests that primarily interventions that confront individuals with morally-charged information tend to reduce cheating (Mazar et al., 2008; Jacquemet et al., 2018; Cagala et al., 2019), although one of the main findings supporting this view has recently come under attack (Verschuere et al., 2018).

While there is, hence, some experimental work from the laboratory on the effectiveness of pledges to comply with no-cheating rules, there is very little evidence from the field, let alone in education-related contexts. Particularly, to the best of our knowledge, there are only two related field-experimental studies from other contexts: Shu et al. (2012) indicate that signatures a no-cheating declaration at the beginning rather than at the end of an insurance self-report increases honesty. By contrast, the Behavioural Insights Team (2012) reports that moving a no-cheating declaration from the bottom to the top of a form to apply for a tax discount may have increased fraud. Thus, it seems as if similar interventions can unfold diverging effects, depending on the context studied. In summary, the literature offers little guidance on whether and how to use pledges to no-cheating declarations to fight academic dishonesty. This observation sets the stage for our paper. Using random assignments to treatments and an objective measure of cheating, we extend both literature strands by studying how requesting students to sign a no-cheating declaration causally affects academic honesty.

The structure of the paper is as follows. Section 2 analyzes the effects of

---

4Recent contributions studying the preference for truth-telling include Kajackaite and Gneezy (2017) and Abeler et al. (2019).
5Verschuere et al. (2018) fail to replicate the finding of Mazar et al. (2008) that reminders of the Ten Commandments reduce misreporting in 19 laboratories, and, hence, question one of the most cited results suggesting that moral reminders are effective.
our intervention on academic cheating, Section 3 discusses possible channels, and Section 4 concludes.

2 The Field Experiment

2.1 Experimental Design

We implemented the field experiment in two written, 60-minute undergraduate exams at the business school of a German university, both of which took place in several lecture halls. The exams covered “principles of economics” (first exam) and “principles of business administration” (second exam, scheduled one week after the first exam). Both exams were compulsory for students in their first semester and were part of the curriculum for a bachelor’s degree. Because of the focus on first-year students, it is unlikely that students noticed the changes in the examination conditions that we introduced with our treatments. The department’s examination board and the lecturers who were responsible for the exams agreed to all the interventions.

As for the design of the examination questions, each exam included 30 multiple-choice problems consisting of four statements. Only one of the four statements was correct. The students’ task was to mark the correct statements on an answer sheet. All multiple-choice problems had the same weight, and the set of exam questions came in only one version. In a given exam, every student answered the same questions appearing in the same order. Marking an incorrect statement did not lead to a penalty. Hence, the rational strategy was to mark a statement, even if a student did not know the correct answer.

Because we are interested in dishonest behavior, in the following, we discuss general elements of the setting that might have affected the students’ decisions to cheat. A first element that was a likely driver of cheating behavior is the expected punishment in case of detection. According to the department’s exam regulations, students who cheat (e.g., by copying answers from neighbors or using mobile phones) fail the exam. It is also part of the exam regulations that supervisors in exams announce standardized examination rules by reading them aloud (see the Online Appendix for a complete list of announcements made before the beginning of an exam). As part of the announcements, supervisors highlight that cheating is prohibited, and that detected cheaters fail the exam. They also emphasize a list of actions counting as cheating attempts, including copying answers from neighbors,

---

6We collected the exam data by scanning and electronically evaluating the multiple-choice answer sheets. This automated procedure ensures that the data are free from corrector bias and measurement error. We linked the exam data to data on student characteristics obtained from administrative records.
using unauthorized materials, and not switching off mobile phones. In the experiment, we made sure that supervisors in all exam halls read the exact same list of announcements. As a result, we believe it is justified to assume that students were similarly aware of the consequences of detected cheating in all the lecture halls. Section 3 presents evidence in line with this notion.

A second essential element affecting cheating behavior is the monitoring level, as it influences the detection probability in case of cheating attempts. Importantly, the setting we study is one in which the level of monitoring is rather low. Commonly, up to 200 students take exams in lecture halls with up to 800 seats, supervised by only two to four members of the university staff (depending on the size of the hall). Moreover, if a supervisor detects a cheating attempt, this leads to a significant hassle during the exam and to additional paperwork with the department’s examination board after the exam. As a result, the supervising staff has little incentive to monitor students effectively. In fact, the records for the two years before the experiment show that no student failed either of the two exams because of attempted cheating.

A third element that might have affected cheating behavior (in particular, copying from neighbors) is the spatial distance between students (as it determines the students’ cheating abilities). In the experiment, the seating arrangement was as follows: Row-wise, a student was sitting in every second seat (i.e., any two students were separated by an empty seat). Column-wise, students were sitting in every second column (i.e., any two rows with students were separated by an empty seat). The fact that the row-wise distance between two students (1.2 meters on average) was smaller than the column-wise distance (1.8 meters) or the diagonal distance (2.2 meters) suggests that students more likely copied answers from neighbors in the same row than from students sitting in the front or the back. As we demonstrate in Subsection 2.4.2, this is, in fact, precisely the spatial pattern of cheating that we find in our data.

Also of note is that the university does not have an honor code. Furthermore, in the years before the experiment, the department did not request students to sign no-cheating declarations before exams.

### 2.2 Treatment and Control Groups

The field experiment’s main purpose is to test how the request to sign a no-cheating declaration affects cheating in exams. To that end, we randomly allocated students from two strata (gender and high-school GPA as a proxy for ability) to one of two groups: A control condition and a signature treatment. All the students in a given hall received the same treatment. We, thus, exclude spillovers between treatments, which substantiates the stable unit treatment value assumption. We also randomly assigned students to seats in
the lecture halls and made sure they took their preassigned seats.\footnote{We informed students before the exam in which lecture hall they would be seated. When arriving at the hall, they looked up their seat number on a list. Once all students took their seats supervisors checked students’ IDs and ensured that the randomized seating order was put into effect.}

The only difference between the control group and the signature treatment was that students in the signature treatment signed a declaration of compliance with the no-cheating rule. We placed this declaration on the cover sheet of the exam materials (see the Online Appendix for details). It read:

“I hereby declare that I will not use unauthorized materials during the exam. Furthermore, I declare neither to use unauthorized aid from other participants nor to give unauthorized aid to other participants.”

The declaration was printed below a form in which students in all treatments had to fill in their names and university IDs. The salient location was meant to direct the students’ attention to the declaration immediately before the beginning of the exam.\footnote{A post-exam check showed that all the students in the signature treatment had signed the declaration of compliance.} Importantly, students in the control group only had to fill in the form (name and ID), but they were not asked to sign it. Students were given enough time to read and complete the form and sign the declaration before the beginning of the actual exam (i.e., all students had precisely 60 minutes to work on the problems).

To further our understanding of the nature of the intervention, two of its aspects are worth noting. First, by letting students sign the declaration, we changed the degree of commitment to an existing no-cheating rule relative to the control group, but neither varied the existence nor the content of the rule itself. In particular, the declaration did not introduce additional information regarding the rule. Instead, the public announcements, which were identical across treatments, laid out the rules by stating that cheating was prohibited and by highlighting the consequences of cheating. Second, the declaration was not morally loaded but neutral in the sense that it did not refer to any ethical norm.\footnote{We used a neutral declaration because practitioners frequently use this type, both in education-related settings and beyond. We do not claim that morally loaded declarations would have had the same effects. In fact, in Cagala et al. (2019), we report evidence from a laboratory experiment and demonstrate that, depending on the type of declaration used, the request to sign a no-cheating declaration can lead to very different responses.}

We also implemented a second treatment with close monitoring of students. In this treatment, the exam materials were identical to those in the control group. Hence, there was no declaration and no signature request. We use the monitoring treatment to substantiate that our methods can identify cheating. In particular, in the spirit of previous work highlighting that
close monitoring can eliminate academic cheating (Levitt and Lin, 2015), we increased the monitoring intensity in the monitoring treatment to a level that we expected would eliminate plagiarism. In the empirical analysis, we then test whether, as expected, this type of treatment variation nullifies or, at least, sharply reduces the amount of cheating detected by our methods.

We implemented close monitoring as follows: In the monitoring treatment, we allocated additional supervisors to the lecture halls such that, on average, one supervisor monitored only 8.4 students. By contrast, in the control group and the signature treatment, on average, there were 44.2 students per supervisor. Importantly, irrespective of the treatment, supervisors in all halls remained at specific predefined spots throughout the exam. In the control group and the signature treatment, supervisors took positions in the front of the hall. In the monitoring treatment, the spots where supervisors located were evenly distributed all-over the hall. Figure A1 in the Online Appendix provides a stylized illustration of the hall setups under baseline and close monitoring.

2.3 Further Details and Implementation

We took several further steps to guarantee that all examination conditions other than the treatment variations were kept constant across all the lecture halls. First, the supervising staff followed a scripted schedule. The script included the exact wording of all the announcements to be made before and after the exam. Second, we equalized monitoring conditions across halls. This includes measures to ensure that the actual student-per-supervisor ratios in all halls were identical to the planned ones. There were also no asymmetries in the number of empty seats between the treatments that would have altered the cheating opportunities of the participating students in an ex-ante, unknown way. Third, we ensured that all the conditions related to the treatment interventions were unobservable to students before the beginning of the exam. In particular, the supervisors entered the hall and went to their preassigned positions only after all the students took their preassigned seats. As a result, on-the-spot decisions on whether or not to take part in the exam should be uncorrelated with the treatment assignment. Indeed, we do not find any systematic differences in the students’ observable characteristics between treatments. Table 1 demonstrates that all characteristics are balanced across the treatments.

Next, we discuss our sampling scheme. Figure 1 presents an overview.

---

10Due to local examination conditions, students could withdraw from the exam until the exam day. To avoid that no-shows affect the student-per-supervisor ratios, we overbooked lecture halls when randomly allocating students to treatments. Due to the overbooking procedure, some students could not be seated in their preassigned hall. We relegated those students to additional halls that were not part of the experiment.
Our overall sample consisted of 1007 students eligible to take the exams. In the first exam, we randomly assigned 432 students to the control group, 265 to the signature treatment, and 310 to the monitoring treatment. The show-up rates did not vary significantly between the treatment groups and ranged between 73% and 78%. In the end, 766 students took the first exam: 333 in the control group, 208 in the signature treatment, and 225 in the monitoring treatment.

The sampling frame for the second exam used only the 432 students from the first exam’s control group. Hereby, we ensured that all the considered students shared a similar treatment history. We did not implement the monitoring treatment in the second exam. Consequently, we assigned the 432 students to the control group or the signature treatment of the second exam. From the sampled students, 353 took the exam (control: 204; signature: 149).\footnote{Differences in the lecture halls’ capacity prevented us from perfectly equalizing the number of students assigned to the different groups.}

### 2.4 Prevalence and Structure of Cheating

This section describes our approach to make cheating in exams observable. Subsection 2.4 presents a method to measure cheating and continue with applying this technique to our data, treatment by treatment. This analysis will give us a sense of whether individuals cheated at all and will inform us about the structure of cheating in our experiment. Subsection 2.5 then describes how we can exploit the method’s underlying idea to estimate how the request to sign the no-cheating declaration affects cheating.

#### 2.4.1 Basic Idea of Tests for Cheating

Our identification approach starts from the idea that plagiarism (i.e., copying the answers of neighbors) leaves detectable traces in the data. If students plagiarize, the similarities in seat neighbors’ answers are higher than in a counterfactual situation without cheating.\footnote{Figure A2 in the Appendix exemplifies spatial patterns in answers by showing students’ answers to one multiple-choice problem in one particular control-group room.} In practice, however, the counterfactual situation is not observable. We must find ways to approximate how the similarities would look like in the absence of cheating.

For that purpose, we propose two tests: A non-parametric randomization test and a regression-based test. Both tests build on a similar core: They approximate the counterfactual world without cheating by creating many artificial counterfactual neighbor pairs who were not sitting side by side in the exam. Counterfactual neighbors could not plagiarize from each other during the exam, providing us with an approximation of the counterfactual
situation. The tests then explore if the similarity in the answers of actual neighbors is statistically higher than that of the counterfactual neighbors.

The identifying assumption for our tests is that plagiarism is the only systematic reason why the similarity in the answers of actual neighbors differs from that in the answers of counterfactual neighbors. Following Manski’s (Manski (1993)) framework for the identification of social effects, the assumption holds under two conditions. The composition of both types of pairs needs to be identical, and, on top of that, both types of pairs must face an identical institutional environment during the exam.

We took two steps to ensure that both conditions were met. First, to guarantee that there were no systematic differences in the composition of pairs, we randomly assigned individuals to seats and also formed counterfactual pairs randomly. Hence, we followed the standard approach in the social-effects literature and exploited randomization schemes to allocate individuals to groups within which interactions may occur (see, e.g., Sacerdote 2001; Falk and Ichino 2006; Kremer and Levy 2008; Guryan et al. 2009). Second, to ensure that actual and counterfactual neighbors faced the same institutional environment, we only used non-neighbors who either sat in the same lecture hall or even in the same row to construct counterfactual neighbor pairs. This nets out lecture hall and row effects, respectively.

Note that our approach identifies cheating in the form of plagiarism only. Other forms of cheating (like, for instance, using crib sheets) stay undetected by our methods. We, therefore, likely understate the actual incidence of cheating. However, our conclusions will hold as long as the treatment effects are uncorrelated with the cheating technology.

2.4.2 Cheating in Exams

In the following, we identify treatment-specific cheating behavior using a spatial randomization test. Randomization testing goes back to Fisher (1922) and is a standard inference tool in the analysis of experiments. The key characteristic of this type of test is that, instead of assuming that the test statistic follows a standard distribution under the null hypothesis, its distribution is generated from the data by resampling.
**Testing Procedure** Our paper exploits a post-experiment randomization scheme to test against the null hypothesis of no above-normal similarity in the answers of actual neighbors. In particular, we examine whether a measure for the similarity in neighbors’ answers (i.e., the test statistic) is unusually high compared to the distribution of this measure in the absence of cheating. Ex-ante, this distribution is not known. However, if the identifying assumption holds, we can approximate the distribution under no cheating by simulating a large number of artificial seat assignments (i.e., by creating counterfactual pairs of neighbors) and then recalculating what the similarity measure would have been if these assignments had been the real ones.\(^\text{17}\)

Our baseline parameterization of the randomization test is one that (a) identifies plagiarism between direct neighbors who were sitting next to each other in a row and (b) randomly reassigns students within halls. Denoting by \(i\) a student in row \(r\) with a left neighbor \(i - 1\) and a right neighbor \(i + 1\), our testing procedure consists of four steps:

1. Calculate the share of all multiple-choice problems \(s_{i,i-1}\) that \(i\) and \(i - 1\) answered identically (correct or incorrect). Do the same for \(i\) and \(i + 1\) to derive \(s_{i,i+1}\). Compute the treatment-specific test statistic as:

\[
\hat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} (s_{i,i-1} + s_{i,i+1}),
\]

where \(N\) is the number of students in the considered treatment.

2. Create counterfactual neighbor pairs by randomly reassigning students within halls to seats (without replacement). Compute the similarity measure for counterfactual neighbors, \(\hat{\Delta}_{c,m} = 1\).

3. Repeat the previous step \(M\) times. This generates a distribution of \(\hat{\Delta}_{c,m}\) with values \(m = 1, \ldots, M\), mean \(\hat{\mu}_{\Delta_c}\), and standard deviation \(\hat{\sigma}_{\Delta_c}\). Intuitively, this distribution corresponds to the distribution of the test statistic under the null hypothesis of no cheating.

4. Calculate the \(p\)-value of a two-tailed test as twice the probability that a draw from this distribution exceeds \(\hat{\Delta}\).

**Results** Pooling observations from both exams, we report the results of our treatment-specific randomization tests in Figure 2 (\(M = 5000\)). For ease of exposition, the figure reports mean-centered values (i.e., it shows \(\hat{\Delta} - \hat{\mu}_{\Delta_c}\) and \(\hat{\Delta}_{c,m} - \hat{\mu}_{\Delta_c}\)). This type of normalization allows us to interpret the test statistic intuitively as the extent by which the share of identical answers

---

\(^{17}\)Our test follows the educational measurement literature in testing for cheating by examining whether the responses of two students are unusually similar (see, e.g., Holland, 1996; Wollack, 1997, 2003, 2006; Wesolowsky, 2000; Sotaridona and Meijer, 2003; van der Linden and Sotaridona, 2006). However, the standard methods in this literature test for cheating of two specific neighbors which educators suspect of cheating. Our methods, instead, test for plagiarism in a large population of possible pairs of students.
among actual neighbors’ differs from the expected share for counterfactual neighbors (in percentage points).

Panel A in Figure 2 shows the results for the control group, while Panel B focuses on the signature treatment. In each panel, the vertical line depicts the mean-centered test statistic. The bell-shaped curves represent the mean-centered counterfactual distributions under the null hypothesis of no cheating.

Two findings emerge from Figure 2. First, Panel A shows that in the control group, the similarity in the answers of actual neighbors is excessively high compared to the counterfactual distribution. The test statistic indicates that the share in actual neighbors’ identical answers is about 1.4 percentage points higher than the expected share for counterfactual neighbors (who cannot cheat). This value is located in the far right tail of the counterfactual distribution, and we can, consequently, clearly reject the null hypothesis of no above-normal similarity in the answers of actual neighbors (p-value = 0.001). This finding is the first piece of direct evidence that in the control group, students copied answers from students sitting next to them in the same row.

Second, and most importantly, Panel B shows that we can also reject the null hypothesis of no cheating in the signature treatment (p-value < 0.001). The test statistic’s value reveals that the average share of identical answers given by neighbors is about 2.4 percentage points higher than the expected share for counterfactual neighbors. We will discuss the impact of the signature treatment at length in Subsection 2.5, but we can already highlight an intermediate result: In our exams, the request to sign the declaration did not eliminate cheating.

To substantiate that our randomization tests identify plagiarism, we also study spatial correlations in the monitoring treatment. The idea of this validation check is simple: It is natural to assume that close monitoring reduced or even eliminated the students’ options to copy answers from neighbors. Hence, if our randomization test successfully identifies plagiarism, we would expect no above-normal spatial correlations in the monitoring treatment. The latter is what we find: As documented in Figure A3 in the Online Appendix, the test statistic for the monitoring treatment is located in the center of the counterfactual distribution, and we cannot reject the null hypothesis of no above-normal similarity in the answers of actual neighbors (p-value 0.999). In sum, Figure A3 reinforces our confidence that the randomization test identifies cheating.

\footnote{Including the specifications that are part of our robustness checks, we use six different neighbor definitions to test for cheating. To guard against spurious findings from multiple testing, we employ a conservative Bonferroni adjustment to correct the reported p-values.}
Robustness Checks  The randomization tests reported in Figure 2 assume that students only copied answers from neighbors in the same row. If students copied answers from other students sitting farther away, it goes undetected by the specification of the randomization tests. Figures A4 and A5 in the Online Appendix present evidence for a variety of alternative specifications (see Panels A). The figures demonstrate that in our context, copying from other students was, indeed, confined to direct neighbors within rows. Another robustness check resamples individuals within treatments (i.e., also across halls) instead of within lecture halls. The Panels B in the Figures A4 and A5 report the respective results and show that our findings are robust. We also performed the same set of robustness checks for the randomization test in the monitoring treatment. Independent of the specification, we consistently find no evidence for cheating under close monitoring.

2.4.3 Students’ Ability and Cheating

The previous subsection established that actual neighbors shared a suspiciously high number of similar answers under baseline monitoring. We expect to detect especially strong traces for cheating if at least one of the two students is of low ability and was, therefore, less likely to know the solutions to the multiple-choice problems. As shown subsequently, we, indeed, find significant cheating only among low-ability students, in line with studies based on student self-reports (Genereux and McLeod, 1995; McCabe and Trevino, 1997). Notably, the evidence for cheating is confined to pairs in which both students are of low ability, and, hence, give many wrong answers. Given that low-ability students tend to copy wrong answers from each other, the similarity in incorrect answers is the most powerful measure of cheating.

Testing Procedure  We suggest a simple approach to study how cheating depends on ability. To that end, we introduce our second test for cheating, the regression-based approach. We exploit regressions as they allow us to study effect heterogeneity, which is impossible with randomization tests.

Let us start by introducing our baseline model. The model also rests on the idea of using counterfactual neighbors as a control group for actual neighbors.

---

19. We prefer to resample individuals within halls because this resampling scheme decreases the probability of false positives by controlling for potential hall effects. The flipside is that this scheme potentially increases the likelihood of false negatives: The counterfactual distribution could pick up other forms of plagiarism (i.e., non-row-wise plagiarism). A randomization scheme that resamples individuals within treatments takes care of this potential problem.

20. In the monitoring treatment, a worsening of the high-school GPA by one standard deviation is, on average, associated with a decrease in the number of correctly solved multiple choice problems by 0.42 standard deviations.
bors sitting in the same row. Defining pairs of students as the unit of observation, we consider the following simple regression without controls:

$$Y_{mp} = \beta_0 + \beta_p \cdot N_p + u_{mp}, \quad (1)$$

where $Y_{mp}$ takes a value of one if both students of a pair $p$ gave the same answer to a particular multiple-choice problem $m$. Note that $p$ can represent actual and counterfactual pairs. Further, $N_p$ indicates whether a pair of students consisted of actual neighbors sitting next to each other in the same row ($N_p = 1$), or not ($N_p = 0$). Because we randomly assigned students to seats, $E(\beta_p)$ is a consistent reduced-form estimate of the average effect of being a pair of actual neighbors (as opposed to counterfactual ones) on the probability that both students gave the same answer. We call this the average neighbor effect (ANE). An ANE significantly larger than zero indicates cheating.

To test how cheating depends on students’ ability, we extend the model to include students’ final high-school GPA. The average high-school grade reflects a student’s performance in the final years in high school and should, therefore, be a reasonable proxy for ability. Using high-school grades, we can flexibly decompose the neighbor effect into a part that depends on the students’ abilities, and a part that does not. Formally, the decomposition reads:

$$\beta_p = E(\beta_p | B_{i,p}, W_{j,p}) + r_p$$

$$= \sum_{i=1}^{I} \beta_{2i} \cdot B_{i,p} + \sum_{j=1}^{I} \beta_{3j} \cdot W_{j,p} + \sum_{i=1}^{I} \sum_{j=1}^{I} \sum_{j \geq i}^{I} \beta_{4ij} \cdot B_{i,p} \times W_{j,p} + r_p. \quad (2)$$

In this equation, $B_{i,p}$ reflects the high-school grade of the one student of a pair $p$ who performed better in school. Specifically, $B_{i,p}$ consists of four dummy variables indicating whether the student’s high-school grade was $A$, $B$, $C$, or $D$. Equivalently, $W_{i,p}$ are indicators for the high-school grade of the student having performed worse in high school, and $r_p$ denotes further pair-specific heterogeneity of the neighbor effect. Note that by construction, we have $E[r_p | B_{i,p}, W_{j,p}] = 0$.

To estimate the neighbor effects for different $B - W$ combinations, we simply plug (2) into (1) and add the following baseline terms to our model: $\beta_1 \cdot N_p$, $\sum_{i=1}^{I} \alpha_{2i} \cdot B_{i,p}$, $\sum_{j=1}^{I} \alpha_{3j} \cdot W_{j,p}$, and $\sum_{i=1}^{I} \sum_{j=1; j \geq i}^{I} \alpha_{4ij} \cdot B_{i,p} \times W_{j,p}$. Our
specification then becomes:

\[ Y_{mp} = \beta_0 + \left[ \sum_{i=1}^l \beta_{2i} \cdot B_{i,p} + \sum_{j=1}^l \beta_{3j} \cdot W_{j,p} + \sum_{i=1}^l \sum_{j=i}^l \beta_{4ij} \cdot B_{i,p} \cdot W_{j,p} \right] \times N_p + \sum_{i=1}^l \alpha_{2i} \cdot B_{i,p} + \sum_{j=1}^l \alpha_{3j} \cdot W_{j,p} + \sum_{i=1}^l \sum_{j=1}^l \alpha_{4ij} \cdot B_{i,p} \times W_{j,p} + \gamma_0 \cdot E_p + \varepsilon_{mp}, \]

with \( \varepsilon_{mp} = u_{mp} + r_p \cdot N_p \). If \( N_p \) is randomized, the OLS estimators of \( \beta_1, \beta_{2i}, \beta_{3j}, \) and \( \beta_{4ij} \) are unbiased and consistent. The OLS estimators of \( \alpha_{2i}, \alpha_{3j}, \) and \( \alpha_{4ij} \) pick up potential correlations between the grade variables and \( u_{mp} \).

We consider observations in the control group and the signature treatment from both exams and all halls. The regressions also include an exam dummy \( E_p \) to control for potential exam effects \( \gamma_0 \).

Two further details of our regression-based approach are worth noting. First, as previously discussed, we expect the issue of how ability affects cheating to be related to the question of whether cheating translates into an above-normal share of identical correct or identical incorrect answers. We, therefore, estimate two different specifications: One that uses an indicator for identical correct answers as the dependent variable and one that considers identical incorrect answers as the outcome. Second, we construct our estimation sample such that it consists of (a) all pairs of actual direct neighbors in the same row and (b) all pairs of counterfactual neighbors (i.e., non-neighbors) who were sitting in the same row. Hence, our regressions identify the neighbor effects by focusing on within-row variation and comparing actual neighbors in row \( r \) with all the counterfactual pairs of students who were not direct neighbors but sat in the same row. This approach has two benefits. One is that it allows us to cluster standard errors at the row level. This clustering is necessary because our randomization tests show that direct neighbors in the same row plagiarized answers from each other. The other benefit is that this approach indirectly controls for row effects as it compares counterfactual neighbors and actual neighbors within rows. Hence, it is more conservative than our previously considered randomization schemes that did not account for possible row-specific differences in cheating behavior.

**Results** Figure 3 presents the primary results of our linear probability model that either uses identical incorrect answers (Panel A) or identical correct answers (Panel B) as the outcome variable. To construct this figure, we estimate model (1) and subsequently calculate the ANEs (2) for all potential grade

\[ 21 \] For simplicity, the regression equation abstracts from the fact that some of the variables drop out when estimating the model due to collinearities.
combinations. The Panels $A1$ and $B1$ show the resulting $ANEs$ dependent on individuals’ ability. The horizontal axis depicts the ability (measured by the high-school grade) of the worse student and the vertical axis that of the better student. The colors in the graph indicate the size of the $ANEs$, ranging from the smallest neighbor effects (blue) to the largest (red). The Panels $A2$ and $B2$ show the associated $p$-values.

The structure of cheating emerging from Panel $A$ is clear-cut: We only find cheating for pairs in which both students are of low ability. To clarify and elaborate on this conclusion, let us consider Figure 3 in detail. Regarding identical incorrect answers, the estimated neighbor effect (as our indicator of cheating) is the largest for pairs in which both students are of low ability. Moreover, the $p$-values indicate that the effect is significantly different from zero only for pairs located in the upper-right part of the colored area.

To get a sense of the effect size, consider two students whose high-school grade was $C$. Counterfactual pairs of this type gave, on average, identical and incorrect answers to 3.7 percent of all multiple-choice problems. Starting from this baseline probability, our model predicts a positive effect of being a pair of actual neighbors of 2.3 percentage points. In absolute terms, the average number of identical incorrect answers for symmetric counterfactual $C$-grade pairs was 1.1 (out of 30). The neighbor effect adds to this baseline another 0.69 identical incorrect answers (on average). If we, instead, consider symmetric pairs of students with even lower ability (high-school grade: $C$-), the average neighbor effect already amounts to 18 percentage points (baseline probability for counterfactual pairs: 3.8 percent). This value corresponds to an additional 5.4 identical incorrect answers, on average. We conclude that cheating among low-ability students significantly increases the likelihood of identical incorrect answers.

Panel $B$ presents the corresponding results for jointly correct answers. The evidence, again, suggests the interpretation that cheating is mainly driven by neighbor pairs in which both students are of low ability. To see this, note that copying the answers of high-ability students should increase the above-normal similarity in jointly correct answers. However, we find that the grade-specific $ANEs$ are widely insignificant (Panel $B2$). In conclusion, we are unable to detect any significant above-normal similarity in correct answers among actual neighbors.

Overall, Figure 3 generates two insights. First, plagiarism between low-ability students, who happened to be seated next to each other, explains the

---

22 We use a thin-plate-spline interpolation to predict values for finer grade steps (e.g., for $A$, $B+$, etc.).

23 Given that the average probability of selecting a correct answer is 72.4% in the control group, it is mechanically harder to detect cheating when students copy correct answers from neighbors. This feature of the data may partly explain why we find more cheating when considering (a) incorrect answers and (b) low ability students who are more likely to give wrong answers.
above-normal similarity in neighbors’ answers under baseline monitoring. Second, when low-ability students, who tend to give many wrong answers, copy from each other, they naturally share an above-normal number of identical incorrect answers. By contrast, there is no evidence of above-normal similarities in jointly correct answers. We conclude that identical incorrect answers are the more powerful indicator of plagiarism in our context. We, therefore, mainly focus on this measure when evaluating the effects of our intervention.

2.5 Treatment Effects

Building on the evidence presented previously, this section examines our primary topic by analyzing how the request to sign the declaration has affected cheating.

Testing Procedure  The regression-based estimation strategy easily extends to the identification of treatment effects. Instead of estimating grade-combination specific neighbor effects as in equation (2), the following regressions account for treatment-specific neighbor effects. The decomposition of the neighbor effect becomes

$$\beta_p = \beta_1 + \beta_2 \cdot S_p + r_p,$$

where $S_p$ denotes an indicator for the signature treatment. As before, we plug (3) into (1), add $\beta_3 \cdot S_p$ to the model, and obtain:

$$Y_{mp} = \beta_0 + \left[ \beta_1 + \beta_2 \cdot S_p \right] \times N_p + \beta_3 \cdot S_p + \epsilon_{mp},$$

with $\epsilon_{mp} = u_{mp} + r_p \cdot N_p$. We then estimate the coefficients with OLS and cluster the standard errors at the row level. As described previously, the regressions use an indicator for identical incorrect answers as an outcome. As a robustness check, Table A1 in the Online Appendix presents the results for regressions that instead rely on all types of identical answers (correct and incorrect).\(^{24}\)

\(^{24}\)Because cheating is confined to pairs in which both students tend to give wrong answers (see the previous discussion), adding identical correct answers to our outcome introduces noise to the dependent variable and makes the identification of the neighbor effect more difficult. However, models that counteract this efficiency loss by adding covariates confirm our results. Note that the covariates leave the size of the main regression coefficients unchanged.
Results Table 2 reports the results of our linear probability models to explore the impact of the intervention on cheating.\textsuperscript{25} Before discussing our results, we highlight that, if two students independently and randomly picked one of the four statements, then the predicted average probability that they would have shared an incorrect answer is 0.0169.\textsuperscript{26} In comparison, the empirical baseline probability that two counterfactual students in the control group shared an identical incorrect answer was higher than under random picking; it amounted to 0.0364. Against this backdrop, we evaluate the coefficients in Table 2.

Beginning with the unconditional estimates in Column (1), we note three points. First, the coefficient of the non-interacted treatment indicator is not statistically significant. The similarity in the answers of counterfactual pairs in the signature treatment was not significantly different from the control group's baseline level of 0.0364. This result is in line with the interpretation that our experimental design successfully eliminated differences across treatments that might have affected non-cheating related correlations between students' answers.

Second, we identify a positive neighbor effect in the control group. The coefficient for the non-interacted actual-neighbors dummy is positive and significant and amounts to 0.0073. Relative to the baseline probability of 0.0364, being a pair of actual neighbors increased the probability of an identical incorrect answer by 20%. We are, hence, able to replicate the finding of the randomization test that students cheated in the control condition.

Finally, we turn to the main result of the field experiment and evaluate the coefficient of the interaction term $\text{Signature} \times \text{Actual Neighbors}$: It equals 0.0088 and is significantly different from zero. Thus, in the signature treatment, the effect of being a pair of actual neighbors on the likelihood of providing identical incorrect answers increased relative to the control group. The increase is not only statistically significant but also substantial in size. We are unable to reject the hypothesis that the interaction effect is equal to the coefficient of $\text{Actual Neighbors}$ ($F$-Test; $p = 0.779$). Put differently, we cannot reject the hypothesis that signing the no-cheating declaration increased the probability of an identical answer to the extent that mirrors the baseline effect of sitting next to each other.

\textsuperscript{25}The number of observations is derived as follows: Denoting with $K$ the number of individuals in one row $r$, we obtain $\frac{K(K-1)}{2}$ unique pairs for this particular row (of which $K - 1$ are unique actual neighbor pairs). The total number of observations is the sum over all pairs (considering all rows).

\textsuperscript{26}This value incorporates that the average probability that a given multiple-choice question was correct equals 70%. Assuming that students not knowing the correct answer randomly pick one of the four options, the probability for someone not knowing the correct answer selecting one specific answer is $0.3/4$. The probability for two students randomly picking the same answer is therefore $(3/40)^2$. With three out of four options being incorrect, the probability for jointly selecting an incorrect answer is $3(3/40)^2$. 

17
A further question that arises from the previous analysis is whether the increase in cheating in the signature treatment led to better grades. The evidence suggests that this was not the case. Regressing the percentage of problems solved correctly on a dummy for the signature treatment, we consistently find across multiple specifications that the treatment indicator’s coefficient is very small and insignificant.\footnote{Pooling both exams and denoting the treatment indicator by $\beta_1$, we find a value of $\beta_1 = -0.181\%$ ($p$-value=0.866) in a regression without controls and $\beta_1 = -0.066\%$ ($p$-value=0.947) if we add strata variables. The $p$-values are for specifications with row clusters. To see that the effects are negligible in size, note that the average student in the control group answered 72.4\% of all multiple-choice questions correctly. We obtain very similar results if we additionally cluster the standard errors at the individual level or run separate regressions in both exams.} This result is in line with the previously reported evidence that pairs of low-ability students copied incorrect answers from each other, as this form of cheating does not affect the percentage of problems solved correctly.\footnote{We also tested if students who signed the declaration in the first exam were more or less likely to take part in the second exam, and if the extent of cheating in the second exam differed relative to students from the first exam’s control group. We neither found evidence for selective attrition nor spillovers to the second exam.}

**Robustness Checks** The estimations reported in Table 2, Columns (2) to (4), provide several robustness checks. Column (2) controls for multiple-choice fixed effects. Hereby, we partial out problem-specific factors that might affect the degree of similarity in neighbors’ answers (like the difficulty of the question) and identify cheating only from the within-multiple-choice problem variation. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and the worse student as well as interactions). Column (4) includes all the control variables. Because we randomly assigned students to seats, there is no a priori reason to expect the controls to affect the coefficients of interest. Indeed, modifying our regressions along these lines leaves the point estimates virtually unchanged.

In the Online Appendix, we report further versions of our estimations and present additional robustness checks. Table A2 shows that all our results hold is we control for row effects, netting out differences in the similarity of answers that are driven, for instance, by variation in the distance to the front of the hall (where the supervising staff located). Table A3 displays the results of regressions that include an indicator variable for each hall to control for hall-specific differences in the similarity of the students’ answers. Again the findings are virtually unchanged. Our results are also robust against estimating logit models instead of linear probability models (see Table A4 in the On-
Finally, Table A5 pools the data across all groups (control, signature, and monitoring) and reports estimations that also include a neighbor effect for the monitoring treatment. The coefficient of the interaction term Monitoring × Actual Neighbors is negative and statistically significant, suggesting that close monitoring reduced cheating in exams. Furthermore, we cannot reject the hypothesis that the similarity in actual neighbors’ answers under close monitoring was equal to the similarity in the responses of counterfactual neighbors in the control group (F-Test; \( p = 0.533 \)). Both results further strengthen our confidence in the methods we have applied to detect cheating.

3 Suggestive Evidence on Channels

In the previous section, we have shown that the signature treatment increased cheating. To assess the external validity of this finding, it is crucial to understand the mechanisms through which this effect operates (Deaton, 2010; Ludwig et al., 2011; Deaton and Cartwright, 2018). However, providing evidence on channels is always challenging, and in our case particularly difficult. Not only is cheating in exams a type of behavior that is, in principle, difficult to measure, but also the motives that lead individuals to cheat are, by nature, unobservable. Besides, the institutional environment of university exams limits our options to collect data that could shed light on channels. In the following, we, nevertheless, take up the challenge of examining the forces that drive the treatment effect and present evidence on mechanisms. The evidence comes from an additional field experiment that we implemented after the main experiment in two consecutive cohorts of students taking the exam on principles in economics. Given the limitations we face (discussed in more detail below), we consider the evidence as being suggestive rather than entirely conclusive.

3.1 Possible Channels

One can think of at least three channels through which the request to sign a no-cheating declaration could increase cheating. First, the request to sign

---

29 Prompted by the article by King and Zeng (2001), one may wonder whether our estimates are biased because of rare-event data; recall that the share of identical incorrect answers is below four percent. However, the underlying problem that causes a rare-event bias is a small number of cases on the rarer of the two outcomes. Because we have almost 5400 observations for this case, our estimations are not subject to this problem.

30 Equation (3) then becomes \( \beta_p = \beta_1 + \beta_2 \cdot S_p + \beta_3 \cdot M_p + r_p \), where \( M_p \) is an indicator for the monitoring treatment.
the declaration may decrease the perceived severity of the sanctions associated with cheating. One situation in which this type of effect could occur is when students overestimate the sanction in the absence of such a request. The demand to sign the declaration may then direct the students’ attention towards the true level of the sanction, decreasing its perceived severity. Second, the request could also lower the perceived detection probability, for example, by signaling that monitoring students is difficult or even impossible. Third, students could perceive the request as a signal that cheating in exams is widespread, weakening the so-called perceived “descriptive norm” of academic integrity. Students with a preference for conformity with an existing social norm (Bernheim, 1994) would then cheat more. While we believe that the aforementioned mechanisms are the most plausible, we cannot rule out that other channels matter as well.

3.2 Experimental Design

Students’ perceptions of sanctions, detection probabilities, and norms are not directly observable. The standard approach to tackle this issue is to elicit students’ perceptions using survey techniques.

Motivation of New Design Two complications forbid an analysis of perceptions using our initial design. First, due to local exam regulations, we were not allowed to ask survey questions during the exam. After we implemented our initial experiment, the department of economics, however, established an online platform used to invite students to participate in on-

---

31 The literature frequently highlights two types of norms (Lapinski and Rimal, 2005). Injunctive norms reflect people’s perceptions about what should be done. Descriptive norms refer to beliefs about what is actually done by others. The considered channel is, hence, one that runs through descriptive norms.

32 For example, one may wonder if the declaration primed students to think about and engage in cheating. However, the announcements made before the beginning of the exam mentioned cheating repeatedly, and those announcements were identical between the control group and the signature treatment. We therefore consider it unlikely that only students who signed the declaration thought about cheating. In a separate paper, we explore psychological reactance as a further possible channel through which the request to sign a no-cheating declaration can lead to more cheating (Cagala et al., 2019). It is difficult to probe the relevance of this channel in our context. However, one observation from the follow-up exams is in line with reactance: In the signature treatment, an excessive share of students did not provide an email address on the cover sheet of the exam materials (treatment 9.7 percent; control 6.2 percent; p-value (t-test) 0.03). As not providing an email address was inconsequential for the grading process, a possible interpretation of an excessive share of missing email addresses is the “related boomerang effect” (Brehm, 1966). In this interpretation, students in the signature treatment tried to restore their behavioral freedom by violating a rule (do provide email address) related to the rule they were requested to follow (do not cheat).
line surveys. This newly established platform allowed us to invite students who took part in the repetition of the field experiment to an online survey that measured perceptions after the exam. Second, the fact that we had to survey students after the exam complicates identification. To understand the potential issue, consider, for example, perceived norms as a channel. If the signature treatment increases cheating, post-exam questions on norms may reflect that students in the signature-treatment halls observe more cheating than students in control-group halls, instead of reflecting shifts in the perceived norm. This would lead us to overestimate the effect of the request on perceived norms. To tackle this issue, when repeating the experiment, we adjusted the sampling scheme relative to the initial experiment. Particularly, we randomly assigned the signature treatment within lecture halls. This design element ensures that, on average, students in both groups experienced the same level of cheating by peers.

Platform and Survey  We recruited students for the survey through an existing online platform at the department. Students registered with this platform receive email invitations to participate in surveys. Participants get a payoff that is communicated in the invitation email and paid via bank transfer.

Survey Design  A total of 1060 students took part in the follow-up exams. Further, a subsample of 233 students was registered with the online platform before the exam. Two hours after the exams, we invited these students via email to participate in an online survey. Answering the survey took about five minutes, and participants received a flat payoff of €3.50. Students who accepted the invitation were redirected to the welcome page of the online survey. This page informed participants that the survey’s goal was to measure “how students perceive exams at the university.” It also asked participants to think about their last exam when answering the questions. To prevent that students foresaw our goal to study the impact of the no-cheating declaration on their survey responses, we did not refer to the previous exam on principles of economics at any point during the survey.

Survey Questions  Table 3 summarizes the survey questions. First, we elicited the perceived sanction for cheating. Particularly, students indicated their belief about the usual sanction for cheating (they choose one sanction out of a list of five). Second, we measured the perceived detection probability. To that end, we elicited beliefs about how many out of 100 cheating

33 There was no other exam scheduled for freshmen students within two days after the exam in principles in economics.
students would have been caught in their last exam. Third, to obtain a measure for descriptive norms, we included several questions in the questionnaire on the subjects’ beliefs about the percentage of peers who cheated in the last exam. Each question came in two versions. The first version referred to cheating in the form of copying answers from neighbors and the second to the use of unauthorized materials like, for example, a mobile phone.

**Treatment and Control Groups**  When repeating the experiment, we evenly split each cohort of students into a control group and a signature treatment. As discussed before, we randomly assigned the treatment status within exam halls. The signature treatment was identical to the treatment in the initial experiment.

Given this design element, one might worry about spillovers from the signature treatment to the control group. Such spillovers would most likely equalize the outcomes between both conditions and, hence, would tend to downward-bias the estimated effect of the request on students’ survey responses. To prevent this downward bias as much as possible, we kept the layout of the cover sheet of the exam materials identical between the signature treatment and the control group. Particularly, instead of the no-cheating declaration, the control group’s cover sheet contained a text of equal length with technical information on how to handle the exam materials (see the Online Appendix). As a result, the exam materials looked very similar in both groups. Furthermore, both in the signature treatment and the control group, the technical information was also printed on the second page of the exam materials. Hence, all students in the experiment received the same set of technical information.

**Drawbacks**  Our design has two potential drawbacks. First, one obvious limitation is the limited sample size. Not all students are registered users of the survey platform, and survey participation conditional on being registered is voluntary. Our survey sample consists of 103 students who completed the survey within two days after the invitation. Second, our design forbids us to reestimate the effect of the request to sign the no-cheating declaration on cheating. The reason is that the treatment status differs between neighbors. Similarities in their answers, thus, reflect a mix of cheating in both conditions.

**Sample**  Across both cohorts, 1060 students participated in the repetition of the field experiment. Table A7 in the Online Appendix shows that the signature treatment and the control group are balanced in observable character-

---

34 As in the original field experiment, we excluded students who had failed the exam previously and were not taking the exam for the first time.
istics. More importantly, the 103 survey participants have similar observable characteristics as non-participants. The only significant sample imbalance is that participants have high-school GPAs which are 0.19 grade points (or 0.31 standard deviations) better than non-participants (see Table A8 in the Online Appendix).\textsuperscript{35} Furthermore, in the sample of participants, the signature treatment ($N = 48$) and the control group ($N = 55$) are well-balanced in all the observable characteristics (see Table A9 in Online Appendix).

\section*{3.3 Treatment Effects}

**Perceived Sanction**  We start our analysis by studying the perceived sanctions for copying answers or using unauthorized materials. The first result is that independent of the treatment, a vast majority of students correctly indicated that cheaters would fail the exam (copying answers: 71.8%; unauthorized materials: 84.5%). Moreover, Table 4 shows no systematic differences in the perceived sanctions between the signature treatment and the control group. Using Fisher’s exact tests, we cannot reject the null hypothesis that the signature treatment did not affect the distribution of answers. Thus, in sum, we not only conclude that most of the survey participants knew the punishment for non-compliance but also that students who signed the request expected very similar sanctions as control-group individuals.

**Perceived Detection Probability**  Our next step is to study the effects of the signature request on the perceived detection probability. To that end, we use the students’ answers to questions $D1$ and $D2$ (see Table 3) as outcome variables of the OLS model

\begin{equation}
Y_{ih} = \gamma_0 + \gamma_S S_{ih} + X_{ih} \gamma_X + \pi_h + u_{ih},
\end{equation}

where $Y_{ih}$ is the stated perception of student $i$ seated in hall $h$, $S_{ih}$ is an indicator for the signature treatment, and $X_{ih}$ is a vector of student controls (age, gender, and high-school GPA). Additionally, the regression includes exam-hall fixed effects $\pi_h$ to absorb exam-hall specific drivers of perceptions. The result tables also report heteroscedasticity-consistent and cluster-robust $p$-values (clusters at exam hall-level).\textsuperscript{36} In addition to the treatment effects

\textsuperscript{35}To probe the robustness of the survey evidence regarding this type of sample imbalance, we also estimated models that reweigh the individual observations such that the first moments of all characteristics are identical to the population estimates. The results are virtually unchanged to those reported subsequently.

\textsuperscript{36}To derive the latter, we use a wild cluster bootstrap-$t$ procedure that accounts for the small number of clusters (Cameron \textit{et al.}, 2008). Across both repetitions, we implemented the second experiment in 17 halls.

23
for the individual outcomes, we also report average standardized effects according to Kling et al. (2004) and Clingingsmith et al. (2009) and exploit Mann-Whitney-U-Tests to non-parametrically test for treatment differences.

Columns (1) to (3) in Table 5 show that the signature treatment neither shifts the perceived detection probability in case of copying nor the one in case of using unauthorized materials. Taken together with the fact that the average student is well informed about the actual sanction, the absence of a significant treatment effect suggests that the signature treatment did not significantly shift the participants’ expected sanction for cheating.

Descriptive Norms  We next analyze how the signature treatment impacts students’ descriptive norm of academic integrity. Our analysis begins by estimating regressions in the spirit of equation (4) that use the participants’ perceived frequency of cheating as an outcome variable (see questions N1 and N2). The point estimates suggest that compared to control-group individuals, students who signed the no-cheating declaration believe that four to five additional peers (out of 100) plagiarized (see Column (4) in Table 5) or used unauthorized materials (see Column (5)). Only the effect of the outcome “unauthorized materials” is statistically different from zero. If we jointly exploit variation in both questions, we find a positive and significant average standardized effect regarding the perceived cheating behavior of other students.

One potential point of skepticism regarding the results on descriptive norms is that the perceived frequency of cheating in the exam may reflect, to some extent, the perceived sanction instead of the underlying descriptive norm (or the perception regarding others’ perception of the sanction). Given the previously reported results on the expected sanction, this is rather unlikely. However, because this result was unknown when designing the experiment, we responded to this measurement concern by including additional questions to our survey, which introduce a hypothetical zero-enforcement scenario (see questions N3 and N4). Column (7) to (9) report the results, again for both cheating technologies. Compared to Columns (4) and (5), we find a much higher level of perceived cheating in the control group, indicating that the perceived sanction, indeed, plays a role. Furthermore, for both outcomes, we confirm that the signature treatment results in a significant shift towards more (perceived) cheating by other students. The average standardized effect on perceived cheating in the zero-enforcement scenario in Column (9) is highly significant. Moreover, Column (10) displays a positive and significant average standardized effect for all four outcomes, capturing the perceived behavior of other students.

In summary, participants who signed the no-cheating declaration expected more cheating. In contrast to this result, we do not find any evidence for a shift in the perceived sanctions. Given the already discussed limitations that
result from our inability to observe perceptions directly (limited sample size, measurement issues, spillovers), the survey evidence cannot ultimately identify the mediating mechanisms. The patterns in the data, however, suggest that the request to sign the no-cheating declaration has weakened the survey participants’ descriptive norms of academic integrity.

4 Conclusion

Academic cheating is a widespread and wasteful illicit activity. Therefore, educators around the world spend considerable resources to uphold academic integrity. One of the frequently used measures to fight academic cheating is to request that students commit to a no-cheating rule. However, there is no causal evidence on how such requests affect academic cheating. We contribute to filling this gap by providing experimental evidence on how the request to sign a no-cheating rule affects plagiarism. The evidence originates from a field experiment at a German university, in which we randomly allocated students to either a control group or a signature treatment. In the latter, they signed a declaration of compliance with the existing no-cheating rule before the beginning of the exam.

In our analysis, we focus on academic cheating by copying multiple-choice answers from neighbors. By randomly assigning students to seats, we ensure that under the null hypothesis of no cheating, the probability of choosing the same answer to a given multiple-choice question for students sitting next to each other should not be different from the probability for students who were not sitting side by side. Comparing the similarity in the answers of actual neighbors to the similarity among counterfactual neighbors allows us to identify plagiarism.

Exploiting our methods, first, we document that students plagiarize by copying answers from their neighbors, both in the signature treatment and the control group. Second, we show that the above-normal similarity in neighbors’ answers reflects copying among pairs of low-ability students. Third, evaluating the request to sign a no-cheating declaration, our findings contradict the conjecture that this type of intervention reduces cheating. To the contrary, the above-normal similarity in neighbors’ answers is significantly higher in the signature treatment relative to the control group. This finding implies that the request backfired and induced more cheating. Fourth, as for channels, we present suggestive evidence that the signature treatment weakened the perceived descriptive norm of academic integrity.

We believe that our findings have implications for the fight against academic cheating. Most importantly, our main result implies that educators around the world should think twice before implementing policies that require students to express their commitment to existing no-cheating rules.
However, more evidence from other settings is needed to clarify if our findings extend to other contexts with their specific set of social norms and policies directed towards upholding academic integrity. Our paper may also speak to additional contexts in which commitment requests are frequently used. Of course, we cannot know whether interventions as the one in our experiment can also backfire in non-educational settings. However, our results certainly warrant further investigations into how pledges to comply with no-cheating rules affect truthful reporting in different contexts.

References


37 For example, when individuals and firms report tax-relevant information to the tax administration, they are commonly requested to sign a declaration confirming the truthfulness of the submitted information. Similarly, individuals applying for social welfare benefits and firms bidding for government contracts must submit declarations of compliance with a host of regulations. Also, all of the Fortune Global 500 corporations have a code of conduct, frequently including a declaration of compliance that newly hired staff have to sign.


## Table 1: Balancing Checks

<table>
<thead>
<tr>
<th></th>
<th>Exam 1</th>
<th></th>
<th></th>
<th>Exam 2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control</td>
<td>Signature</td>
<td>Monitoring</td>
<td>Difference</td>
<td>Signature</td>
<td>Monitoring</td>
</tr>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
</tr>
<tr>
<td>Gender (Female = 1)</td>
<td>0.54</td>
<td>0.56</td>
<td>0.50</td>
<td>0.02</td>
<td>-0.04</td>
<td>0.53</td>
</tr>
<tr>
<td></td>
<td>(0.04)</td>
<td></td>
<td>(0.04)</td>
<td></td>
<td></td>
<td>(0.05)</td>
</tr>
<tr>
<td>High-School GPA</td>
<td>2.47</td>
<td>2.48</td>
<td>2.50</td>
<td>0.02</td>
<td>0.03</td>
<td>2.48</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Math Proficiency</td>
<td>0.75</td>
<td>0.73</td>
<td>0.73</td>
<td>-0.02</td>
<td>-0.02</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>(-0.01)</td>
<td></td>
<td>(-0.01)</td>
<td></td>
<td></td>
<td>(0.02)</td>
</tr>
<tr>
<td>Field of Study (Econ. &amp; Sociology = 1)</td>
<td>0.07</td>
<td>0.06</td>
<td>0.09</td>
<td>-0.01</td>
<td>0.02</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.02)</td>
<td></td>
<td>(0.02)</td>
<td></td>
<td></td>
<td>(0.03)</td>
</tr>
<tr>
<td>Age</td>
<td>19.6</td>
<td>19.6</td>
<td>19.6</td>
<td>-0.04</td>
<td>-0.03</td>
<td>19.7</td>
</tr>
<tr>
<td></td>
<td>(0.10)</td>
<td></td>
<td>(0.10)</td>
<td></td>
<td></td>
<td>(0.11)</td>
</tr>
<tr>
<td>Bavaria</td>
<td>0.81</td>
<td>0.83</td>
<td>0.84</td>
<td>0.02</td>
<td>0.02</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
<td>(0.04)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>333</td>
<td>208</td>
<td>225</td>
<td></td>
<td>204</td>
<td>149</td>
</tr>
</tbody>
</table>

**Notes:** This table shows balancing checks for both exams covered in the field experiment. Columns (1) to (3) report treatment-specific means for Exam 1. Column (4) shows the difference in means between signature and control with standard errors in parentheses. Column (5) reports the difference in means between monitoring and control. Columns (6) to (8) report means and the difference in means for Exam 2. High-School GPA is the grade point average from high school (criterion for university admission), ranging from 1.0 (outstanding) to 4.0 (pass). Math Proficiency is obtained from a university math exam taken prior to the exams studied in the experiment. The proficiency score gives the percentage of total points the student obtained in the math test. Field of Study is a dummy for students with a major in Economics & Sociology, the reference group being students enrolled in Economics and Business Administration. Bavaria is a dummy for students who finished high school in Bavaria. Gender and High-School GPA were used for stratification.
Table 2: Responses to the Signature Treatment: Actual and Counterfactual Pairs

<table>
<thead>
<tr>
<th></th>
<th>Dependent Variable: Indicator for Identical Incorrect Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1) Unconditional Estimates (2) Multiple-Choice Controls (3) Pair Controls (4) All Controls</td>
</tr>
<tr>
<td>Signature</td>
<td>0.0008 [0.7997] 0.0008 [0.7999] 0.0002 [0.9316] 0.0002 [0.9317]</td>
</tr>
<tr>
<td>Actual Neighbors</td>
<td>0.0073*** [0.0012] 0.0073*** [0.0012] 0.0066*** [0.0016] 0.0066*** [0.0016]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.0088** [0.0190] 0.0088** [0.0189] 0.0081** [0.0126] 0.0081** [0.0125]</td>
</tr>
<tr>
<td>Multiple-Choice FE</td>
<td>No Yes No Yes</td>
</tr>
<tr>
<td>Pair Controls</td>
<td>No No Yes Yes</td>
</tr>
<tr>
<td>Share of Identical Incorrect Answers amongst Counterfactual Pairs in Control Group</td>
<td>0.0364</td>
</tr>
<tr>
<td>Number of Clusters</td>
<td>81</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>140,937</td>
</tr>
</tbody>
</table>

Notes: This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical incorrect answers. Estimates are based on linear probability models. Column (1) presents the unconditional estimates. Column (2) controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors are clustered at the row level; p-values in brackets.
Table 3: Post-Exam Survey: Questions

Perceived Sanction
S1 & S2: Imagine the supervising staff in your last exam had witnessed how one participant copies answers from other participants [uses unauthorized materials (like, for instance, a smartphone)]. What do you think would be the likely consequence for this student?

Perceived Detection Probability
D1 & D2: Think back to your last exam, and imagine 100 participants who try to copy at least one answer from other participants [use unauthorized materials (like, for instance, a smartphone) to answer at least one question]. What do you think, how many of those 100 students would have been caught? Please state a number between 0 and 100.

Descriptive Norm
N1 & N2: Think back to your last exam, and imagine a group of 100 participants. What do you think, how many of those have copied at least one answer from other participants [used unauthorized materials (like, for instance, a smartphone) to answer at least one question]? Please state a number between 0 and 100.

N3 & N4: Think back to your last exam, and imagine the supervising staff had left the exam hall for a few minutes. What do you think, how many of 100 participants would have copied at least one answer from other participants in the meanwhile [used unauthorized materials (like, for instance, a smartphone) to answer at least one question in the meanwhile]? Please state a number between 0 and 100.

Notes: This table summarizes how we measure perceptions. Each question has two versions. The first version (S1, D1, N1, N3) refers to cheating in the form of copying answers from neighbors (italics). The second version (S2, D2, N2, N4) concerns the use of unauthorized materials (gray text in brackets). To answer questions S1 and S2, participants select one of the following options: (a) There are no consequences whatsoever. (b) The student receives a verbal warning. No other consequences apply. (c) The student will face a hearing before the examination committee. (d) The committee will decide if the student fails the exam. (e) The student will fail the exam in any case. (f) The student will be relegated from the university. To answer all the other questions, participants state a number between 0 and 100.

Table 4: Post-Exam Survey: Perceived Sanctions

<table>
<thead>
<tr>
<th>Perceived Sanction: Copying</th>
<th>Perceived Sanction: Unauthorized Materials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Control (1)</td>
</tr>
<tr>
<td>No sanction at all</td>
<td>0.0</td>
</tr>
<tr>
<td>Verbal warning</td>
<td>7.3</td>
</tr>
<tr>
<td>Exam committee hears case and decides</td>
<td>20.0</td>
</tr>
<tr>
<td>Student fails exam</td>
<td>72.7</td>
</tr>
<tr>
<td>Student is expelled from university</td>
<td>0.0</td>
</tr>
</tbody>
</table>

p-value, Fisher's exact test | [0.180] | [0.645]
Number of Observations | 55 | 48 | 55 | 48

Notes: This table shows students' expected sanction in case of detected cheating. Particularly, for a list of potential sanctions, the table reports the treatment-specific shares of participants (in percent) who believe that one particular sanction will be implemented in case of detection. Columns (1) and (2) focus on sanctions for copying answers. Columns (3) and (4) focus on sanctions for using unauthorized materials. We also use Fisher's exact tests to explore whether the signature treatment affected the distributions of answers. We report the corresponding p-values [in brackets].
Table 5: Post-Exam Survey: Treatment Effects

<table>
<thead>
<tr>
<th></th>
<th>Detection Probability Copying (1)</th>
<th>Detection Probability Unauthorized Materials (2)</th>
<th>Average Stand. Effect (1)&amp;(2) (3)</th>
<th>% Other Students Using Materials (4)</th>
<th>% Others Using Unauthorized Materials (5)</th>
<th>Average Stand. Effect (4)&amp;(5) (6)</th>
<th>Social Norm Copying (7)</th>
<th>Social Norm Unauthorized Materials (8)</th>
<th>Average Effect (7)&amp;(8) (9)</th>
<th>Average Effect (4),(5),(7),(8) (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of Signature Treatment</td>
<td>-6.0</td>
<td>-2.5</td>
<td>-0.13</td>
<td>4.5</td>
<td>3.6</td>
<td>0.61</td>
<td>14.9</td>
<td>19.9</td>
<td>0.54</td>
<td>0.58</td>
</tr>
<tr>
<td>p-value, robust</td>
<td>[0.400]</td>
<td>[0.727]</td>
<td>[0.489]</td>
<td>[0.228]</td>
<td>[0.010]**</td>
<td>[0.012]**</td>
<td>[0.026]**</td>
<td>[0.004]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
</tr>
<tr>
<td>p-value, hall cluster, wild bootstrap</td>
<td>[0.364]</td>
<td>[0.725]</td>
<td>[0.300]</td>
<td>[0.021]**</td>
<td>[0.052]*</td>
<td>[0.017]**</td>
<td>[0.017]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
</tr>
<tr>
<td>p-value, Mann-Whitney-U-Test</td>
<td>[0.686]</td>
<td>[0.829]</td>
<td>[0.185]</td>
<td>[0.041]**</td>
<td>[0.053]*</td>
<td>[0.029]**</td>
<td>[0.029]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
</tr>
<tr>
<td>Control Group Mean</td>
<td>31.6</td>
<td>31.4</td>
<td>8.6</td>
<td>3.4</td>
<td>52.2</td>
<td>43.4</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
<td>[0.001]**</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
<td>103</td>
</tr>
</tbody>
</table>

Notes: The table reports the effects of the signature treatment on students’ responses in the post-exam survey. The estimates are derived from OLS regressions using gender, age, high-school GPA, and exam-hall fixed effects as additional controls. We report the following types of p-values [in brackets]: (a) heteroscedasticity robust p-values, (b) hall-cluster-robust p-values based on a wild cluster bootstrap-t procedure that accounts for the small number of cluster (Cameron et al., 2008), (c) p-values for Mann-Whitney-U-Tests, and (d) robust p-values for average standardized effects following Kling et al. (2004) and Clingingsmith et al. (2009). Dependent variables in Column (1): Perceived detection probability (in percent) if copying from a neighbor. Column (2): Perceived detection probability (in percent) if using unauthorized materials (like smartphone, etc). Column (4): Perceived share (in percent) of students copying at least one answer. Column (5): Perceived share (in percent) of students using unauthorized materials. Column (7): Perceived share of students (in percent) that would copy at least one answer in the past exam. Column (8): Perceived share of students (in percent) that would use unauthorized materials in the case of no supervision. See the Online Appendix for the exact wording of the survey questions.
Figure 1: Overview of Field-Experimental Design

Exam 1

- **Sample**
  - 1007 students sampled
  - 766 students took exam

- **Control**
  - 432 students sampled
  - 333 students took exam

- **Signature**
  - 265 students sampled
  - 208 students took exam

- **Monitoring**
  - 310 students sampled
  - 225 students took exam

Exam 2

- **Control**
  - 262 students sampled
  - 204 students took exam

- **Signature**
  - 170 students sampled
  - 149 students took exam

Grading / Data Collection

Notes: This figure visualizes the experimental design. The field experiment was implemented in two written exams. Exam 1 comprised a control group and two treatment groups, signature and monitoring. Students assigned to the control group in Exam 1 were also sampled for the intervention in Exam 2, comprising a control group and a signature treatment group. The figure indicates, for each treatment, the number of students assigned to the respective treatment group, and the number of students who actually took the exam. Differences between the two figures are due to the fact that students could postpone participation to later semesters.
Figure 2: Randomization Tests: Cheating by Treatment Group

A: Control

B: Signature

Notes: This figure shows the results for the treatment-specific randomization tests. In each panel, the vertical line represents the test statistic derived from the actual seating arrangement. The bell-shaped curves show the mean-centered distributions of the test statistic under the null of no cheating on the basis of Epanechnikov kernels. For both panels, we obtain $p \leq 0.001$ (two-tailed tests).
Figure 3: Cheating: Heterogeneity with Respect to Students' Ability

**Dependent Variable A: Indicator for Identical Incorrect Answers**

- **A1: Neighbor Effects**
- **A2: p-values**

![Graph A1 Neighbor Effects](image)

![Graph A2 p-values](image)

**Dependent Variable B: Indicator for Identical Correct Answers**

- **B1: Neighbor Effects**
- **B2: p-values**

![Graph B1 Neighbor Effects](image)

![Graph B2 p-values](image)

**Notes:** This figure examines how students’ ability (proxied by high-school performance) relates to their cheating behavior. To construct this figure, we exploit the linear probability model (1) to estimate the effect of being a pair of actual neighbors on the probability that two paired students give identical incorrect answers (Panel A) or identical correct answers (Panel B). We allow for average-neighbor-effect heterogeneity in students’ abilities (see equation 2). Panels A1 and B1 demonstrate this heterogeneity: The horizontal (vertical) axis shows the grade of the worse (better) student of a particular pair, the colors indicating the size of the average neighbor effect for a specific grade combination. High-school grade combinations with the largest (the smallest) neighbor effects are colored in red (blue). The grading scale ranges from A (best grade) to D (worst grade). Panels A2 and B2 show the associated p-values. We use a thin-plate-spline interpolation to predict values for intermediate grades and cluster standard errors at the row level.
Table A1: Responses to the Signature Treatment: All Identical Answers

<table>
<thead>
<tr>
<th></th>
<th>(1) Unconditional Estimates</th>
<th>(2) Multiple-Choice Controls</th>
<th>(3) Pair Controls</th>
<th>(4) All Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature</td>
<td>−0.0029</td>
<td>−0.0029</td>
<td>−0.0001</td>
<td>−0.0001</td>
</tr>
<tr>
<td></td>
<td>[0.8323]</td>
<td>[0.8335]</td>
<td>[0.9896]</td>
<td>[0.9908]</td>
</tr>
<tr>
<td>Actual Neighbors</td>
<td>0.0144***</td>
<td>0.0144***</td>
<td>0.0120**</td>
<td>0.0120**</td>
</tr>
<tr>
<td></td>
<td>[0.0028]</td>
<td>[0.0028]</td>
<td>[0.0205]</td>
<td>[0.0206]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.0105</td>
<td>0.0105</td>
<td>0.0143*</td>
<td>0.0143*</td>
</tr>
<tr>
<td></td>
<td>[0.1561]</td>
<td>[0.1567]</td>
<td>[0.0684]</td>
<td>[0.0683]</td>
</tr>
<tr>
<td>Multiple-Choice FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pair Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Share of Identical Answers amongst Counterfactual Pairs in Control Group 0.5745

Number of Clusters 81
Number of Observations 140,937

Notes: This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical answers (correct or incorrect). Estimates are based on linear probability models. Column (1) presents the unconditional estimates. Column (2) controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors are clustered at the row level; $p$-values in brackets.
<table>
<thead>
<tr>
<th></th>
<th>(1) Unconditional Estimates</th>
<th>(2) Multiple Choice Controls</th>
<th>(3) Pair Controls</th>
<th>(4) All Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature</td>
<td>0.0009</td>
<td>0.0009</td>
<td>−0.0003</td>
<td>−0.0003</td>
</tr>
<tr>
<td></td>
<td>[0.7817]</td>
<td>[0.7819]</td>
<td>[0.9115]</td>
<td>[0.9113]</td>
</tr>
<tr>
<td>Actual Neighbors</td>
<td>0.0073***</td>
<td>0.0073***</td>
<td>0.0065***</td>
<td>0.0065***</td>
</tr>
<tr>
<td></td>
<td>[0.0011]</td>
<td>[0.0011]</td>
<td>[0.0019]</td>
<td>[0.0019]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.0088**</td>
<td>0.0088**</td>
<td>0.0082**</td>
<td>0.0082**</td>
</tr>
<tr>
<td></td>
<td>[0.0197]</td>
<td>[0.0195]</td>
<td>[0.0134]</td>
<td>[0.0133]</td>
</tr>
<tr>
<td>Row FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple-Choice FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pair Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Share of Identical Incorrect Answers amongst Counterfactual Pairs in Control Group</td>
<td>0.0364</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Clusters</td>
<td>81</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td>140,937</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:** This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical incorrect answers. Estimates are based on linear probability models. Column (1) includes row indicators to control for row effects. Column (2) additionally controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors are clustered at the row level; p-values in brackets.
Table A3: Responses to the Signature Treatment: Specifications with Room Effects

<table>
<thead>
<tr>
<th></th>
<th>(1) Unconditional Estimates</th>
<th>(2) Multiple-Choice Controls</th>
<th>(3) Pair Controls</th>
<th>(4) All Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Actual Neighbors</td>
<td>0.0077***</td>
<td>0.0077***</td>
<td>0.0069***</td>
<td>0.0069***</td>
</tr>
<tr>
<td></td>
<td>[0.0003]</td>
<td>[0.0003]</td>
<td>[0.0009]</td>
<td>[0.0009]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.0086**</td>
<td>0.0086**</td>
<td>0.0082**</td>
<td>0.0082**</td>
</tr>
<tr>
<td></td>
<td>[0.0213]</td>
<td>[0.0212]</td>
<td>[0.0107]</td>
<td>[0.0106]</td>
</tr>
<tr>
<td>Room FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Multiple-Choice FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pair Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Share of Identical Incorrect Answers amongst Counterfactual Pairs in Control Group: 0.0364

Number of Clusters: 81
Number of Observations: 140,937

Notes: This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical incorrect answers. Estimates are based on linear probability models. Column (1) includes hall indicators to control for hall effects. Column (2) additionally controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors are clustered at the row level; p-values in brackets.
Table A4: Responses to the Signature Treatment: Logit Models

<table>
<thead>
<tr>
<th></th>
<th>(1) Unconditional Estimates</th>
<th>(2) Multiple-Choice Controls</th>
<th>(3) Pair Controls</th>
<th>(4) All Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Signature</td>
<td>0.0233</td>
<td>0.0241</td>
<td>0.0073</td>
<td>0.0081</td>
</tr>
<tr>
<td></td>
<td>[0.7987]</td>
<td>[0.8001]</td>
<td>[0.9232]</td>
<td>[0.9182]</td>
</tr>
<tr>
<td>Actual Neighbors</td>
<td>0.1896***</td>
<td>0.1977***</td>
<td>0.1697***</td>
<td>0.1792***</td>
</tr>
<tr>
<td></td>
<td>[0.0005]</td>
<td>[0.0005]</td>
<td>[0.0007]</td>
<td>[0.0007]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.1848**</td>
<td>0.1953**</td>
<td>0.1705**</td>
<td>0.1823**</td>
</tr>
<tr>
<td></td>
<td>[0.0226]</td>
<td>[0.0224]</td>
<td>[0.0213]</td>
<td>[0.0206]</td>
</tr>
<tr>
<td>Multiple-Choice FE</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Pair Controls</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

Number of Clusters: 81
Number of Observations: 140,937

Notes: This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical incorrect answers. Estimates are based on logit models. The table shows logit coefficients. Column (1) presents the unconditional estimates. Column (2) controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors are clustered at the row level; p-values in brackets.
Table A5: No-Cheating Declaration and Monitoring: Actual and Counterfactual Pairs

<table>
<thead>
<tr>
<th>Dependent Variable: Indicator for Identical Incorrect Answer</th>
<th>(1) Unconditional Estimates</th>
<th>(2) Multiple-Choice Controls</th>
<th>(3) Pair Controls</th>
<th>(4) All Controls</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring</td>
<td>0.0033</td>
<td>0.0033</td>
<td>0.0036</td>
<td>0.0036</td>
</tr>
<tr>
<td></td>
<td>[0.4445]</td>
<td>[0.4443]</td>
<td>[0.3970]</td>
<td>[0.3965]</td>
</tr>
<tr>
<td>Signature</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0002</td>
<td>0.0002</td>
</tr>
<tr>
<td></td>
<td>[0.7989]</td>
<td>[0.7990]</td>
<td>[0.9524]</td>
<td>[0.9530]</td>
</tr>
<tr>
<td>Actual Neighbors</td>
<td>0.0073***</td>
<td>0.0073***</td>
<td>0.0067***</td>
<td>0.0067***</td>
</tr>
<tr>
<td></td>
<td>[0.0009]</td>
<td>[0.0009]</td>
<td>[0.0014]</td>
<td>[0.0014]</td>
</tr>
<tr>
<td>Monitoring × Actual Neighbors</td>
<td>−0.0095***</td>
<td>−0.0095***</td>
<td>−0.0092**</td>
<td>−0.0092**</td>
</tr>
<tr>
<td></td>
<td>[0.0240]</td>
<td>[0.0245]</td>
<td>[0.0406]</td>
<td>[0.0408]</td>
</tr>
<tr>
<td>Signature × Actual Neighbors</td>
<td>0.0088**</td>
<td>0.0088**</td>
<td>0.0080**</td>
<td>0.0080**</td>
</tr>
<tr>
<td></td>
<td>[0.0176]</td>
<td>[0.0175]</td>
<td>[0.0124]</td>
<td>[0.0123]</td>
</tr>
</tbody>
</table>

| Multiple-Choice FE                                         | No                         | Yes                         | No               | Yes             |
| Pair Controls                                               | No                         | Yes                         | No               | Yes             |

Share of Identical Incorrect Answers amongst Counterfactual Pairs in Control Group: 0.0364
Number of Clusters: 148
Number of Observations: 149,776

Notes: This table reports estimates of the treatment-specific effect of being a pair of actual neighbors on the probability that two paired students provide identical incorrect answers. Estimates are based on linear probability models. Column (1) presents the unconditional estimates. Column (2) controls for multiple-choice fixed effects. Column (3) adds two types of pair-specific variables to our baseline regression: Control variables for gender combinations (a female-female dummy and a male-male dummy) and controls for high-school grade combinations (grade indicators for the better and worse student as well as interactions). Column (4) includes all the aforementioned control variables. All specifications include an exam dummy. Standard errors clustered at the row level, p-values in brackets.
Table A6: Monitoring Intensity by Lecture Hall

<table>
<thead>
<tr>
<th>Hall</th>
<th>Control Students per Supervisor</th>
<th>Signature Students per Supervisor</th>
<th>Monitoring Students per Supervisor</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>51.3</td>
<td>56.5</td>
<td>9.2</td>
</tr>
<tr>
<td>2</td>
<td>49.8</td>
<td>47.5</td>
<td>8.5</td>
</tr>
<tr>
<td>3</td>
<td>38.0</td>
<td>44.5</td>
<td>8.0</td>
</tr>
<tr>
<td>4</td>
<td>29.0</td>
<td>30.0</td>
<td></td>
</tr>
</tbody>
</table>

Treatment-specific Averages

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>46.4</td>
<td>46.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>8.4</td>
</tr>
</tbody>
</table>

**Notes:** This table contains information on the number of students per supervisors in each lecture hall. It also shows the weighted average of the monitoring intensity within the control group, the signature, and the monitoring treatment, respectively (weights: number of students in lecture hall).
### Table A7: Repetition of Field Experiment: Balancing Checks

<table>
<thead>
<tr>
<th></th>
<th>Control (1)</th>
<th>Signature (2)</th>
<th>Difference (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Female = 1)</td>
<td>0.51</td>
<td>0.50</td>
<td>0.02</td>
</tr>
<tr>
<td>High-School GPA</td>
<td>2.57</td>
<td>2.55</td>
<td>0.02</td>
</tr>
<tr>
<td>Math Proficiency</td>
<td>2.73</td>
<td>2.63</td>
<td>0.09</td>
</tr>
<tr>
<td>Field of Study (Econ. &amp; Sociology = 1)</td>
<td>0.12</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Age</td>
<td>21.2</td>
<td>21.4</td>
<td>-0.23</td>
</tr>
<tr>
<td>Bavaria</td>
<td>0.90</td>
<td>0.92</td>
<td>-0.02</td>
</tr>
</tbody>
</table>

**Number of Observations**

- Control: 535
- Signature: 525

**Notes:** This table shows balancing checks for the sample of students participating in the repetition of the field experiment. Column (3) shows the difference in means between signature and control with standard errors in parentheses.
Table A8: Post-Exam Survey: Characteristics of Participants and Non-Participants

<table>
<thead>
<tr>
<th></th>
<th>Non-Participants (1)</th>
<th>Participants (2)</th>
<th>Difference (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Female = 1)</td>
<td>0.51</td>
<td>0.49</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High-School GPA</td>
<td>2.58</td>
<td>2.39</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Math Proficiency</td>
<td>2.70</td>
<td>2.51</td>
<td>0.19</td>
</tr>
<tr>
<td></td>
<td>(0.13)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Field of Study (Econ. &amp; Sociology = 1)</td>
<td>0.12</td>
<td>0.09</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>21.3</td>
<td>21.1</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>(0.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bavaria</td>
<td>0.91</td>
<td>0.90</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(0.03)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of Observations</td>
<td>957</td>
<td>103</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows characteristics of participants and non-participants in the post-exam online survey. Column (3) shows the difference in means between non-participants and participants with standard errors in parentheses. Math proficiency is only available for 692 out of the 957 Non-participants and 83 out of 103 Participants.
Table A9: Post-Exam Survey: Balancing Checks

<table>
<thead>
<tr>
<th></th>
<th>Control (1)</th>
<th>Signature (2)</th>
<th>Difference (3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (Female = 1)</td>
<td>0.45</td>
<td>0.52</td>
<td>-0.07</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.10)</td>
</tr>
<tr>
<td>High-School GPA</td>
<td>2.47</td>
<td>2.30</td>
<td>0.17</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.12)</td>
</tr>
<tr>
<td>Math Proficiency</td>
<td>2.55</td>
<td>2.41</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.24)</td>
</tr>
<tr>
<td>Field of Study (Econ. &amp; Sociology = 1)</td>
<td>0.07</td>
<td>0.10</td>
<td>-0.03</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Age</td>
<td>21.0</td>
<td>21.1</td>
<td>-0.06</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.47)</td>
</tr>
<tr>
<td>Bavaria</td>
<td>0.89</td>
<td>0.90</td>
<td>-0.00</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(0.06)</td>
</tr>
<tr>
<td>Number of Observations</td>
<td>55</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

Notes: This table shows balancing checks for the sample of post-exam online survey participants. Column (3) shows the difference in means between signature and control with standard errors in parentheses. Math proficiency is only available for 96 students (45 in control and 41 in signature).
Figure A1: Monitoring Conditions in the Field Experiment

**Baseline Monitoring**

**Close Monitoring**

**Notes:** This figure is a stylized illustration of baseline monitoring (control group and signature treatment) and close monitoring (monitoring treatment). Gray dots represent students; black squares represent supervisors. The average monitoring intensities were 44.2 students per supervisor under baseline monitoring, and 8.4 students per supervisor under close monitoring.
Figure A2: Responses to a Selected Multiple-Choice Problem in One Lecture Hall

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes: This figure provides an idea of what kind of data patterns our methods exploit. It visualizes the spatial pattern of answers to one multiple-choice problem in a selected control-group room. Each rectangle represents a student, and the shade of the rectangle indicates the student's answer. Because each multiple-choice problem consisted of four statements, there are four different shades of gray in the figure. Many students who sat next to each other provided identical answers. These correlations could reflect a spatial pattern of answers resulting from (some) students copying the responses of a direct neighbor. Such correlations could, however, also arise for other, non-cheating related reasons. For example, there could be a randomly occurring spatial pattern in the smartness of students. To evaluate whether students plagiarized, we would like to test whether the similarities in neighbors’ answers were higher than in a counterfactual situation without any cheating and only randomly occurring similarities. Our tests approximate this unobserved counterfactual by creating artificial neighbor pairs who were not sitting side by side and, thus, could not plagiarize.
Figure A3: Randomization Test: Cheating in the Monitoring Treatment

Notes: This figure shows the result of the randomization test in the monitoring treatment. The vertical line represents the test statistic derived from the actual seating arrangement. The bell-shaped curve shows the mean-centered distributions of the test statistic under the null of no cheating on the basis of Epanechnikov kernels. We obtain $p = 0.999$ (two-tailed test).
Figure A4: Spatial Structure of Cheating and Randomization Schemes

A: Randomization within Rooms

- **A1: Row**
  - (Seats: 7,9)
  - Mean-centered counterfactual distribution after mean-centering (blue circles), the 95% confidence bands for the counterfactual distributions (blue spikes), and the empirical value of the relevant test statistic (red circles).

- **A2: 2nd Order Row**
  - (6,10)

- **A3: Column**
  - (3,13)

- **A4: Diagonal**
  - (2,4,12,14)

B: Randomization within Treatments

- **B1: Row**
  - (Seats: 7,9)

- **B2: 2nd Order Row**
  - (6,10)

- **B3: Column**
  - (3,13)

- **B4: Diagonal**
  - (2,4,12,14)

Notes: This figure examines the spatial structure of cheating (Panel A) and tests the robustness of our results with respect to the randomization schemes (Panel B). The figure also shows a sketch of a representative seating plan, in which the yellow circle represents a particular student (sitting in seat 8) who can copy answers from her neighbors 1 to 15. The Panels A1 and B1 focus on row-wise cheating of direct neighbors (student copies from 7 and 9). The Figures A2 and B2 consider plagiarizing from indirect neighbors (copying from 6 and 10). The Panels A3 and B3 test for column-wise cheating (copying from 3 and 13). The Panels A4 and B4 examine diagonal cheating (copying from 2, 4, 12, and 14). Each of the figures reports the average value of the test statistic in the counterfactual distribution after mean-centering (blue circles), the 95% confidence bands for the counterfactual distributions (blue spikes), and the empirical value of the relevant test statistic (red circles).
Figure A5: Spatial Structure of Cheating and Randomization Schemes

A: Randomization within Rooms

A1: Column Front  
(Seat: 3)

A2: Diagonal Front  
(2,4)

B: Randomization within Treatments  

B1: Column Front  
(Seat: 3)

B2: Diagonal Front  
(2,4)

Notes: This figure examines the spatial structure of cheating (Panel A) and tests the robustness of our results with respect to the randomization schemes (Panel B). The figure also shows a sketch of a representative seating plan, in which the yellow circle represents a particular student who can copy answers from her neighbors 1 to 15. The Panels A1 and B1 assume that the student only copied answers from the student in seat 3. The Panels A2 and B2 examine front-diagonal cheating (i.e., copying the answer of the students 2 and 4). Each of the figures reports the average value of the test statistic in the counterfactual distribution after mean-centering (blue circles), the 95% confidence bands for the counterfactual distributions (blue spikes), and the empirical value of the relevant test statistic (red circles).
Front sheet of exam materials in the field experiment

Answer Sheet for Exam in „Principles of Economics“

<table>
<thead>
<tr>
<th>First Name</th>
<th>Date</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last Name</td>
<td>Semester</td>
</tr>
<tr>
<td>Matriculation Number</td>
<td>Seat Number</td>
</tr>
<tr>
<td>Field of Study</td>
<td>Room</td>
</tr>
<tr>
<td>Email Address</td>
<td></td>
</tr>
</tbody>
</table>

Framed part varied in field experiment: included in Signature, not included in Monitoring and Control

Declaration

I hereby declare that I will not use unauthorized materials during the exam. Furthermore, I declare neither to use unauthorized aid from other participants nor to give unauthorized aid to other participants.

________________________
Signature
I hereby declare that I will not use unauthorized materials during the exam. Furthermore, I declare neither to use unauthorized aid from other participants nor to give unauthorized aid to other participants.

Signature

Please fill in:

Last Name    Date

First Name    Seat Number

Matriculation Number    Room

Email Address

Please carefully read the information provided on the back page!

---

Please carefully read the information provided on the back page!

---

Principles of Economics

Answer Sheet for Exam

CONTROL GROUP

---

Principles of Economics

Answer Sheet for Exam

TREATMENT GROUP

---

Pre-filled

Pre-filled
**List of official announcements to be made before written exam**

**Announcements**
Please read out loud before the exam starts!

1. Bags, folders, etc. need to be set aside such that you cannot access them during the exam.
2. Smoking is prohibited in the lecture hall.
3. Take care to provide legible handwriting. Unreadable parts will not be marked.
4. Cheating is forbidden and any attempt to deceive will lead to failure of the exam; i.e., your exam will be graded with a 5.0.
   Attempts to deceive are:
   - (a) if you are not sitting in your assigned seat
   - (b) if you communicate with your neighbors or copy answers from neighbors
   - (c) if your cellphone is not switched off
   - (d) if you possess or use unauthorized materials during the exam
   Authorized materials are: non-programmable calculator, dictionary of foreign words.

Now is your last chance to hand in unauthorized materials. There will be check-ups during the exams.

5. Please make sure that you received the correct exam materials. Stay in your seats until the exam has ended. The supervisors will collect your answers sheets after the exam. It is your responsibility to hand in the answer sheets.

6. The examination period starts after we have distributed the examination materials (i.e., the problem sets). Don't touch the examination materials until the start of the exam was announced. Questions concerning the problem sets will not be answered.

7. If you feel sick during the exam, you have to report this immediately. After the exam, you cannot claim that you were physically incapable of taking the test.

8. Please only use the provided pen to fill in the answer sheet. This facilitates the automated scanner-based evaluation of the multiple-choice answer sheets. Please make sure that the pen remains at your work desk after the end of the exam. We will collect the pens separately from the exam materials.

9. You now have 5 minutes time to complete the first page of the answer sheet. Instructions how to fill in the multiple-choice answer sheet are provided on the second page.