Measuring Exam Cheating and the Impact of a Reform to Curb Fraud

Gabriel Kreindler

March 2015

PRELIMINARY AND INCOMPLETE
DO NOT CITE OR CIRCULATE
FEEDBACK WELCOME

JEL No: I280, K420

Keywords: exam cheating, forensic economics, anti-fraud crackdown

Abstract. I introduce a measure of the prevalence of cheating during a written exam, defined as copying answers from students seated nearby. The measure is based on the idea that in the absence of cheating there should be no correlation between the scores of students seated next to each other. Unlike existing measures of cheating, which are designed for multiple-choice tests, the measure that I introduce can be used for open answer tests. I analyze how the measure performs in practice using data from nine years of a high-stakes exam in Romania. I find a high correlation between immediate neighbors, which decays quickly in the number of desks between students. I perform several robustness exercises to rule out spurious correlations, notably showing that the GPAs of neighboring students are completely uncorrelated. I also explore cheating heterogeneity in terms of student gender, high school origin, and ability. Finally, I show that the measure of cheating falls significantly one year after a reform intended to curb exam fraud.

1 Introduction

The widespread existence of exam fraud may have negative, important consequence through a multitude of channels. It can distort incentives to acquire human capital leading up to the exam, while benefiting dishonest individuals and rewarding acquisition of cheating abilities (which may be transferable to other contexts). A related potential effect is to shape students’ beliefs about norms and prevalence of incorrect behavior in the public sphere. In the labor market, an exam marred by significant fraud likely constitutes a weaker signal of ability. However, establishing and quantifying the above effects is challenging, in part because measuring exam fraud is difficult. Similarly, education policy-makers have a limited number of indicators to monitor the prevalence of fraud, and to measure the success of anti-fraud interventions.

In this project I describe a method to measure the prevalence of cheating in an exam, more precisely of copying or sharing answers between students seated close to each other. The basic idea is that if students are seated in a quasi-random order at the exam, the correlation, within an exam room or exam center,

---

1 I thank Josh Angrist, Laura Costica, Jerry Hausman, Ben Olken, Parag Pathak, Adrian Troie, and participants at the MIT lunch seminar for useful and ample feedback on various stages of this project.
2 PhD student, Department of Economics, Massachusetts Institute of Technology. Contact: gek@mit.edu.
3 A Romanian education minister compared cheating at the national end of high school exam to “stealing one’s start in life” (Funeriu 2010).
between the scores of students seated next or close to each other is informative about the prevalence of cheating.

I set up a simple model and introduce two indicators of cheating calculated at the level of the exam center. The first indicator $\rho_1$ is the correlation between (properly demeaned) scores of neighboring students, calculated at the exam center level. I also define a more conservative measure of cheating, $\rho_1 - \rho_3$, where $\rho_3$ is the correlation between scores of students sitting three desks away (with two other students sitting between them); this indicator eliminates any factor that may equally affect the scores of a group of several students, including exam room effects, such as a noisy exam room. However, the indicator is also conservative, as it will underestimate cheating if students share answers beyond their immediate neighbors.

I explore this framework in practice using data from the baccalaureate exam from Romania, a high stakes exam at the end of high school that is required for college admission. I document a high and significant correlation between the scores of students sitting next to each other, and I find that the correlation falls quickly with distance.

I perform several robustness tests that address concerns that the identified correlation is spurious (not indicative of cheating). In principle, there may be confounding factors, other than cheating, that would lead to a positive correlation between neighbors’ scores. For example, if seating is alphabetic, as it is in my application, students with similar names may come from similar socio-economic backgrounds. To rule out these scenarios, I run specifications where I drop student pairs with identical last names, and/or I introduce fixed effects for common last names. These changes do not alter the baseline results.

In another exercise, I consider students who are consecutive in the alphabetic order in their high school, and compare such pairs when they also sit next to each other during the exam (most common case) versus when there is a student from a different high school sitting between them. I find that that in the latter case the score correlation within the pair is significantly weaker, which rejects the possibility that common shocks experienced in high school drive the baseline results.

Finally, I use another approach to investigate if previous similarity in ability levels is driving the correlation. I show that middle school GPA and the high school entrance exam are highly predictive of the baccalaureate exam result, yet the correlation between the middle school GPAs of neighboring students is precisely estimated at zero. For the high school admission exam, a very small correlation between neighbors is detectable, yet I argue this may also be due to cheating in that exam. I conclude that the correlation in baccalaureate exam scores is unlikely to be driven by previous similarity in ability levels.

Using personal characteristics, I find that students of the same gender copy more when seated next to each other, and that boys cheat slightly more than girls. The magnitude of these effects is significant: a pair of boys sitting next to each other have on average scores correlated 30% more than a boy and girl. I also show that high ability students (as measured by middle school ability) cheat less. This effect is also quantitatively important.

The Romanian education ministry started a broad campaign to crack down on exam fraud in 2011, and I look at the effects of these changes on cheating and scores. The most robust result is that cheating levels fall dramatically starting from 2012. There is also a smaller decrease in 2010 relative to 2009 in Romanian Language and Literature (RLL), the only mandatory test for all students taking the baccalaureate. Scores

---

4 The high school admission exam is organized in a very similar fashion to the baccalaureate, leading to the possibility that neighbors at the baccalaureate were also sitting next to each other 4 years earlier. If the two students shared information at the high school admission exam, that would show up as a correlation in their scores.
and exam pass rates also fall precipitously starting in 2010 and 2011, and I explore whether this change is more pronounced in exam centers characterized by high baseline cheating levels. I find that exam centers with different baseline cheating levels have different time trends in term of average scores, and once this is taken into account there is no evidence of higher score reductions for exam centers with high baseline cheating.

The contribution of this project is to introduce a measure of the prevalence of cheating in written exams that can be computed with data that is easy to collect (and audit) by exam organizers, and that is designed for open answer tests. As a consequence, it may prove useful to policy-makers interested in monitoring exam fraud and assessing the effects of corrective interventions, as well as to researchers who study the effects and drivers of cheating.

The rest of this paper is organized as follows. Section 1.1 surveys the literature on cheating measures and on the Romanian exam and reform. Section 2 outlines a simple model of cheating from neighbors. Section 3 describes the setting and data in detail. Section 4 presents results and robustness checks. Section 5 analyzes the relationship between the anti-fraud reform and the evolution of the cheating measures, and section 6 present a discussion of the results, possible extensions, and concludes.

1.1 Literature

There is a rich theoretical and practical literature on detecting cheating (Angoff 1974, Frary et al 1977, Holland 1996, Khalid et al 2011). These studies study the problem of detecting individual cheating on multiple-choice tests, by assessing the likelihood of a similar pattern of correct and especially wrong answers between a pair of test takers. In this literature, an important focus is on controlling type I and type II error rates. Romero et al (2014) use the indicators developed in this literature, together with multiple hypothesis testing techniques, to detect massive cheating, defined as exam rooms where at least 60% of students cheated, in the context of a Columbian school exam.

Another strand of related literature studies teacher cheating on school multiple-choice tests (Jacob and Levitt 2003, INVALSI 2010, Angrist et al 2014). The measures in these studies apply to settings where teachers have the opportunity to alter the response sheets of their students. Teachers may choose to alter test results if, for example, their salary depends on their students’ results; alternatively, they may shirk if the task of transcribing and assessing responses is time-consuming. These studies use unusual test score fluctuations, surprisingly high average scores, low within-class variability, and suspicious answer patterns (such as blocks of identical answers) for tests in the same classroom.

Unlike the studies and methods described above, the indicators in this paper apply to tests with open answer questions. Measuring the prevalence of cheating for open answer tests may be a useful step for policy and research, given that many high-stakes tests around the world include or are entirely based on open-answer questions. In addition to the case of the Romanian baccalaureate studied in depth in this paper, other examples of end of high school exams or university admission exams with open answer questions include the National Higher Education Entrance Examination in China, the All Indian Secondary School Examination in India, the National Senior Certificate (NSC) examinations in South Africa, and the Unified State Exam in Russia. The wider applicability of the measures in this paper come at the cost of using finer data on students’ seating arrangement within exam rooms. Collecting this data arguably has a low marginal cost given common exam logistical preparations for a national exam, and the fact that open answer questions need to be hand graded.

The Romanian baccalaureate exam, which is used as a case study in this paper, has been studied in other work. Borcan et al (2015) study the crackdown on exam fraud in Romania starting in 2011, using the same
exam and the same data source as this paper. Specifically, they analyze two components of the reform, due to monitoring (video cameras were installed in exam rooms), and due to improved punishment (a large number of supervising teachers were prosecuted and sentenced for accepting bribes and other fraud starting after the 2010 exam). They use a differences-in-differences approach, taking advantage of the fact that only a fraction of counties used cameras starting in 2011, and the rest in 2012. Their main outcome measures are test scores and exam passing rates, and they find that both punishment (the year effect) and monitoring (the differences-in-differences effect) reduced scores and passing rates at a high stakes test component of the exam. Borcan et al also study the heterogeneous effects according to gender, ability and student SES. Borcan et al (2014) focus on the same exam, and investigate the effects of a 25% pay cut for public employees that occurred in 2010. Using private schools as a control group, they show that test scores for students from public schools increased in 2010, which they interpret at evidence of increased fraud levels. The results in this paper are complementary to those in Borcan et al (2015) and Borcan et al (2014). Indeed, it would be interesting to also study in their setting the measure of prevalence of cheating that I introduce, given that it is a conceptually precise indicator of fraud.

2 A Model of Cheating by Copying from Neighbors

This section introduces a stylized model to illustrate my proposed measures of cheating. The model abstracts from some ways in which cheating can occur; its purpose is to illustrate succinctly why correlation between observed scores can be used to reject the hypothesis of no information sharing. The model features are chosen to match the setting of the Romanian baccalaureate. Specifically, for each test students are ordered alphabetically by last name between and within exam rooms. In other contexts, students may be seated according to random numbers; moreover, the exact seating chart may be known (e.g. an exam room may have three rows of 10 desks). The approach in this section can be extended in a straightforward way to those settings.

Consider an exam center where students take an exam. Students from one or more high schools take the exam in the same center. A student’s alphabetical rank in the exam center is denoted by $i$. Students in an exam center are seated in a long row according to their index $i$, and then divided into exam rooms (e.g. first 20 students are in the first room, etc.). Ignoring class boundaries, $i$ is seated next to $i-1$ and $i+1$.

The basic idea of the model is as follows. Students have a counterfactual (unobserved) score if they took the test alone. I model the expected counterfactual score, and assume that the demeaned counterfactual scores of students sitting next to each other are independent. When students sit in a real exam room, I imagine that they first solve parts of the exam and obtain their counterfactual score; then, they have the opportunity to share responses. The final (observed) score of any student depends on her own counterfactual score, but also depends on the counterfactual scores of her neighbors.

Denote by $y_i$ student $i$’s counterfactual (unobserved) score if she took the exam in isolation and in average room conditions. This is a random variable, and we model the expected counterfactual score of individual

---

5 It is sufficient to assume that the demeaned counterfactual scores are uncorrelated conditionally on any combination of student characteristics.

6 Room conditions refer to factors that impact everyone in the exam room roughly equally, such as if the room is noisy or hot.
\( i \) as the sum of a function of \( i \)'s characteristics \( \chi_i \) (gender, socio-economic status, high school attended, etc.). We can write:

\[
y_i = y(\chi_i) + \epsilon_i
\]

where \( \epsilon_i \) is a zero-mean random noise component. We assume that \( \epsilon_i \) are iid.

The final (observed) score \( x_{ic} \) of student \( i \) in exam room \( c \) will depend on her counterfactual score \( y_i \), on her neighbors’ counterfactual scores \( y_{i-1} \) and \( y_{i+1} \), and on a (zero-mean) exam room effect \( \delta_c \). (I account explicitly for \( \delta_c \) because in the data I observe exam center assignment, but not the exam room. With better data I could partial out exam room fixed effects.) Formally,

\[
(#)
\]

\[
x_{ic} = y_i + py_{i-1} + py_{i+1} + \delta_c
\]

The parameter \( p \) is a measure of copying between immediate neighbors. The objective is to use observed scores and estimate \( p \) or a monotonic function of \( p \).

The first step is to demean \( x_{ic} \). In practice I construct the demeaned score \( \bar{x}_i \equiv x_{ic} - \bar{x}_i \) by regressing the observed scores on high school fixed effects and various functions of \( i \)'s name.\(^7\)

### 2.1 Exam center correlations

The correlation between the scores of two adjacent students across all students in an exam center, ignoring class boundaries, is given by

\[
\rho_1 \equiv \frac{E(\bar{x}_i \bar{x}_{i+1})}{\text{var}(x_{ic})} = \frac{E((\epsilon_i + p\epsilon_{i-1} + p\epsilon_{i+1} + \delta_c)(\epsilon_{i+1} + p\epsilon_{i} + p\epsilon_{i+2} + \delta_{c}))}{E(\epsilon_i^2 + p^2\epsilon_{i-1}^2 + p^2\epsilon_{i+1}^2 + \delta_c^2)} = \frac{\sigma_\delta^2 + 2p\sigma_\epsilon^2}{\sigma_\delta^2 + (2p^2 + 1)\sigma_\epsilon^2}
\]

It is important to note that even in the absence of copying \( (p = 0) \), the correlation will be positive due to the room effect. To isolate copying I also compute the correlation between the scores of students sitting \textit{three desks away} (i.e. who have two other desks or students between them):

\[
\rho_3 \equiv \frac{E(\bar{x}_i \bar{x}_{i+3})}{\text{var}(x_{ic})} = \frac{\sigma_\delta^2}{\sigma_\delta^2 + (2p^2 + 1)\sigma_\epsilon^2}
\]

The difference \( \rho_1 - \rho_3 \) equals zero when \( p = 0 \) and is increasing in \( p \) (provided that \( p \) is less than \( 1/\sqrt{2} \) or not very high relative to \( \sigma_\delta^2/\sigma_\epsilon^2 \)). In words, this measure differences out the room effect.

Note that the measure \( \rho_1 - \rho_3 \) is a conservative measure of cheating. If students share responses beyond their immediate neighbors, then this measure may underestimate the level of cheating. However, this also means that it is also possible that an increase in cheating, e.g. by increasing the number of neighbors with

---

\(^7\) Equation (\#) implies that the mean of \( x_{ic} \) also depends on characteristics of \( i \pm 1 \), not only on characteristics of \( i \). Hence, the correct way to demean \( x_{ic} \) should include these variables as well. In addition, significant dependence of \( x_{ic} \) on \( i \)’s neighbor’s characteristics is in itself tentative evidence of cheating. I will explore this in future work.
which a student shares information, will decrease the measure $\rho_1 - \rho_3$. It follows that the measures $\rho_1$ and $\rho_1 - \rho_3$ should always be analyzed together.

2.2 Regression analysis

In principle, the entire joint distribution of $x_i$ and $x_{i+1}$ is informative about the pattern of cheating. In particular, consider the conditional expectation of $\tilde{x}_i$ given $\tilde{x}_{i+1}$. Assume that $\epsilon_i$ and $\delta_c$ are normally (and independently) distributed, then $\tilde{x}_i$ and $\tilde{x}_{i-1}$ follow a bivariate normal distribution, with mean zero, equal variance and correlation $\rho_1$. A standard formula allows us to write the conditional expectation as

$$E(\tilde{x}_i|\tilde{x}_{i+1}) = \rho_1 \tilde{x}_{i+1}$$

In other words, the conditional expectation function (CEF) is linear in $\tilde{x}_{i+1}$. The results in Section 4.1 show that the empirical CEF is well approximated by a linear function.

With data on student characteristics, it is possible to measure heterogeneity in the prevalence of cheating by these characteristics. For example, consider the gender variable $\text{male}_i$, and assume this is uncorrelated between neighboring students (this can be verified in practice). Then we can estimate the equation

$$E(\tilde{x}_i|\tilde{x}_{i+1}) = \alpha + \rho_1 \tilde{x}_{i+1}$$

separately for each combination of $i$’s and $i + 1$’s genders (e.g. all $i$ such that $i$ is male and $i + 1$ female). This will be informative on whether cheating is more prevalent between male students, between students of different genders, or between female students.

This exercise can be repeated with other characteristics of interest, such as socio-economic status, as well as with continuous variables such as prior educational ability. For a continuous variable $z_i$, for example middle school GPA, we can run the regression:

$$E(\tilde{x}_i|\tilde{x}_{i+1}) = \alpha + \rho_1 \tilde{x}_{i+1} + \beta z_i + \delta \tilde{x}_{i+1} \times z_i$$

The coefficient $\delta$ captures the how students with high versus low values of $z_i$ are involved in cheating.

2.3 Model limitations and extensions

The model presented above makes several simplifying assumptions, which are briefly discussed in this section.

Consider the linear expression for the final score. It may be natural to assume that student $i$’s final score depends non-linearly on the counterfactual scores of $i - 1$, $i$ and $i + 1$. For example, assume that each exam has a unit continuum of questions and a student with counterfactual score $y$ knows how to solve a random subset of measure $y \in [0,1]$ of all questions. If students try to copy answers for questions that they were not able to solve on their own, the final score of $i$ can be expressed as

$$x_{ic} = y_i + (1 - y_i) [py_{i-1} + (1 - py_{i-1})py_{i+1}] + \delta_c$$

It is easy to check that he formulas for $\rho_1$ and $\rho_3$ go through without major change under this specification. (However, the conditional expectation $E(\tilde{x}_i|\tilde{x}_{i+1})$ is no longer exactly linear.)

Alternatively, it may be that questions in an exam are ordered in the same order of difficulty for everyone, so student with counterfactual score $y$ solves the interval $[0, y]$ of questions correctly. A possible specification in this case is
where \( M(y, y') \equiv y + p(y' - y)_+ \).

3 Empirical Setting and Data

I study the baccalaureate (end of high school) exam in Romania between 2006 and 2014. Passing the exam is a requirement for entering higher education, and having the baccalaureate diploma is considered an asset in the labor market, although this relationship has not been studied rigorously.\(^8\) The exam syllabus is closely aligned with the national curriculum.

The exam is organized by the Romanian ministry of education, with the help of county school inspectorates, and high school graduates sit the exam during July after their graduation.\(^9\) One or more high schools take the exam in the same exam center, which is often also a high school. During each test, students are allocated to rooms and seated within rooms in alphabetic order (see section 3.2 for more details). After the tests, the anonymized exam papers are brought to regional grading centers, where they are randomly shuffled and then assigned to graders. This aspect of the exam logistics allays concerns that the correlation in scores may be due to grading.

The exam consists of 5-6 tests taken over a period of two weeks.\(^10\) All participants take the same Romanian language and literature (RLL) test, and the other tests depend on the student’s high school track. Throughout this paper, I mostly focus on the RLL results.\(^11\)

Before 2011, the baccalaureate exam was believed to have high levels of exam fraud of various forms, documented in Borcan et al (2014). Some of the types of fraud will not be detected by the measure I use; for example, if all students in an exam room share the same answers, then this will be absorbed in a (very large) exam room effect. To list another example, if a student bribes an official from the exam organizing committee, this will not be picked up either. However, it is still possible that a significant amount of cheating occurs by informal exchange of information between students, on the day of the exam. The empirical results, and the robustness tests, will reveal the extent to which this is the case.

A Western-educated and reform-minded scientist, Daniel Funeriu, took over as education minister in late 2009 (after the baccalaureate exam was organized) and continued in this position until spring 2012. Funeriu was an outspoken supporter of a “clean” baccalaureate exam, and he introduced changes to that effect. The most visible of these changes were introduced in 2011. That year, the education minister made a visible media campaign to persuade students, parents and supervising teachers that cheating and fraud will not be tolerated during the exam. He threatened to prosecute supervising teachers that allow students to cheat, drawing parallels to a government crackdown on corrupt customs officers earlier that year (Funeriu 2010).

---

\(^8\) Clark and Martorell 2014 use a regression discontinuity approach and find little support for a return to having a high school diploma.

\(^9\) There is a second session, usually organized in August, for students who do not pass in July.

\(^10\) It is important to note that the exam format changed in 2010: the number of tests was reduced from six to five, and the number of tests that enter the calculation of the passing outcome was reduced from six to three. However, the structure of the RLL test and of the two Mathematics tests have not changed significantly in the study period.

\(^11\) The main reason I focus on RLL is that I believe that my estimation of the seating order is more accurate for RLL, because all students take this test. Section 3.2 discusses the seating order estimation method.
The most highly publicized concrete preventive measure was the installation of video cameras in exam classrooms. Borcan et al (2015) use the partial rollout of cameras in 2011 to 25 out of 42 counties in a differences-in-differences study of the impact of the reform on grades and graduation rates. In addition, after the 2010 and 2011 exams, a large number of teachers were prosecuted and some convicted for accepting bribes or other types of fraud.

The exam results in 2010 and 2011 were significantly weaker than in previous years. As Table 1 shows, in 2011 only 44% of the students passed the exam, compared to 66% in 2010 and above 75% until 2009. Borcan et al (2015), as well as the Romanian public, interpret these figures as evidence that there was significantly less cheating compared to previous years. However, lower passing rates could also be caused by more difficult exam questions or tougher grading. Section 4 explores how the measure of copying changed following the reform.

3.1 The data and sample
My main data source consists of publicly available baccalaureate exam results from 2006 to 2014 for the universe of exam takers in Romania, with a total of 1,801,693 observations. The data contains the full name of the candidate, the scores for all exams taken, including whether the student did not show up or was eliminated from the exam, the high school name, and high school specialization (track). A separate data set contains the allocation of high schools to exam centers. Usually, several high schools are allocated to the same exam center. I construct a gender variable by hand-coding the most popular 500 first and middle names, which cover 95% of the entire dataset. For more than half of the students taking the baccalaureate exam in 2009, I obtained their middle school GPA and high school admission exam in 2005.

A crucial part of the data preparation is to ensure that the order that I construct in each exam center using students’ names, closely mirrors the actual seating arrangement on the day of the exam. The analysis using actual seating charts from two exam centers (see section 3.2) reveals that exam centers sometimes, but not always, stratify on certain variables before ordering alphabetically. For example, in one of the two centers part-times students were grouped together with full-time students, while in the other they were ordered alphabetically separately. Similarly, although all students take the same Romanian language and literature test (RLL), in one high school the humanities track students (10.7% of all students) were ordered separately.

The analysis sample is restricted to full-time students, and it does not include exam centers which contain students from high schools split between two centers. For each test, I construct the sample of students who take that test. The mathematics exam has two variants, corresponding to different (sets of) high school tracks. I assume that students taking the same variant of the test are ordered alphabetically. The final sample for the RLL test, which will be used for most of the results, contains 1,281,969 observations.

---

12 Table 1 also highlights the disproportionately high percentage of students achieving exactly the passing grade (5.00 out of 10.00; all grades in increments of 0.05 are possible).
13 This data was downloaded from the ministry’s website by Coman (2014).
14 Exam centers are supposed to have between 250 and 450 exam takers. In a few cases the same high school is allocated to two exam centers; in those cases I drop both exam centers.
15 2,059 observations (or 1.7%) were dropped because the coding for the first and middle name differed.
16 See Pop-Eleches and Urquiola (2013) for more details on the high school assignment system.
17 Some high schools are assigned to two exam centers. Given that I do not know which students go to each exam center, I drop all exam centers that contain at least one school that was split. This eliminates 5.6% of the observations.
3.2 Compliance with alphabetic seating order

The methodology of the baccalaureate exam, published every year by the national education ministry, states that in each exam center, and separately for each test, students should be assigned to exam rooms and desks in alphabetic order by last name. Anecdotal evidence shows that desks are pre-labelled by exam organizers with student’s names, which indicates that this rule is generally followed. However, in practice exam centers may also decide to stratify by variables such as part-/full-time students and student track, before ordering students. In addition, they may sometimes fill up an exam room with students taking different exams, to economize on the number of supervising teachers. Ideally I would use the actual seating arrangement by exam room. Unfortunately, this data is currently not publicly available, and thus for each exam center and each test, within the analysis sample, I order students by last name and use that intended ordering.

Using the intended seating order may lead to two types of errors. The first occurs when in reality two groups were ordered separately and I assume they are ordered as one group. In this case there will be students from different exam rooms who appear to be seated next to each other in my data. This will tend to bias the correlation between neighbors towards zero, although it may also increase the variance. The magnitude of this effect may be very large. The second type of error occurs if my sample selection is more stringent than in reality. In this case students who in reality are seated two or more desks away will appear to be neighbors in my data. Intuitively, this will be less of a problem compared to the first type of error.

For two exam centers, one in 2013 and one in 2009, I was able to obtain the actual assignment to exam rooms. Figure 1 compares the actual seating order with the intended seating order, for the Romanian Language and Literature (RLL), within the analysis sample described in the previous section. In both cases, we notice that the first type of error occurs. Apart from this issue the correspondence is good.

The problems indicated in Figure 1 suggest that I may be underestimating the level of cheating. Nevertheless, in the following section I find both high levels of cheating and a clear temporal pattern. This suggests that, although noisy, using the intended seating order contains valuable information.

---

18 In both cases, this occurs because a high school that was assigned to the exam center does not appear in the actual list published by the exam center. This indicates that the ministry’s dataset indicating the correspondence between high schools and exam centers contains errors.

19 Note that this exercise does not shed light on whether on exam day students actually sit in the order specified by the list. I believe this is generally the case, and any deviation will bias my results downwards.

20 For each exam center, I also ran three variants of the regressions used in Table 2: using analysis sample and intended order, analysis sample and actual order, and actual sample and order. In the third variant I also included exam room fixed effects. The results were qualitatively similar; none of the three cases exhibited a statistically significant correlation between neighbors’ scores.
4 Results

4.1 Relationship between scores of neighboring students

I begin by exploring the conditional expectation function (CEF) of student’s (demeaned) score $\tilde{x}_i$ on her neighbor’s score $\tilde{x}_{i+1}$. I begin by exploring the results at the Romanian Language and Literature (RLL) test. The first step is to model the mean of the observed score $x_i$. I first take out high school-year fixed effects – note that this implicitly removes exam center-year fixed effects as in the analysis sample each high school is assigned to a unique exam center. Secondly, I order students in each county in a given year alphabetically, and regress their scores on a quadratic of their rank.21 The residual from this regression is the main variable $\tilde{x}_i$ in all subsequent analysis.

Figure 2 presents the non-parametrically estimated CEF of $\tilde{x}_i$ on $\tilde{x}_{i+1}$ in 2009 (left) and 2012 (right). Each subfigure plots binned averages of $\tilde{x}_i$ for $\tilde{x}_{i+1}$ in a given bin. The 2009 subfigure shows a linear relationship with slope that appears significantly positive. In 2012, the relationship is noisier, and with a shallower slope. In Table 2 we explore these issues in regression form. Each column in Table 2 reports the results from an OLS regression of $\tilde{x}_i$ on $\tilde{x}_{i+1}$ in a given year, and standard errors are clustered at the exam center level. The coefficients are highly significant in every year.

The results in Figure 2 and Table 2 show that neighboring students’ scores are highly correlated, and there is significant variation in the size of the coefficients across time. This is suggestive evidence for the existence of cheating from neighbors, and also suggestive that changes due to the exam reform had an impact.

An immediate concern is whether these correlations are exclusively due to differences in average scores between different exam rooms. I cannot control directly for exam room effects because I do not observe exam room assignment in the data. To address this concern, Figure 3 presents the OLS regression coefficients from regressions of $\tilde{x}_i$ on $\tilde{x}_{i+d}$ for values of the distance $d$ between 1 and 15. This analysis is done separately by time period: panel A covers years 2006-2009, and panel B covers years 2012-2014; these periods roughly cover the pre- and post-reform years. In both cases the coefficients are precisely estimated and decreasing in $d$. Moreover, coefficients decrease steeply as a function of $d$. For example, in panel A the coefficient $\hat{\rho}_1$ is 38% higher than $\hat{\rho}_2$. By comparison, $\hat{\rho}_2$ is 17% higher than $\hat{\rho}_3$ and $\hat{\rho}_3$ is 6% higher than $\hat{\rho}_4$. This sharp initial decrease is inconsistent with a scenario where exam room effects are driving the entire correlation between students seated next to each other. In panel B, this pattern is even more pronounced. Starting from a lower level ($\hat{\rho}_1 = 0.033$), the first coefficient is roughly double the size of the second, while the coefficients for $d = 2, 3, 4, 5$ are of roughly equal size. In summary, Figure 3 shows that it is unlikely that the correlation between neighbors is entirely due to exam room effects.

4.2 Robustness exercises

In this section I address several specific hypotheses that the estimated correlation is spurious, that is, not caused by cheating.

One concern is whether students with similar names have similar scores for reasons that are independent of cheating. For example, a pair of students with similar or identical names may come from similar ethnic

---

21 The reason to allow $y_i$ to depend on the alphabetical name rank of student $i$ is that students with similar names might come from similar socio-economic background. It may be that students with exactly the same last name are particularly similar ex-ante, and will also be seated in adjacent desks. The fraction of students seated next to someone with identical last name is 16%. Table 3 presents results where I include last name fixed effects and/or exclude pairs with identical last names.
groups, they may have grown up in similar socio-economic environments, or they may have been in contact throughout high school. To address these concerns, I perform two robustness exercises.

In the first robustness check, I run the regression in the first column in Table 2, including common last name fixed effects, and dropping pairs of students with identical names. The common last name fixed effects are constructed as follows. I first determine the last names that appear at least 20 times in 2006 at the national level; these account for half of all observations. For each combination of such a name and one of the 42 counties in Romania, I construct a dummy variable. Table 3 presents the results. Column 1 is identical to the first column in Table 2. The second column includes common last name fixed effects, and we notice a 19% drop in the coefficient, and this difference is statistically significant. It is worth pointing out that including common name fixed effects runs the risk of mechanically downward biasing the OLS coefficient. To see why this may happen, imagine that two students with identical last names sit next to each other at the exam, and they are the only ones in their county with that last name. Including the fixed effects subtracts their mean score, which mechanically introduces a negative correlation between their scores. The third and fourth columns repeat the same analysis restricted to the sample of students i for whom their neighbor $i + 1$ does not have exactly the same last name. The coefficients are larger than in the second column, which suggests that the concern about over-controlling was justified, yet the coefficients are still around 8% smaller than in the first column, and this difference is statistically significant at the 5% level. In summary, common names account for at most a small fraction of the observed correlation.

The second exercise investigates whether neighbor students have similar scores because of common shocks during high school. Specifically, consider a pair of students from the same high school, who are consecutive in the last name alphabetical order in their high school. It is possible that, for example, during high school the two students were seated next to each other, in which case we would expect their scores to be correlated. Most of the time, such a pair of students will also sit next to each other at the exam, which would lead us to observe a correlation that is not due to cheating.

The naïve test for this hypothesis would seem to be to compare the correlation $\rho_1$ calculated for pairs of students from the same high school, and for pairs from different high schools. However, this comparison would be misleading if students from the same high school are more likely to cheat from each other. To disentangle these two effects (common high school shocks and more copying from high school peers), I use the fact that sometimes (high-school) consecutive students are not adjacent during the exam, because another student with a similar name from a different high school sits between them. Hence, I compute the correlation for pairs of students who are adjacent in high school and at the exam, and compare it to the correlation for pairs of students who are adjacent in high school but not at the exam. If the entire relationship between (exam center) adjacent students was driven by common high school shocks, the two correlations should be indistinguishable. To ensure a balanced comparison, the sample for this exercise is chosen as follows: in each exam center it contains all students i such that $i - 1$, $i$ and $i + 2$ are from the same high school, and $i + 1$ is from a different high school. 22

Table 4 presents the results. The results in panel A use the entire sample for 2006, whereas in panel B the sample is the comparison sample described above. Columns (1) and (2) report the bivariate regression coefficients on $\bar{x}_{i-1}$ and $\bar{x}_{i+1}$, respectively. The coefficients are indistinguishable in panel A, whereas in panel B the coefficient on $\bar{x}_{i+1}$ is significantly smaller than the coefficient on $\bar{x}_{i-1}$. Recall that $i$ and $i + 1$ are from the same high school, whereas $i + 1$ is from a different high school. At this point it is not clear

---

22 An alternative specification where $i - 2$, $i$ and $i + 1$ are from the same high school and $i - 1$ is from a different high school produces almost identical results.
whether students from the same high school cheat more, or have similar scores due to a common background.

Column (3) reports the bivariate regression coefficient on \( \hat{x}_{i+2} \). If there was no cheating and score similarities were entirely due to common shocks in high school, we would expect the results in panel B, columns (1) and (3), to be very similar. Instead, the coefficient in column (3) is 48% smaller than the coefficient in column (1), and this difference is highly statistically significant. Column (4) includes all three neighbor’s scores in the same regression; the results are essentially unchanged.

The results in Table 4 support two conclusions. First, we easily reject that the observed correlation is due to common high school background. Second, the evidence suggests that students from the same high school who sit next to each other are more likely to cheat.

The final robustness exercise is to check that neighboring students are balanced in terms of prior ability. Table 5 shows the results for the sample of students from 2009 for whom there is data on middle school GPA and the high school admission exam result from 2005. Columns (1)-(3) replicate the correlation between neighbor’s baccalaureate scores on this restricted sample, and establish the fact that middle school GPA and high school admission exam scores are highly predictive of baccalaureate exam scores, conditional on high school fixed effects. Hence, for students within the same high school, their pre-high school ability is strongly associated with their baccalaureate performance.

Column (4) reports the correlation in middle school GPA for students seated next to each other at the baccalaureate exam. The result is a precisely estimated zero, which shows that it is not prior ability that explains why neighbors have similar baccalaureate scores. Column (5) does the same for high school admission exam scores; here we find a statistically significant coefficient, which is nonetheless roughly eight times lower than the coefficient in column (1). It is quite likely that this correlation is itself due to cheating at the high school admission exam, although more data and analysis are necessary to substantiate this point. Indeed, the high school admission exam is also organized in regional exam centers, and within each exam center students are seated in alphabetical order. Therefore, some of the students seated next to each other in 2009 (at the baccalaureate) also sat next to each other in 2005 (at the high school admission exam). Thus, it is entirely possible that the correlation in column (5) is driven by pairs of students who sat next to each other in 2005, and who also cheated from each other. Unfortunately, I do not have the assignment of students to exam centers in 2005, although I hope to address this point in future work.

Overall, the three robustness exercises show that neither similarity in names, nor common shocks before or during high school can explain the correlation between the scores of neighboring students at the baccalaureate exam. This is evidence in favor of interpreting the latter as evidence of cheating through sharing of information.

### 4.3 Cheating and student characteristics

In section 2.2 I showed that we can use data on student characteristics to measure cheating heterogeneity. This section performs this exercise using the gender of the student and his/her neighbor, as well as using the middle school GPA.

For gender, I estimate the equation \( E(\hat{x}_i | \hat{x}_{i+1}) = \alpha + \rho_1 \hat{x}_{i+1} \) allowing the coefficient \( \rho_1 \) to be different for every pair of genders of \( i \) and \( i+1 \). Table 6 presents the results. Column (1) shows that, conditional of high school fixed effects, the genders of neighboring students are not correlated. Column (2) reports the coefficient \( \rho_1 \) for the four subgroups of neighbors. The results indicate that the genders of both neighboring

23 See the exam organization methodology Education Ministry (2004).
students are differentially associated with the amount of cheating. Students of the same gender cheat more than students of different genders, and boys cheat slightly more than girls. Quantitatively, the coefficient for two neighboring boys is approximately 30% higher than the coefficient for a boy and a girl. This type of analysis can be repeated with other binary or discrete characteristics, such as ethnicity and socioeconomic status.

I also show results using a continuous characteristic: the students’ GPA in middle school. In this case, I run the regression:

$$E(\tilde{x}_i|\tilde{x}_{i+1}) = \alpha + \rho_1 \tilde{x}_{i+1} + \beta z_i + \delta \tilde{x}_{i+1} \times z_i$$

The coefficient $\beta$ measures the direct association between middle school GPA and baccalaureate scores (within a high school). As mentioned in the previous section, we expect $\beta$ to be positive and highly significant. The coefficient $\delta$ measures whether the correlation with the neighbor’s score is higher or lower for students with higher middle school GPA.

Table 7 presents the results for middle school GPA. Column (1) shows the standard correlation in the full sample. Column (2) shows that middle school GPA $z_i$ has a significant, positive effect on the baccalaureate score. Columns (3) includes the interaction $\tilde{x}_{i+1} \times z_i$; the estimated coefficient is negative (and statistically significant). Quantitatively, this implies that a student with middle school GPA one standard deviation above average will have a coefficient $\rho_1$ that is 50% lower than the average $\rho_1$.

The exercises reported in Tables 6 and 7 showed that cheating varies depending on the characteristics of the students sitting next to each other. These relationships have expected signs, and are quantitatively important (30-50% changes in the cheating coefficient $\rho_1$ for different types of students).

5 An anti-fraud exam reform and cheating levels

The results so far establish that there is a strong correlation between the scores of students seated next to each other during the exam, that this correlation is significantly weaker for pairs of students sitting a few desk away, and that the correlation is not driven by predetermined characteristics such as similar names or common shocks experienced in high school.

In this section, I further investigate the changes in cheating across time, in relation to the anti-fraud exam reform in 2011. I use a slightly different method to look at levels and changes in the correlation between neighbors across time. Using the demeaned scores $\tilde{x}_i$ defined in section 4.1, for each exam center and year, I compute the correlation $\rho_1$ between adjacent students, and the correlation $\rho_3$ between students who sit two desks apart. In each year, I then calculate the average $\rho_1$ or $\rho_1 - \rho_3$ in the country, and plot the time series. Standard errors in this exercise are clustered at the county level.

The results are shown in Figure 4. The top two panels show results for RLL, the middle panels show results for the Math I test, and the bottom panels show results for the Math II test. Left panels show results for

24 The middle school GPA has an average of 8.8 with a standard deviation of 0.8. Hence the coefficient $\rho_1$ corresponding to GPA 9.6 is equal to $0.078 = 0.118 - 0.47 \times 0.8$.

25 Results are very similar in a panel regression with county or exam center fixed effects (results not shown). The benefit of not including fixed effects is that the magnitude of the correlation is easier to interpret.

26 Math I is taken by the Math-CS theoretic and the vocational military high school tracks. Math II is taken by the Natural Sciences theoretic and technological tracks. The tracks taking Math I are generally considered higher ability.
\( \rho_1 \), while right panel show results for \( \rho_1 - \rho_3 \). All graphs share the same broad pattern: high cheating levels until 2010-2011 and a sharp and persistent drop in 2012.

For RLL and \( \rho_1 \) (panel A), there is small drop in 2010, no drop in 2011, and a significant and persistent drop in 2012. For RLL and \( \rho_1 - \rho_3 \) (panel B), there is no clear trend in the period until 2011, and then there is a sharp drop in 2012, which seems to reverse slightly in 2013 and 2014. Roughly speaking, the fact that \( \rho_1 \) is decreasing and \( \rho_1 - \rho_3 \) is increasing after 2012 suggests that cheating is both decreasing overall, and becoming more localized to immediate neighbors.

The pattern for Math I and Math II are similar (although the levels are higher for the latter). For \( \rho_1 \) we observe a peak in 2010, a partial decrease in 2011, and further decrease after 2012. This fits with the discussion in Borcan et al (2014) and Borcan et al (2015), who describe the 2010 as the “Xeroxed Exam.” For \( \rho_1 - \rho_3 \), there is no clear pattern until 2010, and a gradual decrease beginning in 2011.

Overall, the results in Figure 4 point to a clear reduction in cheating around the time of the changes instituted by the reform. However, it is interesting to note that while the bulk of the reform occurred leading up to the 2011 exam, and scores and passing rates were significantly lower in 2011, the clear drop in cheating only starts in 2012, one year later. One possible explanation to this puzzle is that scores were lower in 2011 not necessarily because of lower levels of cheating (as measured by \( \rho_1 \) and \( \rho_1 - \rho_3 \)); instead, lower scores might have been due to reductions in other types of fraud, and due to more demanding grading. Furthermore, the decrease in scores in 2011 might have acted as a signal to students, who interpreted this as a signal that cheating will not be tolerated. A related possible explanation is that the punishments that occurred during the 2011 exam, namely the higher number of students eliminated from the exam, alerted the next cohort of students that cheating is more likely to be punished.

I now ask if after the reform, scores fell more in exam centers that had higher baseline levels of cheating. This would be the case, if reductions in scores were in part due to reductions in cheating in places with high levels of cheating. I implement this test through a differences-in-differences approach. I first restrict the sample to exam centers that appear in the data in every year between 2006 and 2014, roughly around 35% of all exam centers. To define baseline cheating I consider the \( \rho_1 \) measure of a given exam center in 2006, and I construct a continuous measure by normalizing this into a z-score. I also define a binary measure that is equal to 1 (or 0) if \( \rho_1 \) is in the top (or bottom) 25% of the distribution in 2006. I run the following regression

\[
x_{ict} = \gamma_c + \delta POST_t + \mu R_c \times POST_t + \varepsilon_{ict}
\]

Above, \( x_{ict} \) is the RLL score of student \( i \) in center \( c \) and year \( t \), \( POST_t \) is a dummy for \( t \geq 2011 \), and \( R_c \) is the continuous or binary baseline cheating measure for exam center \( c \). As a robustness check, I also run the same equation including linear time trends:

\[
x_{ict} = \gamma_c + \delta POST_t + \mu R_c \times POST_t + \psi t + \xi R_c \times t + \varepsilon_{ict}
\]

Figure 5 presents the results graphically. It plots in red the yearly average RLL scores for the group of exam centers with \( \rho_1 \) in the top 25% (high cheaters), and in black the same quantity for the group of exam centers with \( \rho_1 \) in the bottom 25% (low cheaters). The first thing to note is the overall drop in scores starting in 2011, which has already been documented in Table 1. At first sight it looks like the drop is larger for the high cheating group, which supports the idea that the reform (partly) decreases scores by reducing cheating at high cheating exam centers. However, upon closer inspection it becomes visible that the two groups are steadily moving apart, starting in 2007; in other words, it appears that the two groups are on different trend lines. The suspicion of different trend lines is supported by the regression results presented in Table 8.
Columns (1) and (3) show results from the first regression equation above (without time trends) on the period 2007-2014, and in both cases we notice a significant negative effect on the interaction term $R_c \times POST_t$. However, columns (2) and (4) show the results from running the second regression (with time trends), and the effects on the $R_c \times POST_t$ term disappear. Results are similar for the mathematics tests, if I define baseline cheating in terms of 2009 cheating, or as an average of cheating measures between 2006 and 2009, or if I use the $\rho_1 - \rho_3$ measure instead of $\rho_1$. Results also do not change if I define the $POST_t$ dummy starting from 2012.

Taken together, these results suggest that, given the available evidence, we cannot conclude that the reduction in scores following the reform was due to reductions in cheating by copying, as measured by the $\rho_1$ indicator. Indeed, it appears that exam centers that have high levels of cheating were on a more negative trend in terms of their baccalaureate scores even before the reform, compared the exam centers with lower levels of cheating.
6 References

Angrist J., Battistin E. and Vuri, Daniela (2014), In a Small Moment: Class Size and Moral Hazard in the Mezzogiorno, NBER working paper #20173.
Education Ministry (2004), Metodologia de Organizare și Desfășurare a Testelor Naționale în Vederea Accesului Absolvenților clasei a VIII-a în Clasa a IX-a a anului școlar 2005-2006 [The Organization Methodology for the National Tests for High School Admission for 8th Grade Graduates], Available at http://www.edu.ro/index.php/met_rep_reg_ins_etc/118
Romero, Mauricio, Riascos, Alvaro, and Jara, Diego (2012), Answer-copying index comparison and massive cheating detection, working paper.
### Table 1. Descriptive Statistics on the Romanian Baccalaureate Exam, 2006-2014

<table>
<thead>
<tr>
<th>Year</th>
<th>Number of Students</th>
<th>Number of High Schools</th>
<th>Number of Exam Centers</th>
<th>Exam Pass Rate</th>
<th>RRL average score</th>
<th>RRL Pass Rate</th>
<th>RRL Exactly Pass Grade</th>
<th>Math average score</th>
<th>Math Pass Rate</th>
<th>Math Exactly Pass Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>2006</td>
<td>191,371</td>
<td>1,390</td>
<td>568</td>
<td>78.8%</td>
<td>6.5</td>
<td>85.9%</td>
<td>10.8%</td>
<td>6.8</td>
<td>87.6%</td>
<td>5.6%</td>
</tr>
<tr>
<td>2007</td>
<td>198,965</td>
<td>1,402</td>
<td>561</td>
<td>80.3%</td>
<td>6.6</td>
<td>87.9%</td>
<td>9.4%</td>
<td>7.1</td>
<td>88.3%</td>
<td>4.4%</td>
</tr>
<tr>
<td>2008</td>
<td>225,878</td>
<td>1,427</td>
<td>634</td>
<td>76.5%</td>
<td>6.9</td>
<td>89.1%</td>
<td>6.4%</td>
<td>6.5</td>
<td>81.0%</td>
<td>6.0%</td>
</tr>
<tr>
<td>2009</td>
<td>213,389</td>
<td>1,448</td>
<td>614</td>
<td>79.6%</td>
<td>6.7</td>
<td>88.5%</td>
<td>8.0%</td>
<td>6.7</td>
<td>84.1%</td>
<td>5.8%</td>
</tr>
<tr>
<td>2010</td>
<td>210,089</td>
<td>1,447</td>
<td>627</td>
<td>66.7%</td>
<td>6.9</td>
<td>90.9%</td>
<td>5.9%</td>
<td>6.0</td>
<td>75.0%</td>
<td>7.5%</td>
</tr>
<tr>
<td>2011</td>
<td>212,762</td>
<td>1,459</td>
<td>619</td>
<td>43.6%</td>
<td>6.0</td>
<td>76.3%</td>
<td>8.3%</td>
<td>4.6</td>
<td>52.5%</td>
<td>8.7%</td>
</tr>
<tr>
<td>2012</td>
<td>200,187</td>
<td>1,460</td>
<td>611</td>
<td>41.4%</td>
<td>6.0</td>
<td>74.4%</td>
<td>8.0%</td>
<td>4.3</td>
<td>45.2%</td>
<td>7.7%</td>
</tr>
<tr>
<td>2013</td>
<td>188,842</td>
<td>1,518</td>
<td>582</td>
<td>52.3%</td>
<td>6.1</td>
<td>76.0%</td>
<td>8.2%</td>
<td>5.7</td>
<td>64.6%</td>
<td>7.6%</td>
</tr>
<tr>
<td>2014</td>
<td>161,360</td>
<td>1,538</td>
<td>521</td>
<td>56.7%</td>
<td>6.2</td>
<td>78.7%</td>
<td>7.6%</td>
<td>6.2</td>
<td>70.7%</td>
<td>6.5%</td>
</tr>
</tbody>
</table>
## Table 2. Student RLL score as function of neighbor’s score

<table>
<thead>
<tr>
<th>Year</th>
<th>2006</th>
<th>2007</th>
<th>2008</th>
<th>2009</th>
<th>2010</th>
<th>2011</th>
<th>2012</th>
<th>2013</th>
<th>2014</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
<td>(2)</td>
<td>(3)</td>
<td>(4)</td>
<td>(5)</td>
<td>(6)</td>
<td>(7)</td>
<td>(8)</td>
<td>(9)</td>
</tr>
<tr>
<td>RLL Score $\bar{x}_{i+1}$</td>
<td>0.113***</td>
<td>0.141***</td>
<td>0.118***</td>
<td>0.123***</td>
<td>0.086***</td>
<td>0.086***</td>
<td>0.035***</td>
<td>0.038***</td>
<td>0.023***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
</tr>
<tr>
<td>Observations</td>
<td>126,588</td>
<td>126,602</td>
<td>145,276</td>
<td>138,922</td>
<td>139,498</td>
<td>132,765</td>
<td>134,526</td>
<td>130,610</td>
<td>115,024</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.011</td>
<td>0.018</td>
<td>0.011</td>
<td>0.012</td>
<td>0.006</td>
<td>0.006</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Notes: Each column reports the coefficient for $\bar{x}_{i+1}$ from an OLS regression in a given year of the student’s demeaned Romanian language and literature score $\bar{x}_i$ on $\bar{x}_{i+1}$. Observations with $\bar{x}_{i+1}$ below the bottom percentile or above the top percentile of the distribution are dropped. Robust standard errors, clustered at the exam center level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Table 3. Student RLL score as function of neighbor’s score, with last name fixed effects and dropping identical names

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLL score $\tilde{x}_{i+1}$</td>
<td>0.113***</td>
<td>0.095***</td>
<td>0.106***</td>
<td>0.105***</td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td>Drop if $i$ and $i + 1$ have same last name</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Common last name $\times$ county fixed effects</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>126,588</td>
<td>126,588</td>
<td>105,816</td>
<td>105,816</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.011</td>
<td>0.138</td>
<td>0.010</td>
<td>0.157</td>
</tr>
</tbody>
</table>

Notes. Each column reports the coefficient for $\tilde{x}_{i+1}$ from an OLS regression in the year 2006, of the student’s demeaned Romanian language and literature score $\tilde{x}_i$ on $\tilde{x}_{i+1}$. Columns (2) and (4) include county level fixed effects for every common last name (defined as last names which appear at least 20 times nationally). Observations with $\tilde{x}_{i+1}$ below the bottom percentile or above the top percentile of the distribution are dropped from the sample. The sample in columns (3) and (4) also students $i$ that have the same last name with $i + 1$. Robust standard errors, clustered at the exam center level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Table 4. Scores of (high school) adjacent students, who are or are not adjacent at the exam

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: RLL score $\tilde{x}_i$ (school $S$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(1)</td>
</tr>
<tr>
<td>RLL score $\tilde{x}_{i-1}$</td>
<td>0.116***</td>
</tr>
<tr>
<td>RLL score $\tilde{x}_{i+1}$</td>
<td>0.112***</td>
</tr>
<tr>
<td>RLL score $\tilde{x}_{i+2}$</td>
<td>0.076***</td>
</tr>
<tr>
<td>Observations</td>
<td>116,762</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.012</td>
</tr>
</tbody>
</table>

Panel A. Full Sample (2006)

|                  | (1)            | (2)            | (3)            | (4)            |
| RLL score $\tilde{x}_{i-1}$ (school $S$) | 0.150*** | 0.141*** | (0.013) | (0.012) |
| RLL score $\tilde{x}_{i+1}$ (school $S' \neq S$) | 0.094*** | 0.078*** | (0.014) | (0.013) |
| RLL score $\tilde{x}_{i+2}$ (school $S$) | 0.078*** | 0.061*** | (0.012) | (0.011) |
| Observations     | 10,286         | 10,286         | 10,286         | 10,286         |
| R-squared        | 0.019         | 0.008         | 0.005         | 0.029         |

Notes. Panel A reports the relationship between the scores of students seated at distance 1 and 2. Panel B reports the relationship between the scores of (high school) adjacent students, who are or are not adjacent at the exam. Panel A uses the full sample in 2006, and panel B contains all students $i$ such that $i - 1$, $i$, and $i + 2$ are from the same school $S$, and $i + 1$ is from a different school $S' \neq S$. Columns (1)-(3) report bivariate regression coefficients of $\tilde{x}_i$ on $\tilde{x}_{i-1}$, $\tilde{x}_{i+1}$, and $\tilde{x}_{i+2}$, respectively. Column (4) reports the regression coefficients of $\tilde{x}_i$ on $\tilde{x}_{i-1}$, $\tilde{x}_{i+1}$, and $\tilde{x}_{i+2}$. Robust standard errors, clustered at the exam center level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Table 5. Past scores and GPAs of adjacent students

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RRL score $\tilde{x}_{i+1}$</td>
<td>GPA $z_{i+1}$</td>
<td>HSAE $w_{i+1}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRL score $\tilde{x}_{i+1}$</td>
<td>0.129*** (0.007)</td>
<td>0.129*** (0.007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Middle School GPA $z_i$</td>
<td>0.475*** (0.015)</td>
<td>0.476*** (0.015)</td>
<td>-0.003 (0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School Admission Exam Score $w_i$</td>
<td>0.410*** (0.013)</td>
<td>0.409*** (0.013)</td>
<td>0.017*** (0.006)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>High School FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>44,368</td>
<td>44,368</td>
<td>44,368</td>
<td>44,603</td>
<td>44,603</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.068</td>
<td>0.180</td>
<td>0.196</td>
<td>0.373</td>
<td>0.491</td>
</tr>
</tbody>
</table>

Notes. The sample for this table is students who took the Baccalaureate exam in 2009, for whom it was possible to match results from the high school exam entrance in 2005 and middle school GPA. All regressions include high school fixed effects. Robust standard errors, clustered at the exam center level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1
<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>male_{i+1}</td>
<td>ẋ_{i}</td>
</tr>
<tr>
<td>male_{i}</td>
<td>0.001</td>
<td>-0.570***</td>
</tr>
<tr>
<td></td>
<td>(0.001)</td>
<td>(0.007)</td>
</tr>
<tr>
<td>male_{i+1}</td>
<td></td>
<td>0.053***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
</tr>
<tr>
<td>male_{i} * male_{i+1}</td>
<td></td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
</tr>
<tr>
<td>ẋ_{i+1} * female_{i} * female_{i+1}</td>
<td></td>
<td>0.119***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>ẋ_{i+1} * female_{i} * male_{i+1}</td>
<td></td>
<td>0.102***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>ẋ_{i+1} * male_{i} * female_{i+1}</td>
<td></td>
<td>0.106***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
<tr>
<td>ẋ_{i+1} * male_{i} * male_{i+1}</td>
<td></td>
<td>0.135***</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
</tr>
</tbody>
</table>

High School FE | Yes | Yes |
Observations   | 640,691 | 627,147 |
R-squared      | 0.055 | 0.058 |

Notes. The sample for this table is all exam results up to 2010 for students with first names among the most popular 500 first names (which cover 95% of all students). All regressions include high school fixed effects. Robust standard errors, clustered at the exam center level, are reported in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Table 7. Cheating Heterogeneity by Middle School GPA

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>RLL score $\bar{x}_{i+1}$</td>
<td>0.118***</td>
<td>0.529***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.005)</td>
<td>(0.046)</td>
<td></td>
</tr>
<tr>
<td>Middle school GPA $z_i$</td>
<td>0.656***</td>
<td>0.657***</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.015)</td>
<td>(0.015)</td>
<td></td>
</tr>
<tr>
<td>$\bar{x}_{i+1} \times z_i$</td>
<td></td>
<td>-0.047***</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td></td>
</tr>
<tr>
<td>High School FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>142,977</td>
<td>76,634</td>
<td>76,634</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.014</td>
<td>0.124</td>
<td>0.139</td>
</tr>
</tbody>
</table>

Notes. The sample in column (1) is all observations in 2009, and in columns (2) and (3) only the students who were successfully matched to middle school GPA records. Observations with $\bar{x}_{i+1}$ below the bottom percentile or above the top percentile of the distribution are dropped from the sample. All regressions include high school fixed effects. Robust standard errors, clustered at the exam center level, in parentheses. *** p<0.01, ** p<0.05, * p<0.1
Table 8. RLL Scores in Exam Centers with High/Low Baseline Levels of Cheating

<table>
<thead>
<tr>
<th></th>
<th>Dependent variable: RLL score $x_{ict}$</th>
<th>$R_c$ continuous measure</th>
<th>$R_c$ discrete measure</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(1)</td>
<td>(2)</td>
</tr>
<tr>
<td>$POST_t$</td>
<td>-0.744***</td>
<td>-0.815***</td>
<td>-0.610***</td>
</tr>
<tr>
<td></td>
<td>(0.047)</td>
<td>(0.046)</td>
<td>(0.089)</td>
</tr>
<tr>
<td>$t$</td>
<td>0.018</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_c \times POST_t$</td>
<td>-0.174***</td>
<td>-0.038</td>
<td>-0.407***</td>
</tr>
<tr>
<td></td>
<td>(0.046)</td>
<td>(0.047)</td>
<td>(0.123)</td>
</tr>
<tr>
<td>$R_c \times t$</td>
<td>-0.034***</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exam center c FE</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Observations</td>
<td>379,203</td>
<td>379,203</td>
<td>204,947</td>
</tr>
<tr>
<td>R-squared</td>
<td>0.260</td>
<td>0.260</td>
<td>0.244</td>
</tr>
</tbody>
</table>

Notes. Each column in each panel reports the coefficients from a regression of RLL score $x_{ict}$ of student $i$ in exam center $c$ in year $t$ on exam center fixed effects, a $POST_t$ dummy equal to one from $t = 2011$ onwards, and the interaction between $POST_t$ and a measure of cheating in exam center $c$ in 2006, $R_c$. Columns (2) and (4) also contain linear time trends and their interaction with $R_c$. In columns (1) and (2), the cheating measure $R_c$ is the z-score of the exam center $\rho_{i,2006}^c$ measure in 2006; in columns (3) and (4) $R_c$ is equal to 1 if $\rho_{i,2006}^c$ is in the top 25% of the distribution in 2006, equal to 0 if $\rho_{i,2006}^c$ is in the bottom 25% of the distribution, and missing otherwise. The sample is restricted to the 35% of exam centers that appear in every year between 2006 and 2014. Robust standard errors, clustered at the exam center level, in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
Figure 1. Intended and Actual Seating order in Two Exam Centers

Notes: Each figure presents the correspondence between the actual seating order (provided by two exam centers) and the implied seating order (computed on the sample used in the analysis), for two high schools. The analysis sample is defined in section 3.1. The X axis reports the position in the actual seating order, and the Y axis reports the implied seating order. Students who appear in the latter but not in the actual seating list are ordered alphabetically, and placed after all the students who appear in the list.
Notes: Each figure plots averages of $\bar{x}_{i+1}$ for $\bar{x}_i$ in the corresponding bin. The bars indicate 95% confidence intervals around the estimated conditional mean ($\pm 1.96$ SE), not clustered. The score $\bar{x}_i$ is obtained from the observed score $x_i$ net of high school-year fixed effects, and a separate quadratic function in alphabetical rank in each county and year.
Figure 3. Relationship between student scores by seating distance, 2006-2009 and 2012-2014

Panel (A) Years 2006-2009

Panel (B) Years 2012-2014

Notes: In each panel, the $d^{th}$ bar reports the value and 95% confidence interval around the coefficient for $\tilde{x}_{i+d}$ from an OLS regression of the student’s demeaned Romanian language and literature score $\tilde{x}_i$ on $\tilde{x}_{i+d}$. The sample in panels A and B is years 2006-2009, and years 2012-2014, respectively. Observations with $\tilde{x}_{i+d}$ below the bottom percentile or above the top percentile of the distribution are dropped. The number of observations varies between 537,388 ($d = 1$) to 507,959 ($d = 15$) in panel A, and between 380,160 and 358,987 in panel B. Confidence intervals are constructed using robust standard errors clustered at the exam center level.
Figure 4. Exam center level correlation and relative correlation

Notes: For each exam center and year, \( \rho_d \) is the correlation between (demeaned) RLL scores of students sitting \( d \) desks away. Panels (A), (C) and (E) plot mean \( \rho_1 \) across time, and panels (B), (D) and (F) plot mean \( \rho_1 - \rho_3 \) across time. Robust standard errors are clustered at the county level.
Notes: This figure shows RLL test scores in high schools that were in the top 25% of the cheating distribution in 2006 (red) and those in the bottom 25% (black). Gray error bars represent 95% confidence intervals derived from robust standard errors clustered at the county level.