



**HAL**  
open science

# The role of prior warnings when cheating is easy and punishment is credible

Marc Humbert, Eric Villard, Xavier Lambin

► **To cite this version:**

Marc Humbert, Eric Villard, Xavier Lambin. The role of prior warnings when cheating is easy and punishment is credible. *Information Economics and Policy*, 2022, 58, pp.100959. 10.1016/j.infoecopol.2021.100959 . hal-04321116

**HAL Id: hal-04321116**

**<https://hal.science/hal-04321116>**

Submitted on 22 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

# The Role of Prior Warnings When Cheating Is Easy and Punishment Is Credible

Marc Humbert<sup>1</sup>, Xavier Lambin<sup>2</sup>, Eric Villard<sup>1</sup>  
*Preliminary working document. Please do not share.*

## Abstract

During the COVID-19 sanitary crisis, many exams were hastily moved to online mode. This revived a much needed debate over the privacy issues associated with online proctoring of exams, while the validity and fairness of unproctored exams were increasingly questioned. With a randomized control trial, we estimate the effectiveness of prior warnings as a means of discouraging academic dishonesty in exams. We use original, non-intrusive technologies to surreptitiously identify cheating in a series of unproctored assignments and send a targeted warning to half of the students who were identified as cheaters. We then compare their cheating behavior on the final exam with the behavior of the group of unwarned cheaters. The warning proves effective but does not completely eliminate cheating, as some students' cheating strategies become more sophisticated following issuance of the warnings. We conclude that switching traditional exams to online mode should be accompanied by proctoring. When proctoring is not possible, credible and effective anti-cheating technologies should be deployed together with adequate warnings.

---

<sup>1</sup> Grenoble Ecole de Management, 12 rue Pierre Semard, F-38000 Grenoble, France.

<sup>2</sup> ESSEC Business School and THEMA, 3 Av. Bernard Hirsch, B.P. – 50105, Cergy, 95021, France. Please send correspondence to : [lambin@essec.edu](mailto:lambin@essec.edu)

**Keywords:** Education, Online Exams, Unproctored Exams, Prior Warnings, Nudges, Randomized Control Trial

**JEL codes:** D8, I20, I21, I23.

## 1- Introduction

Online education has experienced sustained growth in recent decades. The 2020 global sanitary crisis caused by the COVID-19 virus suddenly rendered remote learning ubiquitous and paved the way for even more extensive use in the future. Naturally, these abrupt developments stimulated active debates over the benefits of online teaching and the associated risks, in particular the issue of academic dishonesty on distance exams. The stakes go beyond the already crucial issue of fairness in education, as several authors have noted a strong correlation between academic and professional dishonesty (Becker et al, 2006; Brodowsky et al, 2019).

Following the observation that unproctored online exams result in extensive cheating (Norris, 2019; Holden et al. 2020), several strategies have been proposed. The randomization of questions (see e.g. Li, 2020; Nizam, 2020), when implementable, provides satisfying results but raises issues of fairness between students facing distinct sets of questions. This process also has technical limits, because an examiner may not be able to find enough variations of a given question to avoid repetition. Online proctoring is also a popular solution (Biometric: Chirumamilla, A., & Sindre, G. (2019); Mellar et al, 2018; Bawartith et al, 2017; Idemudia, 2016; lockdown browsers : Halbehr et al, 2014) but faces strong public opposition reflecting concerns over students' access to the necessary technologies (such as a webcam or a stable internet connection) and, most importantly, privacy.

We took advantage of the specific conditions under which online exams were administered during the COVID-19 crisis, where cheating was predictably widespread, to develop a new strategy to discourage cheating. The strategy, based on friendly warnings sent to students, is respectful of student privacy and does not require specific equipment. In this research, we seek an answer to the following research question “can targeted warnings discourage students from cheating on exams?”. We restrict our analysis to cheating in the form of illicit sharing of information between students.<sup>3</sup> To that aim, we surreptitiously analyze similarities between copies of pre-exam assignments and various trick questions to identify (illicit) collaborations. Students are not aware of these techniques, which makes very accurate identification of cheaters and their strategies possible.<sup>4</sup> To a randomly selected subgroup of students identified as cheaters on the assignments, we send a friendly warning stating that their copies were suspicious and reminding them that cheating on the final exam is prohibited. Mean comparisons indicate that only 9% of the members of our treated group (assignment-cheaters who received a warning) cheated on the final exam, down from 40% in the control group (unwarned assignment-cheaters). Controlling for various variables, the warning sent to cheaters decreases the cheating rate by 23 percentage points. By design of the experiment, this effect is causal. A warned cheater has a probability to cheat on the final

---

<sup>3</sup> In more general settings, cheating can be defined as “a rule-breaking behavior that is exhibited with the intention of gaining an unfair advantage over a party or parties with whom the cheater is associated through a norm-governed relationship” (Green, 2004, p140).

<sup>4</sup> The surreptitious method has the advantage of not relying on students' candidness but makes cheating difficult to prove with certainty. Further, it may seem like entrapment. For this reason, the similarity analysis was used only for the purpose of this article and did not result in consequences for students.

exam 4 percentage points *lower* than a similar student who was not identified as assignment-cheater. We conclude that warnings are effective in discouraging cheating, insofar as warned cheaters behave similarly to non-cheaters. Cheating is, however, not eliminated entirely. We observe a similar effect when we restrict the analysis to “leaders” (students who gave their exam copies to other students). This effect is, however, subject to measurement errors, and properly testing leadership status would require a larger dataset to gain significance.

To the best of our knowledge the present paper is the first to exploit observational methods to analyze the dynamics of cheating in a series of distance exams. It is the first to analyze how suspected students respond to a targeted nudge warning. The experiment was pre-registered at the AEA’s Social Science Registry.<sup>5</sup> The rest of the paper is organized as follows. Section 2 reviews the literature. In Section 3 we describe our experimental setting. In Section 4 we report our results. Section 5 provides a discussion of our results and we conclude in Section 6.

## 2- Literature Review

This paper lies at the nexus between three very active streams of the literature.

First, it sheds some light on the prevalence of cheating in higher education. All aspects of the notorious “fraud triangle” (Ramos 2003; Becker et al 2006; King, Guyette, & Piotrowski, 2009) are present in our experimental setting and strengthened by the strict lockdown of

---

<sup>5</sup> RCT ID : AEARCTR-0005696. Available at <https://www.socialscienceregistry.org/trials/5696>

Spring 2020 in France. The fraud triangle outlines three components that contribute to increasing the risk of fraud. The first component is the “opportunity”. Cheating was exceptionally easy in the particular circumstances of the COVID-19 sanitary crisis. In Spring 2020, the entire country was under a severe lockdown, preventing any non-motivated trips. All institutions and in particular schools and universities were closed. Within that context all courses and exams in the institution where the experiment took place were moved to online mode, with no possibility of proctoring. Despite the physical distance between students who were all strictly confined at home, all channels of communication such as online messaging were accessible. This makes collaboration unusually easy. The second component “incentives” was also particularly salient: grades are a strong determinant of exchange opportunities in following years. These opportunities were likely to become scarcer as a result of anticipated travel restrictions, making cheating an appealing option. Third component is “rationalization”. Anecdotal feedback from students support the claim of a moral disengagement (Bandura, 2013) where cheating is rationalized by the fact that “everyone else does it”. This claim that was also widely supported by specialized and general media during the sanitary crisis. Finally, Becker et al. (2006) and McCabe and Treviño (1995) have reported that business students, which is our population, are consistently near the top of the rankings of students who are most likely to cheat.

In light of the fraud triangle, the reasons why students may cheat are quite clear. However, the mechanism through which students react to our warning is less evident. Our methodology does not allow attribute causality to a specific mechanism, but the literature described a few plausible explanations: the warning may make the threat more realistic (Bing

et al. 2011), but also have a normative effect on student's perception of rules (Cialdini, 2004), or focus their attention on the abnormality of their behavior (Cialdini, 2003).

Second, the paper offers a methodological contribution to a rich and expanding literature on curbing academic dishonesty in higher education. Academic dishonesty may take many forms and taking account of this variety goes beyond the scope of this report. A comprehensive overview of the main mechanisms of cheating may be found in McCabe (2001, 2016). We focus on cheating in the form of seeking outside help during an exam. Such cheating has become more prevalent and harder to detect over the past decades due to increased availability of information on the internet (Scanlon, 2003). The issue is likely to be even more acute when exams are taken at a distance without proctoring, as several studies have found that participants are less honest when they interact online than when they interact face-to-face (Rockmann and Northcraft, 2008; Van Zant and Kray, 2014). Reviews of literature from Norris (2019) and Holden et al. (2020) reveal that between 60% and 90% of students admit to have cheated on online exams. Contrary to most previous studies that are based on anonymous surveys administered after exams, we use original technologies to reveal cheating behavior. This brings about several improvements over methodologies that are commonly found in the existing literature. First, cheating is not stated (as in self-reports) but revealed, which eliminates the strong declarative biases of surveys, as reported in general settings by Sudman and Bradburn (1974) and Kerkvliet and Sigmund (1999). The bias is particularly strong in the case of cheating (Bing et al., 2011). With observational data, interesting statistical approaches at a group-level such as those found in Harmon and

Lambrinos (2008), Arnold (2016), D'Souza (2017), and Fendler et al. (2018) reduce this bias but fail to explain the cheating mechanisms.

In the present paper, cheating status is attributed to specific individuals, which allows us to analyse the cheating strategies in considerable detail.

Third, our paper contributes to the literature on the effectiveness of “nudges” (Thaler and Sunstein, 2008). Nudges have been widely analysed in the context of consumer choice (Allcott and Mullainathan, 2010; Ferraro et al., 2014; Beshears et al. 2011) but also showed encouraging results in education. As an example, Zamir et al. (2016) show that adequate deadline management by professors can encourage students self-benefiting behavior such as limiting procrastination and forgetfulness. The careful design of the nudge is a key determinant of its success, as is the prior understanding of the behavioral mechanism behind cheating (Damgaard and Nielsen, 2018). Auferoth (2020) shows that the effectiveness of reminders during exam preparation highly depends on the backgrounds of the students. Most studies on the effect of nudges of the propensity to cheat on exams, concern the “code of honor”. This rich literature shows strikingly mixed results (McCabe et al., 1999; Arnold et al., 2007, Konheim-Kalkstein et al, 2008; Mazar, 2008), making the “honor-code” at best a partial solution. Indeed, in a review of 63 economic and psychological studies, Rosenbaum et al. (2014) conclude that there “appear[s] to be a consistent proportion of unconditional cheaters and noncheaters [. . .], with the honesty of the remaining individuals being susceptible to a range of variables, most notably monitoring and intrinsic lying costs” (Rosenbaum et al. 2014, p. 194).



In the present paper, we reinforce the monitoring pressure on students suspect of cheating and implement an original nudge in the form of a warning prior to the exam. The warning informs some of the students that their professors suspect they cheated on the assignments and reminds them that cheating on the final exam will be penalized. Close to our work, Bing et al. (2012) and Corrigan-Gibbs et al. (2015) show that pre-exam warnings produce significantly greater reduction in cheating than “honor codes”. However, the warning is untargeted. In our study, members of the treated group receive individual warnings, which makes the threat of being caught considerably more realistic. And indeed we observe that our simple and inconsequential but targeted warning is remarkably more effective than the untargeted warnings of Bing et al. (2012) and Corrigan-Gibbs et al. (2015). We finally note that the high benefits (in the form of curbing cheating) and relatively low costs (in the form of designing our detection technology) of our strategy, is one of the main practical appeals of nudges (Madrian and Shea, 2001; Thaler and Benartzi, 2004).

### **3- Experimental setting**

Our data consist of exam copies for 644 undergrad students in a French business school. We examine their performance in a series of five tests in a programming class in Spring 2020. These tests account together for a very small share of the final grade (10%). They are used for pedagogical and participation purposes as well as in preparation for a final exam that accounts for most of the final grade (90%). The form of the test and exam are similar and consist in writing small pieces of code on an online platform. They differ only in length and topics. The exam covers the whole course and lasts 1 hour and 30 minutes, while the tests

cover course chapters and last about 60 minutes each. While students can do the tests any time over a 1-week time window, the final exam is synchronous (same time slot for all students). In the present paper we assimilate cheating to collaboration, i.e. situations when two or more students take the examination together or exchange answers.<sup>6</sup> Appendix A1 reports the communications between the professors and the students throughout the experiment: collaboration between students is very explicitly disallowed on both the tests and the exam.

We used copies of the last assignment before the exam (“Test 5”), that was very close in format to the final exam, to categorize students as assignment-cheaters or assignment-non-cheaters.<sup>7</sup> We used a totally non-intrusive technology to identify collaboration: we analysed the syntax to identify suspicious similarities between submitted copies. The method, which is based on exact text matching, Gestalt Pattern Matching (for textual syntax similarity) and Syntax Analysis (for abstract code similarity) and random questions, is described in more detail in Appendix A2. The method is probabilistic but we were able, with a high degree of confidence, to identify 233 assignment-cheaters of the 644 students. Between this last assignment and the final exam, a standard email was sent to all students reminding them of the rules of the exam as well as the sanction policy for cases of cheating. In addition to this

---

<sup>6</sup> We therefore restrict our analysis to cheating in the form of illicit sharing of information between students and exclude other types such as impersonation or using external help.

<sup>7</sup> As is described below, the identification of potential cheaters is based on Test 5 only. This is because collaboration was allowed in Tests 1 to 3. A few inattentive students might have kept collaborating in Test 4. Between Test 4 and Test 5 students were very explicitly warned that collaboration was forbidden (see Appendix A1). They were also notified that Test 5 would be longer and a bit harder than Tests 1 to 4, so as to better prepare them for the final exam. This explains why the average grade is lower in Test 5 than in previous tests.

email, half of the cheaters (117 students selected randomly from among the 233 cheaters) also received a warning stating that they had been identified and put on a watch list. There was no sanction applied at this point, but the warning reminded students that such behavior at the exam would be sanctioned. This is our treated group. The other half of the assignment-cheaters were not warned and received the same standard information as any other student. This constitutes our control group. The students who were not identified as cheaters on Test 5 constitute a “baseline” group, which we use as a benchmark. The standard email was sent two days after Test 5 and five days before the final exam. The treatment (sending the warning to half of the cheaters) was applied a few minutes after the standard email. The use of several distinct identification techniques for Test 5 and the exam (where we added a few “trap questions” described in Appendix A2) made it possible to circumvent sophisticated cheating strategies (Bachore 2016). The messages sent to students and more detail regarding our identification techniques can be found in Appendices A1 and A2, respectively. Table 1 displays our descriptive statistics. Though we believe the labels are transparent, a detailed description may be found in Appendix A3.

*Table 1: Descriptive Statistics*

	Full sample			Non-cheaters (benchmark group)			Unwarned cheaters (control group)			Warned cheaters (treated group)			Control/treatment comparison	
	obs	mean	sd	obs	mean	sd	obs	mean	sd	obs	mean	sd	t-stat	p-value
Exam cheater	644	0.18	0.38	411	0.14	0.35	116	0.40	0.49	117	0.09	0.29	5.70	0.00
Test 5 cheater	644	0.36	0.48	411	0.00	0.00	116	1.00	0.00	117	1.00	0.00	NaN	NaN
Test 5 leader	644	0.09	0.29	411	0.00	0.00	116	0.23	0.42	117	0.28	0.45	-0.86	0.39
Nb missed assignments	644	0.37	0.79	411	0.41	0.88	116	0.26	0.56	117	0.33	0.63	-0.96	0.34
Absent at Test 5	644	0.04	0.19	411	0.06	0.23	116	0.00	0.00	117	0.00	0.00	NaN	NaN
Grade test1	644	81.94	22.05	411	81.80	20.93	116	82.53	23.62	117	81.81	24.36	0.23	0.82
Grade test2	644	76.71	25.70	411	75.37	26.79	116	79.44	24.15	117	78.74	22.97	0.22	0.82
Grade test3	644	78.12	32.97	411	77.19	33.87	116	80.45	29.48	117	79.07	33.12	0.34	0.74
Grade test4	644	82.88	28.40	411	81.16	29.35	116	88.15	23.02	117	83.68	29.34	1.30	0.20
Grade Test 5	644	58.54	22.87	411	52.62	23.40	116	67.61	16.40	117	70.34	18.73	-1.18	0.24
Female	644	0.47	0.50	411	0.50	0.50	116	0.41	0.49	117	0.44	0.50	-0.60	0.55
Admission grade (written)	644	12.81	1.66	411	12.77	1.63	116	12.88	1.70	117	12.89	1.73	-0.07	0.94
Admission grade (oral)	644	16.68	2.44	411	16.70	2.31	116	16.43	2.58	117	16.85	2.70	-1.21	0.23
Grade in other classes	644	14.56	0.96	411	14.64	0.99	116	14.50	0.92	117	14.36	0.83	1.26	0.21
Did classe prepa	644	0.72	0.45	411	0.74	0.44	116	0.71	0.46	117	0.67	0.47	0.66	0.51
Scientific high school	644	0.47	0.50	411	0.45	0.50	116	0.53	0.50	117	0.47	0.50	0.85	0.40
High school honors	644	0.35	0.48	411	0.39	0.49	116	0.28	0.45	117	0.30	0.46	-0.25	0.81

A few observations are in order. First, we observe that some degree of Exam-cheating (14%) was observed even within the benchmark group of students who did not cheat on Test 5. This is not surprising, as the stakes of the final exam are much higher than those of the assignments and student were allotted more time to organize for collaboration. We refer to this level of cheating as the “baseline” level. Second, not all unwarned assignment-cheaters cheated on the final exam (40%). Again, this is not surprising because informational spillovers might have happened: an assignment cheater may have received a warning and told other students that cheating was credibly traced. This means that some students in the control group may have adjusted their cheating behavior in reaction to the warning sent to their classmate (e.g. with less or more subtle cheating). One may expect such informational spillovers especially within cheating group. Similarly a student who was warned might have decided to stop sharing information with their cheating group at the final exam. Some of

(unwarned) members of this group might have stopped cheating too, for lack of a source of information.

The last two columns of Table 1 confirm that our control and treated groups are very similar. The main purpose of this paper is to analyze the response of the treated group to the treatment with respect to cheating behavior associated with the final exam. We can also analyze the behavior of students who passed their answers to other students in Test 5 (the “leaders”), as opposed to students who received these answers (the “followers”). A leader is defined as a student who first submitted answers in the Test 5—which students could hand-in over a 1-week period—that were then also submitted by other students. Leaders handed-in their Test-5 on average 6.5 hours earlier than followers. The secondary purpose of the study is to analyze how leaders respond to the warning. The random selection of treated students was clustered by leadership status, enabling us to treat half of the 60 leaders and keep the other half in the control group.

To best account for the binary nature of the outcome variable, we use a logistic regression model. The classic model assumes that student’s binary decision to cheat or not is governed by a latent (unobserved) variable  $exam\_cheat_i^* = \beta_0 \tilde{X}_i + \epsilon_i$ , where  $\epsilon_i$  is an error distributed by the standard logistic distribution, and  $\tilde{X}_i$  is a vector of individual-specific explanatory variables. The probability of experiencing the event is in this model the probability that the latent variable  $exam\_cheat_i^*$  is larger than zero. As such, the logistic regression can be understood simply as finding the  $\beta_0$  parameters that best fit:

$$exam\_cheat = \begin{cases} 1 & \text{if } exam\_cheat_i^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

, where  $exam\_cheat$  is the observed data. More specifically, our base specification is :

$$exam\_cheat_i^* = Intercept + \alpha_1 was\_cheater_i + \alpha_2 was\_cheater\_got\_warning_i + \beta X_i + \epsilon_i$$

where  $was\_cheater_i$  takes the value of 1 if a student was found cheating on Test 5 and 0 otherwise.  $was\_cheater\_got\_warning_i$  corresponds to the treatment: it takes the value of 1 if a student was found cheating on Test 5 and received a warning and 0 otherwise. Depending on the specification, we may add various controls  $X_i$  such as student gender, group fixed effects (students were taught in separate groups of 40), academic achievement in other classes and age. In some specifications, we add variable  $was\_leader_i$  to the right-hand side, which takes the value of 1 if a student collaborated illicitly on Test 5 and was identified as a leader and 0 otherwise.  $was\_leader\_got\_warning_i$  takes the value of 1 if the leader is in the treated group a zero otherwise.  $\epsilon_i$  is an idiosyncratic error term distributed by the standard logistic distribution. Because the assignment to the “was\_cheater” category is obviously non-random, coefficient  $\alpha_1$  relates to the increment in the likelihood to cheat on the final exam for assignment-cheaters relative to assignment non-cheater. This denotes a correlation between having cheated in the past and cheating in the future, but this does not necessarily account for a causal effect. In contrast, because assignment to the control (no warning) and test groups (warning) is by design random,  $\alpha_2$  does represent the causal effect of our warning.

## 4- Results

The main results of our experiment are summarized in Table 2. Because the course was taught in classes of around 40 students, we use robust standard errors clustered at the class

level. Specification (1) does not include any controls. Specification (2) controls for the majors of the students, their gender and whether they handed-in the Test 5. Specification (3) also controls for class fixed effects. Specifications (4) and (5) correspond to specifications (2) and (3), respectively, but also control for leadership status. For interpretability, the table reports average marginal effects. We use “non-cheaters” as a baseline, which means all the coefficients should be interpreted as the marginal effect of the parameter relative to the cheating rate among “non-cheaters”.

Table 2: Effects of Warnings on Cheating. Dependent Variable: Student Cheats at Final Exam.

	(1)	(2)	(3)	(4)	(5)
Cheated in assignment	0.188*** (0.037)	0.191*** (0.036)	0.228*** (0.030)	0.174*** (0.036)	0.201*** (0.033)
Cheated in assignment and got warned	-0.251*** (0.057)	-0.246*** (0.048)	-0.265*** (0.042)	-0.232*** (0.046)	-0.246*** (0.036)
Absent at test 5		0.117* (0.062)	0.103* (0.053)	0.122** (0.061)	0.097* (0.053)
Female		-0.017 (0.022)	-0.022 (0.023)	-0.023 (0.023)	-0.027 (0.023)
Age		-0.018 (0.026)	-0.018 (0.025)	-0.022 (0.025)	-0.021 (0.025)
Admission grade written		0.002 (0.018)	-0.005 (0.017)	-0.002 (0.018)	-0.003 (0.017)
Admission grade oral		0.005 (0.009)	0.004 (0.008)	0.002 (0.008)	0.003 (0.007)
Grade other classes		-0.032* (0.019)	-0.037** (0.017)	-0.031* (0.018)	-0.037** (0.016)
Led cheating group in assignment				0.097** (0.045)	0.115** (0.046)
Led cheating group in assignment and got warned				-0.083 (0.099)	-0.077 (0.091)
Constant	-1.806	4.491	5.206	5.741	5.849
Curriculum FE	no	yes	no	no	no
Class FE	no	no	yes	yes	yes
Nationality FE	no	yes	yes	yes	yes
High school type FE	no	yes	yes	yes	yes
High school honors FE	no	yes	yes	yes	yes
Admission procedure FE	no	yes	yes	yes	yes
Observations	644	644	644	644	644
Log Likelihood	-281.650	-253.494	-243.574	-254.111	-241.559
Akaike Inf. Crit.	569.300	590.988	591.149	592.222	591.118

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Robust standard errors, clustered at the class level



Our preferred specification is Specification (3), as it controls for class fixed effects and excludes the leader/follower distinction, which lacks significance and may be used only as suggestive evidence for future work.

These regressions enable us to assess the effects of a warning on cheaters. In all specifications, the magnitude of the coefficient (in absolute terms) is greater for *was\_cheater\_got\_warning<sub>i</sub>* than for *was\_cheater<sub>i</sub>*: having collaborated illicitly on assignments increases the probability of cheating on the final exam by 23% (coefficient of “Cheated on assignment”, Specification 3). Being warned more than offsets this effect, as the marginal effect of “cheated on assignment and got warned” is -27%. This means that warned cheaters have a 4% *lower* probability of cheating relative to the baseline level (that of students who did not cheat on Test 5). This suggests that prior warnings are highly effective at discouraging cheating.

Even though there is no statistical significance, the effect seems slightly more pronounced for the cheaters who were leaders of their cheating group (specifications 4 and 5). To ascertain this, further work with a larger sample and more accurate identification of cheating clusters would however be needed.

Because our identification methods are probabilistic, we acknowledge that other specifications could have been used, as well as alternative criteria for distinguishing between cheaters and non-cheaters. Appendix A4 offers a few robustness checks that confirm that our results convey to a broad range of sensible criteria.

## 5- Discussion

As we have noted, our approach is based on a statistical analysis of exam responses. A natural question is the extent to which our methodology can be applied in different exam settings and formats. Indeed, our approach is facilitated by the structure of our exam, which is based on evaluating competencies such as code writing and code understanding. The questions are open-ended, which enables us to identify suspicious similarities confidently, as opposed to multiple-choice questions where the questions are by nature closed-ended and leave little room for student creativity. Because we tailored our detection strategies to our specific exam format, the methodology may not readily apply to any other exam format. Yet, the increased availability and democratization of sophisticated machine learning techniques heralds the possibility to address many more exam formats in the future. More importantly, we believe our general methodology (stage 1: identify likely cheaters in assignments, stage 2: send personalized warning to likely cheaters prior to the exam) could be used in most settings, provided that cheaters can be identified with reasonable accuracy.

Our solution is, however, only partial. We restricted our technologies to the analysis of similarities between copies. However, as Karnalim (2017) shows, cheating takes almost infinite forms, as do the detection avoidance strategies. Replicating a similar protocol in more controlled conditions would allow a more precise representation of the mechanisms at play. In particular, we acknowledge that our study may include a few “false negatives” (true cheaters identified as non-cheaters). This may occur in both Test 5 (which allows us to identify assignment-cheaters) and the final exam (where classify students as ‘exam-cheaters’ or not). First, very good programmers might have anticipated some of principles that

underlie our detection technology, and acted in such a way that they evade detection. Because the class was only an introductory course, we believe such obfuscation techniques were far beyond the reach of most of the students. Further, obfuscation would be a superfluous effort for good programmers, for whom this class was very easy. Also, there is still a possibility that cheaters casually used the copy of another student with the exact same versions of the three “trap questions”. Because there were 5 versions of each of these questions, there is a very low probability that a given student who always cheats is unidentified. Still, some students who cheated only in a few questions might have been classified as non-cheaters. Students who cheated only in standard questions would also escape detection.

Another caveat we want to highlight is that our criteria for assignment into cheating categories slightly differs between Test 5 and the exam. The reason for this is that we anticipated that some students may try to escape detection, in case they formed educated guesses about the possible techniques implemented in Test 5. This is why we also added “trap questions” in the final exam, which were arguably very difficult to anticipate. Also, the motivations for cheating may have been different for the exam where stakes were higher. Cheating might have been harder too, due to the synchronous timing. For all these reasons, one cannot compare Test-5 and exam cheating rates. This is why we focus our attention to the effect of the warning at the exam only.

We also have to keep in mind that the main objective is to eliminate or at least limit cheating. The two main levers of action are prevention and repression. We show that targeted warnings make prevention significantly more effective than traditional warnings. However,

prevention may not be enough. Repression in the form of sanctions is often necessary because sanctions increase confidence in the results of exams and limit the feeling of injustice in those who do not cheat. Our statistical analysis provides an indication only of the *probability* of fraud but rarely provides sufficient evidence to trigger a sanction, and we acknowledge that the proof mechanisms should be improved. The “trap random questions” that we used and present in Appendix A2 represent a promising avenue.

An important issue in research on academic dishonesty is the difficulty of estimating the prevalence of cheating. With the increasing prevalence of online exams, we believe the need to examine methods that allow for creativity and at the same time are robust to cheating will only increase in the coming years. This paper proposes solutions to enhance this robustness. Because of a relatively small sample size and inaccurate identification of cheating groups inherent to our probabilistic technique, we leave the more detailed study of the leader/followers dynamics for future research. We believe this could uncover many of the fundamental mechanisms behind cheating and, in turn, options to curb this behavior.

## 6- Conclusion

The COVID-19 crisis gave digital technologies a unique opportunity to showcase how they can contribute to our education systems. There is overwhelming evidence that these technologies will occupy an ever-growing place in education not only in times of crisis but also as a new teaching standard. The crisis has also revealed, however, some of the limits of online education, not the least of which involves the viability of online exams. Digital tools

meant to address these limits, such as online exam platforms, were also questioned, as they support exam-taking but at the same time facilitate cheating.

In this paper we develop an original and effective method that can be used to identify illicit collaborations in a probabilistic way. We first show that cheating can be massive when exams are given at a distance without proctoring. We then demonstrate with a randomized experiment that a credible and effective mechanism could be deployed to discipline students and restore the validity and fairness of exams. Prior warnings are particularly effective in inducing honest behavior. In light of these results, we argue that the wise use of warnings constitutes a promising alternative to proctoring, when supervision is impossible for practical or ethical reasons.

Our work could be extended fruitfully in at least two directions. First, a larger sample size could allow to uncover some interesting mechanisms behind group formation and disbandment, and in particular the relative effectiveness of warnings sent to specific students (leaders or not), or warnings sent to all members of a group relative to only part of it. Second, the ability to control intra-group and inter-group communication between students between the treatment and the final exam would allow to better understand the mechanism through which the warning disciplines students.

## Bibliography

Allcott, A., Mullainathan, S., (2010). Behavior and Energy Policy. *Science*, 327: 1204-1205

Arnold, R., Martin, B., Jinks, M., & Bigby, L. (2007). Is there a relationship between honor codes and academic dishonesty? *Journal of College and Character*, 8, 1- 20

Auferoth, F. (2020) Who Benefits from Nudges for Exam Preparation? An Experiment. Working paper

Bachore, M. M. (2016). The nature, causes and practices of academic dishonesty/cheating in higher education: The case of Hawassa University. *Journal of education and practice*, 7(19), 14-20.

Bandura, A. (2002). Selective moral disengagement in the exercise of moral agency. *Journal of Moral Education*, 31: 101–119.

Bawarith, R., Basuhail, A., Fattouh, A., Gamalel-Din, S. (2017) “E-exam Cheating Detection System” *International Journal of Advanced Computer Science and Applications(IJACSA)*, 8(4), 2017

Beshears, J., James J. Choi, David Laibson, Brigitte C. Madrian, and Katherine L. Milkman. The Effect of Providing Peer Information on Retirement Savings Decisions. NBER Working Papers 17345, National Bureau of Economic Research, Inc, August 2011. URL <http://ideas.repec.org/p/nbr/nberwo/17345.html>.

Becker, D., J. Connolly, P. Lentz and J. Morrison. (2006). Using the Business Fraud Triangle to Predict Academic Dishonesty among Business Students. *Academy of Educational Leadership Journal*, 10 (1), 37–54.

Bing, M., Davison, H., Vitell, S., Ammeter, A., Garner, B., & Novicevic, M. (2012). An Experimental Investigation of an Interactive Model of Academic Cheating Among Business School Students. *Academy of Management Learning & Education*, 11(1), 28-48. Retrieved July 8, 2020, from [www.jstor.org/stable/23100455](http://www.jstor.org/stable/23100455)

Brodowsky, G.H., Tarr, E., Ho, F.N., & Sciglimpaglia, D. (2020). Tolerance for Cheating From the Classroom to the Boardroom: A Study of Underlying Personal and Cultural Drivers. *Journal of Marketing Education*, 42, 23 - 36.

Cialdini, R. B., & Goldstein, N. J. 2004. Social influence: Compliance and conformity. *Annual Review of Psychology*, 55: 591–621.

Cialdini, R. B. (2003). Crafting normative messages to protect the environment. *Current Directions in Psychological Science*, 12: 105–109.

Corrigan-Gibbs, H., Gupta, N., Northcutt, C., Cutrell, E., & Thies, W. (2015). Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction*, 22(6). <https://doi.org/10.1145/2810239>

D'Souza, K. A., & Siegfeldt, D. V. (2017). A conceptual framework for detecting cheating in online and take-home exams. *Decision Sciences Journal of Innovative Education*, 15(4), 370-391.

Damgaard, M. T., & Nielsen, H. S. (2018). Nudging in education. *Economics of Education Review*. <https://doi.org/10.1016/j.econedurev.2018.03.008>

Ferraro P.J, Juan Jose Miranda, and Michael K Price. The Persistence of Treatment Effects with Norm-Based Policy Instruments : Evidence from a Randomized Environmental Policy Experiment The Persistence of Treatment Effects with Norm-Based Policy Instruments : Evidence from a Randomized Environmental Policy. *American Economic Review: Papers & Proceedings*, 101:318-322, 2014. ISSN 0002-8282. doi: 10.1257/aer.101.3.318.

Fendler, R. J., Yates, M., & Godbey, J. (2018). Observing and Deterring Social Cheating on College Exams. *International Journal for the Scholarship of Teaching and Learning*. <https://doi.org/10.20429/ijstl.2018.120104>

Green, S. (2004). Cheating. *Law and Philosophy*, 23: 137–185.

Holden, O., Kuhlmeier, V. A., & Norris, M. (2020, June 10). Academic Integrity in Online Testing: A Research Review. <https://doi.org/10.31234/osf.io/rjk7g>

Karnalim, O. (2017). Python Source Code Plagiarism Attacks on Introductory Programming Course Assignments. *Themes in Science and Technology Education*, 10, 17-29.

Kerkvliet, J., & Sigmund, C. L. (1999). Can we control cheating in the classroom? *Journal of Economic Education*. <https://doi.org/10.1080/00220489909596090>

Konheim-Kalkstein,Y, Stellmack, M, Shilkey,M (2008) Comparison of Honor Code and Non-Honor Code Classrooms at a Non-Honor Code. *University Journal of College & Character* VOLUME IX, NO. 3,

Madrian, B. C. and D. F. Shea (2001): "The Power of Suggestion: Inertia" in 401 (k) Participation and Savings Behavior," *The Quarterly Journal of Economics*, 116, 1149{1187.

Mazar, N., On A., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *J. Marketing Res.* 45, 6, 633–644.

McCabe , D, Klebe, L, Butterfield (1999) Academic Integrity in Honor Code and Non-Honor Code Environments A Qualitative Investigation

McCabe, D., & Treviño, L. (1995). Cheating among business students: A challenge for business leaders and educators. *Journal of Management Education*, 19: 205–218.

McCabe, D. L., Treviño, L. K., & Butterfield, K. D. (2001). Cheating in academic institutions: a decade of research. *Ethics & Behavior*, 11(3), 219–232.

McCabe D (2016) Cheating and honor: lessons from a long -term research project. In: Bretag T (ed) *Handbook of academic integrity*, pp 187–198. doi:10.1007/978-981-287-098-8\_22

Nizam, N.I.; Gao, S.; Li, M.; Mohamed, H.; Wang, G. (2020) Scheme for Cheating Prevention in Online Exams during Social Distancing. *Preprints 2020*, 2020040327 (doi: 0.20944/preprints202004.0327.v1).

Norris, Mark (2019) *University Online Cheating -- How to Mitigate the Damage* [2019] *Research in Higher Education Journal*, v37 Nov 2019. 20 pp.

Rockmann K, & Northcraft, G. (2008). To be or not to be trusted: The influence of media richness on defection and deception. *Org. Behav. Hum. Decision Processes* 107, 2, 106–122.

Sudman, S., and N. Bradburn. 1974. *Response effects in surveys: A review and synthesis*. Chicago, Ill.: Aldine

Thaler, R., and Sunstein, C. (2008) *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New York: Penguin Books

Thaler, R. H. and S. Benartzi (2004): "Save More Tomorrow : Using Behavioral Economics to Increase Employee Saving," *Journal of Political Economy*, 112, 164-187.

Van Zant , A. and Kray, L. (2014). "I can't lie to your face": Minimal face-to-face interaction promotes honesty. *J. Exp. Soc. Psychol.* 55, 1, 234–238.


Zamir, E., Lewinsohn-Zamir, D. and Ritov, I. (2017), It's Now Or Never! Using Deadlines as Nudges. *Law Soc Inq*, 42: 769-803. <https://doi.org/10.1111/lisi.12199>



## Appendix

### A1. Messages sent to students

The students received two generic emails. The first was sent after Test 4 and before Test 5. It stated that Test 5 and the Exam were individual and reminded students of the consequences of verified cheating. It also notified the students that Test 5 would be a bit different from the previous tests, so as to better match the format of the final exam.

 LAMBIN Xavier  
python : information test 5

À PGE-1A-Parc.Mgt\_19-20; PGE-1A-Parc.Ulysse\_19-20; PGE-1A-Parc.ET\_19-20  
Cc HUMBERT Marc

---

As indicated, Test 5 will open on Monday, April 20 and will close on Sunday, April 26 at midnight. In order to get closer to the final conditions of the exam it will be done under the following conditions:

- Test 5 is a little longer than the previous ones. It is open for 1 hour from the moment you start it (the exam lasts 1h30).
- Like the exam, it consists of three parts:
  - Quiz (course questions)
  - Code understanding questions
  - Code writing questions
- Like the exam, this test is now strictly individual..
  - **No form of collaboration during the test is allowed. Any cheating will be severely penalized.**
  - However, the following are allowed:
    - Course materials, personal notes, and resources recommended in the course.
    - The use of IDLE or other development environment to test your code before submitting it.

The second message was sent between Test 5 and the exam. It stated that some cheating had been observed to have occurred on Test 5 and again reminded students of the rules of the exam.

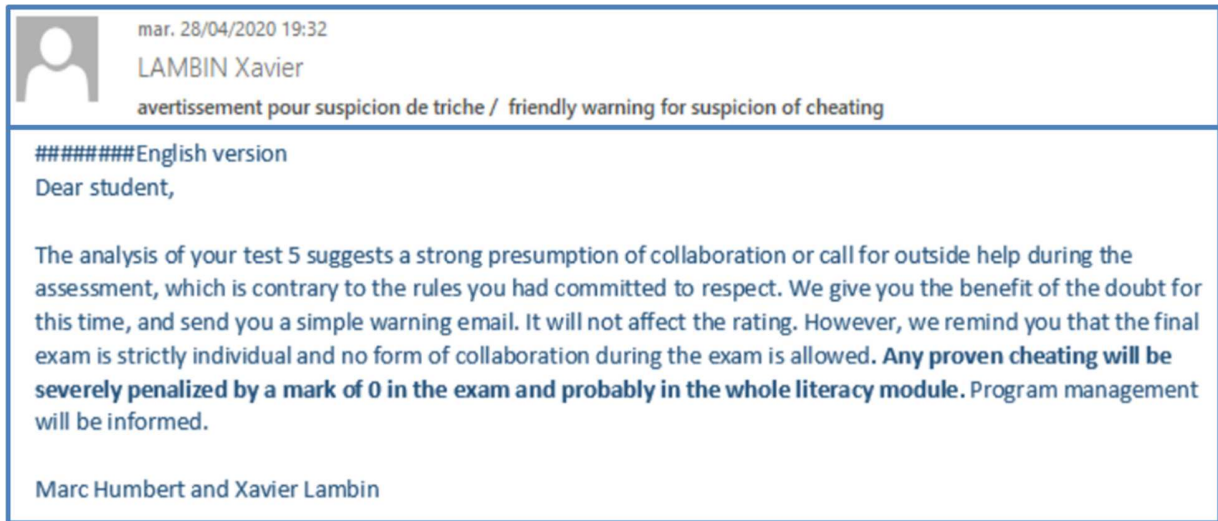
**Rules:**

- As with Test 5, this exam is strictly individual
- **No form of collaboration during the review is permitted.**

We have observed a number of cases of collaboration/external help in Test 5. Fortunately, a majority of you respected the rules after having accepted them. Those suspect of fraud have been notified.

We left the benefit of the doubt for Test 5, but **any proven cheating on the exam will be penalized by a mark of 0 on the exam and probably on the entire literacy module. Program management will be notified.** In case of contestation, an individual oral examination can be arranged very quickly.

In addition to these generic emails, we sent the half of the students we identified as cheaters a warning email.



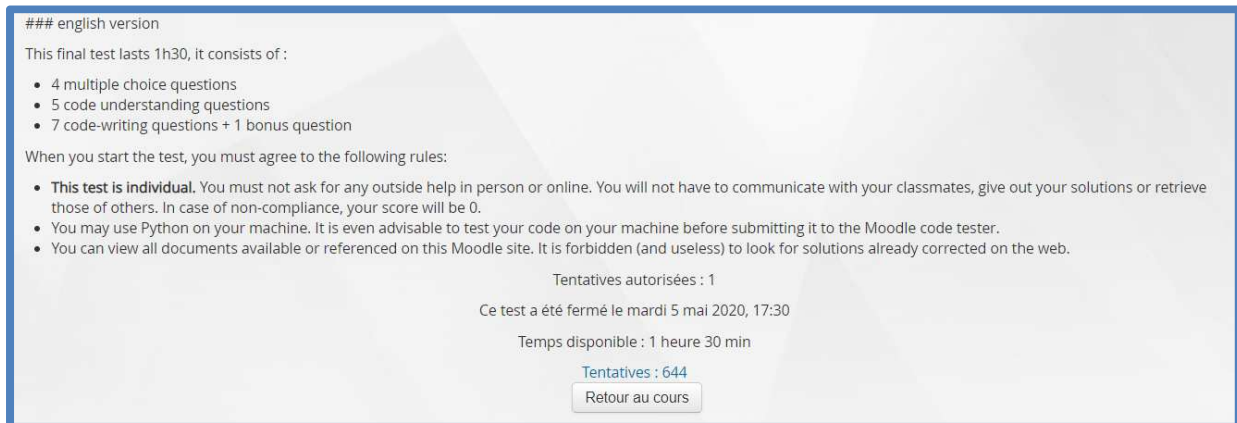
mar. 28/04/2020 19:32  
LAMBIN Xavier  
avertissement pour suspicion de triche / friendly warning for suspicion of cheating

#####English version  
Dear student,

The analysis of your test 5 suggests a strong presumption of collaboration or call for outside help during the assessment, which is contrary to the rules you had committed to respect. We give you the benefit of the doubt for this time, and send you a simple warning email. It will not affect the rating. However, we remind you that the final exam is strictly individual and no form of collaboration during the exam is allowed. **Any proven cheating will be severely penalized by a mark of 0 in the exam and probably in the whole literacy module.** Program management will be informed.

Marc Humbert and Xavier Lambin

The preamble and first question on both Test 5 and the exam require students to adhere to a no-cheating “honor code”:



### english version

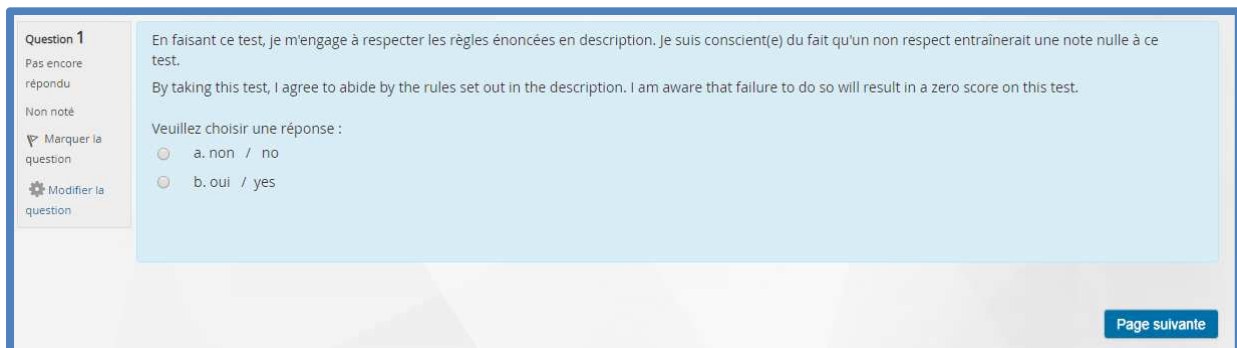
This final test lasts 1h30, it consists of :

- 4 multiple choice questions
- 5 code understanding questions
- 7 code-writing questions + 1 bonus question

When you start the test, you must agree to the following rules:

- **This test is individual.** You must not ask for any outside help in person or online. You will not have to communicate with your classmates, give out your solutions or retrieve those of others. In case of non-compliance, your score will be 0.
- You may use Python on your machine. It is even advisable to test your code on your machine before submitting it to the Moodle code tester.
- You can view all documents available or referenced on this Moodle site. It is forbidden (and useless) to look for solutions already corrected on the web.

Tentatives autorisées : 1  
Ce test a été fermé le mardi 5 mai 2020, 17:30  
Temps disponible : 1 heure 30 min  
Tentatives : 644  
Retour au cours



**Question 1**  
Pas encore répondu  
Non noté  
Marquer la question  
Modifier la question

En faisant ce test, je m'engage à respecter les règles énoncées en description. Je suis conscient(e) du fait qu'un non respect entraînerait une note nulle à ce test.  
By taking this test, I agree to abide by the rules set out in the description. I am aware that failure to do so will result in a zero score on this test.

Veillez choisir une réponse :

- a. non / no
- b. oui / yes

Page suivante

## A2. Plagiarism detection technologies: details

### **Cheaters:**

We used three distinct detection technologies. The first two technologies were used to identify cheaters on Test 5. To defeat potential hacking (students understanding our technologies and intentionally avoiding them during the exam), we used the third technology only on exam results.

The first two technologies consist in textual analysis of copies. Because some simple hacking techniques are well known, we did more than use exact matching of copies (technology 1). Previous analysis of tests 1 through 4 showed that students quickly learn to add spaces, change variable names, etc., to avoid matching exam copies perfectly.

Following the spirit of many plagiarism-detection algorithms such as the one used by Karnalim (2017), our second technology analyses text similarity between copies. To account for the fact that our texts are typically very short (consisting of a few lines as opposed to several pages for most plagiarism-detection services), we use Gestalt Pattern Matching (Ratcliff and Obershelp, 1983). This enables us to identify copies that “look similar” to each other even if they are not perfect matches. Crucially, we complement Gestalt Pattern Matching with analysis of the abstract syntax of the codes. A given copy’s abstract syntax tree is extracted and compared with all other abstract syntax trees. For both concrete and abstract syntax analysis, a distance matrix between all copies is computed. Copies are then clustered with a DBSCAN (Density-Based Spatial Clustering of Applications with Noise) algorithm. Copies that are closely packed together are considered clusters of cheaters. The members who constitute such a cluster are labelled as “cheaters”. It is useful at this stage to note that this strategy yields only a probabilistic estimation of cheating behavior. It is therefore an

effective prevention tool but is of little help when it comes to applying sanctions (see discussion); hence the need for a third technology.

The third technology consists in a classical approach, the use of random questions, to which we added an original twist. For each specific question (displayed to all students as, say, “question 8”), we randomly assigned a slightly different version of the question to each student (student A gets question 8A, student B gets question 8B, etc.). A cheater will therefore give an answer that is incorrect but corresponds to the correct answer of another version of the question. We designed our versions such that it was extremely unlikely that a student could have genuinely come up with the answer to another version of a question. Further, differences between versions are visible only to particularly attentive eyes or to students expecting such a strategy to be deployed. Given that there is no precedent for such “trap random questions” in the institution, this eventuality is very unlikely to materialize. A cheater will be classified as such if s/he gives an answer that corresponds to the correct answer of another version of the question. We use this third technology for exam copies but not for the assignments, so students cannot learn to avoid detection.

As we discussed, the criteria that discriminates cheaters from non-cheaters is different in Test 5 and the Exam. This is due to the third technology (random questions) that was kept aside for the exam, and the fact the Exam was longer and harder. We note however, that this difference in criteria does not impact the validity of our randomized control trial: all students of the study group (test 5-cheaters) have been assigned to an assignment-cheating status according to the first criteria. Assignment to test and control groups (warned and unwarned cheaters) was random, and the identification of exam cheating was done according to the

same criteria in both of these groups. We provide below the detail of our criteria for the assignment of the cheating status.

**Test 5** : questions 11 to 14 allow for significant variety in the answers. As such they can be used to identify anomalous similarities. We classify a student as a cheater is (s)he meets at least *two* of these criteria:

- Exact same answer to Q11 as another copy
- Exact same answer to Q12 as another copy
- Exact same answer to Q13 as another copy
- Exact same answer to Q14 as another copy
- Gestalt Pattern Matching applied to Q11-Q14 shows similarity with another copy
- Abstract syntax tree applied to Q11-Q14 shows similarity with another copy

**Exam** : questions 12 to 18 allow for significant variety in the answers. We use a similar criteria to that of Test 5. We classify a student as a cheater is (s)he meets at least *three* of these criteria:

- Exact same answer to Q12 (Q13...Q18) as another copy
- Gestalt Pattern Matching applied to Q12-Q18 shows similarity with another copy
- Abstract syntax tree applied to Q11-Q18 shows similarity with another copy
- The student gave the correct answer to another version of either of the trap questions Q9, Q10, Q13.

## Leaders :

We have access to the exact time at which Test 5 was handed in by each student. We define a “leader” as a student who belongs to a cheating cluster and was the first to hand in her/his copy.

## A3. Details descriptive statistics

We provide here some details on the variables displayed in Table 1.

- **Exam cheater:** the student has been identified as a cheater at the final exam
- **Test5 cheater:** the student has been identified as a cheater at the last assignment before the exam.
- **Test5 leader:** the student has been identified as a the leader of a cheating group at the last assignment before the exam. That student would also be labeled as a “Test5 cheater”
- **Nb missed assignments:** Number of assignments, out of 5, that the student did not hand in.
- **Absent at test5:** Dummy variable that takes value 1 if the student did not hand in the last assignment.
- **Grade test1, Grade test2, Grade test3, Grade test4, Grade test5:** grade received at the various pre-exam assignments
- **Female :** takes value 1 if the student is a female, 0 otherwise.
- **Admission grade (written):** grade received at the competitive exam for admission at the school (written examination).
- **Admission grade (oral) :** grade received at the competitive exam for admission at the school (written examination)
- **Grade in other classes:** average grade received in the other classes in 1<sup>st</sup> year at the school

- **Did classe prepa:** takes value 1 if the student followed "Classes préparatoires aux grandes écoles" (classes to prepare students for the entrance examinations to the top-ranking higher education establishments), 0 otherwise.
- **Scientific high school:** takes value 1 if the student went to a scientific high school (as opposed to Economic, literary or technical), 0 otherwise
- **High school honors:** takes value 1 if the student got "high honors" ("très bien") or better at the French "Baccalauréat".

*Table 3: Additional Descriptive statistics. Focus on the Leader/Follower Distinction*

	Full sample			Cheaters			Unwarned leaders (control group)			Warned leaders (treated group)			Control/treatment comparison	
	obs	mean	sd	obs	mean	sd	obs	mean	sd	obs	mean	sd	t-stat	p-value
Exam cheater	644	0.18	0.38	233	0.24	0.43	27	0.52	0.51	33	0.09	0.29	3.87	0.00
Test 5 cheater	644	0.36	0.48	233	1.00	0.00	27	1.00	0.00	33	1.00	0.00	NaN	NaN
Test 5 leader	644	0.09	0.29	233	0.26	0.44	27	1.00	0.00	33	1.00	0.00	NaN	NaN
Nb missed assignments	644	0.37	0.79	233	0.30	0.60	27	0.22	0.58	33	0.27	0.57	-0.34	0.74
Absent at Test 5	644	0.04	0.19	233	0.00	0.00	27	0.00	0.00	33	0.00	0.00	NaN	NaN
Grade test1	644	81.94	22.05	233	82.17	23.95	27	89.18	19.41	33	84.89	24.80	0.75	0.46
Grade test2	644	76.71	25.70	233	79.09	23.51	27	73.52	29.85	33	79.61	24.42	-0.85	0.40
Grade test3	644	78.12	32.97	233	79.76	31.30	27	85.31	21.40	33	85.30	25.84	0.00	1.00
Grade test4	644	82.88	28.40	233	85.90	26.43	27	91.35	20.15	33	85.09	28.73	0.99	0.33
Grade Test 5	644	58.54	22.87	233	68.98	17.62	27	74.76	15.54	33	72.20	19.03	0.57	0.57
Female	644	0.47	0.50	233	0.42	0.50	27	0.48	0.51	33	0.48	0.51	-0.03	0.98
Admission grade (written)	644	12.81	1.66	233	12.88	1.71	27	12.41	1.43	33	12.76	1.62	-0.89	0.38
Admission grade (oral)	644	16.68	2.44	233	16.64	2.64	27	17.22	2.25	33	16.29	2.98	1.38	0.17
Grade in other classes	644	14.56	0.96	233	14.43	0.88	27	14.47	0.88	33	14.40	0.72	0.37	0.71
Did classe prepa	644	0.72	0.45	233	0.69	0.46	27	0.74	0.45	33	0.73	0.45	0.12	0.91
Scientific high school	644	0.47	0.50	233	0.50	0.50	27	0.59	0.50	33	0.55	0.51	0.36	0.72
High school honors	644	0.35	0.48	233	0.29	0.46	27	0.22	0.42	33	0.42	0.50	-1.69	0.10

#### A4. Robustness checks

One key technical difficulty of the present paper is to classify students into cheating and non-cheating categories. Indeed, the cheating status is by definition concealed by the students and, therefore, difficult to observe. To achieve this, we designed a few original identification techniques described in Appendix A2. None of them is perfect and the criteria for classification into cheating categories has by nature an arbitrary component.

Table 4 displays a few robustness checks related to the classification in the cheating categories. The set of controls is the one of the preferred specification of the main text (specification 3 in Table 2). The seven specifications represented in Table 4 differ by the dependent variable. The dependent variables of specifications (1) to (4) are the result of either of our detection strategies:

- Specification 1: Do the answers show strong textual similarity with at least another copy? (Gestalt Pattern Matching). (1 if yes, 0 otherwise)
- Specification 2: Does the abstract syntax of the answers show strong similarity with at least another copy? (DBSCAN). (1 if yes, 0 otherwise)
- Specification 3: Did the student give the correct answer to at another version of at least one question? (trap questions) (1 if yes, 0 otherwise)
- Specification 4: How many of the answered questions exactly match the set of answers of at least another copy? (exact matching)(numeric variable)

We observe that all of these criteria result in the same qualitative insight: the nudge has a strong effect on students propensity to cheat. Yet, none of these detection techniques is perfect. Some students might for example avoid exact matching by changing variable names or marginally re-ordering the code, which would generate false negatives. These techniques can also generate “false positives” as discussed in Section 5.

This is why we built a custom numeric variable, the “cheat score”, which is the sum of all tests (assuming a positive test equals 1, 0 otherwise). In the main text we classify as cheater any student who has a cheat score greater or equal to 3. As a robustness check, Table 4 uses



thresholds of 2 (Specification 5) and 4 (Specification 6). Specification (7) simply takes the cheat score as a dependent variable. Again we verify that our results are qualitatively maintained. Expectedly, when the criteria is too narrow (Specifications 1 to 4) or too strict (specification 6), our results are attenuated. We note finally that contrary to the specifications of Table 2, Specifications (4) to (7) in Table 4 use standard optimum least squares, owing to the numeric nature of the dependent variable.

Table 4: Robustness checks

Estimator:	Dependent variable:						
	Text similarity	Abstract similarity	Trap questions	Number of exact matches	Cheat score $\geq 2$	Cheat score $\geq 4$	Cheat score
	<i>logistic</i>	<i>logistic</i>	<i>logistic</i>	<i>OLS</i>	<i>logistic</i>	<i>logistic</i>	<i>OLS</i>
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Was cheater	0.093*** (0.024)	0.047** (0.019)	0.023 (0.021)	0.512*** (0.130)	0.244*** (0.065)	0.078*** (0.016)	0.955*** (0.205)
Was cheater and got warning	-0.104*** (0.025)	-0.057* (0.031)	-0.035** (0.016)	-0.607*** (0.161)	-0.215** (0.093)	-0.144*** (0.034)	-0.997*** (0.247)
Female	-0.032* (0.017)	0.027 (0.023)	0.010 (0.014)	-0.128* (0.070)	-0.018 (0.026)	-0.049** (0.022)	-0.132 (0.081)
Absent at Test 5	-0.023 (0.069)	0.034 (0.037)	-0.012 (0.038)	0.264 (0.222)	0.112** (0.051)	-0.034 (0.030)	0.344 (0.258)
Age	-0.003 (0.015)	-0.023 (0.023)	-0.005 (0.012)	0.014 (0.065)	0.011 (0.029)	-0.002 (0.018)	-0.0002 (0.102)
Admission grade written	-0.010 (0.012)	0.011 (0.014)	-0.009 (0.010)	0.060 (0.047)	-0.010 (0.022)	0.011 (0.007)	0.036 (0.058)
Admission grade oral	-0.004 (0.006)	0.001 (0.004)	-0.001 (0.003)	-0.002 (0.021)	0.006 (0.008)	-0.001 (0.005)	-0.003 (0.028)
Grade other classes	-0.028** (0.011)	-0.002 (0.012)	-0.026* (0.016)	-0.186*** (0.052)	-0.071*** (0.020)	-0.028* (0.015)	-0.246*** (0.078)
Constant	6.902	-1.416	-7.322	3.754	36.318	3.820	5.281
Curriculum FE	no	no	no	no	no	no	no
Class fixed FE	yes	yes	yes	yes	yes	yes	yes
Nationality FE	yes	yes	yes	yes	yes	yes	yes
High school type FE	yes	yes	yes	yes	yes	yes	yes
Admission procedure FE	yes	yes	yes	yes	yes	yes	yes
Observations	644	644	644	644	644	644	644
R <sup>2</sup>				0.153			0.217
Adjusted R <sup>2</sup>				0.098			0.150
Log Likelihood	-145.054	-201.407	-98.623		-356.546	-112.383	
Akaike Inf. Crit.	370.108	482.814	277.247		817.092	328.765	

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01