Tulane Economics Working Paper Series

Do Academic Honesty Statements Work?

James Alm
Tulane University
jalm@tulane.edu

Patrick Button
Tulane University

Christine P. Smith
Tulane University

Toni Weiss
Tulane University

Working Paper 2510
November 2025

Abstract

Many colleges have attempted to deal with student cheating by using "academic honesty statements," or statements that students must read and acknowledge that they will follow. In this paper, we conduct a randomized controlled experiment that investigates the impact of academic honesty statements on college student examination performance, using an objective measure of student examination performance as a proxy for student cheating. Overall, we find no statistically significant differences in the test performance of students who are given the academic honesty statements and students who are not given these statements. These results indicate that academic honesty statements do not affect student performance in a significant way, so that their use is unlikely to be a reliable tool in reducing cheating. However, other explanations are possible.

Keywords: Student cheating; academic misconduct; academic integrity; nudges; priming; randomized control trial
JEL codes: A22, I21, C93

Do academic honesty statements work?

James Alm[a], Patrick Button[b], Christine P. Smith[c], and Toni Weiss[d] *

[a] Department of Economics, Tulane University, New Orleans, LA USA
[b] Department of Economics, Tulane University, New Orleans, LA USA
[c] Freeman School of Business, Tulane University, New Orleans, LA USA
[d] Department of Economics, Tulane University, New Orleans, LA USA

ABSTRACT
Many colleges have attempted to deal with student cheating by using "academic honesty statements," or statements that students must read and acknowledge that they will follow. In this paper, we conduct a randomized controlled experiment that investigates the impact of academic honesty statements on college student examination performance, using an objective measure of student examination performance as a proxy for student cheating. Overall, we find no statistically significant differences in the test performance of students who are given the academic honesty statements and students who are not given these statements. These results indicate that academic honesty statements do not affect student performance in a significant way, so that their use is unlikely to be a reliable tool in reducing cheating. However, other explanations are possible.

KEYWORDS
Student cheating; academic misconduct; academic integrity; nudges; priming; randomized control trial.

JEL CODES
A22; I21; C93.

CONTACT
James Alm, jalm@tulane.edu, Professor Emeritus, Department of Economics, Tilton Hall, Tulane University, New Orleans, LA 70118 (jalm@tulane.edu).

Most academic institutions have honor codes of various types, which deal with multiple forms of academic integrity issues like plagiarism, fabrications, misrepresentations, and the like. An especially common issue is student cheating on examinations. Concern with student cheating is a perennial problem in colleges and universities, and these concerns have increased significantly in recent years, especially with advances in testing technology, increased online testing during and after the pandemic, and expanded use of generative artificial intelligence (AI) like ChatGPT (or "Generative Pre-trained Transformer"). Exam cheating may occur by copying from another student's exam, by using the internet in an unauthorized way to find or communicate with others, or by using cell phones or other communication devices to acquire information from others. The motivation for such cheating is the obvious one: students who cheat expect to receive higher scores on the examination, without being caught and penalized for their cheating.

One way that many institutions have attempted to deal with cheating is by using an "academic honesty statement." These statements typically appear at the beginning of an examination, and students are required to read and to acknowledge that they will comply with these statements.[1] A typical academic honesty statement from the University of Montana reads:

> "I have not received, I have not given, nor will I give or receive, any assistance to another student taking this exam, including discussing the exam with students in another section of the course. I will not use any non-instructor approved electronic device to assist me on an exam. I will not plagiarize someone else's work and turn it in as my own. I will not remove the exam from this room, either on test day or the day it is reviewed in class. I understand that acts of academic dishonesty may be penalized to the full extent allowed by the University of Montana Student Conduct Code, including receiving a failing grade for the course. I recognize that I am responsible for understanding the provisions of the University of Montana Student Conduct Code as they relate to this academic exercise."[2]

Statements such as these often appear on many other forms, including financial statements, mortgage applications, insurance applications, or individual income tax returns. These forms

usually require the filer to sign a statement attesting they have answered all questions honestly.[3] However, the effects of these honesty statements on the veracity of any individual declarations are decidedly mixed and inconclusive.

As discussed in detail later, there is in fact a large literature that examines the prevalence of student cheating on examinations, going back at least to work by Bowers (1964). There is also work on how this cheating can be mitigated via academic honesty statements. Some early research suggested that these statements often reduced cheating, sometimes by significant amounts, depending on the specific features both of the statement and of the institution (McCabe, Treviño, and Butterfield 2001). More recently, these effects have been increasingly questioned (Tatum 2022). One constant challenge in this research has been the reliance on indirect and self-reported measures of cheating, typically drawn from student responses to survey questions that ask them whether they cheated on an examination. These measures obviously rely upon the honesty of students, and it is likely that they significantly underreport cheating, especially if the survey is provided by instructors or others who are responsible for enforcing student honor codes. Which students report their cheating could also be a function of student background, such as sex and nationality. Overall, as recently emphasized by Zhao et al. (2022), there is no firm consensus on whether academic honesty statements have any significant impact on student cheating.

In this paper, we conduct a randomized controlled experiment that quantifies the impact of academic honesty statements on student performance on examinations. We do this by examining the performance of undergraduate students in two large introductory economics and accounting courses at a large private university during a recent, post-pandemic academic semester (2022). We begin by randomly assigning students from these classes to treatment and

control groups. Before assigning students to treatment and control groups, and before our intervention, we collect baseline data by observing the performance of these students on a first online test.[4] For the subsequent online test several weeks later, we require that students in the treatment group read and agree to an academic honesty statement at the start of the exam, while the control group gets an identical examination but without this academic honesty statement.

To quantify the impact of the academic honesty statement, we then compare the treatment and control groups based on their performance on the second exam relative to their performance on the first (baseline) exam. We use exam performance as our proxy measure for student cheating because cheating directly affects exam performance and, most importantly, using exam performance avoids the problems with measuring cheating using self-reported and subjective measures. Our hypothesis is that the treatment group will be less likely to cheat than the control group and thus the treatment group will exhibit lower average test scores than the control group; that is, students who cheat on an examination will achieve higher scores on the test, so that reducing such cheating via academic honesty statements will lead to lower exam scores.

Our study makes two main methodological contributions. First, our use of test scores as a proxy measure for student cheating avoids the pitfalls of using self-reported and subjective measures of cheating. Second, we conduct a randomized experiment, which is the "gold standard" for reliably estimating and identifying the effects of interventions. This methodology helps to ensure that the "treatment group" (e.g., those students who are given the honesty statement) and the "control group" (e.g., those students who are not given the statement) are drawn from the same subject pool, and who on average are identical.

Overall, we find no statistically significant differences in the examination performance of students who are given the academic honesty statements and students who are not given these

statements. We also test for heterogeneous treatment effects to see if the effects of academic honesty statements vary by student or course characteristics, and we do not find effects for any sub-groups. Our results suggest that academic honesty statements do not affect student performance in a significant way, so that their use does not have a statistically significant effect on reducing student cheating on examinations. However, other explanations are possible.

Previous research: Student cheating and academic honesty statements

There is little doubt that student cheating on examinations exists and is widespread across colleges of all types. The academic study of student cheating goes back at least a century (Barnes, 1904), but the systematic measurement of cheating began largely with the pioneering work of Bowers (1964).[5] He surveyed students across nearly 100 U.S. colleges and universities, asking students whether they engaged in various acts of academic honesty, and he found that roughly 75 percent of the 5,000 students who were surveyed admitted to one or more acts of dishonesty. A follow-up student survey by McCabe and Treviño (1997) for a small subset of these same institutions found similar overall results for more recent time periods, along with significant increases in cheating on examinations; in contrast, Brown and Emmett (2001) found no significant increases in student survey responses of cheating over this same period of study, and they explained the differences in results between studies by the use of different measures of cheating across studies, illustrating the sensitivity of any results to the self-reported measures of student cheating. Indeed, until recently, most studies of student cheating have followed Bowers (1964) by using student self-reports of cheating; that is, students are asked to fill out an anonymous survey asking if they have cheated and if they have observed cheating. The accuracy of such self-reports is of course uncertain.[6]

Much subsequent research has examined specific factors that affect student self-reported cheating in surveys, such as sex, grade point average, student perceptions of other students' behaviors, and student perceptions of institutional policies (including tolerance of and penalties for cheating). See McCabe, Treviño, and Butterfield (2001) for a somewhat dated but still useful survey of this literature and Tatum (2022) for a more recent and comprehensive review; for a review that focuses specifically on online courses, see Chang, Zhu, and Yu (2022). Lang (2013) has added that the design of the course itself can contribute to issues of academic integrity, such as when performance is prioritized over process, a semester grade relies almost exclusively on the results of one or two high stakes exams, or when students have a low expectation of success.

An additional and important factor that affects cheating may be the testing environment. Recent research by Alessio et al. (2017) has examined the differences in test scores between environments in which cheating is more likely, such as online versus in-person examinations or proctored versus unproctored exams. Alessio et al. (2017) found higher test scores in those environments in which cheating was more likely. For example, there was a 17 point differential in test scores between students who used a proctoring software versus students who did not.

Much research has also attempted to identify the effects of specific tools by which cheating might be reduced. It is here where academic honesty statements enter. The notion that these statements might discourage student cheating on examinations is based on several factors. The presence of an academic honesty statement might serve as a "nudge", as a "priming", or even as a "moral reminder" that cheating is unacceptable behavior.[7] The statement might also demonstrate an institutional commitment to ensuring academic integrity (including perhaps the institution's potential use of penalties for cheating). Additionally, the statement could send a

signal to an individual student that other students will resist cheating, and such peer effects have been shown to be an important factor in discouraging cheating.

Indeed, there are many studies that have found that academic honesty statements are associated with reduced cheating on examinations, at least in some contexts (McCabe 1993; McCabe, Butterfield, and Treviño 2003; McCabe and Treviño 1993, 1997; McCabe, Treviño, and Butterfield 1999, 2002). As noted by McCabe, Treviño, and Butterfield (2001) in their review of this work, all of these students are based on surveys of students across institutions of various sizes and types in which the students are asked about how prevalent is their cheating behavior and what factors influence their behavior.[8]

More recent research has demonstrated the fragility of this result. This research has generally found no consistent correlation between the presence of academic honesty statements and student self-reports of cheating (Roig and Marks 2006; Arnold et al. 2007; Konheim-Kalkstein, Stellmack, and Shilkey 2008; Mastin, Peszka, and Lily 2009; LoSchiavo and Shatz 2011). However, this work has also found that specific contextual factors may affect the impact of academic honesty statements on cheating (e.g., courses that combined online teaching with face-to-face teaching, courses that were taught by instructors with a strong perceived commitment both to their students and to academic integrity). There is also recent evidence generated from laboratory experiments (Mazar, Amir, and Ariely 2008; Shu et al. 2012; Corrigan-Gibbs et al. 2015). These experiments have the advantage of using direct measures of honesty and cheating, although they do not measure cheating in a typical classroom context (e.g., the laboratory environment is not "in the field"). Therefore, these experiments present concerns about external validity; that is, would the effects of the academic honesty statement differ

between the lab experiment and a real-world classroom? Overall, these studies have found somewhat mixed and inconclusive evidence of honesty statements.[9]

In sum, this research demonstrates that academic honesty statements have mixed and inconclusive results in encouraging student honesty on examinations. This is due in large part to widespread variations in the form of the honesty statements and to the many important but uncontrolled differences in institutional environments in which these academic honesty statements have been used and investigated. An especially important issue in this research is its reliance on indirect measures of student cheating that are derived from student self-reports of cheating, which makes it impossible to determine the direct impact of academic honesty statements on actual student cheating. The next section discusses our methods for resolving some of these issues.

Experimental design

As noted earlier, we conducted a randomized controlled field experiment that investigates the impact of academic honesty statements on student performance by measuring their effect on undergraduate student performance in two large introductory economics and introductory accounting courses at a large private university during a recent post-pandemic academic semester. These introductory classes consisted of 368 students in two sections of "Introductory Macroeconomics" (ECON 1020) and 231 students in three sections of "Financial Accounting" (ACCN 2010), all offered in the Fall 2022 semester.

One contribution of our paper is our use of a randomized controlled experiment. This experiment involves giving one group of randomly assigned students (the "treatment group") a series of academic honesty statements that the students must read and then agree to; a second

group of randomly assigned students (the "control group") is not given these statements. We then compared the test performance of the treatment group with the control group. Our maintained hypothesis is that students who received the academic honesty statements will be less inclined to cheat on the exam than students who did not receive these statements, and so the average test scores of the treatment group will be lower than the average scores of the control group, given that the treatment and control groups are drawn randomly from the same student population. Note that another important contribution of our paper is the use of an objective measure of student examination performance (or the test score) as a proxy measure of student cheating, rather than a measure based on student responses (and the honesty of these student responses) to a survey question that asks students whether they cheated on the exam. To provide more detail on our methods, we now list the specific steps in our analysis.

Our first step was to randomly assign students in each of the classes to one of two separate groups, following standard randomization procedures.[10] Our expectation is that the randomization will generate groups with control variables that are statistically indistinguishable from one another across the groups. Note that the initial randomization resulted in some differences between the treatment and the control groups, and the randomization step was repeated until few differences were detected.[11]

Table 1 presents the demographics of the students in the overall testing population, with treatment group and control group student demographics and academic characteristics reported separately; Table 1 also presents the results from difference-in-means tests for all these variables. The treatment and control groups were nearly identical across 26 student characteristics, with a few exceptions that would be normally expected.[12]

[INSERT TABLE 1 ABOUT HERE]

Second, following this randomization, we administered the first test (the "baseline" test) to both classes. This first exam for both classes consisted of an online examination of multiple-choice questions; these tests obviously differed between the economics and the accounting classes, but the test given to each class was identical. The first test had the attributes that it was completed during a single 75-minute class period, questions were randomly selected from a pool of questions by the online testing platform, question responses were shuffled for each individual question, and questions were administered one at a time, although students were allowed to move forward and backward between questions. Importantly, on this first examination there were no academic honesty statements for either the treatment group or the control group. The main purpose of the initial randomization was to ensure that the randomization of the student assignments generated identical test performance of the students in the two groups on the baseline (first) test.

Third, we administered the second test in both classes several weeks later. This test was also an online examination consisting of multiple choice questions that were identical for the treatment group and the control group and that followed the same testing protocols outlined above. However, on this second test we presented a series of academic honesty statements to the treatment group of students, while the control group did not receive these statements. We then compared the performance of the treatment group and the control group on this second test, with the hypothesis that the treatment group will exhibit lower average test scores than the control group because the treatment group will be less likely to cheat than the control group.

The academic honesty statements that were given to the treatment group of students on the second test consisted of the following four statements:

> *I understand that if someone other than me were to take this assessment on my behalf, or offer assistance to me while I took it, this would be considered academic dishonesty.*

I understand that if I take photos or copy any portion of this assessment either for my own private use or to upload to social media, any website, or the internet, this would be considered academic dishonesty.

I understand that if I copy any answers from someone else while taking this assessment, provide anyone else taking the assessment my answers, or give anyone who has not yet taken the assessment information in order to improve their performance on the assessment, that would be considered academic dishonesty.

I understand that if I use my computer to look at any resource from any website, from any document saved on my computer, or on Canvas other than the assessment I am taking, that would be considered academic dishonesty.

These statements were derived from an "Academic Integrity Assignment" published by John Wiley & Sons, Inc., as part of the Instructor Resources for the textbook *Financial Accounting – Tools for Business Decision Making* (9th Edition), written by Paul D. Kimmel, Jerry J. Weygandt, and Donald E. Kieso. There were eleven statements in the original assignment, as reported in the Appendix. From these eleven statements, we chose and revised four statements that were most relevant to our testing environment. Each of the four academic integrity statements that we used prompted the students to respond via a drop-down menu with a "Yes" or "No" response, and the Yes/No response choices were randomized by question and by student. These procedures ensured that the students presented with these statements had to be intentional while reading and responding to the questions. These statements closely resemble those used by many other institutions.

Several other test procedures are worth noting. First, students were seated physically in a classroom, and they completed the tests on student-owned computers, in the presence of multiple faculty and teaching assistant proctors. Given the size and configuration of the classrooms and the number of students taking the examinations, it was impossible to closely monitor the screens of all students simultaneously. Even so, the proctors were visible and active in observing the

students. Second, about 10 percent of the students received special accommodations that allowed them to take the examination in a separate physical testing facility, largely to allow them additional time for the test and/or to allow them to take the test in a less crowded environment. These students were closely monitored by academic personnel in this testing facility, but not by faculty or teaching assistants; faculty or teaching assistants were not allowed in this facility, and they had no direct control over this testing environment. Third, there was some overlap in the students in the two courses, with 35 students who took both the accounting and the economics courses. In order to reduce any potential bias in the results from students who were, say, in the treatment group in the accounting class and in the control group in the economics class, we dropped these students from our analysis. Our empirical results are based on student samples that do not include students who took both classes, although our results are largely unchanged even if we include these students.

The next section presents the data generated from this experimental design.

Data

Table 1 presents our data by control group and treatment group. As discussed earlier, the control and treatment groups were nearly identical. Most students were U.S. citizens (96.4 percent), were White and non-Hispanic (67.4 percent), and were sophomores (49.9 percent), followed by freshmen (32.9 percent). Almost half of the students had at least two majors (45.1 percent), with about three quarters having at least one business major (77.0 percent). Our sample was roughly evenly split between women (48.1 percent) and men (51.9 percent),[13] although women were somewhat underrepresented compared to the average student body, which is almost two-thirds women.

Statistical methodology

Since we conducted a randomized control experiment, we can analyze these data by simply taking the average difference in test scores between test 2 (or the test after treatment assignment) and test 1 (or the baseline, initial test) between the treatment and control groups to generate an unbiased estimate of the effect of the academic honesty statements on test scores. However, we improve on this simple approach in two ways. First, we include control variables to improve precision and to show that our estimates are not sensitive to which control variables are included, as would be the case in a randomized controlled experiment. Second, we use our baseline test 1 data to determine whether there is a difference in the *change* in test scores (test 2 – test 1) between the treatment and control groups. This allows us both to net out average differences in test scores between the treatment and control group (although, as discussed later, we found no differences, as would be the case in an experiment) and to improve precision by better controlling for the variability of test scores because test scores have a higher variance than a change in test scores.

To include control variables and to leverage our baseline test data, we conduct a difference-in-differences (DiD) regression.[14] Table 2 illustrates how the DiD is calculated (without controls), along with presenting the means for each group and test. We have two dimensions in our data, the group (treatment group, control group) and the test number (1, 2), providing four combinations. Let A, B, C, and D denote the mean test scores for these four group-by-test combinations, where each represents the mean score on the respective test (test 1 versus test 2) by the respective student group (with academic honesty statements versus no academic honesty statements). The change in test scores for the treatment group is given by (D-

B). We then net out the change in test scores that would have happened anyway (e.g., test 2 is slightly harder than test 1) by subtracting from this the change in test scores for the control group, given by (C-A). The resulting DiD estimate is therefore calculated as [(D-B)-(C-A)], or [D-B-C+A]. For example, Table 2 shows that the treatment group had a decrease in test scores of 2.06 percentage points (D-B), while the control group had a decrease of 2.58 percentage points (C-A), giving a DiD estimate of 0.52 percentage points (with a standard error of 2.42). Note that the DiD estimate can also be computed by differencing first by test and then by treatment (or [D-C)-(B-A)] to give again [D-C-B+A]. See also Figure 1, which shows graphically the very small and statistically insignificant differences in mean test scores for the treatment and the control groups.

[INSERT TABLE 2 ABOUT HERE]

[INSERT FIGURE 1 ABOUT HERE]

To estimate this DiD while also including control variables, we estimate regressions of the form:

$$S_{ij} = \alpha + \gamma_1 \, Treatment_i + \gamma_2 \, Test\ 2_j + \varphi \, [Treatment_i * Test\ 2_j] + X_i\beta + \varepsilon_{ij}, \qquad (1)$$

where S_{ij} is the score of student i on test j (j=1,2), X_i is a matrix of characteristics for student i (e.g., year in school, sex, major) with corresponding coefficient vector β, *Treatment* is an indicator variable for student i being assigned to the treatment group (=1), 0 otherwise, with corresponding coefficient γ_1, *Test 2* is an indicator variable for test 2 (=1), 0 otherwise, with corresponding coefficient γ_2, and ε_{ij} is the error term. Since we observe students over time, and since test scores for the same student are not independent, we cluster our standard errors at the student level.

The coefficient that provides our unbiased estimate of the effect of the treatment is the DiD coefficient φ, or the coefficient on the interaction term [*Treatment * Test 2*]. This coefficient measures the impact of the academic honesty statements on the scores of the treatment group of students relative to the control group of students on the second test relative to the first test. Our expectation is that this estimated coefficient will be negative; that is, students who receive the academic honesty statements will cheat less and thus score lower relative to students who do not receive the statements, on the second examination relative to the first examination.

Results

Table 3 reports our DiD regression estimates. We start by estimating the DiD regression without controls (column [1]) and then gradually add more controls in various specifications. Column [1] excludes all controls; column [2] includes section fixed effects, which control for average differences between course sections; column [3] includes section fixed effects and also test group fixed effects, the latter of which controls for average differences between test groupings (or versions of the same test but with different randomized question orderings); column [4] includes student demographic characteristics (e.g., sex, race, ethnicity, non-citizen status, and if the student has an exam writing accommodation) and high school GPA (or the quadratic of high school GPA), along with an indicator variable for whether or nor the high school GPA is missing; column [5] includes student classification (e.g., freshman, senior) fixed effects and program characteristic controls (e.g., economic major, business major, double major, triple or quadruple major, and indicator variables for having a major in each school); column [6] includes all control variables from [2] to [5]; and column [7] includes student-by-section fixed effects, providing the most saturated level of controls that control for all differences

between students and for all differences between the same student in different courses. Note that these fixed effects control for all the time-invariant controls from column [6].

[INSERT TABLE 3 ABOUT HERE]

Our main result is that, regardless of which controls we include, we find very small differences between the treatment and the control groups, ranging from a 0.329 percentage point increase (column [7]) to a 1.410 percentage point increase (column [2]) for the treatment group relative to the control group. Importantly, none of these differences are statistically significant; that is, there is no difference in the test performance of students who received the academic honesty statements and students who did not receive the statements.[15]

In addition to estimating average treatment effects, we also test if treatment effects differ by sub-group or by student or course characteristics. We estimate our preferred regression in column [6] with an added interaction between [*Treatment* * *Test 2*] and another variable (e.g., an ECON course indicator variable). The coefficient on this additional interaction shows if treatment effects differed by this variable, such as women reacting differently to the honesty statement relative to men. We test if treatment effects differ by sex, race, and ethnicity (e.g., women versus men, white non-Hispanic versus all others, and white non-Hispanic men versus all others), by student classification (freshman versus all others), by high school GPA, by course (ACCN versus ECON), by major (ECON or business school majors versus all others), and by whether or not the student had an accommodation so that they took their exam in a separate room.

Conclusions

Our results provide little evidence that academic honesty statements have any significant impact on student performance, at least in the specific institutional context that we examine here.

Our approach here is not without some limitations. Our interpretation of these results is that the academic honesty statements do not reduce cheating. However, other explanations are possible. For example, there could be little cheating on the examinations, so that the academic honesty statements have little room for any effect. Another possibility is that the academic honesty statements could in fact reduce cheating, but the reduced cheating does not affect student performance. We cannot rule out these alternative explanations.

Relatedly, while we have an objective proxy measure of student *performance*, our measure of student *performance* may not fully capture actual student *cheating*.[16] We infer student cheating by comparing the performance of students in the treatment group relative to students in the control group, while controlling for as many other possible determinants of student performance that are allowed by our data set. We believe that our measure is a distinct improvement over subjective measures of student cheating that assume students will self-report their cheating honestly. Even so, we admit that we cannot verify this assumption. In any event, we believe that our use of this alternative measure instead of the self-reported measure is an innovation that merits additional study.

In addition, it is possible that the treatment itself could affect grades for reasons other than the academic honesty statements. For example, if the statements increase student anxiety (even if they do not cheat), this might negatively affect assessment performance. Also, the additional time required to read and agree to the statements, albeit minimal, reduces the time the students have to complete the exam, and, relatedly, results could be skewed if students who were cheating took time during the exam to access outside resources. However, given that time did not

appear to be a significant constraint for the students in the two groups, these time issues seem unlikely to have affected our results.

Regardless, we cannot be certain that these types of treatment effects had no impact on our results. Thus, if academic honesty statements have no impact on student honesty, what other variables should be considered that have a greater impact on improving academic integrity? For example, as discussed earlier, our testing design had four specific attributes: the test was completed during a single 75-minute class period, questions were randomly selected from a pool of questions by the online testing platform, question responses were shuffled for each individual question, and questions were administered one at a time. Additional research could include isolating each of these four attributes in a similarly designed controlled study.

In any event, our results indicate that academic honesty statements, in and of themselves, have no significant impact on student performance, so that their use seems unlikely to be a reliable tool in reducing student cheating on examinations. However, these results do not necessarily mean that such statements, or any of the other mentioned testing design variables, cannot have some impact when they are part of a larger, coordinated, and sustained effort to address cheating. Indeed, the same academic studies that indicated these statements may, or may not, affect student cheating as a *single* tool often find that using them as part of a *multi-faceted* approach could be successful in reducing student cheating, as suggested by research by Kerkvliet and Sigmund (1999), McCabe, Trevino, and Butterfield (2001), and, more recently, Pieters (2024). This multi-faceted approach could also address the overall institutional climate and the design of courses, not just individual test assessments.

Indeed, the most effective strategies for reducing student cheating seem to be those that implement multi-dimensional policies that affect the many motivations for cheating. For

example, students need to recognize that their cheating may be caught and penalized, which requires clear communication both of policies regarding academic integrity by their faculty, and, more importantly, of policies regarding the academic administration of the institution. In addition, however, the design of a course should make it easier for students to behave ethically. By clearly communicating expectations and policies (especially those regarding behavior that constitutes appropriate conduct and behavior that constitutes cheating), prioritizing learning over output, and providing multiple opportunities for students to demonstrate such learning, a faculty member can create a class culture that helps facilities academic integrity of students. Finally, faculty and their academic administration should create an environment in which students trust that they will be treated fairly and consistently.

It is especially here where a student honor code – and academic honesty statements – may still play a role in creating trust in the individual classroom and trust in the overall institutional environment. Indeed, improving trust may be the most effective way of building academic integrity – and of reducing student cheating.

Disclosure statement
The authors report that there are no conflicts of interest to declare.

ORCiD
James Alm https://orcid.org/0000-0003-4008-1637
Patrick Button https://orcid.org/0000-0001-5414-9730

Acknowledgments
This study was approved by Tulane University's Institutional Review Board (#2019-1516). We are grateful to Elliott Isaac for many helpful comments. We are also grateful to the editors and to an anonymous referee for many useful suggestions.

References

Alessio, Helaine M., Nancy Malay, Karsten Maurer, A. John Bailer, and Beth Rubin. 2017. Examining the effect of proctoring on online test scores. *Online Learning* 21 (1).

Alm, James, Jay A. Soled, and Kathleen DeLaney Thomas. 2023. Multi-billion dollar tax questions. *Ohio State Law Journal* 84 (4): 895-952.

Arnold, Ivo J. M. 2016. Cheating at online formative tests: Does it pay off? *The Internet and Higher Education* 29 (1): 98-106.

Arnold, Rodny, Barbara N. Martin, and Linda Bigby. 2007. Is there a relationship between honor codes and academic dishonesty? *Journal of College & Character* 8 (2): 1-20.

Babbie, Earl R. 2021. *The Practice of Social Research* (15th Edition). Boston, MA: Cengage Publishing.

Barnes, Earl. 1904. Student honor: A study in cheating. *International Journal of Ethics* 14 (4): 481-488.

Bowers, William J. 1964. *Student Dishonesty and its Control in College.* New York, NY: Bureau of Applied Social Research, Columbia University.

Brown, Bob S., and Dennis C. Emmett. 2001. Explaining variations in the level of academic dishonesty in studies of college students: Some new evidence. *College Student Journal* 35 (4): 529-538.

Burns, Douglas N., Steven B. Caudill, and Daniel M. Gropper. 1992. Crime in the classroom: An economic analysis of undergraduate student cheating behavior. *The Journal of Economic Education* 23 (3): 197-207.

Chang, Feng-Kuang, Dan Zhu, and Wenhao Yu. 2022. A systematic review of academic dishonesty in online learning environments. *Journal of Computer Assisted Learning* 38 (4): 907-928.

Corrigan-Gibbs, Henry, Nakull Gupta, Curtis Northcutt, Edward Cutrell, and William Thies. 2015. Deterring cheating in online environments. *ACM Transactions on Computer-Human Interaction* 22 (6): Article No. 28.

Cox, D. R., and Nancy Reid (2000). *The Theory of the Design of Experiments*. Boca Raton, FL: Chapman & Hall/CRC, Monographs on Statistics and Applied Probability.

Fask, Alan, Fred Englander, and Zhaobo Wang (2014). "Do Online Exams Facilitate Cheating? An Experiment Designed to Separate Possible Cheating from the Effect of the Online Test Taking Environment." *Journal of Academic Ethics*, 12 (1), 101-112.

Hall, Teresa L., and George D. Kuh (1998). "Honor Among Students: Academic Integrity and Honor Codes at State-assisted Universities." *NASPA Journal*, 36 (1), 2-18.

Harrison, Glenn W., Morton I. Lau, and E. Elisabet Rutström (2009). "Risk Attitudes, Randomization to Treatment, and Self-selection into Experiments." *Journal of Economic Behavior & Organization*, 70 (3), 498-507.

Kerkvliet, Joe. 1994. Cheating by economics students: A comparison of survey results. *The Journal of Economic Education* 25 (2): 121-133.

Kerkvliet, Joe, and Charles L. Sigmund. 1999. Can we control cheating in the classroom? *The Journal of Economic Education* 30 (4): 331-343.

Kimmel, Paul D., Jerry J. Weygandt, and Donald E. Kieso. 2018. *Financial Accounting – Tools for Business Decision Making* (9th Edition). Hoboken, NJ: John Wiley & Sons, Inc.

Konheim-Kalkstein, Yasmine L., Mark A. Stellmack, and Margaret L. Shilkey. 2008. Comparison of honor code and non-honor code classrooms at a non-honor code university. *Journal of College & Character* 9 (3): 1-13.

Kristal, Ariella S., Ashley V. Whillans, Max H. Bazerman, Francesca Gino, Lisa L. Shu, Nina Mazar, and Dan Ariely. 2020. Signing at the beginning versus at the end does not decrease dishonesty. *Proceedings of the National Academy of Sciences* 117 (13): 7103-7107.

Lang, James M. 2013. *Cheating Lessons: Learning From Academic Dishonesty*. Cambridge, MA: Harvard University Press.

List, John A., Sally Sadoff, and Mathis Wagner. 2011. So you want to run an experiment, now what? Some simple rules of thumb for optimal experimental design. *Experimental Economics* 14 (4): 439-457.

Lohr, Sharon L. 2021. *Sampling: Theory and Design* (3rd Edition). Boca Raton, FL: CRC Press/Taylor & Francis Group, a Chapman & Hall Book.

LoSchiavo, Frank M., and Mark A. Shatz. 2011. The impact of an honor code on cheating in online courses. *Journal of Online Learning and Teaching* 7 (2): 179-182.

Martinelli, César, Susan W. Parker, Ana Cristina Pérez-Gea, and Rodimiro Rodrigo. 2018. Cheating and incentives: Learning from a policy experiment. *American Economic Journal: Economic Policy* 10 (1): 298-325.

Mazar, Nina, On Amir, and Dan Ariely (2008). "The Dishonesty of Honest People: A Theory of Self-concept Maintenance." *Journal of Marketing Research*, 45 (6), 633-644.

McCabe, Donald L. 1993. Faculty responses to academic dishonesty: The influence of student honor codes. *Research in Higher Education* 34 (5): 647-658.

McCabe, Donald L., Kenneth D. Butterfield, and Linda Klebe Treviño. 2003. Faculty and academic integrity: The influence of current honor codes and past honor code experiences. *Research in Higher Education* 44 (3): 367-385.

McCabe, Donald L., and Linda Klebe Treviño. 1993. Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education* 64 (5): 522-538.

McCabe, Donald L., and Linda Klebe Treviño. 1997. Individual and contextual influences on academic dishonesty: A multicampus investigation. *Research in Higher Education* 38 (3): 379-396.

McCabe, Donald L., Linda Klebe Treviño, and Kenneth D. Butterfield. 1999. Academic integrity in honor code and non-honor code environments: A qualitative investigation. *The Journal of Higher Education* 70 (2): 211-234.

McCabe, Donald L., Linda Klebe Treviño, and Kenneth D. Butterfield. 2001. Cheating in academic institutions: A decade of research. *Ethics & Behavior* 11 (3): 219-232.

McCabe, Donald L., Linda Klebe Treviño, and Kenneth D. Butterfield. 2002. Honor codes and other contextual influences on academic integrity: A replication and extension to modified honor code settings. *Research in Higher Education* 43 (3): 357-378.

Mixon, Franklin G., Jr. 1996. Crime in the classroom: An extension. *The Journal of Economic Education* 27 (3): 195-200.

Pieters, Gina C. 2024. Designing effective assessments in economics courses: Guiding principles. *The Journal of Economic Education* 55 (1): 63-76.

Roig, Miguel, and Amanda Marks. 2006. Attitudes toward cheating before and after the implementation of a modified honor code: A case study. *Ethics & Behavior* 16 (2): 163-171.

Shu, Lisa L., Nina Mazar, Francesca Gino, Dan Ariely, and Max H. Bazerman. 2012. Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end. *Proceedings of the National Academy of Sciences* 109 (38): 15197-15200.

Stack, Steven. 2015. The impact of exam environments on student test scores in online courses. *Journal of Criminal Justice Education* 26 (3): 273-282.

Tatum, Holly A. 2022. Honor codes and academic integrity: Three decades of research. *Journal of College & Character* 23 (1): 32-47.

Thaler, Richard H., and Cass R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven, CT: Yale University Press.

Vandehey, Michael, George Diekhoff, and Emily LaBeff. 2007. College cheating: A twenty-year follow-up and the addition of an honor dode. *Journal of College Student Development* 48 (4): 468-480.

Wuthisatian, Rattaphon. 2020. Student exam performance in different proctored environments: Evidence from an online economics course. *International Review of Economic Education* 35 (2): Article No. 100196.

Zhao, Li, Haiying Mao, Brian J. Compton, Junjie Peng, Genyue Fu, Fang Fang, Gail D. Heymand, and Kang Lee. 2022. Academic dishonesty and its relation to peer cheating and culture: A meta-analysis of the perceived peer cheating effect. *Educational Research Review* 36 (7): Article No. 100455.

Table 1. Pre-treatment means of student demographic characteristics by assignment (control group versus treatment group).

	Mean		Difference	
	Control (N=256)	Treatment (N=269)	Difference	p-value
Test 1 Score (%)	66.300	67.746	-1.446	0.4119
High School GPA	3.088	3.336	-0.248	0.0115
Missing High School GPA	0.129	0.078	0.051	0.0554
Sex				
Female	0.441	0.520	-0.079	0.0703
Male	0.559	0.480	0.079	0.0703
Race and Ethnicity				
White	0.648	0.699	-0.050	0.2184
Black or African American	0.055	0.037	0.018	0.3378
Hispanic	0.148	0.138	0.011	0.7221
Asian	0.023	0.045	-0.021	0.1833
Multi-Racial	0.055	0.026	0.029	0.0942
Unknown or Not Reported	0.031	0.022	0.009	0.5257
Non-Citizen	0.039	0.033	0.006	0.7317
Has Exam Accommodation	0.105	0.112	-0.006	0.8240
Classification				
Freshman	0.316	0.342	-0.026	0.5337
Sophomore	0.531	0.468	0.063	0.1506
Junior	0.125	0.175	-0.050	0.1116
Senior	0.023	0.011	0.012	0.2793
Major Characteristics				
Double Major	0.406	0.494	-0.088	0.0425
Triple or Quadruple Major	0.039	0.041	-0.002	0.9150
Economics Major	0.035	0.026	0.009	0.5437
Finance Major	0.098	0.123	-0.025	0.3616
School(s) of Major(s)				
Liberal Arts	0.266	0.223	0.043	0.2569
Business	0.766	0.773	-0.008	0.8364
Science and Engineering	0.129	0.167	-0.038	0.2173
Public Health	0.020	0.019	0.001	0.9317
Professional Advancement	0.016	0.011	0.004	0.6559
Architecture	0.031	0.074	-0.043	0.0280

Notes: Student demographic and program data are from the Office of the Registrar. Students are deemed in a school if any of their majors is housed in that school, so that the school variables do not sum to one. To match the analysis, we exclude the few students with no Test 1 exam scores, although results are nearly identical if they are included.

Table 2. Mean exam scores by test and group and difference-in-differences estimate,

		Test		
		Test 1	**Test 2**	**Difference**
Group	**Control group**	No statement [A] 66.30% (1.23) N = 256	No statement [C] 63.72% (1.14) N = 257	[C-A] -2.58 (1.68)
	Treatment group	No statement [B] 67.75% (1.25) N = 269	With statement [D] 65.69% (1.20) N = 268	[D-B] -2.06 (1.73)
	Difference	[B-A] 1.45 (1.76)	[D-C] 1.96 (1.65)	**DiD** 0.52 (2.42)

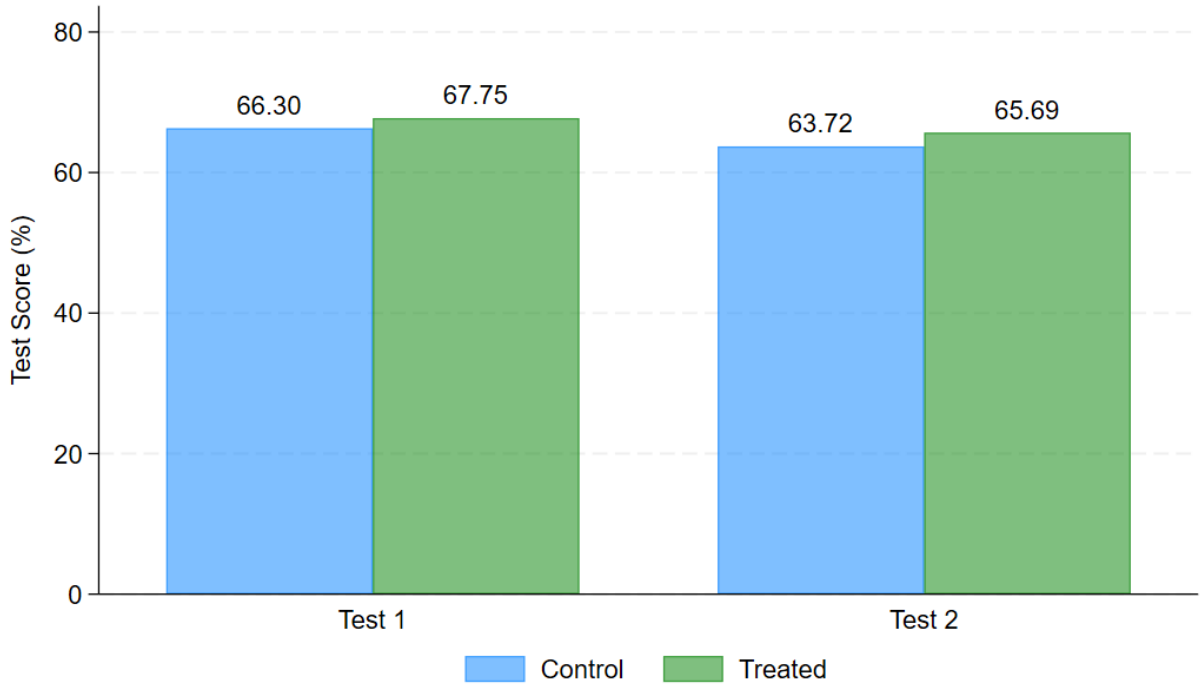
Notes: DiD = [(D-B)-(C-A)] = [D-B-C+A] = [D-C-B+A] = [(D-C)-(B-A)]. Standard errors are in parentheses.

Table 3. Treatment effect estimates: Difference-in-differences estimation.

	Specification						
Variable	**[1]**	**[2]**	**[3]**	**[4]**	**[5]**	**[6]**	**[7]**
Treatment X Test 2	0.517	1.410	0.454	0.482	0.395	0.398	0.329
	(1.120)	(1.124)	(1.120)	(1.130)	(1.126)	(1.134)	(1.577)
Test 2	-2.578***	-2.987***	-2.499***	-2.539***	-2.486***	-2.446***	-2.383**
	(0.829)	(0.850)	(0.833)	(0.837)	(0.831)	(0.842)	(1.159)
Treatment	1.446	…	…	-0.115	1.360	…	…
	(1.760)			(1.708)	(1.679)		
Section Fixed Effects?		Yes	Yes			Yes	Yes
Test Group Fixed Effects?			Yes			Yes	Yes
Student Demographics?				Yes		Yes	Yes
High School GPA?				Yes		Yes	Yes
Student Classification?					Yes	Yes	Yes
Program Characteristics?					Yes	Yes	Yes
Student-by-Section Fixed Effects?							Yes
N	1,050	1,050	1,050	1,050	1,050	1,050	1,050
Adjusted R^2	0.0026	0.4746	0.4751	0.0980	0.1390	0.5232	0.7905

Notes: Regression estimates are based on equation (1). *Section* refers to the five separate course sections in ECON (two sections) and ACCN (three sections). *Test Group* refers to versions of the same test but with different randomized question ordering. *High School GPA* is included as a quadratic term, along with an indicator variable for not having a high school GPA reported. *Student Demographics* include fixed effects for sex, race, or ethnicity, non-citizen status, and if the student has an exam writing accommodation. *Program Characteristics* include indicator variables for having an economics major, a business major, a double major, a triple or quadruple major, and indicator variables for having a major in each school. Standard errors, clustered on student, are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure 1. Mean test scores by test number and treatment status.



Notes: No mean comparisons are statistically significantly different from each other at conventional levels (10% or lower).

Appendix

1. If I take photos of any quiz or examination, group project or homework, then distribute it to any third party, no matter what my purpose is, this [ ] be considered academic dishonesty.

2. If I obtain any document which contains any portion of a prior or current quiz or examination, group project or homework, no matter how innocent my purpose is, this [ ] be considered academic dishonesty.

3. If I share any quiz or examination, group project or homework with someone else, this [ ] be considered academic dishonesty.

4. Unless it has been previously approved by my professor, if I submit anything for our course which is not the product of my own work, this [ ] be considered academic dishonesty.

5. If someone other than me were to take any quiz or examination on my behalf, this [ ] be considered academic dishonesty.

6. If I am aware of any instance of academic dishonesty which may be occurring in our class, and I do not report this to my professor or TAs, then this [ ] be considered academic dishonesty.

7. If I collaborate with anyone else on any homework or group project which has not been previously approved by my professor or TA, this [ ] be considered academic dishonesty.

8. If I obtain any document which contains any portion of a prior or current quiz or examination, this [ ] be considered academic dishonesty.

9. If I copy any answers from anyone else on a quiz, examination, group project or homework, this [ ] be considered academic dishonesty.

10. If I upload or post any quiz or examination or course materials (including my professor's PowerPoints or outlines) to social media, any website or the Internet, this [ ] be considered academic dishonesty.

11. If I upload or post any quiz or examination, group project or homework to social media, any website or the Internet, this [ ] be considered academic dishonesty.

Notes

[1] Some institutions build into their honestly statements the requirement that students report on each other if they witness a violation (Lang 2013)

[2] See https://www.umt.edu/business-ethics/academic-honesty.php.

[3] For example, Form 1040 in the individual income tax includes a statement at the end of the form that requires an individual signature, stating that "Under penalties of perjury, I declare that I have examined this return and accompanying schedules and statements, and to the best of my knowledge and belief, they are true, correct, and complete."

[4] This pre-treatment test allows us to test for differences in the outcome variable between the treatment and control group before treatment occurs, which is the most important balance test for randomized control trials to show that the randomization was successful. We also use the pre-treatment data to estimate if the treatment results in a change in test scores (from test one, or the pre-treatment test, to test two, or the post-treatment test), rather than if post-treatment test scores differ. This improves statistical power and is akin to a "difference-in-differences" regression methodology, although with a randomly assigned treatment.

[5] Student cheating has also been examined by the "popular press". For example, see: Jessica Cheung, "The Fading Honor Code", *The New York Times*, 13 April 2014, available online at https://www.nytimes.com/2014/04/13/education/edlife/the-fading-honor-code.html?smid=em-share; Christian B. Miller, "Just How Dishonest Are Most Students?", *The New York Times*, 13 November 2020, available online at https://www.nytimes.com/2020/11/13/opinion/sunday/online-learning-cheating.html?smid=em-share; and Suzy Weiss, "Dishonor Code: What Happens When Cheating Becomes the Norm?", *The Free Press*, 23 February 2013, available online at https://www.thefp.com/p/dishonor-code-what-happens-when-cheating?utm_source=post-email-title&publication_id=260347&post_id=105297296&isFreemail=false&utm_medium=emailhttps://www.thefp.com/p/dishonor-code-what-happens-when-cheating?utm_source=post-email-title&publication_id=260347&post_id=105297296&isFreemail=false&utm_medium=email.

[6] See Bunn, Caudill, and Gropper (1992), Kerkvliet (1994), and Mixon (1996), who examine the role of incentives in student cheating using student self-reports of cheating in randomized surveys.

[7] The use of "nudges" as a way of encouraging certain types of behavior has been popularized by Sunstein and Thaler (2008), and their use is now widespread across many dimensions of behavior. For a specific application of nudges to another type of cheating, or individual income tax compliance, see Alm, Soled, and Thomas (2023).

[8] For example, McCabe and Treviño (1993) used as their measure of student cheating "…a composite measure consisting of twelve types of self-reported academic dishonesty: using crib notes on a test; copying from another student during a test; using unfair methods to learn what was on a test before it was given; copying from another student during a test without their knowledge; helping someone else to cheat on a test; cheating on a test in any other way; copying material and turning it in as your own work; fabricating or falsifying a bibliography; turning in work done by someone else; receiving substantial, unpermitted help on an assignment; collaborating on an assignment when the instructor asked for individual work; copying a few sentences of material from a published source without footnoting it", with students then asked "…to identify the frequency with which they had engaged in each of these behaviors on a four-

point Likert scale (from never = 1 to many times = 4)". This approach has been the common and the dominant method for measuring student cheating.

[9] Notably, the study by Shu et al. (2012) has also been criticized for fabricating and manipulating the data, as first discovered by researchers at the Data Colada blog, and the Shu et al. (2012) paper has now been retracted by Kristal et al. (2020). For further discussion, see various posts available online at: https://datacolada.org/98, https://datacolada.org/109, https://datacolada.org/110, https://datacolada.org/110, and https://datacolada.org/112. See also various popular press reports, such as Gideon Lewis-Kraus, "They Studied Dishonesty. Was Their Work a Lie?", *The New Yorker*, 9 October 2023, available online at https://www.newyorker.com/magazine/2023/10/09/they-studied-dishonesty-was-their-work-a-lie.

[10] There is an enormous literature on randomization. For useful discussions of randomization methods, see Cox and Reid (2000) and Babbie (2021), and Lohr (2021); see also the especially detailed discussion in List, Sadoff, and Wagner (2011). Based on the design suggestions of these works, we used the following procedures in the randomization:

1. We used a random number generator to give each student a random number from the uniform distribution [0, 1].
2. We sorted the students by the resulting random number values.
3. We assigned the students with the lowest 50 percent of the random numbers as the treatment group and the students with the highest 50 percent of the random numbers as the control group.
4. We checked the balance of two groups by computing means of various control variables available to us (e.g., declared major, year in school, students who receive testing accommodations) and performing standard t-tests between the treatment and control group means for each variable.

[11] Again following Cox and Reid (2000), Babbie (2021), Lohr (2021), and List, Sadoff, and Wagner (2011), we re-randomized the two groups if more than 20 percent of the t-tests for a control variable were significantly different for the two groups at the 5 percent level. As noted later, for our final samples, only 3 of the 26 control variables were statistically different across the two groups.

[12] For the 26 variables in Table 1, we would expect there to be on average one or two significant differences at the 5 percent level. In our case, we have three significant differences at the 5 percent level. Having at least three significant differences at the 5 percent level is uncommon but not unusual. Indeed, this occurs 8.87 percent of the time due to random chance, calculated as: $P(X \geq 3) = 1 - P(X = 0) - P(X = 1) - P(X - 2) = 1 - C(26,0) * (0.05)^0 * (0.95)^{26} - C(26,1) * (0.05)^1 * (0.95)^{25} - C(26,2) * (0.05)^2 * (0.95)^{24} = 0.0887$. Harrison, Lau, and (2009) show that circumstances like this do not cause bias in randomized control trials.

[13] The data we use from the registrar included information on legal sex, and only included female and male as possible categories.

[14] Note that our use of a DiD approach does not face the typical concerns around identifying assumptions for the DiD to provide causal estimates; that is, because our treatment is randomized, the DiD assumptions (especially the parallel trends assumptions) are satisfied.

[15] Note that we also administered a third test in which we repeated the first test procedure where neither group was given any academic honesty statements. This third test was intended to determine whether the academic honesty statements for the treatment group on the second test had any carryover, or long-lasting, effects. In fact, we did not find any effects of the academic

honesty statements either on the second test (e.g., the short-run impact) or on the third test (e.g., the long-run impact). We do not report the results from the third test, but all results are available upon request.

[16] For a recent innovative study that uses direct measures of student cheating, see Martinelli et al. (2018). Their approach was not feasible here because their method involves the statistical detection for all possible pairs of students whose response patterns are unusually, and statistically unlikely, similar.