Measuring and Maximizing the Effectiveness of Honor Codes in Online Courses

Henry Corrigan-Gibbs
Stanford University
Stanford, CA, USA
henrycg@stanford.edu

Edward Cutrell
Microsoft Research India
Bangalore, Karnataka, India
cutrell@microsoft.com

Nakull Gupta
Microsoft Research India
Bangalore, Karnataka, India
t-nakgup@microsoft.com

William Thies
Microsoft Research India
Bangalore, Karnataka, India
thies@microsoft.com

Curtis Northcutt
MIT EECS
Cambridge, MA, USA
cgn@mit.edu

L@S 2015, March 14–18, 2015, Vancouver, BC, Canada.
ACM 978-1-4503-3411-2/15/03.
http://dx.doi.org/10.1145/2724660.2728663

Abstract
We measure the effectiveness of a traditional honor code at deterring cheating in an online examination, and we compare it to that of a stern warning. Through experimental evaluation in a 409-student online course, we find that a pre-task warning leads to a significant decrease in the rate of cheating while an honor code has a smaller (non-significant) effect. Unlike much prior work, we measure the rate of cheating directly and we do not rely on potentially inaccurate post-examination surveys. Our findings demonstrate that replacing traditional honor codes with warnings could be a simple and effective way to deter cheating in online courses.

Author Keywords
cheating; honeypot; honor code; online course; MOOC

ACM Classification Keywords
K.3 [Computers and Education]; J.4 [Social and behavioral sciences]: Psychology.

Introduction
Honor codes are surprisingly effective at discouraging cheating [9] and encouraging truth-telling [6] in physical encounters. In online environments, however, prior work has not found that honor codes have a discernible impact on cheating behavior [7, 8]. In spite of the lack of

evidence supporting the use of traditional honor codes on the Web, online education platforms are adopting and deploying conventional honor codes for online use [3, 4].

In this work, we use a randomized controlled trial to rigorously evaluate the effectiveness of a conventional honor code at deterring cheating in an online course. In addition, we evaluate a potential alternative to traditional honor codes: a stern pre-task *warning*. By priming test-takers with a warning that details the potential consequences of cheating, we aim to increase the perceived costs and decrease the expected benefits of cheating. We evaluated the effectiveness of our pre-task honor code and warning in the context of the final examination for an online course in India.

Along the way, we introduce a new a method for directly measuring rates of cheating in online courses that does not rely on potentially inaccurate *post hoc* self-reporting. To measure cheating, we created a Google-indexed "honeypot" website containing key words and phrases from the online exam. Using a tracking cookie, we were able to identify students who visited our honeypot website while taking the exam—in violation of the rules of the exam. We also manually analyzed answers to a free-response question to detect copying from peers and online resources (other than the honeypot).

Prior work on physical-world cheating informs our approach. Mazar et al. theorize that people decide to cheat based on (1) the expected payout (estimated costs and benefits) of cheating and (2) "the manner in which the act of" acting dishonestly makes a person "perceive themselves" [9]. Honor codes target the latter mechanism: by reminding participants of their ethical standards, an honor code makes it more difficult to maintain a positive self concept and still cheat. The warnings we used in this work target the former mechanism: by informing participants of the consequences of cheating, we aimed to increase the perceived costs and decrease the expected benefits of cheating.

Our primary finding is that *a pre-task warning led to a statistically significant decrease* in the rate of cheating in a real online examination in India. Application of the warning decreased the rate of cheating to 15.5% from a baseline of 34.4%. The rate of cheating under an honor code was 25.5%, suggesting a potential benefit; however, consistent with prior work [7, 8], we did not find this benefit to be statistically significant. Our results indicate that a simple warning may be more effective at deterring would-be cheaters than a traditional honor code. With the application of appropriate warnings, administrators of online courses may be able to effectively promote honest test-taking behavior at negligible cost.

Related Work

Prior work has investigated the prevalence of cheating in unsupervised tasks and the question of how to influence cheating behavior by modifying the task. In particular, past studies find pre-task honor codes effective in *physical encounters*. Mazar et al. demonstrated that having students recall the Ten Commandments or sign an honor code before completing a self-graded task reduced the rate of cheating in a classroom setting [9]. Pruckner and Sausgruber demonstrated that placing an appeal to honor on the price tag of a newspaper box decreased the rate of newspaper theft [10].

In contrast, there is scant evidence that honor codes are effective *online*. Mastin et al. did not observe any effect of an honor code in the context of an online psychology experiment [8]. LoSciavo and Shatz were unable to find

Honor Code: Please show that you will respect the rules of the exam by typing the following text into the box below:

> I promise not to visit other websites or take help from other people during the exam.

Figure 1: Honor Code.

Warning: **Do not** visit other websites or take help from other people during the exam. If we discover that you did either of these things, we may:

- Cancel your exam.
- Cancel your account on MEC.
- Notify your institution.

Please show that you understand the rules of the exam by typing the following text into the box below:

> I understand the consequences of visiting other websites or taking help from other people during the exam.

Figure 2: Warning.

any effect of an honor code on *self-reported* cheating in an online exam [7]. Other work found an honor code ineffective in by-mail interactions [5].

This paper demonstrates that a simple warning can have an effect on rates of cheating, even in situations in which an honor code does not offer commensurate benefits. There is some precedent for the idea that increasing the perceived risks of cheating can decrease the cheating rate. For example, Braumoeller and Gaines find that overt use of plagarism detection software may deter cheating in the classroom [2].

Methods

In the spring of 2014, we conducted a free online course ("The Design and Analysis of Algorithms") targeted at undergraduate engineering students in India. We gave students who completed the online lectures and activities the option of taking a final exam to qualify for a certificate. We offered proctored in-person versions of the exam in five different cities and offered an online version of the exam for students who were not able to travel to a test site.

The exam consisted of fifteen multiple-choice questions[1] and one free-response question. All questions were original, required critical thinking, and would not easily benefit from third-party reference materials.

We required that students taking the exam: (1) not consult other materials (books, notes or other websites) while taking the test, and (2) neither give nor receive aid from other people during the exam. To explore the effect of honor codes and warnings on students' compliance with these rules, we randomly assigned all students taking the online exam to one of three conditions:

- **No additional instructions.** This was identical to exams provided to students in the proctored setting.
- **Honor code.** For this condition, students were asked to read and type out an honor code at the beginning of the exam (Figure 1). This code was designed to appeal to their sense of integrity to follow the rules of the exam.
- **Warning.** For this condition, students were asked to read and type out a warning statement at the beginning of the exam (Figure 2). This statement emphasized the negative consequences of breaking the rules by cheating.

To ensure that students in the latter two conditions actually read the honor code or warning, we asked students to type one sentence of the honor code or warning into a text box. We displayed the relevant text as an embedded image to prevent students from copying-and-pasting the sentence.

We employed two different techniques to estimate the prevalence of cheating:

1. Examination of "free response" questions. The free response question was one of the harder questions on the exam. It required students to design a graph algorithm and describe it informally in a few sentences (our solution was 38 words long). To catch students who copied their answer directly from an Internet website, we identified responses with idiosyncratic language or symbols and entered them into an Internet search engine.

[1] Due to an ambiguity in one question, only 14 multiple-choice questions were graded.

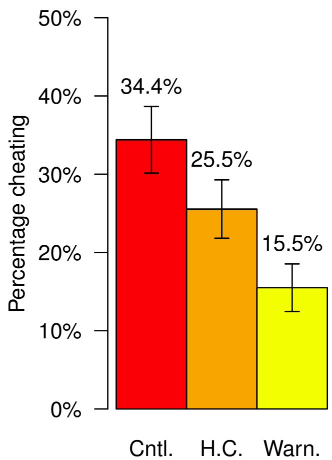

Figure 3: Rates of cheating on the online exam. Error bars indicate the standard error of the proportion.

We labeled students who submitted responses that exactly matched the text on a website as cheaters. To identify students who copied answers from other students, we performed manual comparison of responses. We ranked all response pairs by longest common substring, and also by longest common subsequence, and inspected pairs with high scores. In addition, we sorted all responses alphabetically, and also by total length, and examined adjacent entries. Three of the authors, all blind to treatment condition, jointly examined similar responses for cheating. For each pair that evidenced cheating, we labelled both students as cheaters.

2. "Honeypot" website. To help detect students who consulted the Web to seek help on the exam, we placed all of the exam questions on a public-facing website (the "honeypot") that was indexed by Google. If students searched for the exact text of any exam question, our website was the first hit returned. The website did not include the answers to the questions, but it did include a button ("Click to show answer") for each question on the test; when that button was pressed, the website paused and simulated a timeout, without showing the answer. We instrumented the honeypot website to check for a cookie set by our online exam platform. The cookie allowed us to identify the students who tried to download the exam answers from the honeypot page.

Ethical Issues

Prior to the exam we sent an email to all participating students. The email explained that we would be performing a research study during the final exam to test features designed to improve the fairness and reliability of future online examinations. Students had the opportunity to opt out of the study, in which case they received a default version of the exam (following other MOOCs, the default version used the honor code). When we communicated the scores to students, we debriefed them on the purpose of the experiment and on the methods we used to conduct it.

We did not take any disciplinary actions against students who we labeled as cheaters on the exam. Our experimental protocol received approval from our institutions' ethics review boards.

Participants

There were 409 students who took the online exam (and 674 students who took the in-person exam). We exclude from our analysis two students who opted out of the experiment, and three students who did not advance past the exam's instructions page. Thus we analyze data for 404 students.

Results

Overall, we classified 24% of students taking the exam as cheaters. The breakdown of cheaters by experimental condition is shown in Figure 3. Cheating was highest in the baseline condition (34% of students), followed by the honor code condition (25% of students), followed by the warning condition (15% of students). The difference between the baseline condition and the warning condition was significant[2] ($\chi^2(1, 267) = 11.9$, $p < 0.001$). The difference between the honor code and the warning has borderline significance without correction ($\chi^2(1, 279) = 3.74$, uncorrected $p = 0.053$) but is not significant with correction ($p = 0.11$).

Of the 100 students who we classified as cheaters, 84% were connected to plagiarism on the free response

[2]We present p values for the χ^2 tests as adjusted using the Bonferroni-Holm correction.

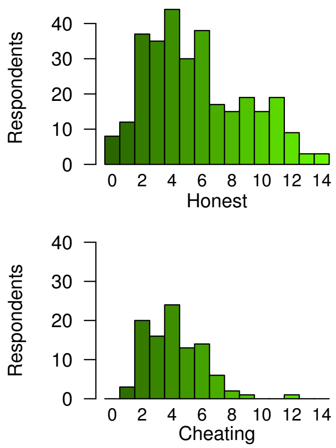

Figure 4: Histogram showing the number of correct exam answers for honest and cheating participants in the online exam.

question, 18% visited the honeypot, and 2% were connected to both plagiarism and the honeypot. Of the 23 unique IP addresses that visited the honeypot site during the exam, there were 5 that we could not associate with any student. These may represent visitors who were not taking the exam, or may correspond to additional devices used by exam participants to access the honeypot but not to access the exam site.

Among the plagiarized responses, most (80%) showed similarity to another student's response, 42% showed similarity to an Internet website, and 21% fell into both categories. The vast majority of repeated answers were submitted by exactly two students, though six responses were submitted by 3-5 students and one answer (copied from Wikipedia) had overlap between 8 students. Often the responses were identical, though sometimes they were reworded slightly, suggesting that students may have intentionally tried to avoid detection by a plagiarism check. For example, one student wrote "find the path though all safe edges using dijikshraw's algorithm" while another wrote "find the path using safe edges only by dijikshraw's algorithm". In addition to similarity in sentence structure, both of these responses had a unique misspelling of "Dijkstra's algorithm".

The set of online resources copied by students included Wikipedia, a tutorial, a course book, and peer-reviewed publications. None of the plagiarized responses were correct and many were nonsensical. For example, two students submitted a response copied from a paper published in the *Wilson Journal of Ornithology*. The response begins, "edges are often associated with a high risk of brood parasitism by Brown-headed Cowbirds". We later discovered that this paper is the top search hit for "risky edges", a phrase that appeared in the question.

Figure 4 illustrates the distribution of exam scores for cheaters and non-cheaters. Despite their efforts to cheat, the average score for cheaters (30% correct, $SD = 14\%$) was lower than the average score for non-cheaters (40% correct, $SD = 23\%$). This difference is statistically significant ($t(402) = 4.18, p < 0.0001$). We conjecture that the weaker students felt more inclined to cheat, but due to the questions (and random variations) on the exam, cheating did not offer large benefits to their scores. These results are in concert with prior work, which found that students who abuse digital learning systems perform poorly on subsequent assessments [1].

Discussion and Future Work

Although our results demonstrate that warnings can deter cheating online, further study is needed to fully understand the implications for online courses. One open question is how the effectiveness of a warning changes over time. If users see the same pre-task warning repeatedly, they may become numb to the warning's threat of negative consequences. In addition, if the consequences are difficult to impose or if it is difficult to detect cheating behavior, users may cheat in spite of the warning. Studying how the effect of online warnings changes over time presents an interesting challenge for future work.

A second issue with adopting warnings as an alternative to honor codes is that they may intimidate or alienate well-intentioned users of an online education platform. If the warning deters cheating at the cost of "scaring off" honest students, it may not be a beneficial solution overall. It would be valuable to study the effect of an honor code versus a warning on drop-out rates in the context of an online course.

Finally, we would like to study if and how the effect of an honor code or warning varies with demographic factors (country, age, education level, etc.). Our study targeted undergraduate engineering students living in India. Replicating the experiment with a more diverse participant population would give insight into how demographics affect our results.

Conclusions

We measure the rates of cheating in a 409-student online course using an online "honeypot" combined with analysis of a free-response text question. We found that displaying a pre-task warning that focused users on potential negative outcomes of dishonesty deterred cheating in the context of an online exam in India. Consistent with prior studies, we did not find a significant benefit of a traditional honor code (though we did see a trend that may suggest smaller benefits). Our results indicate that pre-task warnings may be an effective and easy-to-implement alternative to honor codes in online examinations.

Acknowledgements

We would like to thank Srinath Bala, Andrew Cross, Viraj Kumar, Madhusudan Parthasarathy, Shambwaditya Saha, Sumit Gulwani, and the whole MEC team for help creating and administering the online exam. We would like to thank David Nemer for helpful conversations about the study design and we thank the reviewers for their suggestions for future work. The first author received support from an NSF Graduate Research Fellowship under Grant No. DGE-114747 and from an NDSEG Fellowship. This work was also supported by the Catalyst Center for Sustainable Development at Microsoft Research India.

References

[1] Baker, R. S., Corbett, A. T., Koedinger, K. R., and Wagner, A. Z. Off-task behavior in the cognitive tutor classroom: When students "game the system". In *CHI*, ACM (2004), 383–390.

[2] Braumoeller, B. F., and Gaines, B. J. Actions do speak louder than words: Deterring plagiarism with the use of plagiarism-detection software. *Political Science & Politics 34*, 04 (2001), 835–839.

[3] Coursera Student Support Center: What is the honor code? Retrieved 20 September 2014. http://help.coursera.org/customer/portal/articles/1164381-what-is-the-honor-code-.

[4] edX honor code pledge. Retrieved 20 September 2014. https://www.edx.org/edx-terms-service.

[5] Fellner, G., Sausgruber, R., and Traxler, C. Testing enforcement strategies in the field: Threat, moral appeal and social information. *J. of the European Economic Association 11*, 3 (2013), 634–660.

[6] Jacquemet, N., Joule, R.-V., Luchini, S., and Shogren, J. F. Preference elicitation under oath. *J. Environ. Econ. and Mgmt. 65*, 1 (2013), 110–132.

[7] LoSchiavo, F. M., and Shatz, M. A. The impact of an honor code on cheating in online courses. *MERLOT Journal of Online Learning and Teaching 7*, 2 (2011).

[8] Mastin, D. F., Peszka, J., and Lilly, D. R. Online academic integrity. *Teaching of Psychology 36*, 3 (2009), 174–178.

[9] Mazar, N., Amir, O., and Ariely, D. The dishonesty of honest people: A theory of self-concept maintenance. *J. Marketing Research 45*, 6 (2008), 633–644.

[10] Pruckner, G. J., and Sausgruber, R. Honesty on the streets: A field study on newspaper purchasing. *J. Euro. Econ. Assoc. 11*, 3 (2013), 661–679.