

# Dishonesty and Selection into Public Service: Evidence from India

By REMA HANNA AND SHING-YI WANG

## Online Appendix A: Model Extension

Thus far, we have assumed that ability only matters in determining the wage returns in the private sector. However, many developing countries use civil services exams to screen for high-ability candidates and taking these exams is costly: individuals in India spend substantial amounts of time, even years studying, and considerable amounts of money to pay for preparation assistance.<sup>1</sup> Assume that high-ability individuals need to exert less effort to pass the exam: the cost of taking the exam,  $e(A_i)$ , depends on ability such that  $e'(A_i) < 0$  and  $e(A_i) \geq 0$ . Individuals will take the exam if and only if:

$$(1) \quad k + g(P_i, C_i) - e(A_i) > f(A_i).$$

Under these assumptions, the earlier prediction that among those that will apply for a government job, higher  $A_i$  corresponds with higher levels of  $P_i$  or  $C_i$  now holds only if  $f'(A_i) > -e'(A_i)$ , i.e. if the wage returns to ability are greater than the degree to which ability helps in the exam process. If  $f'(A_i) = -e'(A_i)$ , then among those who prefer a government job, there will be no correlation between ability and these other characteristics, pro-social preference and propensity for corruption. Finally, if  $f'(A_i) < -e'(A_i)$ , we would expect that high-ability candidates would have relatively lower  $P_i$  and  $C_i$ .

An alternative interpretation of the function  $e(\cdot)$  is that it represents the returns to ability in the government sector rather than the cost of effort of preparing for an exam. In this case, the earlier predictions about the types of individuals who apply for jobs in the government sector may not hold if the returns to ability in the government sector are greater than or equal to the returns to ability in the private sector.

This relationship has significant policy implications: it implies that the propensity for corruption of those who want to enter public service will be determined in part by the relative

<sup>1</sup> For example, see Mohanty (2013).

returns of ability in the private sector versus taking the civil service exams. Moreover, depending on this relationship, hiring exams that only screen on ability—within the given candidate pool—may ultimately exacerbate or mitigate this propensity for corruption. The impact on corruption among civil servants depends on the sign and the degree of correlation between ability and propensity for corruption.

## **Online Appendix B: Experimental Methods and Data Collection (Extended Section III from Paper)**

We conducted a series of surveys and lab experiments with both college students and government nurses in Karnataka, India, to measure the propensity for corruption, pro-social preferences, and ability. We describe the procedures below. Survey documents are also available upon request to the authors.

### *A. Student Sample*

As we want to examine individual behaviors prior to entering the civil service, our sample is drawn from university students. We recruited seniors from seven large, mid-tier universities within the city of Bangalore in Karnataka, India. Six were mixed gender schools and one was a women's college. We obtained permission from each university to recruit subjects from classrooms and from recruitment booths on campus; our project team spoke to the rector at each university so that they understood the study and agreed to take part. We obtained IRB approval from Harvard and New York University for the study, and then subsequently from the University of Pennsylvania.

We chose to recruit from classes comprised of seniors in majors where both government and private sector jobs were viable options. To maximize power, we avoided majors in which all of the students within the major would enter one sector (public or private). This allows us to make within-major comparisons of career preferences.<sup>2</sup> We informed students that the sessions would explore the “cognitive skills, aspirations, background, and personality characteristics of

<sup>2</sup> To identify target majors, we conducted polling in classrooms prior to the recruitment stage to ask students whether they preferred government or private sector jobs. In the end, about 80 percent of the survey respondents were in the Commerce Stream, while the remaining were in Science. We did not survey Arts students, as few entered government service.

graduating students,” that the sessions would take about one hour and that they would be paid INR 20 (about USD 0.45) upon arrival to the session and up to an additional INR 392 (USD 8.71) depending on the session tasks; the average payment was INR 216 (USD 4.80).<sup>3</sup>

In August and September 2012, 669 students from 7 colleges participated in 28 sessions (Appendix Table 1).<sup>4</sup> These schools comprise about 3,215 students (Appendix Table 1). We designed the sessions to be close to the university and to not conflict with class times. In total, 1,081 students signed up to attend a session, but not everyone attended. While we do not have data on the students who did not attend, we can compare them to the student body population (in Appendix Table 5): the sampled students appear comparable in terms of economic background to those of students in Karnataka as a whole, with the caste distribution nearly identical across both groups. However, the sampled students are less likely to be male (40 percent versus 53 as a whole). And, the students in our sample are more likely to be from the commerce major than science major, which is an artifact of our sampling strategy. We do not observe any heterogeneity in the main findings based on these two factors, suggesting that these characteristics are not driving observed effects in the sample.<sup>5</sup>

We made a large effort to make it easy for recruited students to attend the sessions. We conducted the sessions at rooms at the university or in restaurants and other event spaces close by, and at any given time, there were up to four separate rooms in use for each session. We placed the times at the sessions when students were free from class (after class times or weekends). When students arrived at the sessions, we first checked them in. Since friends often attended the sessions together, we tried to separate them into different survey rooms within the sessions. The students were then organized into groups; each group had an enumerator team that ran that group’s session (there were 9 possible enumerator teams). All enumerator teams were monitored by a project manager who was present at all time at the survey site. Informed consent procedures were first conducted. Then, the survey commenced. The subjects filled out the

<sup>3</sup> We designed the financial incentives to be in a range that would appeal to students to participate, but not too large that it would be coercive. For comparison, the price of a ticket to a high end movie theater is about INR 400.

<sup>4</sup> These schools comprise about 3,215 students (Appendix Table 1). We designed the sessions to be close to the university and to not conflict with class times. In total, 1,081 students signed up to attend a session, which implies that 61 percent of those who signed up attended one. As Appendix Table 2 shows, the sessions ranged from 6 to 39 students; the final sessions tended to have lower attendance due to university protests and a city-wide transit strike.

<sup>5</sup> Note that in order to increase the generalizability of the sample, we recruited students at 7 different schools. The show-up rate differed across these different universities, so we can examine the heterogeneity of the results by the fraction of students who showed up. We find no significant differences (in fact, if anything, the findings are stronger in schools where more people showed up). Thus, the incomplete show-up rates does not drive the findings.

surveys forms on their own (with enumerators providing oral directions for the tasks). We provided the subjects with cardboard folders as dividers across students to ensure additional privacy as they filled out the survey forms.

We had initially planned to conduct the surveys and tasks in electronic form (and even programmed the survey module). However, during piloting, we realized that this was not feasible, due to electricity issues and internet issues at the sites we chose to run the sessions in. We could run the sessions in our office, which would have eliminated these problems, but we were afraid that fewer students would travel to attend the sessions. Thus, we switched to pen-and-paper surveys. Note that all documents were available in both Kannada and English, after piloting suggested that there was heterogeneity in preferences for language.

The survey questions covered demographics, work experience and post-graduation plans, preferences and expectations. We asked questions covering several psychology measures including locus of control (Rotter 1966). We included some commonly used survey questions to assess attitudes about cheating and corruption, such as what percent of individuals in the classroom would cheat during an exam and whether they thought that most businesses paid bribes. We also inquired about actual corrupt behavior, such as hiring an illegal agent who facilitates bribes to obtain a government service. Finally, we collected extensive contact information for the students, their relatives, and their friends in order to be able to track them in several years in the future to ascertain their ultimate job outcomes.

Ideally, we would have randomized the order of the survey questions to ensure that certain questions on the survey did not prime answers on other questions. For example, we would not want explicit corruption questions to change how people played the various games and tasks. However, given the limitations of conducting laboratory games in rented rooms near the colleges, we were not able to randomize the ordering. The data were collected in the following order:

- background information (age, academic and work experience)
- memory task
- family employment history
- personality questions
- dice task
- demographics (marriage and children)

- message game (sending messages)
- career preferences and expectations
- civil service exam questions
- cheating scenarios and corruption beliefs,
- locus of control questions
- continuation of the message game (receiving messages and making choices)
- questions on actual corrupt behavior (using agents)
- pro-social preferences game
- risk aversion questions.

Note that we had specifically had them complete the dice task and the key (sending) part of the message game before answering questions about their job preferences to ensure that stated work preferences did not affect the behavior on the task and game that we were the most interested in.

The crux of the surveys was a series of laboratory experiments designed to measure honesty, pro-social behaviors, and ability. Each experimental measure is outlined below:

*The Dice Task.*—To obtain an individual measure of dishonesty, we asked each participant to privately roll a six-sided die 42 times and to record the outcome of the die after each roll. For each value of 1 reported, we paid the participants INR 0.5; the payment increased by INR 0.5 for each higher value on the die, up to INR 3 for each reported roll of 6. Thus, the minimum possible payment is INR 21 and it occurs if the participant reports rolling all 1's, while the maximum payment (for all 6's) is INR 126. The monetary amounts were chosen to fit inside our total expected budget for the project. Appendix Figure 1 provides the task directions.

We ensured privacy: in addition to the cardboard folders, we instructed the survey team to either exit or be on the opposite side of the room during this task. Thus, participants could be assured that it would impossible for us to know *for certain* if they lied. However, we can determine how far the distribution of each individual's outcomes is from the uniform distribution.<sup>6</sup> Thus, even though we cannot say with *certainty* who cheated, this provides a measure that is strongly correlated with doing so.

<sup>6</sup> To detect non-random shifts in the value of the dice rolls per individual, we conducted power calculations using the effect sizes observed in Fischbacher and Föllmi-Heusi (2013), which ranged from 0.7 to 0.15. Using a conservative effect size of 0.7, along with a power level of 0.8 and an alpha equal to 0.05, the one-sided required sample size was 37. We rounded up to 42 as it was evenly divisible by 6.

This task is adapted from Fischbacher and Föllmi-Heusi (2013). One key difference is that they asked participants to roll the dice only once. This allows them to make statements about the group of individuals in each session they conduct, but not about each individual in the group. Their methodology is well suited to the laboratory context, where one can induce variation in treatments at the group level and the outcomes in lab behavior can be observed at the aggregate level. The innovation in our approach is important in that it allows us to examine the empirical relationship between an individual's revealed levels of dishonesty with real world outcomes and choices.

*The Message Game.*—Another experimental approach to examining lying behavior is to implement a cheap talk sender-receiver game where individuals with private information have the choice of whether to send an honest or dishonest message to another player. We implement a game that was developed by Gneezy (2005) and that has also been used, for example, by Sutter (2009) and Hurkens and Kartik (2009). We present the sender in the game with two possible pay-offs associated with a binary choice made by the receiver (see Appendix Figure 2). The sender then has a choice of two messages to send:

Message 1: “Option A will earn you more money than option B.”

Message 2: “Option B will earn you more money than option A.”

The sender is told that the receiver will not see the actual pay-offs associated with each choice, but will only see their message. The sender can choose either to send an honest message that indicates the choice that will give the receiver more money and the sender less, or a dishonest message that indicates the opposite. We implement three rounds with variation in the pay-offs as shown in Appendix Table 3. We stressed that neither party will ever know who they were paired with, although they did know that it was someone from within their session and that our enumerators saw their choices (which was an artifact of conducting the sessions on paper rather than on computers).

We are interested in whether the sender chooses to lie during the game. Every participant plays the role of the sender first. This is a slight departure from previous studies where half of the participants are senders and the rest are receivers. Our method ensures that we have outcome data for all subjects, thereby increasing our ability to correlate the key outcome with individual

preferences. Later in the session, each participant also plays the role of the receiver, mainly to ensure that the payoffs are realistic.

*The Pro-Social Preferences Game.*—We used a dictator game to measure willingness to give to others (see Camerer, 2003, for an overview). We instructed participants that they can divide INR 50 between themselves and a charity of their choice from among seven well-known, respected charities (UNICEF, Child Rights and You, Being Human, Help Age INDIA, CARE India, Red Cross and Save the Children).<sup>7</sup> For each rupee that they donated rather than kept for themselves, the amount given to the charity was doubled. The appropriate charitable donations were made.

For this measure to be interpreted as a measure of “pro-social behavior,” the subjects must believe that the charities are doing good work, rather than being seen as incompetent or a waste of money. India has a well-respected non-profit sector, and according to 2014 data from the Charities Aid Foundation, citizens donate much more to charity than in countries of comparable income levels. Moreover, in order to choose charities with good reputations, we chose the listed charities by asking local students and staff for their opinions on the most respected charities in Bangalore in order.

*Cognitive Ability Measures.*—We employed two incentivized tests to measure cognitive ability:

- We administered a digit span memory test in which participants listened to a series of digits and, after ten seconds, were asked to write down the number. We conducted five rounds, where the first round contained 5 digits and each subsequent round increased the number of digits by 2. The students were paid INR 2 for each correct round.
- We adapted a test of cognitive ability from Ariely, Gneezy, Loewenstein and Mazar (2009). We gave participants a set of matrices, with 12 numbers displayed in each matrix (Appendix Figure 3). They were asked to identify the two numbers in each matrix that add to 10. Participants were given 12 matrices to solve within 3 minutes, and received INR 2 for each correct answer.

<sup>7</sup> In the classic version of this type of dictator game, player 1 chooses how much of an endowment to keep for themselves or to share with other participants in the session, and the outcome is determined only by player 1’s actions. The subsequent adaptation to giving the money to a charity rather than other individuals is also fairly common in the literature (e.g. Eckel and Grossman, 1996; Carpenter, Connolly and Myers, 2008).

## *B. Nurse Sample*

We aimed to validate the dice task in a real world setting. Being able to conduct surveys with a large number of government officials is hard to gain permission for. It is even harder to then systematically measure a form of corruption or leakages that is directly attributable to a specific government employee, and link it back to collected survey data. Absenteeism is, thus, an attractive form of leakage to study since in addition to having real effects on service provision and outcomes, it can also be traced back to an individual employee using random check methods. As one of the coauthors of this paper (Rema Hanna) had another project where she was (1) interviewing government staff and (2) measuring their attendance behavior as part of a collaboration with a state government in India, it was a great opportunity to try to validate the dice task that we may not otherwise had.

We administered the dice task to government nurses within the context of a broader experiment that is described in detail in Dhaliwal and Hanna (2013), where we had a real measure of corruption. The experiment spanned 333 primary health centers (PHC) across five districts in Karnataka and focused on understanding whether increased attendance monitoring of health care workers through the use of a biometric device improved access to medical services.

We focus on absenteeism, a pervasive form of corruption both in India and in the developing world in general (Chaudhury et al., 2006). Government employees have a particular number of days that they are allowed to be absent and they have to record their absences for these days. If they are absent more than the prescribed number of days, their salary should be deducted for each additional absence, and they can cash out a certain number of their “sick days” if they do not take them. In practice, very few individuals report absence days and many cash out their “sick days,” despite *very* high absence rates.<sup>8</sup>

As Banerjee, Hanna, and Mullainathan (2012) point out, bureaucratic absenteeism is an attractive form of corruption to study because one can measure, by cross-checking, whether the

<sup>8</sup> Staff often claim that they are missing since they are “in the field” tending to the sick (even when their job responsibility requires them to be in the office). However, most accounts suggest this is not the case: for example, Banerjee, Deaton and Duflo (2004) did a tracer study with *sub-center* nurses who were absent during their random checks and found that they were only in their assigned villages 12 percent of the time that they were absent. The rate is likely to be much lower for staff nurses in our study because sub-center nurses do have job duties in the field while staff nurses do not.



bureaucrat is fraudulently collecting a paycheck for a day not worked and it has real implications on health. Dhaliwal and Hanna (2013) conducted this cross-checking: they implemented 9 rounds (two baseline, 7 post-intervention) of independent random checks of the PHC staff between July 2010 and November 2012.

PHCs within the same sub-district were generally surveyed at the same time; we randomly assigned the time of day that PHCs were checked so that no PHC was always checked at the same time of day. Note that although they were infrequent checks, there was a concern that the monitoring associated with the random checks could affect attendance as well; therefore, 50 percent of the sample was randomly selected to be visited in every other follow-up survey round. Dhaliwal and Hanna (2013) show that the monitoring frequency does not impact the absence rate. Some larger PHCs had multiple staff nurses; however, for budgetary reasons, we only interviewed one nurse per PHC. We tried to interview the nurses who were typically staffed during the day to correspond to the time when Dhaliwal and Hanna (2013) conducted the random checks. In many cases, the doctor gave us permission as to which nurse we could talk to at his or her PHC. In Appendix Table 4, Column 2, we regress the attendance rate on an indicator variable for being surveyed, PHC fixed effects, and the survey controls. We find no difference in the attendance rates between those nurses that were interviewed with the other nurses within their PHC.

The random checks proceeded as follows: the enumerator conducts a surprise visit to the PHC and records the staff attendance at the moment of arrival; if the PHC was closed on arrival, everyone is considered absent. Individuals who were transferred or resigned were subsequently dropped from the sample from then on.

Between November 2012 and January 2013, they conducted a series of endline surveys with the health center staff for their experiment. For the staff nurses, we obtained permission from the government to add the dice task and the memory test to their survey. The sample consisted of nurses in the 185 PHCs where the position was not vacant. Unlike the random checks, we made appointments to ensure that the nurses would be present and conducted revisits when possible if the nurse was absent. We interviewed staff nurses at 165 PHCs; Appendix Table 4, Column 1, shows that there is no significant difference between the attendance rates of nurses at PHCs that

we were able to interview and those that we were unable to do so (either because they left the PHC or because we could not secure an appointment with them).<sup>9</sup>

We aimed to design the nurses' tasks to be comparable to those of the students, but there were several differences. Most importantly, we could not pay government workers in cash for ethical reasons. Instead, we obtained permission from the government to pay them in *candy* (Appendix Figure 4). One piece of candy is worth about Rs 1, and therefore, we offered double the amount for the nurses for the tasks than for the students, but in candy rather than cash.

In addition, the memory test differed from the student test in three ways. First, while the students' test started with a 5 digit sequence, piloting with non-sample nurses informed us that this was already quite difficult. Therefore, we started with a 3 digit sequence for the nurses. Second, we gave the students five rounds of number sequences, but increased it to nine rounds for the nurses in order to increase the measured variation in ability. Third, because the enumerators read out the numbers to the students as a group and then the students individually filled out their forms, the students were asked all five rounds; for the nurses, the test ended as soon as they incorrectly remembered a sequence.

Since we surveyed nurses at work, we had to be cognizant of both time and logistical factors. Thus, we could not administer all of the experimental tasks that we gave to the students. We administered the dice task since it was our key indicator of interest. However, we did not administer the message game because it requires two players, and takes a long time to play. Piloting informed us that the nurses found the matrices test too difficult and thus we did not administer it.

Note that in terms of the ordering, the attendance measures were collected prior to the tasks (since they were in the endline). It is possible that nurses who were absent on the attendance measures would have felt more comfortable cheating on the tasks. However, had it been the opposite, those who clearly identified themselves as cheaters on the tasks may have thus felt more comfortable being absent at work. Thus, for this and other reasons, the analysis should be taken as descriptive, and not causal.

<sup>9</sup> Some larger PHCs had multiple staff nurses; however, for budgetary reasons, we only interviewed one nurse per PHC. We tried to interview the nurses who were typically staffed during the day to correspond to the time when Dhaliwal and Hanna (2013) conducted the random checks. In many cases, the doctor gave us permission as to which nurse we could talk to at his or her PHC. In Appendix Table 4, Column 2, we regress the attendance rate on an indicator variable for being surveyed, PHC fixed effects, and the survey controls. We find no difference in the attendance rates between those nurses that were interviewed with the other nurses within their PHC.

Finally, the survey also included questions on the nurses' basic demographic characteristics and on their beliefs on statements such as "It is possible to operate a business in India without bribing" and "Promotions should be based primarily on job performance rather than seniority."