CSCC Working Papers 01/18

Little lies
Math performance and cheating
in primary schools in Congo

Mario A. Maggioni[*] and Domenico Rossignoli[*]

[*]*DISEIS and CSCC, Universitá Cattolica del Sacro Cuore, Milano*

CSCC Working Papers 01/18

Little lies
Math performance and cheating in primary schools in Congo

Mario A. Maggioni[*] and Domenico Rossignoli[*]

[*]*DISEIS and CSCC, Universitá Cattolica del Sacro Cuore, Milano*

October, 2018

Abstract

This paper provides a novel contribution on the relation between school performance, cheating behaviour and pro-social attitudes by analyzing a sample of 170 pupils in 10 primary schools located in the outskirts of Goma (Congo, DR). Children were administered a questionnaire - that included a Dictator Game (DG) and a modified Dice Rolling Task (DRT), while information on their school performance was obtained through the collection of school reports - in two subsequent school years. Exploiting this research design, we analyzed whether cheating (measured through DRT) could be explained by school performance (measured by Math, and Total, scores) and altruism (measured by DG) when controlling for individual (such as age, sex, and previously recorded cheating attitudes) and background (such as class, school, interviewer) characteristics. Our results show that cheating is positively associated with school performance (measured by both total and math scores), supporting the hypothesis that the development of cognitive skills affects the propensity to act opportunistically. This relation is robust to the inclusion of altruism as an explanatory variable, which negatively relates to cheating, as if children under analysis considered lying as an anti-social behaviour *per se* even when their lies do not explicitly harm other similar individuals. We also show that, while pupils' cognitive skills are a good predictor of cheating, the opposite - cheaters recording higher marks because of their deviant behaviour - does not hold. Finally we give evidence that only when we limit our measure of performance to Math Score, there is a significant relation between cheating behavior and the size of the reward arising from cheating.

Keywords: Cheating; Altruism; Dice Rolling Task; Dictator Game; School Performance; Children

JEL Classification: C91, C93, D91

1 Introduction

In everyday life we often observe dishonesty and cheating: citizens evade taxes, drivers park in forbidden spaces, viewers illegally download content from the Internet, students copy in written exams, employees call in sick when they are not ill, users enjoy public transportation as free riders. In an effort to limit dishonesty, governments apply fiscal inspections, city councils hire parking inspectors, media companies and content distributors implement technological innovations to hinder copyright infringements, professors employ Ph.D. students as invigilators during exams, official doctors may check up on sick leave, and bus companies hire ticket checkers[1].

Cheating and lying are generally seen as anti-social behaviors since being unable to trust others' word and behavior bears substantial economic and social costs and destroys the social fabric of human coexistence. In this perspective, cheaters are perceived as social outcasts which try to compensate their lack of talent or effort with unfair behaviors, thus cheating can be better understood within a cost-benefit framework (Gibson et al., 2013; Gneezy, 2005).

However, there are also plenty of counterexamples: on the one hand, there are situations and micro-cultures in which cheating is sometimes seen as a proxy of smartness, such as gang and street culture (Burke, 2013) or even business culture (Cohn et al., 2014); on the other hand, psychological empirical research (Vasek, 1986; Evans et al., 2011; Evans and Lee, 2013) shows that lying in children is correlated with the development of cognitive skills, therefore it can be a good predictor of future school achievements.

An extensive experimental literature studies dishonesty in Psychology (Glätzle-Rützler and Lergetporer, 2015; Wilson et al., 2003; Polak and Harris, 1999) and Economics (Pate, 2018; Kajackaite and Gneezy, 2017; Ruffle and Tobol, 2017; Ariely et al., 2014; Fischbacher and Föllmi-Heusi, 2013; Mazar et al., 2008; Gneezy, 2005). Within both fields, a subset of papers (Cantin et al., 2016; Maggian and Villeval, 2016; Glätzle-Rützler and Lergetporer, 2015; Evans and Lee, 2013; Ding et al., 2014; Bucciol et al., 2011; Bucciol and Piovesan, 2011; Talwar et al., 2007) explicitly focuses on children lying behavior, investigating the influence of different factors such as: age, gender, social preferences and second-order belief understanding.

Within this literature a vast and detailed taxonomy of lies has been established and a panoply of different tests, tasks and situations have been devised in order to study every different shades of lies. Erat and Gneezy (2012) through a dice rolling experiment distinguish between *Altruistic white lies* (when the lie harm the liar but help the other person), *Pareto white lies* (when both sides earn more as a result of the lie[2]), *Selfish black lies* (when the lie help the liar at the expenses of the other person) and *Spiteful black lies* (when both sides looses as a result of the lie).

Our paper delves within this debate by providing a novel contribution on the relation existing between school performance (with a specific focus on math scores),

---

[1]For extensive surveys on behavioral mechanisms and psychological causes affecting dishonesty, see Mazar and Ariely (2006) and Jacobsen et al. (2018).

[2]Maggian and Villeval (2016), along the line of Gneezy (2005), correctly points out that the gain for both sides is not needed in order to define a *White* lie as a *Pareto* one since any lie which increases the utility of the "partner" without diminishing the utility of the "player" will perfectly fit the definition.

cheating and pro-social attitudes among primary school children, by adopting a slightly modified version of the dice-rolling task[3], in which no counterpart in the game is explicitly mentioned and the subject is required to report the outcome (thus possibly lying) under the direct scrutiny of the interviewer.

Through a cross-section analysis of 170 children randomly selected[4] in 10 primary schools in the outskirts of Goma, we are able to show that *cheating behavior* is a stable and specific characteristic of the sample and that it is strongly predicted by both *Math* and *Total* scores, as in school reports, and negatively correlated with an experimental measure of *Altruism*. Finally we give evidence that math-skilled pupils tend to cheat more when the rewards from cheating are higher, thus suggesting that these pupils are better equipped to rapidly identify costs and benefits of cheating and act consequently.

To achieve this aim, the paper exploits data produced in a Lab-in-the-Field Experiment we conducted, with two other colleagues[5], in the Democratic Republic of Congo (March 2016 and May 2017), originally devised to measure the effectiveness of a distance adoption program implemented by AVSI[6], an Italian NGO. The nature of the dataset allows us to introduce a twofold original contribution: on the one hand, we can account for possible unobserved heterogeneity in cheating levels *before* Total and Math performance were recorded (i.e. in the previous school year); on the other hand, we can control for a measure of altruism (also recorded in the previous school year), to account for individual heterogeneity in pro-social preferences.

Since our original experimental framework involves a dice rolling task (DRT, henceforth) to be performed by a primary school child under the direct sight of an adult interviewer, in contrast with the original setting proposed by Fischbacher and Föllmi-Heusi (2013), we are in the best position to test whether cheating is a prominent behaviour within our sample. In other words if we are able to detect a significant level of cheating in our experiment, *a fortiori* we would have observed an even larger level if the experiment were performed in the standard format.

Furthermore, since in our experimental protocol the child is aware that in the DRT no counterpart (i.e a fellow pupil, as it is the case, conversely, in the Dictator Game) is harmed by his/her lie, the only effect being that the interviewer has to disburse a larger number of biscuits, we are able to test whether cheating is conceived as an anti-social behaviour *per se*.

Finally, by studying cheating patterns (i.e. the relations between cheating behaviour and the gain from cheating in each individual dice draw) we are able to test whether math related school performance significantly differs in this respect from generic school performance (as measured by *Total score*).

The paper is organized as follows: Section 2 presents the research design and experimental methods, Section 3 outlines the estimation techniques, Section 4 provides main results and robustness checks, Section 5 deals with potential reverse causation, Section 6 investigates cheating patterns, Section 7 discusses main findings and concludes the

---

[3]As in Ariely et al. (2014).

[4]See Section 3.2 for a description of sampling procedures. Please note that due to some missing information in a handful of school reports, the actual number of observations in the regression tables can be smaller than 170.

[5]Names removed in the anonymous version for referees.

[6]See https://www.avsi.org/en/.

paper.

2   Research Design

2.1   The sample

The experimental procedure involves the administration of a questionnaire, including a set of incentivized tasks, on a sample consisting of 170 children from 28 different classes, across ten primary schools in the outskirts of Goma, a city located in the troubled North Kivu province of the Democratic Republic of Congo. The questionnaire has been administered twice, always at the end of the second term of two school years (2015/16 and 2016/17)[7]. Thanks to the time structure of the data collection protocol, we are able to exploit the time interval (around 13 months) between the two observations and account for potential ex-ante individual heterogeneity in cheating behaviour and to investigate the existence of reverse causation.

The sample for this paper consists of a number of "control" children that have been randomly chosen to match, w.r.t. to school, class, age and gender, a given number of children supported by AVSI distance adoption program (SAD), on the basis of a "vulnerability scale" computed (taking into account both the children and the household situation) by the NGO. For each treated child, we asked each school to randomly select two children matching age and gender and attending the same class, to provide a control group for the evaluation of the program. Since SAD children are randomly assigned to different classes by school headmaster, for the purpose of this paper, we well may consider that no influence of the support program comes into play and our sample of "control" children to be representative of the entire population of non-treated children in the selected 10 primary schools.

From the initial resulting number of 188 children, 18 children have been excluded from the analysis since they failed the class and had to repeat the same grade in 2016/2017. Therefore, the final sample consists of 170 children. Descriptive statistics of our sample's individual characteristics are shown in Table 2.

2.2   Experimental procedures

Incentivized tasks in behavioral and experimental economics are usually performed by using money as incentive. However, both ethical concerns, given the subjects' age (ranging from 6 to 16), and the geopolitical conditions of the area suggested not to handing out (even small amount of) money to the interviewed children. Following the suggestions of a number of NGOs, working in the area, we resorted to use packets of biscuits as incentive goods[8].

---

[7]The questionnaire was administered in paper-and-pencil, the first time in late March 2016 and the second in early May 2017.

[8]NGO's staff identified the biscuits brand and type "Cremica Glucose Biscuits" whose packaging is best known and which is most appreciated by children. Local sources informed us that these biscuits are also used as means of exchange among school children, thus minimizing possible satiation issues involved when using food as a reward in experiments.

Parents provided active consent after being adequately informed before the experiment took place. Moreover, both the interviewers and the good used to reward the participants (biscuits) were presented in advance to all the children and their parents by the school headmasters. The questionnaires have been administered by twenty-seven interviewers purposely recruited among students of the local university in Goma. Interviewers were independent from both the schoolteachers and the NGO and were previously unknown to children.

Pupils were interviewed one at a time, by one interviewer sitting at a desk in front of one pupil, in order to ensure that instructions were fully understood. Nobody else was allowed to stay in the room during the experiment. After introducing him/herself to the child, the interviewer set the table by putting a picture of schoolchildren (taken in a different school in Congo) in the middle of the table, as a priming for the child to visualize his/her anonymous partner for the Dictator Game.

Thanks to the cooperation of schools' headmasters, children involved in the experiment were gathered together in a courtyard. After a child completed the questionnaire, he/she was allowed to return home, thus avoiding that he/she could talk about the experiment to other children still waiting to be interviewed. All the tasks included in the questionnaire were played in an anonymous double blind setting: children were randomly assigned a code; the NGO staff held records about the matching between individual names and codes, but could not access individual outcome data; the research team could access individual outcome data, matched with anonymous codes, but could not access individual names[9].

Each task yields a payoff in terms of packets of biscuits, depending on children's choices. At the end of the questionnaire, only one of the incentivized tasks is drawn (through a die roll) and actually rewarded: in this way, children are supposedly putting the same effort on every task.

2.3 Incentivized tasks: Dice Rolling and Dictator

The incentivized tasks relevant to this paper are described below:

- *Dice rolling task*. This task, originally developed by Fischbacher and Föllmi-Heusi (2013) and applied in different contexts (e.g. Ariely et al., 2014), exploits the statistical properties of random dice rolls to make inference about mind cheating (i.e. misreporting of chosen outcomes) in both children and adults. In our experiment, we slightly modified the original version of the task in order to make it easier for children to understand. Instead of using a single die (with the ex-ante choice being between the top or bottom side), the child is provided with a couple of fair dice (one red and one blue) and asked by the interviewer to perform twenty rolls. Before every roll, the child has to decide in his/her mind which of the two dice he/she will choose (either the red or the blue one). After observing the outcome of the roll in the questionnaire form, he/she communicates his/her own choice to the interviewer that takes note of the choice. At the end of all 20 rolls, one is chosen and selected for being rewarded (in case the dice task is drawn as the task to be rewarded) at the end of the interview). The choice

---

[9]For further details on the experimental procedures see also [Authors 2017].

is not declared before the roll, therefore the child has always the incentive and the opportunity to deviate from his/her original choice and simply choose the highest outcome between the two dice. If the child's reporting is sincere, the average outcome of the chosen dice should approximate the expected value of the series[10], since any ex ante choice strategy is independent of the outcome. If, on the contrary, the observed mean of the chosen outcomes exceeds this value, it is likely that the child may have misreported his/her choices (thus cheating) to maximize their payoff. Clearly, the statistical properties of this task hold on average for the sample but not necessarily at the individual level because of the small number of rolls (equal to 20)[11].

Thus, the first measure of cheating (*MeanRoll*) for an individual child is the difference between the average of the 20 dice values he/she has chosen and the total average computed over of his/her series of 20 rolls of two (one blue, one red) dice [12].

As a robustness check, we also provide an alternative indicator, *(MaxChoice)*, based on the proportion of maximum values in a given roll chosen by children over the total number of rolls (net of ties). In fact, once the child has decided which is the selected die, he/she has 50% chances that the chosen die reads a lower outcome than the unchosen one. Therefore, an average proportion of maximum choices that exceeds 50% of the throws (net of ties), signals that children, on average, are likely to have misreported their choices in order to maximize their outcomes (i.e. to have lied).

As already stated in the introduction, our experimental framework differs from other papers using DRT (Shalvi et al., 2011; Fischbacher and Föllmi-Heusi, 2013; Ariely et al., 2014; Hao and Houser, 2017; Gächter and Schulz, 2016) in two main respects: firstly the task was performed by a primary school child under the direct sight of an adult interviewer, thus imposing a rather high "psychological cost" of lying given on the subject due to the fear of being caught; secondly, and differently from what happen in the Dictator Game, we make the child explicitly aware that in the DRT no counterpart (i.e a fellow pupil) is harmed by his/her lie[13]. Through these two small changes in the protocol we are therefore able to test whether cheating is a prominent behaviour within our sample and whether it is conceived as an anti-social behaviour *per se*, irrespective of the consequences.

- *Dictator Game*. The questionnaire includes a modified version of the Dictator Game (DG, henceforth) (Kahneman et al., 1986) in which the interviewed child

---

[10]It must be noted that because our subjects (children) roll two fair dice, the outcomes of the blue die is independent from the outcome of the red die. For this reason, differently from Ariely et al. (2014); Fischbacher and Föllmi-Heusi (2013) each child records a possibly different expected value of his/her dice rolls series.

[11]Please refer to Ariely et al. (2014) for further details on the statistical properties of the DRT.

[12]Please note that the total average can be different from one child to another one: therefore, our indicator adjusts for individual variation providing a more accurate aggregate indicator of cheating behavior.

[13]The only effect being that the interviewer has to disburse a larger number of biscuits, thus in a sense we may consider our experimental protocol to target a very specific sub-type of *Pareto White* lie.

acts as a Proponent, being endowed with five packets of biscuits and matched to an anonymous child who has received no endowment[14]. The task requires the child to choose if and how to split the packets of biscuits he/she received between him-/herself and the other anonymous child. Within a pure game theoretical framework with self-interested agents, the Proponent is expected to retain the entire endowment for him/herself. Deviations from the *selfish* equilibrium solution in the DG can thus measure empathy, altruism and/or pure generosity (Forsythe et al., 1994; Camerer, 2003; Guala and Mittone, 2010). We use the proportion of the initial endowment of packaged biscuits shared with the other child as an indicator of *Altruism*.

2.4 School reports

In this paper we employ Math score as our main explanatory variable, while using Total school performance as a robustness check.

We collected school reports for two school years (2015/16 and 2016/17). The school year in Congo begins in October and ends in June. Results are recorded at the end of each term; in this paper we used 2nd terms' results because they better matched the timing of administration of the questionnaire. School reports in Congo include a large number of scores, splitting subjects into topics and terms in smaller sub-periods[15]. Aggregate scores, by subject and term, are obtained by the sum of all the relevant sub-components. In order to make Total and Math scores comparable across grades and school years, we harmonized the scores w.r.t. the maximum points achievable for each subject in any given grade. This procedure ensures that our scores are fully comparable across school grades being expressed as the share of the maximum outcome available per each subject in each year.

3 Empirical analysis

3.1 Estimation techniques

We estimate the relation between school performance and cheating behaviour through a cross-section analysis in which the main outcomes, namely the value of one of the cheating indicators in school year 2016/17, are predicted by the main independent variable, i.e. Math score. To estimate the effect on cheating (alternatively measured by either MeanRoll or MaxChoice) we estimate the following model:

---

[14]Children are primed, through a photo and a statement, "You have been matched with one of these children portrayed in the photo", that the Respondent is an unknown child from a primary school, in the same region, not included in the analysis.

[15]School reports for primary schools in Congo are very complex and exhaustive documents reporting for each child two intermediate marks plus a final score per each term in six main subjects/areas: Religion and Civic education, National Languages, French Language, Mathematics, Sciences, Arts. Math and Total score included in the analysis provided in this section both refer to the same term of 2016/17.

$$Cheat_i = \alpha + \beta Perf_i + \gamma Altruism_i + \delta Cheat(baseline)_i$$

$$+ \sum_{k=1}^{K} \rho_k Z_{ki} + \sum_{j=1}^{J} \mu_j Dummies_{ji} + \epsilon_i \qquad (1)$$

where i identifies pupils; $Cheat$ is the value of the cheating indicator in school year 2016/2017, $Perf$ is the main explanatory variable for school performance, alternatively measured by Math and Total score, $Altruism$ is the main explanatory variable for pro-social attitudes, Z is a set of K individual characteristics, such as gender and age, as well as the individual payoff obtained in the first experimental session in 2015/16, $Dummies$ is a set of J dummies to control for invariant characteristics, such as pupils' school and grade of attendance, peer effects (as defined in Table 1) and to control for the interviewer that administered the questionnaire; α, β, γ, δ, ρ, μ are the parameters to be estimated, while ϵ_i is the usual error term.

When the dependent variable is MeanRoll, we estimate the model through OLS, including robust standard errors. Conversely, when we estimate the effect of Math on the alternative cheating indicator, MaxChoice, we adopt a more consistent estimator, namely GLM for the binomial family with a logit link function, to account for the fact that MaxChoice represents a share, hence it is bounded within 0 and 1 (Papke and Wooldridge, 1996).

Finally, in Section 6 we investigate the pattern of dice outcome choices through the whole roll series, therefore our dependent variable is binary. In this case, to predict the probability of choosing the maximum outcome as a function of school performance ($Perf$), conditional on the observed difference between the two dice, we implement the following (interacted and fully saturated) logit model:

$$Y_{i,r} = \begin{cases} 1 & \text{if pupil } i \text{ chooses maximum outcome between the two dice in roll } r \\ 0 & \text{otherwise} \end{cases}$$

$$log\left(\frac{\pi_{i,r}}{1 - \pi_{i,r}}\right) = \alpha + \beta Perf_i + \gamma_d Perf_i \times Diff_{dir} + \delta_d Diff_{dir}$$

$$+ \sum_{k=1}^{K} \rho_k Z_{ki} + \sum_{j=1}^{J} \mu_j Dummies_{ji} \qquad (2)$$

where $\pi_{i,r}$ is the probability that $Y_{i,r}$ equals 1, Y is the dependent variable, $Diff_{dir}$ is one of the possible (absolute) differences between the two dice observed by pupil i in roll r, with $1 \leq d \leq 4$, Z and $Dummies$ are defined as in Equation 1 and α, β, γ_d, δ_d, ρ, μ are the parameters to be estimated. Standard errors are clustered to the individual level.

3.2 Data and variables

Data for our outcome indicators (MeanRoll and MaxChoice) as well as for our main explanatory variables (Math and Total score) and control variables (Female, Age, Payoff)

refer to the school year 2016/17. Altruism and "baseline" cheating indicators (used as covariates) are measured in school year 2015/2016, i.e. roughly one year earlier. Further, Math and Total scores are collected in the second term of each school year and their values have been re-scaled in relative term to allow inter-class/inter-term comparability as explained in Section 2.4[16].

All variables included in the analysis are briefly described in Table 1, while Table 2 provides descriptive statistics.

The final sample included in the analysis presented in the next Section accrues to 170 pupils, randomly chosen in 28 different classes, across 10 primary schools in Goma.

[Table 1 about here.]

[Table 2 about here.]

4 Results

4.1 Descriptive analysis

Despite the experimental framework was unfavourable to cheating - children had to perform the task and report their choice in front of an adult - cheating is the prevailing behaviour in our sample as shown by the sample means of both indicators(MeanRoll and MaxChoice) in Table 2. A t-test on the mean value of MeanRoll show that its value is positive and significantly different from zero (p<0.001). The same analysis performed on the mean value of MaxChoice shows that it is significantly different from 0.5 (p<0.001), which is the aggregate probability of randomly choosing ex-ante the die showing the highest value of the couple[17]. These results are graphically depicted by the histograms in Figure 1, where the distributions of both MeanRoll and MaxChoice are skewed to the right of the dashed line (one centred on 0, the other on 0.5), representing an hypothetical random (i.e. non cheating) behaviour. On the contrary, both Math Score and Total Score display an almost normal distribution centred on values around 0.6 (see Figure 2).

[Figure 1 about here.]

[Figure 2 about here.]

Looking at possible determinants of individual variation in cheating behaviour, and in particular at the role played by cognitive ability (here proxied by school performance, with specific reference to Math), we may start with a simple graphical description. From a quick inspection of Figure 3 - showing a scatterplot of cheating behaviors (elicited through MeanRoll in DRT) and Math scores (as recorded in schools reports in May 2017) - a clear pattern emerges: children better performing in Math are more likely to misreport their outcome in the DRT.

---

[16]W.r.t. the experiment administration, school year 2015/16 refers to the first wave (Wave 1); school year 2016/17 refers to the second wave (Wave 2).

[17]The same results hold also in the baseline cheating indicators, recorded in s.y. 2015/16.

[Figure 3 about here.]

We further investigate whether this relation is stable across all school grades, since the developmental psychology literature stresses the influences of age on social preferences in children, attitudes and behaviors. The four panels of Figure 4 show the existence of a positive correlation between cheating behavior (as proxied by MeanRoll) and Math scores across all grades in our sample[18].

[Figure 4 about here.]

Obtaining exact data about children age in South Saharan Africa is rather difficult and one may also think that "social age" is more important than "biological age" in explaining a child's deviant behavior, due to the relevance of peer effects. We dealt with this issue through a twofold strategy: on the one hand we double checked self-reported age with official school records; on the other hand, we used *grade* as a proxy of social age and peer effects. To control for possible peer effects arising from interaction at the school level among pupils attending the same grade we included in our analysis a dummy variable, labeled "Peer Effects", capturing simultaneously school and grade of pupils[19].

4.2   Regression analysis

In the regression analysis we took therefore into account age, seniority and peer effects by using the biological age of children as an explanatory variable and alternatively including school, grade or peer effect dummies to the models.

Further, to take into account the experimental features of the data and to control for possible biases introduced by the experimental setting, we also include in all our model specifications a set of dummies to control for interviewer-specific effects, as well as the value of the payoff earned by each child in the participation to the experiment in March 2016.

The main results of the paper are displayed in Table 3, which shows the effects of Math scores on cheating (here measured by MeanRoll). The top panel, from columns (1) to (6) displays the base model in which higher Math scores are always associated with a higher level of cheating. The only other significant coefficients are those associated with gender (females cheat less than males) and age (older children cheat more than younger ones), when school dummies are used instead of peer effect dummies[20].

---

[18]Since the original experiment implied two distinct observations with a time lag of about one year, the first available data in our sample refers to children attending grade 1 in s.y. 2015/2016 thus grade 2 in 2016/17. The eldest children in our sample attended grade 5 in s.y.2016/17 since there was no child, attending grade 5 in s.y. 2015/16, supported by the NGO program.

[19]Ideally we would like to be able to include a set of proper "class" dummies in our analysis to control for peer effect. However, the allocation to classes, within same grade, is not always stable in our schools: in fact, due to the large number of pupils attending, they are often moved from one class to another during their school career. However, it is more likely that social interactions within the school occur by cohorts attending same grades, since they share most of their curricular activities. Therefore, we resorted to a less precise, though still sensible, measure to capture peer effect, that is also more parsimonious in terms of reduction of degrees freedom and hence more suited for empirical analysis.

[20]Given the structure of the sample (see Section 2.1) schools are more balanced w.r.t. to age and genders, than classes.

[Table 3 about here.]

In the bottom panel of Table 3 we included a measure of Altruism (share of initial endowment sent to respondent in the DG), as a further explanatory variable in the model, to account for previous studies highlighting the relation existing between integrity (an inverse measure of cheating) and other social preferences at the aggregate level: indeed, although heterogeneity in cheating and in social preferences are by now well documented, relative little is known about the relationship between these two variables at the individual level (Kerschbamer et al., 2017, p. 2).

The results of columns (1) to (6) in the bottom panel of Table 3 provide two important insights. Firstly, Altruism is negatively related to our measure of cheating and the coefficient is (weakly) statistically significant in all but one reported specifications. These results suggest that pro-social attitudes, such as altruism and sincerity, are likely to be mutually reinforcing at the individual level even if no explicit harm to a school-mate counterpart is caused by the act of lying, thus suggesting that the children in our sample have already internalized a value judgment on lying as an anti-social behaviour *per se*. Secondly, and most important for our analysis, the inclusion of Altruism does not affect our main result, since the coefficient of Math is still positive and significant, despite being slightly reduced in magnitude. Results for all other other regressors are consistent with those shown in the top Panel of the table. Finally, both the top and bottom panels of Table 3 report the outcome of the analysis when the baseline value of MeanRoll is included. As the table shows, even controlling for potential ex-ante heterogeneity in individual cheating attitudes, the sign, magnitude and statistical significance of Math score is substantially unaffected.

Table 4 presents the results of the same analysis provided in Table 3, with the only exception of Total score been used instead of Math score as an alternative indicator of school performance and, indirectly, of cognitive abilities. This aggregate performance indicator, which may be a better proxy of the whole educational performance of the child, determines, at the end of each school year, the pupil's passing or failing. Table 4 shows that a more comprehensive indicator of school performance is even a better predictor of cheating behavior (as compared to Math Score) in our sample, thus probably hinting that Total Score is a better proxy for the overall child smartness allowing him/her to realize that the interviewer has no possibility of detecting his/her lies, thus lowering significantly the cost of lying.

[Table 4 about here.]

5   Inspecting reverse causality

The analyses performed in the previous section show the existence of a robust positive correlation between Math scores and cheating behavior in primary school Children in Goma (and of a negative correlation between altruistic attitudes and cheating behavior). However, for the moment we are unable to distinguish between the case in which Math ability cause lower integrity from the opposite case in which a higher cheating propensity (as detected by the DRT in the experiment) signal a deviant individual attitude at school of the pupil who may compensate his/her lack of talent and or effort

with unfair behaviors (such as copying during exam papers). In this last case, cheating may well explain a higher Math score, since it is obtained by deception.

From a pure statistical perspective we need to check whether our models do not suffer from "reverse causality". We perform this task by running a regression in which the dependent variable is Math scores in 2016/17 and we insert, among the covariates, the cheating behavior as measured in 2015/16.

Table 5 shows no significant effect of cheating attitudes on pupil's school performance. This result holds for both cheating indicators: MeanRoll in the upper panel and MaxChoice in the lower panel.

[Table 5 about here.]

The coefficient associated with past level of cheating is not significantly different from zero, both in the benchmark model, as in columns (1) to (3), and when controlling for past schools performance in Math, as in columns (4) to (6). This result holds both for the MeanRoll (top panel) and MaxChoice (bottom panel) cheating indicators.

6 Cheating patterns

In the economic literature on cheating - specifically those papers using DRT as eliciting device (such as Fischbacher and Föllmi-Heusi, 2013; Ariely et al., 2014) - once shown that cheating behaviour is prevalent in a given sample, it is customarily required to investigate the relation existing between the "temptation to cheat" - i.e. the difference in the values displayed by the dice in each given roll, as a measure of the "reward associated with lying" - and the actual cheating behavior.

Before dealing with the analysis of these specific cheating patterns, it is useful to look for all regularities and patterns possibly emerging from our experimental framework.

6.1 Testing for learning effects

Since we asked children in our sample to be interviewed twice, once in 2015/16 (Wave 1) and a then in 2016/17 (Wave 2), and we used lagged variables in the regression analyses, as a preliminary analysis, we test whether lying behavior changes when the action is repeated in time. In other words we are interested in testing whether our interviewed subjects show some sort of learning behavior. Differently from Fischbacher and Föllmi-Heusi (2013) we cannot compare the behavior of inexperienced participants with experienced ones in a panel data set. However we can compare both the individual and the aggregate choice in Wave 1 and Wave 2 searching for possible relations hinting at some systematic differences in the subjects' behavior.

Figure 5 shows that no systematic differences can be detected in the aggregate choice of children when comparing the first with the second wave of interviews.

[Figure 5 about here.]

Table 6, through a model in which both cheating indicators are alternatively used as dependent variable while a time dummy, as well as the same set of controls included in previous model specifications, is used as independent variable, shows that both MeanRoll and MaxChoice cheating measures are stable over time in our sample, thus excluding that time-specific unobserved effects are affecting our main results. Thus, differently from Fischbacher and Föllmi-Heusi (2013), we do not observe any significant "learning" effect leading children to cheat more in Wave 2 as compared to Wave 1[21].

[Table 6 about here.]

Further, by focusing only on Wave 2, i.e. school year 2016/17, we also test the existence of path dependence dynamics within the roll series w.r.t the chosen outcome: for instance, it could be assumed that for later rolls the probability of choosing a higher outcome is larger than for earlier rolls, as if the temptation to cheat grows near the end of the task. To check whether such effects are operating in our sample, we perform a set of unit-root tests specific for panel data, both accounting for auto-regressive parameters that are common across panels (such as the Levin-Lin-Chu and the Harris-Tzavalis unit-root tests) and for panel-specific auto-regressive parameters (Im-Pesaran unit-root test). These tests allow to reject the null hypothesis that all panels contain a unit-root at the highest conventional level of statistical significance ($p<0.001$): therefore, we can exclude that the choice made by pupils in each outcome depends on the structure of the roll series. This result is made evident in in Figure 6 in which no chosen outcome (out of the 6 possible alternatives) display an increasing or decreasing trend along the series of 20 dice rolls.

Figure 6 provides further evidence of the aggregate cheating behavior described in section 4.1. If every single outcome had been chosen at random by children in our sample, all series of (connected) dots in the diagram would have been laying around the dashed line representing 1/6 (i.e. the expected value for each one of the six outcomes in a random roll series). This is clearly not the case in our sample, with higher outcomes (6, 5, 4) more likely to be chosen than lower outcomes (1, 2, 3). Outcome-wise proportion tests against the null hypothesis that the proportion of each chosen outcome is equal to 1/6 show that while the null cannot be rejected for outcome 3 and only weakly significantly ($p<0.10$) for outcome 4, it is strongly rejected ($p<0.01$) both for lower outcomes, 1 and 2, that are chosen significantly less than 1/6 of the times, and for higher outcomes, 5 and 6, that are chosen significantly more than 1/6 of the times.

[Figure 6 about here.]

6.2 Testing for "temptation" effects

Finally, in order to assess the existence of systematic cheating behaviors due to the difference of the scores shown by the red vs. the blue die (or *vice versa*) - i.e. whether cheating behavior is proportionally influenced by the size of gains arising from cheating itself in each specific dice roll - we explicitly model the probability of picking the

---

[21]On the same line we can interpret the insignificant coefficients of Payoff, obtained in 2015/16 and not affecting cheating behavior in 2016/17.

highest value between each pair of rolled dice, by estimating the model outlined in Equation 2, in which Math score is interacted with the possible difference dummies in the observed dice values and where observations refer to non-tied dice rolls in the experiment performed in Wave 2[22]. The results of this analysis are shown in Table 7.

[Table 7 about here.]

While the coefficient of Math Score is positive and significant, when we account for a measure of social age (column 2) or peer effects (column 3), the only interactive coefficients that are positive and significant are those relative to the highest differences (4 and 5) in column (3), i.e. the best specified version of the model, confirming that math-skilled pupils tend to cheat more when the difference between the dice (thus the reward form cheating) is higher [23].

It is worthwhile noting that while this result does not hold when the cheating pattern analysis is performed by using Total score, rather than Math score (see Table 8) as main explanatory variable.

We may thus jointly interpret these results and suggest that, while Total score, better approximating the overall cognitive abilities of the children (if not a more advanced theory of mind), provides a better prediction of cheating behaviour, by lowering the psychological costs implied by the fear of being discovered by the interviewer while lying, Math Score and its interaction terms, better approximating pure logical cognitive skills, are the only significant regressors able to explain the relationships between the likelihood to cheat and the size of advantages deriving from lying.

[Table 8 about here.]

In the economists' jargon[24] we may suggest that math-skilled pupils are better endowed to immediately spot when it is more convenient to cheat, thus maximizing variable cheating rewards in specific dice rolls while costs remain constant.

Based on the estimation results of the model shown in column (3) of Table 7, we also computed the marginal probabilities of choosing the die showing the maximum face value conditional on the difference between the dice values (labeled "temptation") at increasing levels of Math score. In particular, we focus on three alternative levels of Math score: the sample average value (equal to 0.64), and two other values obtained by adding and subtracting one standard deviation to the sample average. The resulting values are 0.49 and 0.79. The margins are plotted in Figure 7 showing that, while a clear monotonic pattern does not emerge, when temptation is stronger, i.e. for larger Diff values (such as: 3, 4, and 5), the probability of choosing the die showing the maximum value is always higher for Dice Diff= 6 as compared to Dice Diff= 1 and that this difference is higher for more math-skilled children. This outcome strengthens the evidence that cheating patterns are conditional on Math skills.

---

[22]The differences between the two dice face values may record 6 possible outcomes: 0, 1, 2, 3, 4, 5. Zero is irrelevant for our analysis since when two dice show the same value there is no possibility to cheat; thus, for this reason, ties are excluded from this analysis. Further, to avoid perfect collinearity, one difference dummy must be excluded and used as reference category for the included ones. We chose Diff= 1 as reference category, that's why it is excluded from the model.

[23]This result is weaker when we only control for social age (column 2) and it is not present in column (1), where only school dummies are included among fixed effects.

[24]Along the line of a "rational liar" *à la* Becker (1968).

[Figure 7 about here.]

This cheating pattern may be compatible with a mental model shared by math skilled pupils in which, while the "psychological cost" of cheating under the sight of the interviewer (in terms of fear of being discovered) is independent from the actual values shown by the dice, the advantages of cheating are directly proportional with Dice Diff. Students better at math may then decided to cheat especially (if not only) when it is worth the risk.

7 Conclusions

This paper provides a novel contribution into the investigation of the relation between school performance and cheating behavior. By exploiting data from a Lab-in-the-Field Experiment, we tested, on a sample of 170 primary school children in Goma (DRC), whether cheating behavior is influenced by cognitive abilities (measured by Math performance and overall performance) as recorded in school reports).

Our analysis shows that cheating behavior, elicited from a modified Dice Rolling Task (DRT), is positively and significantly correlated with better school performance as measured by higher Math and Total scores. Results are robust to the inclusion of a set of individual characteristics and experimental features, as well as to the inclusion of the baseline outcome of a DRT performed one year earlier.

Results are also robust to the inclusion of a measure of altruism (the outcome of a Dictator Game, DG). The coefficient associated to Altruism, when significant, displays a negative sign even if our experimental framework involved no damage to other fellow pupils, thus suggesting that children under analysis interpret lying as an anti-social behaviour even when their lies do not explicitly harm others.

Further, we also show that, while pupils' cognitive skills are a good predictor of cheating, the opposite - cheaters recording higher marks because of their deviant behaviour - does not hold.

Finally we give evidence that only when we limit our measure of school performance to Math Score, there is a significant relation between cheating behavior and the size of the reward arising from cheating. Thus math-skilled children not only do cheat more than their classmates but that they tend to cheat more when the reward from cheating is larger, thus suggesting that they are more able to identify a variation in the benefits (or revenues) deriving from cheating in specific dice rolls, contrary to costs which are constant across all rolls, and act accordingly.

All the above encourages further investigations into the mechanisms that link higher cognitive skills to a more advanced theory of mind and to cheating behaviour in primary school children.

References

Ariely, D., Garcia-Rada, X., Hornuf, L., and Mann, H. (2014). The (true) legacy of two really existing economic systems. *Munich Discussion Paper No. 2014-26*.

Becker, G. S. (1968). Crime and punishment: An economic approach. In *The economic dimensions of crime*, pages 13–68. Springer.

Bucciol, A., Houser, D., and Piovesan, M. (2011). Temptation and productivity: A field experiment with children. *Journal of Economic Behavior & Organization*, 78(1-2):126–136.

Bucciol, A. and Piovesan, M. (2011). Luck or cheating? a field experiment on honesty with children. *Journal of Economic Psychology*, 32(1):73–78.

Burke, R. H. (2013). *An introduction to criminological theory*. Routledge.

Camerer, C. (2003). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press, Princeton.

Cantin, R. H., Gnaedinger, E. K., Gallaway, K. C., Hesson-McInnis, M. S., and Hund, A. M. (2016). Executive functioning predicts reading, mathematics, and theory of mind during the elementary years. *Journal of Experimental Child Psychology*, 146:66–78.

Cohn, A., Fehr, E., and Maréchal, M. A. (2014). Business culture and dishonesty in the banking industry. *Nature*, 516(7529):86.

Ding, X. P., Omrin, D. S., Evans, A. D., Fu, G., Chen, G., and Lee, K. (2014). Elementary school children's cheating behavior and its cognitive correlates. *Journal of experimental child psychology*, 121:85–95.

Erat, S. and Gneezy, U. (2012). White lies. *Management Science*, 58(4):723–733.

Evans, A. D. and Lee, K. (2013). Emergence of lying in very young children. *Developmental psychology*, 49(10):1958.

Evans, A. D., Xu, F., and Lee, K. (2011). When all signs point to you: Lies told in the face of evidence. *Developmental Psychology*, 47(1):39.

Fischbacher, U. and Föllmi-Heusi, F. (2013). Lies in disguisean experimental study on cheating. *Journal of the European Economic Association*, 11(3):525–547.

Forsythe, R., Horowitz, J. L., Savin, N. E., and Sefton, M. (1994). Fairness in simple bargaining experiments. *Games and Economic Behavior*, 6(3):347–369.

Gächter, S. and Schulz, J. F. (2016). Intrinsic honesty and the prevalence of rule violations across societies. *Nature*, 531:496.

Gibson, R., Tanner, C., and Wagner, A. F. (2013). Preferences for truthfulness: Heterogeneity among and within individuals. *American Economic Review*, 103(1):532–48.

Glätzle-Rützler, D. and Lergetporer, P. (2015). Lying and age: An experimental study. *Journal of Economic Psychology*, 46:12–25.

Gneezy, U. (2005). Deception: The role of consequences. *American Economic Review*, 95:384–394.

Guala, F. and Mittone, L. (2010). Paradigmatic experiments: the Dictator Game. *The Journal of Socio-Economics*, 39(5):578–584.

Hao, L. and Houser, D. (2017). Perceptions, intentions, and cheating. *Journal of Economic Behavior Organization*, 133(C):52–73.

Jacobsen, C., Fosgaard, T. R., and Pascual-Ezama, D. (2018). Why do we lie? a practical guide to the dishonesty literature. *Journal of Economic Surveys*, 32(2):357–387.

Kahneman, D., Knetsch, J. L., and Thaler, R. (1986). Fairness as a constraint on profit seeking: Entitlements in the market. *The American Economic Review*, 76(4):728–741.

Kajackaite, A. and Gneezy, U. (2017). Incentives and cheating. *Games and Economic Behavior*, 102:433–444.

Kerschbamer, R., Neururer, D., Gruber, A., et al. (2017). Do the altruists lie less?

Maggian, V. and Villeval, M. C. (2016). Social preferences and lying aversion in children. *Experimental Economics*, 19(3):663–685.

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

Mazar, N. and Ariely, D. (2006). Dishonesty in everyday life and its policy implications. *Journal of Public Policy and Marketing*, 25:117–126.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, 11(6):619–632.

Pate, J. (2018). Temptation and cheating behavior: Experimental evidence. *Journal of Economic Psychology*.

Polak, A. and Harris, P. L. (1999). Deception by young children following noncompliance. *Developmental Psychology*, 35:561–568.

Ruffle, B. J. and Tobol, Y. (2017). Clever enough to tell the truth. *Experimental Economics*, 20(1):130–155.

Shalvi, S., Dana, J., Handgraaf, M. J., and De Dreu, C. K. (2011). Justified ethicality: Observing desired counterfactuals modifies ethical perceptions and behavior. *Organizational Behavior and Human Decision Processes*, 115(2):181–190.

Talwar, V., Gordon, H. M., and Lee, K. (2007). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. *Developmental psychology*, 43(3):804.

Vasek, M. E. (1986). Lying as a skill: The development of deception in children. *Deception: Perspectives on human and nonhuman deceit*, pages 271–292.

Wilson, A. E., Smith, M. D., and Ross, H. S. (2003). The nature and effects of young children's lies. *Social Development*, 12:21–45.
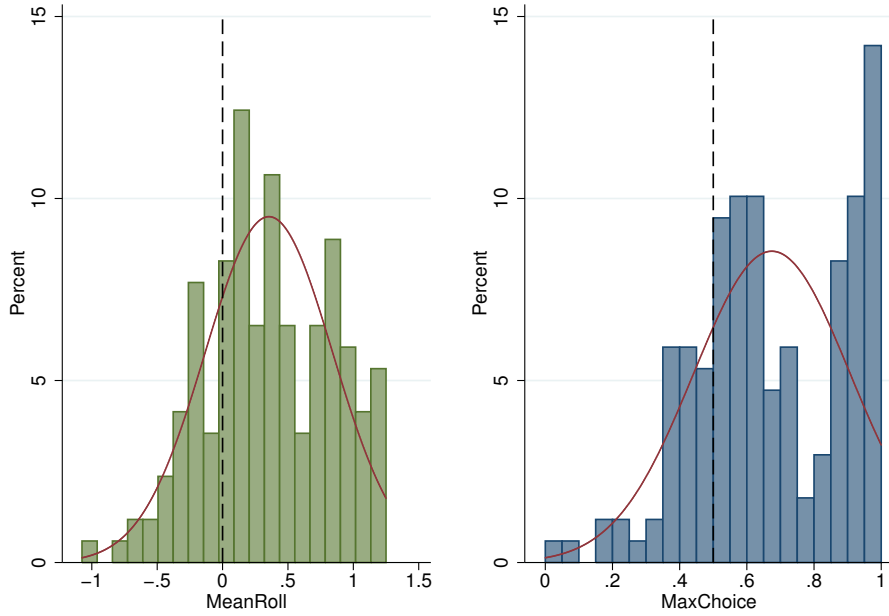
Dashed line set at 0, for MeanRoll and 0.5, for MaxChoice, i.e. values representing a random choice.
Normal density plot included in both graphs.

Figure 1: Distribution of MeanRoll and MaxChoice indicators, s.y. 2016/17.
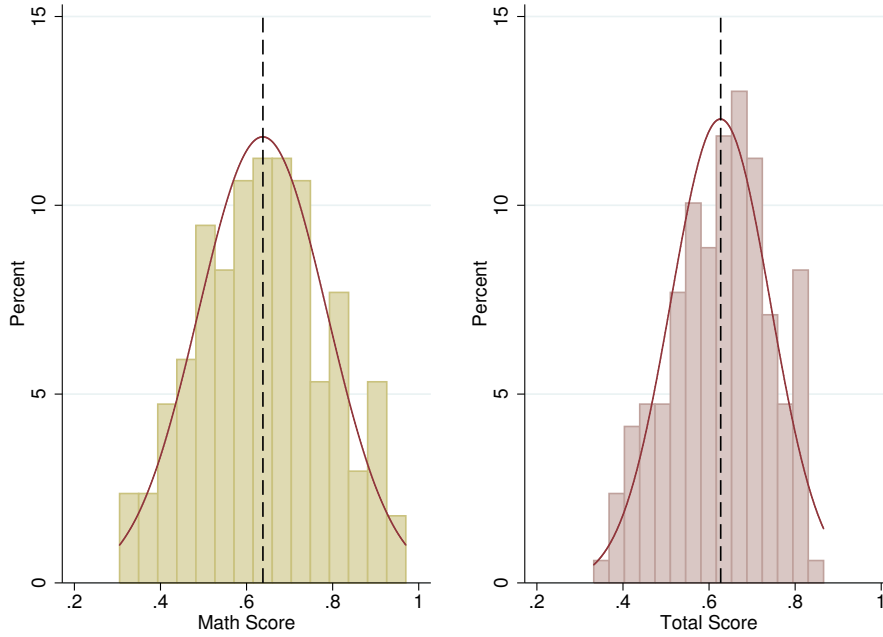
Dashed line at sample average; normal density line plot included in both graphs.

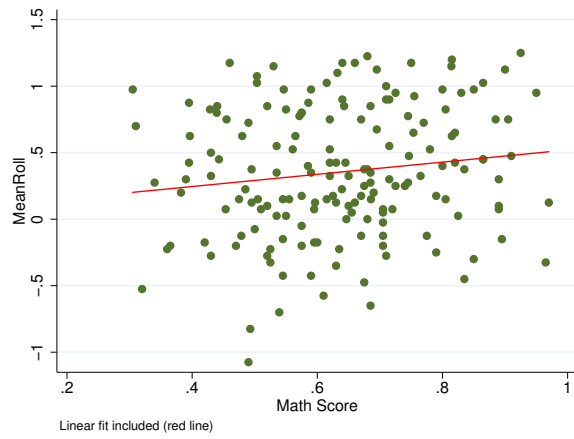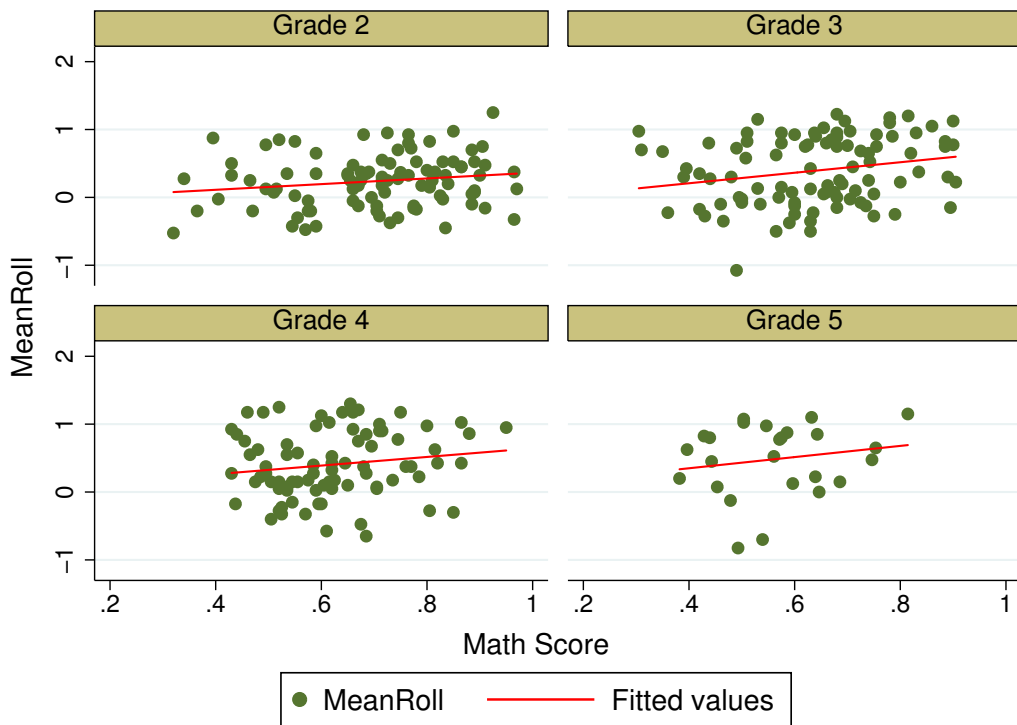Figure 2: Distribution of Math and Total score outcomes, s.y. 2016/17.

Figure 3: MeanRoll and Math scores, s.y. 2016/17

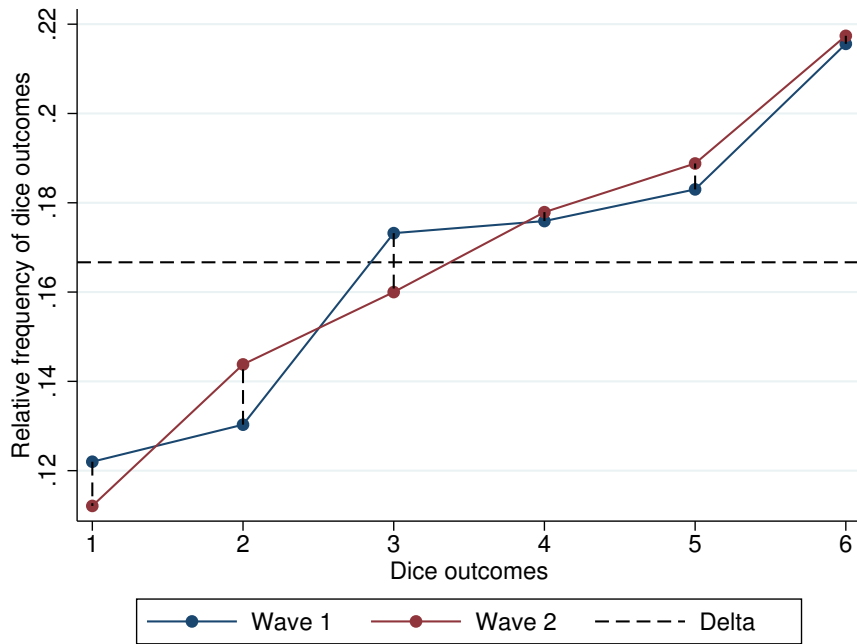Figure 4: MeanRoll and Math scores, by grade, s.y. 2016/17

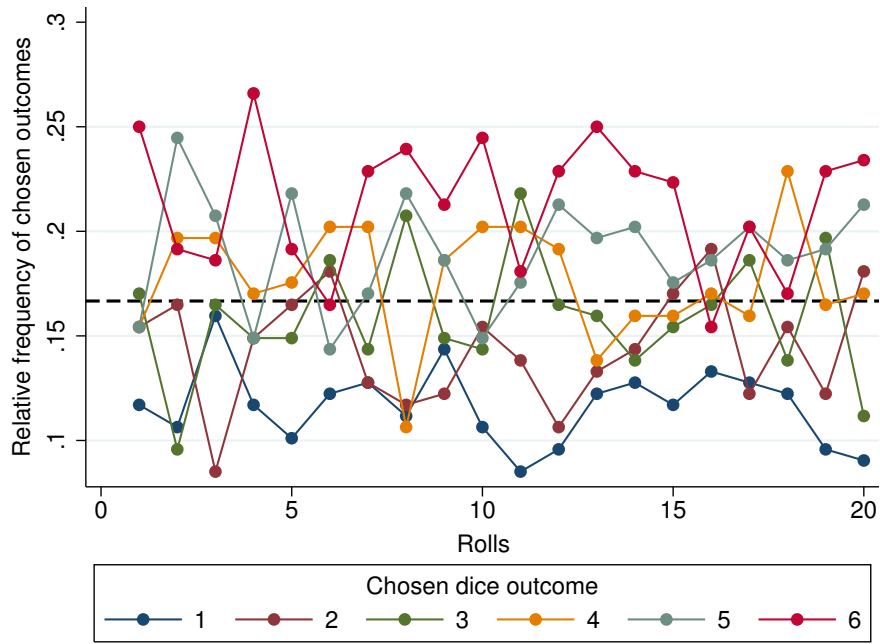Figure 5: Comparison of proportions of chosen outcome, by observed Dice Diff, between wave 1 and wave 2

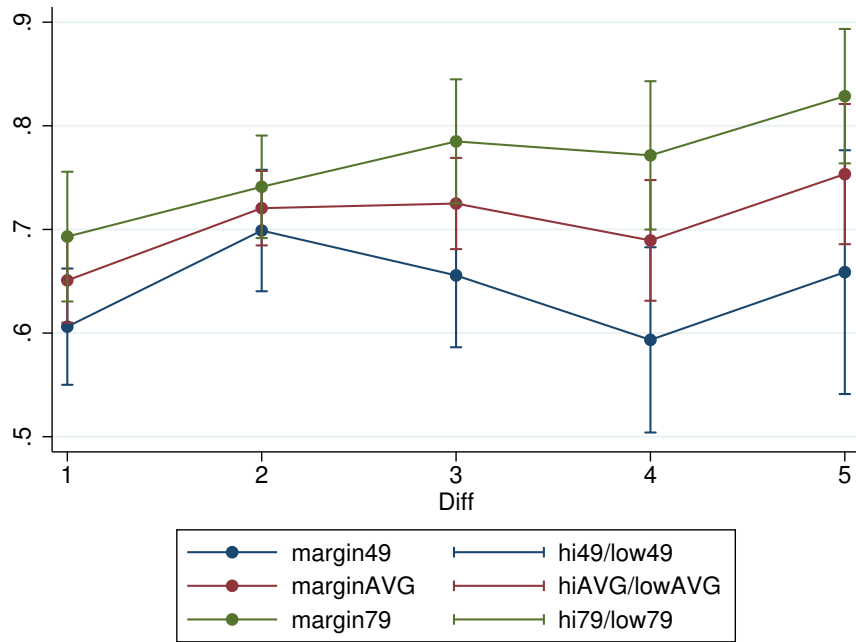Figure 6: Relative frequency of chosen outcomes, by roll

Figure 7: Predicted margins of MaxChoice, at alternative levels of Math score, by observed Dice Diff

Table 1: Brief description of the variables included in the analysis

Variable	Description
Behavioural indicators	
MeanRoll	Difference between average reported outcome and expected values of individual dice rolls
MaxChoice	Proportion of maximum values in a given roll chosen by children over the total number of throws (excluding ties)
Altruism	Proportion of packets of biscuits sent to the anonymous respondent
School performance indicators	
Math score	Score in Math in second term, harmonized by dividing "raw scores" resulting from official school reports by the maximum achievable points for term 2
Total score	Total score in term 2 (arithmetic sum of all individual subject scores, harmonized (same procedure described above)
Dice roll patterns	
Dice Diff=k	Absolute difference between red and blue dice observed outcomes in individual roll series, for k possible values, with $0 \leq k \leq 5, k \in \mathbb{N}$
Individual characteristics	
Age	Self-reported age of the child
Female	Dummy variable equals to 1 if the child is a female
Experiment characteristics	
Payoff	Payoff earned by child at the end of the first wave (September 2017)
Dummies:	
- Schools	Dummy variables identifying pupil's school
- Grades	Dummy variables identifying pupil's school
- Peer effects	Dummy variables identifying pupil's grade within a school
- Interviewers	Dummy variables identifying interviewers

Table 2: Summary statistics, school year 2016/2017

	Mean	*SD*	*Min*	*Max*	*Obs*
Behavioural indicators					
MeanRoll	0.356	0.487	−1.075	1.250	170
MeanRoll, baseline	0.296	0.455	−0.775	1.300	170
MaxChoice	0.676	0.234	0.000	1.000	170
MaxChoice, baseline	0.648	0.217	0.176	1.000	170
Altruism, baseline	0.412	0.167	0.000	0.800	170
School performance indicators					
Math score	0.638	0.150	0.305	0.970	169
Total score	0.627	0.116	0.332	0.866	169
Control variables					
Female	0.500	0.501	0.000	1.000	170
Age	8.729	1.459	6.000	13.000	170
Payoff	2.147	1.442	0.000	6.000	170

Notes: Baseline refers to school year 2015/2016. The sample includes only children that were not repeating their grade due to failure in school year 2015/2016.

Table 3: MeanRoll, Altruism and School performance: Math score

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
Math score	0.678**	0.772***	0.813**	0.665**	0.722**	0.769**
	(0.289)	(0.284)	(0.328)	(0.288)	(0.281)	(0.331)
MeanRoll, baseline				0.202**	0.217**	0.151
				(0.097)	(0.090)	(0.116)
Female	−0.139*	−0.138*	−0.137	−0.136*	−0.128	−0.137
	(0.081)	(0.082)	(0.096)	(0.079)	(0.082)	(0.094)
Age	0.105***	0.073	0.067	0.087***	0.071	0.074
	(0.032)	(0.047)	(0.063)	(0.031)	(0.046)	(0.063)
Payoff	0.015	0.013	0.007	0.011	0.010	0.007
	(0.033)	(0.033)	(0.038)	(0.033)	(0.032)	(0.038)
Constant	−1.036**	−0.745*	−0.514	−0.861**	−0.686	−0.635
	(0.444)	(0.442)	(0.691)	(0.432)	(0.439)	(0.717)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.056	0.055	0.000	0.079	0.084	0.007
Obs	168	168	168	168	168	168
AIC	261	257	273	258	253	273
BIC	386	363	442	386	362	445
	(1)	(2)	(3)	(4)	(5)	(6)
Math score	0.589*	0.700**	0.718**	0.578*	0.654**	0.668*
	(0.308)	(0.293)	(0.338)	(0.305)	(0.290)	(0.340)
Altruism, baseline	−0.402*	−0.447*	−0.475	−0.397*	−0.429*	−0.493*
	(0.236)	(0.251)	(0.293)	(0.227)	(0.249)	(0.292)
MeanRoll, baseline				0.201**	0.212**	0.161
				(0.093)	(0.085)	(0.111)
Female	−0.131	−0.114	−0.108	−0.128	−0.105	−0.107
	(0.081)	(0.085)	(0.098)	(0.079)	(0.084)	(0.096)
Age	0.101***	0.071	0.068	0.083***	0.070	0.075
	(0.031)	(0.045)	(0.064)	(0.031)	(0.045)	(0.063)
Payoff	0.008	0.006	0.001	0.004	0.003	0.000
	(0.034)	(0.033)	(0.039)	(0.034)	(0.033)	(0.039)
Constant	−0.762	−0.541	−0.350	−0.592	−0.492	−0.395
	(0.482)	(0.448)	(0.694)	(0.478)	(0.443)	(0.756)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.067	0.069	0.016	0.090	0.097	0.025
Obs	168	168	168	168	168	168
AIC	260	255	271	256	251	270
BIC	388	365	443	388	363	445

Notes: Dependent variable: MeanRoll. OLS estimation. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 4: MeanRoll, Altruism and School performance, robustness check: Total score

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
Total score	1.012∗∗∗	1.204∗∗∗	1.048∗∗	1.022∗∗∗	1.152∗∗∗	1.039∗∗
	(0.358)	(0.364)	(0.413)	(0.368)	(0.368)	(0.418)
MeanRoll, baseline				0.210∗∗	0.217∗∗	0.174
				(0.097)	(0.088)	(0.118)
Female	−0.130	−0.120	−0.134	−0.127	−0.111	−0.134
	(0.081)	(0.081)	(0.095)	(0.079)	(0.080)	(0.093)
Age	0.106∗∗∗	0.070	0.071	0.088∗∗∗	0.068	0.078
	(0.032)	(0.043)	(0.061)	(0.031)	(0.042)	(0.061)
Payoff	0.016	0.013	0.008	0.012	0.010	0.007
	(0.033)	(0.032)	(0.038)	(0.033)	(0.032)	(0.038)
Constant	−1.226∗∗∗	−0.965∗∗	−0.759	−1.059∗∗	−0.903∗∗	−0.786
	(0.462)	(0.449)	(0.676)	(0.452)	(0.443)	(0.726)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.069	0.078	0.002	0.094	0.108	0.014
Obs	168	168	168	168	168	168
AIC	259	253	273	255	248	272
BIC	384	359	442	383	358	443
	(1)	(2)	(3)	(4)	(5)	(6)
Total score	0.909∗∗	1.096∗∗∗	0.917∗∗	0.922∗∗	1.051∗∗∗	0.904∗∗
	(0.374)	(0.381)	(0.428)	(0.381)	(0.383)	(0.432)
Altruism, baseline	−0.393∗	−0.395	−0.461	−0.385∗	−0.376	−0.476
	(0.236)	(0.255)	(0.302)	(0.226)	(0.251)	(0.300)
MeanRoll, baseline				0.208∗∗	0.212∗∗	0.180
				(0.093)	(0.084)	(0.111)
Female	−0.122	−0.100	−0.106	−0.119	−0.092	−0.105
	(0.080)	(0.083)	(0.097)	(0.078)	(0.082)	(0.095)
Age	0.102∗∗∗	0.069	0.072	0.084∗∗∗	0.067	0.079
	(0.032)	(0.042)	(0.061)	(0.031)	(0.041)	(0.061)
Payoff	0.009	0.007	0.002	0.006	0.004	0.001
	(0.034)	(0.033)	(0.039)	(0.034)	(0.033)	(0.039)
Constant	−0.949∗	−0.764	−0.643	−0.790	−0.712	−0.534
	(0.501)	(0.467)	(0.778)	(0.497)	(0.459)	(0.770)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.079	0.087	0.017	0.104	0.116	0.030
Obs	168	168	168	168	168	168
AIC	258	252	271	254	247	269
BIC	386	361	443	385	360	444

Notes: Dependent variable: MeanRoll. OLS estimation. Robust standard errors in parentheses.
∗ $p < 0.10$, ∗∗ $p < 0.05$, ∗∗∗ $p < 0.01$

Table 5: Inspecting reverse causality: effect of Cheating on Math performance

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
MeanRoll, baseline	0.007	0.025	0.034	0.014	0.025	0.018
	(0.032)	(0.030)	(0.033)	(0.028)	(0.027)	(0.030)
Math, baseline				0.507***	0.494***	0.575***
				(0.089)	(0.093)	(0.097)
Female	−0.006	0.014	0.001	−0.017	−0.003	−0.009
	(0.027)	(0.024)	(0.027)	(0.023)	(0.022)	(0.024)
Age	−0.007	0.024*	0.026	−0.006	0.011	0.015
	(0.012)	(0.014)	(0.021)	(0.010)	(0.013)	(0.015)
Payoff	−0.005	−0.004	−0.006	0.001	0.000	−0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	0.615***	0.360**	0.586***	0.239	0.085	−0.049
	(0.163)	(0.146)	(0.221)	(0.174)	(0.138)	(0.228)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.119	0.120	0.168	0.317	0.300	0.389
Obs	168	168	168	164	164	164
AIC	−147	−152	−154	−188	−188	−204
BIC	−22	−46	14	−61	−79	−34

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
MaxChoice, baseline	−0.016	0.022	0.033	0.002	0.028	0.002
	(0.065)	(0.063)	(0.065)	(0.058)	(0.056)	(0.059)
Math, baseline				0.507***	0.497***	0.584***
				(0.089)	(0.092)	(0.096)
Female	−0.007	0.013	0.001	−0.017	−0.003	−0.009
	(0.027)	(0.024)	(0.027)	(0.023)	(0.022)	(0.025)
Age	−0.006	0.024*	0.025	−0.005	0.011	0.014
	(0.011)	(0.013)	(0.021)	(0.010)	(0.013)	(0.015)
Payoff	−0.005	−0.004	−0.006	0.001	0.000	−0.003
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Constant	0.611***	0.347**	0.561**	0.226	0.067	−0.024
	(0.153)	(0.143)	(0.219)	(0.166)	(0.135)	(0.229)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Adj. R-sq.	0.119	0.116	0.161	0.315	0.296	0.386
Obs	168	168	168	164	164	164
AIC	−147	−151	−153	−188	−187	−203
BIC	−22	−45	16	−60	−79	−33

Notes: Dependent variable: Math score. OLS estimation. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 6: MeanRoll and MaxChoice stability over time

	MeanRoll[1]			MaxChoice[2]		
	(1)	(2)	(3)	(4)	(5)	(6)
Time dummy	−0.266	−0.192	−0.352	−0.128	−0.073	0.198
	(0.189)	(0.183)	(0.737)	(0.088)	(0.082)	(0.154)
Female	−0.085	−0.071	−0.080	−0.033	−0.023	−0.031
	(0.064)	(0.062)	(0.076)	(0.027)	(0.027)	(0.029)
Age	0.098 ***	0.077 **	0.050	0.045 ***	0.035 **	0.016
	(0.024)	(0.034)	(0.044)	(0.011)	(0.017)	(0.021)
Payoff	0.019	0.016	0.013	0.011	0.011	0.009
	(0.023)	(0.024)	(0.027)	(0.010)	(0.011)	(0.010)
Constant	−0.635 **	−0.362	0.327			
	(0.306)	(0.433)	(0.945)			
Dummies:						
- Schools	*Yes*	*No*	*No*	*Yes*	*No*	*No*
- Grades	*No*	*Yes*	*No*	*No*	*Yes*	*No*
- Peer effects	*No*	*No*	*Yes*	*No*	*No*	*Yes*
- Interviewers	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Adj. R-sq.	0.100	0.065	0.035			
Pseudo R-sq.				0.083	0.068	0.107
Obs	284	284	284	284	284	284
AIC	407	414	438	362	354	421
BIC	623	607	759	577	547	757

Notes:
[1] Dependent variable: MeanRoll, OLS estimations, robust standard errors in parentheses.
[2] Dependent variable: MaxChoice, GLM for the binomial family estimations (Logit link), robust standard errors in parentheses.
The time dummy is equal to 1 if the observation refers to s.y. 2016/17; 0 if it refers to s.y. 2015/16.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: Cheating patterns: the effect of Math conditional on Dice Diff

	(1)	(2)	(3)
Math score	0.201	0.239*	0.255*
	(0.130)	(0.127)	(0.132)
Dice Diff=2 × Math score	−0.105	−0.109	−0.115
	(0.132)	(0.135)	(0.133)
Dice Diff=3 × Math score	0.175	0.184	0.178
	(0.154)	(0.158)	(0.152)
Dice Diff=4 × Math score	0.291	0.296	0.303*
	(0.188)	(0.191)	(0.183)
Dice Diff=5 × Math score	0.352*	0.382*	0.355*
	(0.205)	(0.208)	(0.195)
Dice Diff=2	0.129	0.126	0.139
	(0.084)	(0.087)	(0.085)
Dice Diff=3	−0.048	−0.056	−0.045
	(0.099)	(0.102)	(0.099)
Dice Diff=4	−0.149	−0.155	−0.159
	(0.119)	(0.121)	(0.118)
Dice Diff=5	−0.128	−0.154	−0.129
	(0.137)	(0.139)	(0.130)
Female	−0.052	−0.049	−0.049
	(0.033)	(0.034)	(0.036)
Age	0.049***	0.033	0.021
	(0.014)	(0.024)	(0.031)
Payoff	0.005	0.004	0.001
	(0.014)	(0.014)	(0.014)
Dummies:			
- Schools	*Yes*	*No*	*No*
- Grades	*No*	*Yes*	*No*
- Peer Effects	*No*	*No*	*Yes*
- Interviewers	*Yes*	*Yes*	*Yes*
Adj. R-sq.	0.018	0.012	−0.002
Obs	2741	2741	2725
AIC	3332	3352	3300
BIC	3610	3595	3666

Notes: Dependent variable: MaxChoice. Margins from GLM for binomial family estimations (Logit link). Robust standard errors in parentheses, clustered at children level.
The analysis is performed on all reported dice rolls; dice rolls in which Diff=0 (ties) are not included in the analysis. Reference category (omitted) is Dice Diff=1.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: Cheating patterns: the effect of Total score conditional on Dice Diff

	(1)	(2)	(3)
Total score	0.421 ***	0.538 ***	0.428 ***
	(0.161)	(0.157)	(0.164)
Dice Diff=2 × Total score	−0.109	−0.120	−0.153
	(0.184)	(0.187)	(0.183)
Dice Diff=3 × Total score	0.071	0.052	0.052
	(0.179)	(0.179)	(0.175)
Dice Diff=4 × Total score	0.207	0.187	0.226
	(0.260)	(0.263)	(0.255)
Dice Diff=5 × Total score	0.325	0.345	0.355
	(0.311)	(0.312)	(0.301)
Dice Diff=2	0.131	0.134	0.161
	(0.115)	(0.118)	(0.115)
Dice Diff=3	0.019	0.028	0.036
	(0.112)	(0.113)	(0.110)
Dice Diff=4	−0.091	−0.082	−0.105
	(0.158)	(0.161)	(0.157)
Dice Diff=5	−0.106	−0.123	−0.123
	(0.201)	(0.202)	(0.195)
Female	−0.047	−0.041	−0.048
	(0.033)	(0.034)	(0.035)
Age	0.050 ***	0.029	0.021
	(0.014)	(0.023)	(0.030)
Payoff	0.005	0.004	0.001
	(0.014)	(0.014)	(0.014)
Dummies			
Schools	*Yes*	*No*	*No*
Grades	*No*	*Yes*	*No*
Peer effects	*No*	*No*	*Yes*
Interviewers	*Yes*	*Yes*	*Yes*
Adj. R-sq.			
Obs	2741	2741	2725
AIC	3325	3336	3299
BIC	3603	3579	3666

Notes: Dependent variable: MaxChoice. Margins from GLM for binomial family estimations (Logit link). Robust standard errors in parentheses, clustered at children level.
 The analysis is performed on all reported dice rolls; dice rolls in which Diff=0 (ties) are not included in the analysis. Reference category (omitted) is Dice Diff=1.
 * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Appendix

[Figure A1 about here.]

[Figure A2 about here.]

[Table A1 about here.]

[Table A2 about here.]

S1 Supplementary materials

S1.1 Experimental procedures

The instruction to the children were read in Swahili. In the next sections below we report the English translation. Since the interviewer could read and speak both Swahili and French, therefore the instruction for the interviewer in the original paper questionnaire have been provided in French, while the lines to be read to the children were written in Swahili.

To perform the incentivized tasks, interviewers were provided with plastic cards representing packets of biscuits. The child was asked to perform his/her choice by handling the cards. When the task involved the allocation of biscuits between him/herself and the partner, the child would put the cards into envelopes: a white envelope for the biscuits that he/she was going to keep for him/herself; a yellow envelope for the biscuits eventually given to the anonymous partner. If the task involved no partner (as in the dice rolling task), only the white envelope was provided. Once all the tasks had been performed and all the envelopes appropriately filled with biscuits cards, the payment of the payoffs was administered according to the following protocol. The child randomly chose the incentivized tasks to be rewarded (the dice game involved also the selection of the individual throw to be rewarded). Then, the interviewer opened the corresponding envelope, counted the packet cards inside it and accordingly rewarded the child with real biscuits.

In both school years the survey took place between the end of the second term and the beginning of the third term, so to be temporally matched with official second term evaluations.The experimental design with two distinct observations (time span being around 13 months), allowed us to account for potential ex-ante heterogeneity in cheating behavior and to investigate the existence of reverse causation.

S1.2 Dice Rolling Task instructions (English translation)

In this section the child is asked to perform a die-throwing task. He/she will roll a couple of physical dice 20 times. You shall report in the table below the outcomes of the rolls and encircle the choice made by the child.

Interviewer: In this situation you are asked to throw two dice 20 times, in order to obtain some packets of biscuits. Every point you earn is equivalent to 1 packet of biscuits. The two dice have different colors, one being blue, the other one being red. Before rolling the dice, you must decide in your mind (without telling to anyone) whether you will choose the blue or the red die. Keep your choice in your mind without telling it to anyone.

You will earn some points (hence some packets of biscuits) corresponding to the number shown on the die you have chosen.

Let's start!

The interviewer, once the child has thrown the dice, takes note of the outcome and fills in a table. He/she will also report the choice of the child, by encircling the number corresponding to the choice made. Each throw corresponds to a letter from A to T in the table.

Let's make an example. Before throwing the dice (throw A), the child makes his/her choice in his/her mind ("RED" or "BLUE") without telling to anyone. After throwing the dice, the red die reads 5 and the blue die reads 2. The interviewer writes "5" and "2" in the corresponding cells in the table (column A). The child declares his/her choice: «RED» and the interviewer encircles the number "5" in the first cell of the table (column A). Other example: the child makes his/her choice. After throwing the dice, the red die reads 5 and the blue die reads 2. The interviewer writes "5" and "2" in the corresponding cells in the table (column A). The child declares his/her choice: «BLUE» and the interviewer encircles the number "2" in the first cell of the table (column A). And so on and so forth for all the 20 throws.

The verbal interaction between the interviewer and the child should follow this scheme:

- Did you think to a color (red or blue)?

- Throw the dice The child throws the dice item Tell me your choice

The child throws the dice, the interviewer fills in the table and encircles his/her choice

S1.3   Dictator Game instructions (English translation)

Before starting, the interviewer puts on the table the picture representing some children.

Interviewer: In this situation you are matched with one of the children you see in the picture, whose identity will be never revealed. The same rule applies to the other child: he/she will never know who you are.

You receive 5 packets of biscuits. The other child has received none. Your choice is about the number of packets of biscuits to send to the other child. At the end of this situation, you will get the number of packets of biscuits you have received, less the packets of biscuits you have sent to the other child. The other child will receive the number of packets of biscuits you have decided to send him/her.

How many packets of biscuits do you choose to send to the other child?

The child answers the question and the interviewer writes the number (between 0 and 5) in the box provided in the questionnaire.
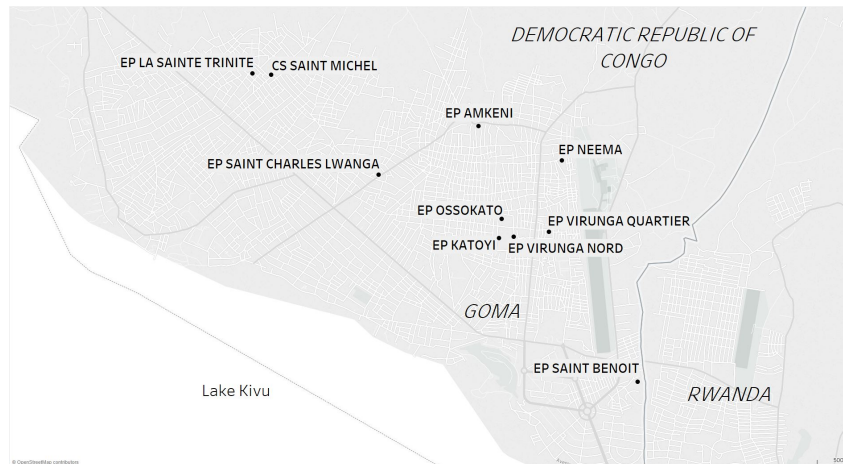
Figure A1: Goma, Democratic Republic of Congo

Figure A2: Position of the sample primary school in the outskirts of Goma. Source: [Authors 2017]

Table A1: MaxChoice, Altruism and School performance: Math score

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
Math score	0.285 **	0.324 ***	0.349 ***	0.285 **	0.310 ***	0.335 ***
	(0.117)	(0.115)	(0.120)	(0.118)	(0.115)	(0.121)
MaxChoice, Baseline				0.212 **	0.209 ***	0.172 *
				(0.085)	(0.079)	(0.093)
Female	−0.046	−0.043	−0.040	−0.043	−0.037	−0.040
	(0.034)	(0.035)	(0.037)	(0.034)	(0.035)	(0.036)
Age	0.049 ***	0.033	0.019	0.040 ***	0.032	0.022
	(0.014)	(0.025)	(0.032)	(0.014)	(0.025)	(0.032)
Payoff	0.006	0.005	0.003	0.003	0.003	0.001
	(0.014)	(0.015)	(0.014)	(0.014)	(0.014)	(0.014)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-sq.	0.078	0.066	0.103	0.085	0.073	0.107
Obs	168	168	168	168	168	168
AIC	225	215	255	225	215	256
BIC	346	318	430	350	322	434
	(1)	(2)	(3)	(4)	(5)	(6)
Math score	0.241 **	0.289 **	0.304 **	0.243 **	0.267 **	0.289 **
	(0.122)	(0.118)	(0.120)	(0.123)	(0.116)	(0.120)
Altruism, baseline	−0.193 **	−0.207 **	−0.238 **	−0.188 **	−0.190 **	−0.240 **
	(0.092)	(0.100)	(0.101)	(0.088)	(0.087)	(0.099)
MaxChoice, Baseline				0.207 **	0.209 ***	0.174 *
				(0.081)	(0.077)	(0.089)
Female	−0.042	−0.032	−0.027	−0.039	−0.028	−0.026
	(0.034)	(0.036)	(0.037)	(0.033)	(0.034)	(0.037)
Age	0.047 ***	0.032	0.022	0.039 ***	0.039 ***	0.024
	(0.014)	(0.025)	(0.033)	(0.014)	(0.015)	(0.033)
Payoff	0.002	0.002	−0.001	−0.000	0.000	−0.003
	(0.014)	(0.015)	(0.014)	(0.014)	(0.015)	(0.014)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-sq.	0.082	0.070	0.108	0.089	0.077	0.112
Obs	168	168	168	168	168	168
AIC	226	216	256	227	211	257
BIC	351	322	434	355	311	438

Notes: Dependent variable: MaxChoice. Margins from GLM for binomial family estimations (Logit link). Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table A2: MaxChoice and School performance, robustness check: Total score

	Benchmark models			Including baseline outcome		
	(1)	(2)	(3)	(4)	(5)	(6)
Total score	0.482 ***	0.585 ***	0.469 ***	0.482 ***	0.566 ***	0.468 ***
	(0.147)	(0.147)	(0.149)	(0.151)	(0.150)	(0.151)
MaxChoice, Baseline				0.211 **	0.202 ***	0.186 **
				(0.085)	(0.078)	(0.094)
Female	−0.041	−0.035	−0.040	−0.038	−0.029	−0.039
	(0.033)	(0.034)	(0.036)	(0.033)	(0.034)	(0.035)
Age	0.049 ***	0.029	0.020	0.041 ***	0.027	0.022
	(0.014)	(0.023)	(0.030)	(0.014)	(0.023)	(0.030)
Payoff	0.007	0.005	0.003	0.004	0.003	0.001
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-sq.	0.083	0.074	0.105	0.090	0.081	0.109
Obs	168	168	168	168	168	168
AIC	224	213	254	225	214	256
BIC	346	316	429	350	320	434
	(1)	(2)	(3)	(4)	(5)	(6)
Total score	0.434 ***	0.537 ***	0.406 ***	0.436 ***	0.522 ***	0.405 ***
	(0.151)	(0.152)	(0.149)	(0.154)	(0.154)	(0.150)
Altruism, baseline	−0.181 **	−0.175*	−0.227 **	−0.176 **	−0.166*	−0.228 **
	(0.091)	(0.098)	(0.103)	(0.087)	(0.096)	(0.100)
MaxChoice, Baseline				0.207 **	0.198 ***	0.187 **
				(0.081)	(0.075)	(0.090)
Female	−0.038	−0.027	−0.027	−0.034	−0.022	−0.026
	(0.033)	(0.035)	(0.036)	(0.033)	(0.034)	(0.036)
Age	0.048 ***	0.029	0.022	0.040 ***	0.027	0.025
	(0.014)	(0.023)	(0.031)	(0.014)	(0.023)	(0.031)
Payoff	0.003	0.002	−0.001	0.000	−0.000	−0.003
	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)	(0.014)
Dummies:						
- Schools	Yes	No	No	Yes	No	No
- Grades	No	Yes	No	No	Yes	No
- Peer effects	No	No	Yes	No	No	Yes
- Interviewers	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R-sq.	0.087	0.077	0.109	0.109	0.084	0.113
Obs	168	168	168	168	168	168
AIC	225	215	256	226	216	257
BIC	350	321	434	354	325	438

Notes: Dependent variable: MaxChoice. Margins from GLM for binomial family estimations (Logit link). Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$