



Munich Personal RePEc Archive

To pay or not to pay: Measuring risk preferences in lab and field

Brañas-Garza, Pablo and Estepa Mohedano, Lorenzo and Jorrat, Diego and Orozco, Víctor and Rascon-Ramirez, Ericka

LoyolaBehLab, Universidad Loyola Andalucía, Spain, ETEA-Development Institute, LoyolaBehLab, Universidad Loyola Andalucía, Spain, LoyolaBehLab, Universidad Loyola Andalucía, Spain, The World Bank, US, Middlesex University London, UK

23 September 2020

Online at <https://mpra.ub.uni-muenchen.de/103088/>
MPRA Paper No. 103088, posted 29 Sep 2020 09:37 UTC

To pay or not to pay: Measuring risk preferences in lab and field*

Pablo Brañas-Garza

LoyolaBehLab, Universidad Loyola Andalucía, Spain

Lorenzo Estepa

ETEA-Development Institute, LoyolaBehLab, Universidad Loyola Andalucía, Spain

Diego Jorrat

LoyolaBehLab, Universidad Loyola Andalucía, Spain

Victor Orozco

The World Bank, US

Ericka Rascón Ramírez

Middlesex University London, UK

September 23, 2020

Abstract

Measuring risk preferences in the field is critical for policy, however, it can be expensive and may generate unequal payoffs *due to bad luck*. For instance, the commonly used measure of Holt and Laury (2002) relies on a dozen of lottery choices and payments which makes it time consuming and costly, but also raises moral concerns as a result of the unequal payments generated by the lotteries. We propose a short version of the Holt and Laury (2002) which produces in the lab (Spain) the same results as the long HL. Using the short HL in the field (Honduras and Nigeria), we observe that paying or not for the measurement of risk preferences produces the same findings. Our low-cost approach makes the measurement of risk preferences simpler, faster and cheaper in the lab and field.

Keywords: Risk preferences, Holt Laury, Field Experiments, Monetary Payoffs, Incentives.

JEL codes: D81, C91, C93.

*We thank Avner Ben-Ner, Jordi Brandts, Antonio Cabrales, Rosemarie Nagel, Juan Camilo Cardenas, and participants at ESA2020 and SABE2020 for their comments. We also thank A. Amorós, J. F. Ferrero, E. Mesa and A. Nuñez who conducted the lab experiment. Data collection in Honduras and Nigeria was coordinated by Asociación PILARH and Hanovia Limited, respectively. We also thank Laura Costica and Edwin Daniel (Nigeria) and Gladis Gonzales (Honduras) for their excellent field coordination. We are grateful to the World Bank (WB-TF0A7752), The Spanish National Science Foundation (PGC2018-093506-B-I00), Excelencia Junta (PY-18-FR-0007), IdD Fundación (ETEA) 2019 and Middlesex University Business School Grants (RFF 2018-2019) for their financial support. Data collection in Honduras followed the sample selection used in Nigeria. The Nigerian data collection was part of a large-scale intervention conducted and funded by The World Bank (DIME), Norad and the Nigerian Government to inform baseline and follow-up data collection. Further details of this intervention can be found in <https://www.socialscisearch.org/trials/3619>.

Contents

1	Introduction	3
2	Experimental Design	4
2.1	The reduced Holt-Laury	4
2.2	Dimensions of Risk Preferences	5
2.3	Experimental Protocol	5
2.4	Use of enumerators	6
3	<i>Study 1: The lab experiment</i>	6
3.1	Sample size and balance	7
3.2	Ethics committee and pre-registration	7
3.3	Results: The impact of payment schemes	7
3.4	Testing the reduced Holt-Laury	9
3.5	Discussion	10
4	<i>Study 2: Field in Nigeria</i>	11
4.1	Sample size and balance	11
4.2	Ethics committee and pre-registration	11
4.3	Results	12
4.4	Discussion	13
5	<i>Study 3: Field in Honduras</i>	13
5.1	Sample size and balance	14
5.2	Ethics committee and and pre-registration	14
5.3	Results	14
5.4	Discussion	16
6	Equivalence Tests	17
7	Conclusion	18

1 Introduction

Measuring risk preferences is fundamental for modeling and predicting how people make choices. Risk preferences matters for human capital investment and career decisions (Weiss, 1972), technology adoption (Liu, 2013), financial portfolios (Schubert et al., 1999), insurance (Friedman, 1974) and migration (Katz and Stark, 1986); as a result, there is a growing interest by social scientists in measuring these preferences in observational and experimental studies.

While there is a common practice of using monetary incentives in lab experiments when eliciting risk preferences, in the field, the use of such incentives is less common. In particular, when data collection involves thousands of observations as in national representative household surveys and impact evaluation data, researchers tend to use hypothetical measures of risk preferences.

Standard non-hypothetical measures of risk preferences use incentive compatible devices to elicit individual preferences for risk. This approach entails three important implications: *i*) subjects make real choices (hence, risk preferences are not self-reported as in hypothetical measures), *ii*) choices have real monetary consequences, and *iii*) subjects' final earnings depend on their choices and nature. Therefore, the distribution of payoffs is unequal across participants given the chances of getting high or low payoffs after choosing between lotteries.

This inequality in payoffs may also create a sense of unfairness among participants even when they are fully aware of what they can get and the chances of getting it when making choices. Knowing what they can get and in which circumstances is true for different devices such as those suggested by Holt and Laury (2002); Gneezy and Potters (1997); Eckel and Grossman (2008) and Crosetto and Filippin (2013).

As a consequence, moral concerns emerge in the lab and field when using incentive compatible mechanisms. If payments are real and unequal, experiments using lotteries generate *ex-post* inequality among participants. This inequality is inevitable since any alteration of the randomness of the lottery would imply deception to the participants. Hence, paying experimental subjects real money when using lotteries generates inequality. This issue raises further concerns in the field when unequal payments create tensions among neighbours, inhabitants living in the same communities, or even between experimenter and field participants. Because of these potential risks, it is unclear whether using lotteries is acceptable for only research purposes.

Paying subjects with real money –the gold standard in lab experimental economics– instead of using hypothetical choices, as in observational studies and many impact evaluations, is therefore an issue. There are two reasons why incentives have been acclaimed to be necessary: *i*) apparently, when payoffs are hypothetical and subjects do not risk their own money, they are more likely to be more risk loving; and *ii*) in the absence of monetary incentives, subjects do not put enough effort on the task and consequently, they make random choices¹.

Our study tests to what extent the use of hypothetical payoffs has an impact on the measurement of risk preferences. This study has two important features. *First*, we ran a series of lab experiments in Spain (April, 2019). A reduced version of the Holt and Laury (2002) task with only 5 choices (hereafter HL5) was introduced to reduce the number of choices and time that participants are exposed to. The original HL considers 10 choices. We ran a trial in the lab to test whether the reduced HL5 differs from the original HL in terms of risk aversion and inconsistency. We also ran a series of experiments where participants were randomly assigned to three treatment arms with probability 1/3. Each treatment differs from the other in payment schemes: real payment, paying 1 out of 10 (BRIS)² and no payment at all (hypothetical). We show that HL5 does not differ from HL regardless of the payment scheme.

¹However, choices are not always random. For instance, in Holt and Laury (2002) subjects select the choice A and do not change their choice across the entire Multiple Price List (MPL) - even when the last choice is strictly dominated. See Amador et al. (2020) for a detailed analysis of inconsistencies in Holt and Laury tasks.

²*Between Random Incentive Subjects – BRIS*.

Second, we bring our reduced form of the Holt-Laury test (HL5) to the field. We ran two field experiments in Kano (Nigeria, April 2019) and Copán (Honduras, May 2019). We tested the HL5 in two developing countries as these are the main target of development agencies to promote long-term investments (e.g., human capital, technology adoption, saving) and for which risk preferences are a key determinant of such outcomes. For these countries, we tested whether paying or not to the participants produces different profiles of risk preferences. As in the lab experiment, all participants were randomly assigned to a real payment, payment with probability 1 out of 10 and no payment at all. To compare risk preferences across treatment arms, we contrast three dimensions of the measure: number of inconsistent subjects, number of safe options (under consistency) and response time.

We find that hypothetical and probabilistic payments have no impact on consistency, number of safe choices or response time. This means that when field experimentalists elicit risk preferences using hypothetical measures, they can trust that these are consistent across payment schemes (i.e., real, probabilistic and hypothetical) and that this can be done faster (and therefore cheaper) by using a short version of the Holt-Laury approach. However, paying only a fraction of the sample (probabilistic payment) introduces noise in the elicitation of risk preference in the lab: subjects become more risk loving when only a fraction of them is paid than when all of them are paid. A possible explanation is the idea of a “hot hand”: participants may assume that if they were lucky in the first lottery, they will also be lucky in the second one (Miller and Sanjurjo, 2018).

While the question about the effect of incentives on risk elicitation is not new in the lab, the evidence is somehow mixed and even scarce about their effect in the field. Wiseman and Levin (1996) show that subjects make the same risky decisions under real and hypothetical consequences. In the same line, Kühberger et al. (2002) show that hypothetical choices match real choices for small as well as for large payoffs. Conversely, Holt and Laury (2002, 2005) find that increasing the size of real payoffs leads to more risk averse behavior than hypothetical payments. Etchart-Vincent and l’Haridon (2011) and Barreda-Tarrazona et al. (2011) find that real and hypothetical choices significantly differ in the gain domain.

Our study provides robust evidence on the elicitation of risk preferences from two different settings: lab and field experiments. Our lab exercise took place in Spain and this was replicated in two developing countries: Honduras and Nigeria. Using our data collected in the three countries, we compare the three different payment schemes widely used in economics and psychology.

We now proceed to describe our experimental design, as well as the details of the three experiments encompassing this study. Our last section concludes and discusses some recommendations for future data collection.

2 Experimental Design

2.1 The reduced Holt-Laury

When eliciting risk preferences in the field, observational and experimental studies commonly face the pressing issue of the time spent by enumerators when visiting households. This issue becomes more salient when participants’ attention and time are limited as a result of domestic/job-related chores. To address this issue, we propose a short version of the measure suggested by Holt and Laury (2002) using 5 choices (HL5) instead of 10 (HL), which allows the elicitation of risk preferences in less time.³

³We decided to elicit 5 choices, over other possibilities, to keep the extreme choices as in the standard HL ($q = 0.1 \cdot \text{High} + 1 - q = 0.9 \cdot \text{Low}$ and the complementary case $q = 0.9 \cdot \text{High} + 1 - q = 0.1 \cdot \text{Low}$). Our selection of middle points ($q = 0.4, 0.5, 0.6$; $1 - q = 0.6, 0.5, 0.4$, respectively) was based on previous evidence where the majority of participants switch between options at these points.

The use of a short (trimmed) version of the HL might have a direct impact on the elicitation of risk preferences. To assess this possible effect, in *Study 1*, we ran a lab experiment to compute differences between HL and HL5. Section 3.4 shows that there are no differences between the short and long measures.

Subsequently, we use in *Study 1, 2* and *3* the HL5 measure. As in HL, every subject is asked to choose between two lotteries A and B. Both A and B offer a low and high payment with probability q and $(1-q)$. For the HL5, the set of choices is the following:⁴

1st: for $q = 0.1 \rightarrow$ A: $0.1*100N + 0.9*80N$; B: $0.1*200N + 0.9*2N$.

2nd: for $q = 0.4 \rightarrow$ A: $0.4*100N + 0.6*80N$; B: $0.4*200N + 0.6*2N$.

3rd: for $q = 0.5 \rightarrow$ A: $0.5*100N + 0.5*80N$; B: $0.5*200N + 0.5*2N$.

4th: for $q = 0.6 \rightarrow$ A: $0.6*100N + 0.4*80N$; B: $0.6*200N + 0.4*2N$.

5th for $q = 0.9 \rightarrow$ A: $0.9*100N + 0.1*80N$; B: $0.9*200N + 0.1*2N$.

Observe that the first choice is trivial since the expected value of A is much higher than B (82N vs 21.8N) while the opposite happens in decision 5 (98N vs 180.2N). In decisions 2, 3 and 4 the expected values of A and B are much closer (88N vs 81.2N, 90N vs 101N, 92N vs 120.8N, respectively) therefore subjects preferences for risk –rather than expected value– might play a critical role in the choice between A and B. Under risk neutrality, subjects should select: A, A, B, B, B . Under consistency, A, B, B, B, B implies risk loving and B, B, B, B, B strong risk loving behaviour. On the contrary, A, A, A, B, B is indicative of slight risk aversion, A, A, A, A, B risk aversion and A, A, A, A, A reflects strong risk aversion.

2.2 Dimensions of Risk Preferences

In *Study 1, 2* and *3*, we compare three dimensions of the elicitation of risk preference under the three payment schemes. The first one is *consistency*, that is, whether subjects make inconsistent choices. A typical example of inconsistency is multiple switching from A to B and then from B to A. Any switch from B to A is indeed another example of inconsistency. The second dimension is the *number of safe choices* (A instead of B) under consistency. Finally, we compute *response time*, that is, the duration of the task. This last dimension was only measured in *Study 2* and *3*. Here we summarise our dimensions:

C_i Consistency: whether the subject makes consistent choices or not.

RA_i Risk aversion: number of safe choices (A) taking values from 0 to 5, where 5 refers to extremely risk averse.

RT_i Time: number of seconds spent by the subject to complete the task.

2.3 Experimental Protocol

Reduced vs. Long HL

Using a sample of participants in Spain (*Study 1*), we consider two experimental arms with 1/2 probability to be selected in one or the other. Subjects are exposed to the reduced HL (5-item HL – HL5) or the long version (10-item HL – HL10).

⁴This example is expressed in Nairas, the local currency in Nigeria.

Payment Schemes

Study 1, 2 and *3* consider three experimental arms with $1/3$ probability. Subjects are exposed to payment schemes under the first two arms and to no payment scheme under the third one. The difference across treatments is the probability of being paid, p :

$$T_R: p = 1; T_B: p = 1/10; T_H: p = 0$$

Under T_R , all subjects are certain about receiving a payment, whereas under T_H subjects are certain of receiving no payment. For each study, the entire sample was randomly assigned to one out of three treatment arms: T_R refers to real payments, T_B to 1 out of 10 and T_H to hypothetical payments. Subjects were fully aware of their payment scheme before the elicitation of their preferences.

2.4 Use of enumerators

Both field experiments (*Study 2* and *3*) were conducted by enumerators who were trained by the authors. Using enumerators implied that subjects did not self-manage the instructions and these were read and explained by the enumerator. Enumerators used computer-assisted personal interviewing (CAPI) questionnaires in Nigeria and paper-based in Honduras. In both cases, enumerators received a list of households they had to visit, including the type of questionnaire they had to apply (i.e., T_R , T_B or T_H questionnaire). The authors conducted the random allocation of treatments prior to the visit to the communities and the enumerators did not have any influence on such selection. To be sure the enumerators were applying the corresponding questionnaire to the households, a field coordinator monitored the correct use of the lists created by the researchers in the ground. Prior to running both experiments, the authors piloted the risk preference questionnaire with around 20 subjects to ensure the translations in Hausa (Nigeria) and Spanish (Honduras) were appropriate to the context. All questionnaires and instructions were originally written in English.

For both field experiments, enumerators conducted all face-to-face interviews in the households of the participants. Only one experimental subject was interviewed per household.

In sharp contrast, *Study 1* did not use enumerators for data collection of risk preferences. Subjects self-managed the instructions in a lab.

3 *Study 1*: The lab experiment

This lab experiment was run in the School of Economics and Business at the University of Sevilla, the largest public university in Andalusia (Southern Spain). The experiment used paper-based questionnaires and was conducted the last week of April 2019.

In the majority of lab experiments, participants are university students who are self-selected into the experiment and hold high levels of education, see Exadaktylos et al. (2013). The main advantage of lab experiments is the absolute control of the conditions faced by the subject: a) participants cannot interact among them, unless interaction is required by the experiment; and b) there are no external distractions. For these reasons, a lab experiment is the cleanest approach to test our main research questions: *i*) do monetary incentives make a difference in the elicitation of risk preferences? and *ii*) can we use a simplified version of the HL measure and still being informative as the original 10-item HL?

The HL5 used in the lab is identical to the multiple price list (MPL) shown in subsection 2.1 but adjusting the monetary values according to the Spanish context (amounts were adjusted to be meaningful to Spanish participants and expressed in Euros). Thus, for *Study 1* lottery A offers 5 euros with q probability and 4 euros with $(1 - q)$; lottery B offers 10 and 0.1 euros with probabilities q and $(1 - q)$, respectively.

The rest of the section is organised as follows. First, we show sample sizes and balance tests for all treatments considered in *Study 1*. Then we provide details on the ethical approval and pre-registration for this experiment. Next, we present results addressing our two main research questions and finally, we discuss the implications of this study for field experiments.

3.1 Sample size and balance

The entire sample consisted of $n=178$ Spanish subjects⁵. Every subject was randomly assigned to one out of three payment schemes (T_R : 60, T_B : 57, T_H : 62). Table 1 shows the balance between sub-samples.

Table 1 shows no differences across treatments regarding age, gender, cognitive reflection test (CRT) and courses (freshmen, sophomore, etc). Participants included in the BRIS treatment are ($p < 0.1$) more likely to belong to year 1 (freshmen). Applying a Bonferroni correction to the p-values, the coefficient is not longer significant.

Table 1: Balance across treatments in *Study 1*

	<i>obs.</i>	<i>mean_R</i>	<i>R - H</i>	<i>p</i>	$ B - H $	<i>p</i>
Age	178	21.567	0.154	0.806	-0.163	0.770
Male	178	0.500	-0.008	0.806	-0.026	0.778
Courses	178	2.967	0.017	0.938	-0.353	0.089*
CRT	178	0.800	-0.090	0.539	-0.133	0.370

Note: H refers to hypothetical, R real payment and B probabilistic payment with 1 out of 10 chance of winning (BRIS). Inference was made using robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. R refers to Real, H to Hypothetical and B to BRIS

3.2 Ethics committee and pre-registration

The study was approved by the Ethics Committee of Loyola Andalucía University. All participants signed an informed consent prior to starting the experiment.

The lab experiment was pre-registered in April 10, 2019 and made public in April 25, 2019.⁶

3.3 Results: The impact of payment schemes

Table 2 shows the impact of payment schemes on the risk preference measure. The top of the table focuses on consistency (C_i) where n_c corresponds to the number of consistent subjects and the second row shows the number of safe choices (RA_i) under consistency.

Columns 1-3 show different statistics for each treatment. The percentage of subjects making consistent choices is shown in row 1. These percentages are not significantly different between treatments. Row 2 shows the mean and the standard deviation of the number of safe choices. Again, we do not observe significant differences either.

Finally, columns 4 and 5 provide the regression coefficients of each treatment for the two outcomes variables: consistency and risk aversion responses. For C_i , we estimate a linear probability model (LPM) and for RA_i a negative binomial regression model. All regressions control for age, gender, CRT and session fixed effects.

⁵We invited $n=298$ subjects: Approximately 60% subjects entered in the study of the three different payment schemes ($n=179$) while around 40% ($n=119$) entered in the experiment where we compare the reduced HL5 vs HL10, see subsection 3.4.

⁶More information is available at *AsPredicted* webpage: <https://aspredicted.org/2ar2u.pdf>.

Column 5 in Table 2 shows that risk preferences measured under real payments do not differ from those under hypothetical payment. This is true for the number of consistent individuals (C_i) and the level of risk aversion (RA_i).

However, column 4 shows that paying 1 out 10 (T_B) has an impact. While there is no difference in the number of consistent individuals, the number of safe choices is fewer under T_B . Subjects who are paid under $p = 1/10$ are less likely to be risk averse, that is, they are more risk loving. The coefficient is significant at the 1% significance level. Since T_B involves two lotteries (to be selected for payment and the HL), a possible explanation might be that subjects consider that both lotteries are correlated⁷

Figure 1 confirms the results found on the level of risk aversion (RA_i). The lines represent the cumulative distribution for the number of safe choices by treatments. While there is no difference between T_H and T_R (blue and black lines, respectively), the cumulative distribution for T_B is above the other two. The latter suggests that paying 1 out 10 generates a higher concentration of subjects choosing a lower number of safe choices than the other payment mechanisms. We performed a Kolmogorov-Smirnov test for equality of distributions for the three treatments. The results confirmed that Hypothetical and Real treatment present the same distribution while the BRIS treatment not (T_H vs T_R , $p = 0.80$; T_H vs T_B , $p = 0.03$; T_R vs T_B , $p = 0.02$).

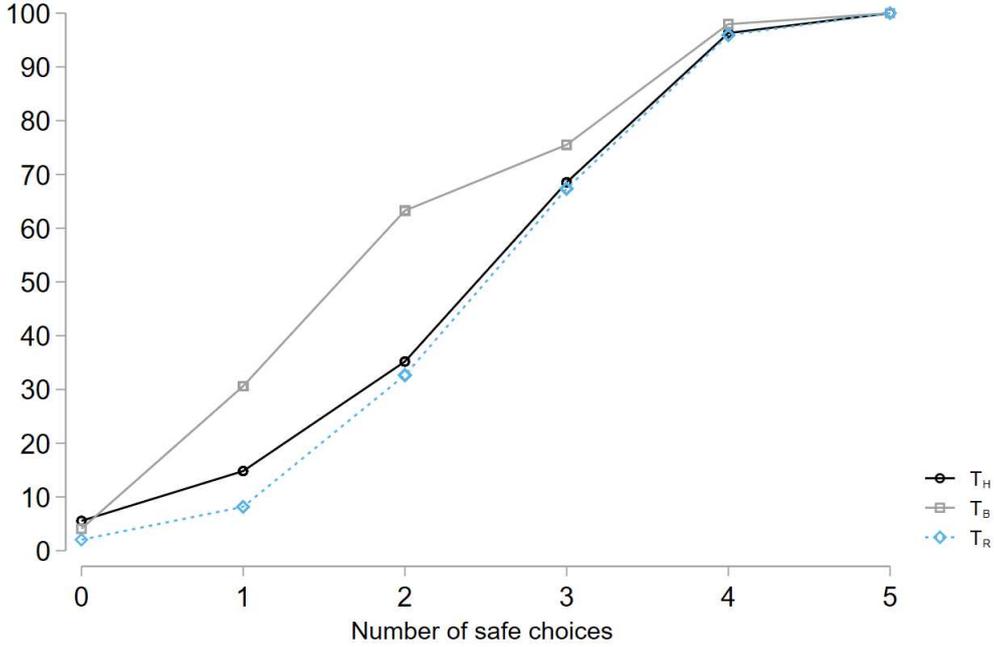
Table 2: Summary results in the Lab – *Study 1*

	(1)	(2)	(3)	(4)	(5)
	T_R	T_B	T_H	$\hat{\beta}_{T_B}$	$\hat{\beta}_{T_R}$
C_i					
n_c	49	49	53	0.003	-0.053
%	81.6	85.9	86.8	(0.958)	(0.435)
RA_i					
mean	2.938	2.796	2.285	-0.537**	0.135
SD	1.068	1.219	1.258	(0.050)	(0.541)
Obs.	60	57	62		

Note: Column 4 and 5 display the coefficients of regressing each outcome variable (row) in the Real and BRIS treatment (column). We use the treatment with hypothetical money as a reference group, that is, the coefficients in column 4 and 5 estimate the difference between T_B and T_R payment with respect to the Hypothetical treatment (T_H).

⁷Under the idea of a “hot hand”, participants may assume that if they were lucky in the first lottery, they will also be lucky in the second one (Miller and Sanjurjo, 2018).

Figure 1: Cumulative distribution for the number of safe choices.



As robustness check, in the Appendix we show that these results are hold using different specifications⁸. From the findings of *Study 1*, we conclude:

Result 1: *Compared to hypothetical payments in the lab: paying all the subjects have no impact in consistency or number of safe choices.*

Result 2: *Compared to hypothetical payments in the lab: paying 1/10 generates no differences in terms of consistency but decreases the level of risk aversion.*

3.4 Testing the reduced Holt-Laury

As noted before, our field set-up uses a reduced version of the HL test. In here, we test whether this trimmed version has any potential impact on the individual choices (consistency and risk aversion). The second lab experiment was conducted with 40% of the original pool of participants and took place in the same place as the previous experiment on payment schemes, see footnote 5. In this study, subjects were randomly assigned to one out of two experimental arms (by treatments, T_R : 60, T_H : 59). Observe that we do not have T_B in this sample.

Table 3 addresses two main questions: *i*) whether paying or not has an impact on risk choices in the HL10 and *ii*) whether short (HL5) and long (H10) lottery tasks generate similar risk preference profiles.⁹

First and second columns show real and hypothetical HL10: $HL10_R$ and $HL10_H$. Rows 1 and 2 show the number of consistent individuals (n_c) and the number of safe choices (RA_i) under consistency.

⁸Table A1 shows the results for C_i . Column 1 reports the results with no controls: $\hat{\beta}_{T_B} = -0.009(p = 0.885)$ & $\hat{\beta}_{T_R} = -0.052(p = 0.434)$. Column 2 shows the results with controls but without FE: $\hat{\beta}_{T_B} = 0.003(p = 0.958)$ & $\hat{\beta}_{T_R} = -0.053(p = 0.435)$. Finally, column 4 shows the results with controls, without FE but clustering errors by session (bootstrapped): $\hat{\beta}_{T_B} = 0.003(p = 0.959)$ & $\hat{\beta}_{T_R} = -0.053(p = 0.317)$. Table A2 studies the number of safe choices, RA_i : *i*) No controls: $\hat{\beta}_{T_B} = -0.540(p = 0.038)$ & $\hat{\beta}_{T_R} = 0.133(p = 0.525)$. *ii*) With controls but without FE: $\hat{\beta}_{T_B} = -0.553(p = 0.032)$ & $\hat{\beta}_{T_R} = -0.156(p = 0.449)$. *iii*) With controls without FE but clustering errors by session (bootstrapped): $\hat{\beta}_{T_B} = -0.553(p = 0.006)$ & $\hat{\beta}_{T_R} = 0.156(p = 0.469)$.

⁹Hereafter we use HL10 to refer to the 10-item HL to emphasise the number of choices under the original Holt-Laury measure.

Table 3: Testing the reduced HL version

	(1)	(2)	(3)	(4)	(5)
	$HL10_R$	$HL10_H$	$\hat{\beta}_{T_H}$	$\hat{\beta}_{5R 10R}$	$\hat{\beta}_{5H 10H}$
C_i					
n_c	53	49	-0.022	0.082	-0.067
%	88.3	83.0	(0.773)	(0.269)	(0.305)
RA_i					
mean	2.849	2.530	-0.316	-0.332	-0.285
SD	1.276	1.156	(0.258)	(0.178)	(0.283)
Obs.	60	59			

Note: Column 3 reports the coefficients of regressing each outcome variable (rows) in the hypothetical treatment in the HL10. Column 4 and 5 compare HL5 and HL10 in terms of the outcome variables, maintaining constant the payment scheme. P-values are in parentheses.

Columns 3-5 show the regression analysis. As before, regressions control for age, gender, CRT and session fixed effects.

Column 3 replicates Result 1 for HL10. This means that hypothetical payments have no impact in consistency or number of safe choices using the long version HL.

Column 4 and 5 addresses whether HL5 produces different risk preference profiles from HL10 or not.¹⁰

Column 4 – $\hat{\beta}_{5R|10R}$ – compares HL5 and HL10 under *real* payments. Regression uses the same controls and fixed effects noted before. No significant differences are found.

Column 5 – $\hat{\beta}_{5H|10H}$ – compares HL5 and HL10 under *hypothetical* incentives. Regression also uses the same controls and fixed effects. We do not find significant differences between HL5 and HL10 when using hypothetical payments.

From this second part of *Study 1*, we conclude:

Result 3: *Compared to the HL10 in the lab: The HL5 produces similar outcomes in terms of consistency and number of safe choices regardless the payment scheme.*

3.5 Discussion

Study 1 analyses risky decision-making in the lab. We address two main questions: *i*) whether the payment scheme impacts the measurement of risk preferences and *ii*) whether the instrument (the reduced HL5 task) affects such measurement.

Result 1 shows that paying all the subjects or none have no significant effect on consistency or number of safe choices. This is a critical issue for field data collection since samples (and therefore costs) are substantial when household visits are required.

Result 2, on the contrary, shows that BRIS payments may have an impact on risk taking (although do not alter inconsistency).

Result 3 shows that the HL5 we used in the field has no impact on the measurement of risk preferences: neither in consistency nor in number of safe choices. This is also relevant for field studies where time constraints are frequently present.

Overall, *Study 1* served for two purposes: to test whether incentives are important (and apparently are not) and to test whether our “fast” instrument is informative. In the following two sections, we will test Result 1 and 2 in the field – Nigeria and Honduras – using the short version HL5.

¹⁰We use the samples of $HL5_R$ and $HL5_H$ from the previous study.

4 Study 2: Field in Nigeria

This study was conducted in Kano (Nigeria) in April 2019. We ran the field experiment in three different villages in Kano: Dorayi, Ja'en and Gidan Maharba where 360 households were randomly selected. Our sample in each village was selected according to the eligibility criterion of having at least one child between 6 and 9 years old.

The experiment was composed of four modules: social norms (coordination games), subjective expectations, time and risk preferences. The HL5 was the last module we elicited. Random assignment to T_R , T_B or T_H treatments remained the same throughout the entire experiment. Response time (RT_i) was collected using CAPI for all four modules.

To elicit risk preferences, we used the same HL5 as in the lab experiment. We adapted the currency to Nairas (see the example in subsection 2.1). On average, subjects earned 126 Nairas in the HL5 section¹¹.

The rest of the section is organised as follows: First we show the sample size and balance across treatments. Then, we describe the ethical approval and pre-registration. Last, we show results and discuss the implications of our main findings for field experiments.

4.1 Sample size and balance

Sample size is $n = 360$ (by treatments, T_R : 120, T_B : 124, T_H : 116). Table 4 shows the balance between subsamples.

In this sample, 25% has no education, 8% has completed primary education and 32% has completed the secondary education. Around 40% of the sample is female and the average age is 39 years old. This sample is quite different from the sample used in the lab, but reflects common characteristics of the populations studied in Development Economics and who are subject to interventions aiming at increasing financial and human capital investment.

Table 4 shows the balance across treatments. We observe significant differences in age, where subjects in the real and BRIS treatment (T_R and T_B) have 2.7 years more than those in the hypothetical treatment (T_H). Another significant difference appears in the number of children at age 6-9, where those in the BRIS treatment (T_B) have a higher number than in T_H . However, all these differences were no longer significant after adjusting p-values using Bonferroni corrections.

Table 4: Balance across treatments in Nigeria field

	<i>obs.</i>	<i>mean_H</i>	<i>R - H</i>	<i>p</i>	<i>B - H</i>	<i>p</i>
Age	360	39.605	2.695	0.049**	2.807	0.032**
Female	360	0.403	-0.003	0.957	0.033	0.607
Education	360	9.245	-0.154	0.860	0.548	0.508
Sufficient ⁺	360	0.816	-0.007	0.885	-0.006	0.902
# children	360	1.625	-0.037	0.704	0.232	0.029**

Note: Inference was made using robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sufficient⁺ refers to sufficient money to eat. R refers to Real, H to Hypothetical and B to BRIS.

4.2 Ethics committee and pre-registration

Study 2 was approved by the Ethics Committee of Middlesex University London. All participants signed an informed consent.

¹¹We planned payments in order to cover one-day average wage for the entire experiment. This daily average wage was equal to 1080 Nairas (3 US\$). We paid (on average) 350N in coordination, 405N in time discount and 126N in risk.

The field study was pre-registered in April 15, 2019 and made public in April 25, 2019.¹²

4.3 Results

Table 5 reports a summary of results. The top of the table shows consistency (C_i), where n_c and % reflects the number and percentage of consistent subjects, respectively. The second line focuses on number of safe choices or risk aversion (RA_i) and the last one reflects the response time (RT_i). The number of observations per treatment appears at the bottom of the table. Columns 1, 2 and 3 report consistency and the number of safe choices by experimental arms T_R , T_B and T_H . The last columns show the regression coefficients of outcome variables on the treatment variables T_B and T_R .

		(1)	(2)	(3)	(4)	(5)
		T_R	T_B	T_H	$\hat{\beta}_{T_B}$	$\hat{\beta}_{T_R}$
C_i	n_c	117	119	111	-0.004	0.005
	%	97.5	94.4	97.4	(0.835)	(0.758)
RA_i	mean	3.085	2.899	3.135	-0.119	-0.027
	st. dev.	1.534	1.607	1.615	(0.459)	(0.867)
RT_i	mean	54.558	58.198	58.877	5.243	14.646*
	st. dev.	69.322	80.773	67.321	(0.535)	(0.082)
Obs (n)		120	126	114		

Note: Column 4 and 5 reflect the regression coefficients for each outcome variable (rows) in the BRIS and Real treatments (columns). We use the treatment with hypothetical money as a reference group. The time variable is expressed in seconds. Robust p-values are presented in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$.

As shown in Table 5 the fraction of consistent individuals, the mean of safe choices and the response time are nearly the same across treatments, which implies that there are no significant differences.

To test whether there are statistical differences between treatments we run different regression models: a linear probability model for C_i , a negative binomial regression for RA_i and an OLS for RT_i . All models control for enumerator effects, age, gender, education level and income. The last 2 columns of Table 5 show the results of these regressions: column 4 focuses on the BRIS treatment and column 5 on the real case. The reference group is the hypothetical payment in both cases.

For C_i (first row), we do not find any significant effect between T_B and T_H . The estimated impact of paying 1/10 (vs hypothetical payment) is not significant ($p = 0.835$). Similarly the estimated impact of paying all subjects (real) is not significant ($p = 0.388$). Table B1 in the Appendix show that these results are robust to different model specifications.

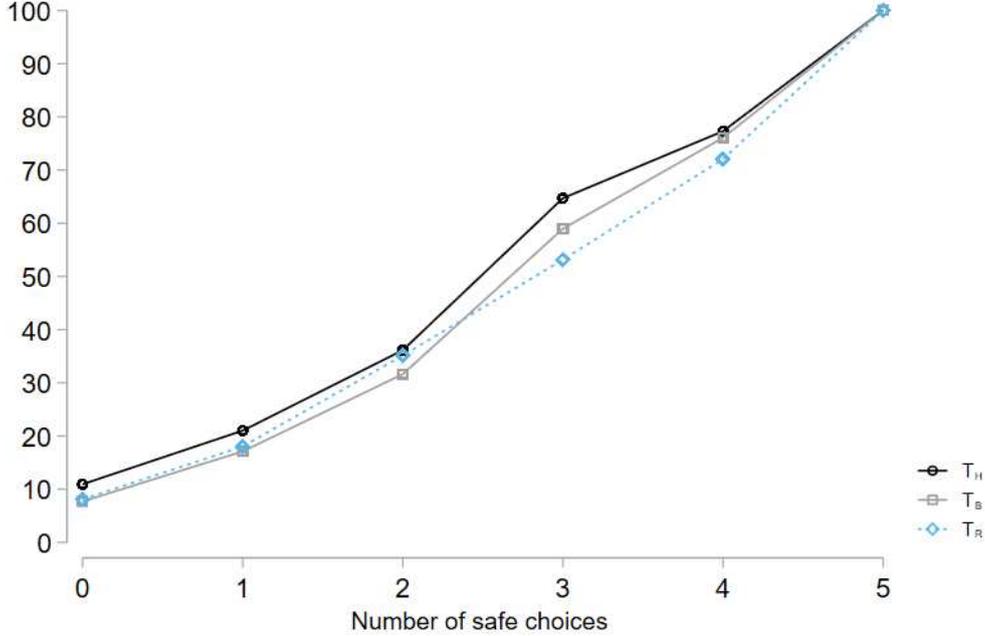
For RA_i (second row) no significant differences are found between treatments either. The estimated coefficients T_B and T_R are not significant ($p = 0.459$ and $p = 0.867$). These results are also robust to different model specifications (see Table B2 in Appendix).

Finally, RT_i (last row) offers some differences across treatments: While there is no effect for T_B ($p > 0.1$) we find that T_R is positive but weakly significant ($p = 0.082$). Indeed subjects making incentivized choices (with probability equal one of received a payment) are 15 seconds slower than those making hypothetical choices. However, these results are not robust to different model specifications (see Table B3 of in Appendix).

¹²Further information can be found in: <https://aspredicted.org/6qh4a.pdf>.

Finally, Figure 2 shows the cumulative distribution of the number of safe choices for the three treatments. The three lines are similar to each other and they cross for some values of safe choices. A Kolmogorov-Smirnov test also provides evidence that the three treatments present the same distribution (T_H vs T_R , $p = 0.99$; T_H vs T_B , $p = 0.43$; T_R vs T_B , $p = 0.99$).

Figure 2: Cumulative distribution for the number of safe choices in Nigeria.



In this study, we find that the payment scheme has no impact on consistency and the number of safe choices. For the last outcome, it also shows no difference in distributions between payment schemes. However, we find a slightly significant effect of real payments over the response time.

Here we summarise our main findings for *Study 2*:

Result 4: *Compared to hypothetical payments in the field: paying all subjects have no impact in consistency or number of safe choices.*

Result 5: *Compared to hypothetical payments in the field: Paying 1/10 payments have no impact in consistency, number of safe choices or response time.*

Note that the fact that subjects are (weakly) faster in hypothetical choices is not included in R4 since its not consistent.

4.4 Discussion

Results provided in *Study 2* show that elicitation of risk preferences with monetary incentives in the field provides the same results as the hypothetical ones.

5 *Study 3*: Field in Honduras

The main purpose of this experiment is to test whether Results 4 and 5 replicate in a different location (Banerjee et al., 2017; Peters et al., 2018; Al-Ubaydli et al., 2017). We ran a replication of the Nigerian

experiment in Santa Rosa de Copán (Honduras) between May 1 and 14, 2019. Our experiment took place in four different districts in Santa Rosa de Copán: Osorio, El Carmen, Prado Alto and Santa Teresa, where 360 households were randomly selected based on the eligibility criterion of having a child between 6 and 9 years old, following the criterion used for the field experiment in Nigeria.

As in Nigeria, the entire experiment consisted of four tasks (coordination, expectations, risk and time preferences). The main difference is that the risk preference task took place in the 3rd position, not in the last one as in *Study 2*. Subjects assignment to T_R , T_B or T_H remained the same for the entire experiment. RT_i was recorded by the enumerator for the entire block of risk preferences.

To elicit risk preferences, we used the same reduced HL task (HL5) as in the lab experiment and in *Study 2*. We just adapted the currency to Lempiras (see the example in subsection 2.1). On average, subjects earned 18 Lempiras in the HL5, instead of paying 100N or 80N in lottery A we paid 50L or 40L. In lottery B we changed 200N and 2N to 100L and 1L¹³.

The rest of the section is organised as follows: First we show the sample size and balance. Then, we give details on the Ethics committee and the pre-registration. Last, we show results and discuss their implications on field experiments.

5.1 Sample size and balance

The total sample consist of 360 subjects. Every subject was randomly assigned to 1 out of 3 arms resulting the following distribution. T_R : 109, T_H : 126, T_B : 125. Table 6 shows the balance across treatments. We observe significant differences in age, where the subjects allocated to the real payment (T_R) had 3.3 years less than those allocated to the hypothetical payment (T_H). However, this difference is not economically important (9% of the average age).

5.2 Ethics committee and and pre-registration

Study 3 was approved by Loyola Andalucía Ethics Committee. All participants signed an informed consent.

The field study was pre-registered in April 15, 2019 and made public in April 25, 2019. ¹⁴

Table 6: Balance across treatments in Honduras field

	<i>obs.</i>	<i>mean_R</i>	$T_B - T_H$	<i>p</i>	$T_R - T_H$	<i>p</i>
Age	359	35.750	-1.059	0.406	-3.310***	0.006
Male	360	0.144	0.015	0.746	0.034	0.437
Education	358	9.048	0.170	0.763	-0.002	0.997
Sufficient ⁺	360	0.768	0.009	0.854	0.003	0.962

Note: Inference was made using robust standard errors. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Sufficient⁺ refers to sufficient money to eat. R refers to Real, H to Hypothetical and B to BRIS.

5.3 Results

Table 7 shows the summary of results. As in *Study 2*, the top of the table refers to consistency (C_i), the second line to risk aversion (RA_i) and the last one to response time (RT_i). The number of observations per treatment appears in the bottom of the table. Columns 1, 2 and 3 respectively show the relevant

¹³The daily wage of unskilled workers in this rural community is about 150L (6 US\$)

¹⁴Further details can be found in: <https://aspredicted.org/6qh4a.pdf>.

values for T_R , T_B and T_H . The last 2 columns show the estimated impact of the payment scheme using a negative binomial model with controls: column 4 focuses on the BRIS payment and column 5 on the real case (where the reference group is the hypothetical payment).

For consistency C_i (first row), both treatments T_B and T_R are no significant ($p = 0.563$ and $p = 0.416$, respectively). These results are robust to different specifications (see Table C1).

Finally, our payment schemes have no impact on risk aversion and response time. It should be notice, that the coefficients of both treatments in RT_i are positive. This suggests that paying all subjects or 1 out 10 increases the response time in comparison to hypothetical payments, however, the difference is not economically relevant (between 11 and 14 seconds).

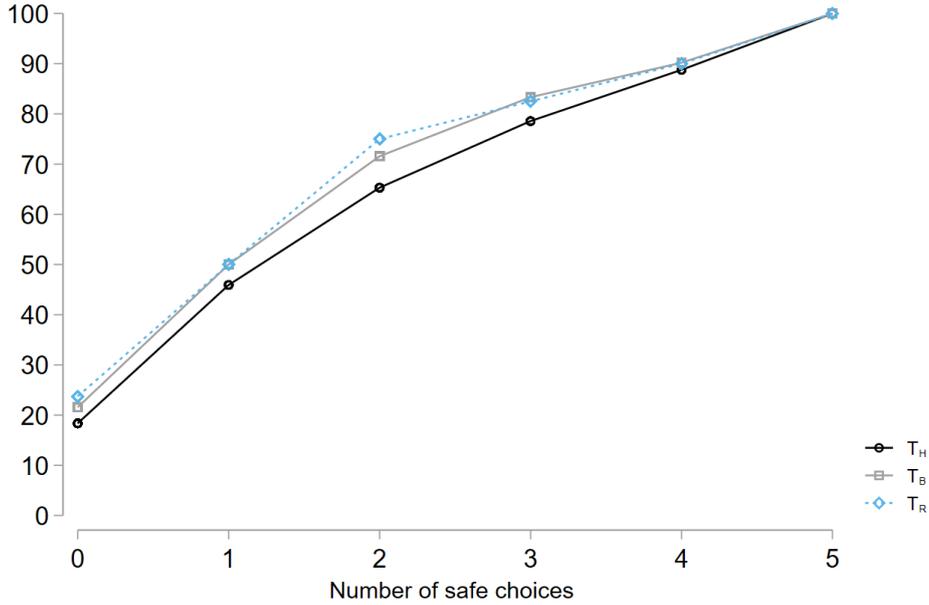
As before, Figure 3 shows the cumulative distributions for the three treatments and confirms the results found on the level of risk aversion (RA_i). While there is no difference between T_R and T_B lines, T_H line is below the other two lines. However, a Kolmogorov-Smirnov test provides evidence that the three treatments have no significant differences in their cumulative distributions (T_H vs T_R , $p = 0.80$; T_H vs T_B , $p = 0.99$; T_R vs T_B , $p = 0.98$).

Table 7: Summary results in Honduras field

	(1)	(2)	(3)	(4)	(5)
	T_R	T_B	T_H	$\hat{\beta}_{T_B}$	$\hat{\beta}_{T_R}$
<i>C_i</i>					
n_c	80	102	98	0.047	-0.051
%	73.4	80.9	78.2	(0.364)	(0.384)
<i>RA_i</i>					
mean	1.787	1.833	2.030	-0.151	-0.230
st. dev.	1.572	1.554	1.608	(0.456)	(0.317)
<i>RT_i</i>					
mean	185.523	190.516	170.772	14.625	11.549
st. dev.	105.606	132.818	112.466	(0.275)	(0.348)
Obs (<i>n</i>)	109	126	125		

Note: Column 4 and 5 reflect the regression coefficients for each outcome variable (rows) in the hypothetical and BRIS treatment (columns). We use the treatment with real money as a reference group, that is, the coefficients in column 4 and 5 estimate the difference between real and BRIS treatment with respect to the hypothetical treatment. The time variable is expressed in seconds. Robust p-values are presented in parentheses. Significance levels: * $p < 0.10$, ** $p < 0.05$ and *** $p < 0.01$.

Figure 3: Cumulative distribution for the number of safe choices in Honduras.



Our main findings show that the payment scheme has no impact on consistency, number of safe options or response time. We also find that there are no differences in the distributions of the number of safe choices among payment schemes.

Here we summarise our main findings for *Study 3*:

Result 6: *Result 4 is replicated in Honduras: Compared to hypothetical money, the use of real incentives has not impact on consistency, or risk taking.*

Result 7: *Result 5 is replicated in Honduras: Compared to hypothetical money, paying 1 out of 10 subjects has not impact on consistency, risk taking or response time.*

Observe that, compared to Nigeria, in Honduras time RT_i is not longer significant. However, Table C3, shows that the treatment variables are significant only for one specification – when controlling for school fixed effects and clustering standard errors at the enumerator level.

5.4 Discussion

Our findings from *Study 3* reinforce our conclusion from *Study 2*. The elicitation of risk preferences with monetary incentives in the field provides the same results as in the hypothetical scheme. Following Andersen et al. (2008); Apesteguia and Ballester (2018); Apesteguia et al. (2019), we also estimated a Random Utility Model (RUM) to analyse our data. Using the whole sample or restricting our analysis to only consistent subjects, our findings are the same in the lab, Nigeria and Honduras.

Our findings in the field are important for two reasons: *First*, existing hypothetical measures gathered in the field are indeed informative. *Second*, future measurements of risk preferences in the field do not need to use monetary incentives to get accurate proxies of risk preferences.

6 Equivalence Tests

In this section we test whether our estimates are different across treatments comparing estimates within a range instead of with respect to a point estimate.

A possible alternative is to explore whether or not the observed effect is large enough to be deemed worthwhile. This procedure is called *equivalence* (Lakens, 2017; Wellek, 2010). This procedure consists in testing whether the observed effect falls within or outside of an equivalence interval, defined by two exogenous bounds: the lower ($-\gamma_L$) and the upper (γ_U). To test for equivalence, a two one-sided test (TOST) approach is applied. Two composite null hypotheses are tested: $H_{01} \rightarrow \gamma \leq -\gamma_L$ and $H_{02} \rightarrow \gamma \geq \gamma_U$. When both null hypotheses are rejected, we can conclude that $-\gamma_L < \gamma < \gamma_U$ or that the observed effect falls within the equivalence bounds and it is *close enough to zero to be practically equivalent* (Lakens, 2017).

Our objective and exogenous bounds are defined based on Holt and Laury (2005) where they found a difference of 1 in the average number of safe choices between real and hypothetical incentives. Hence, our equivalence level is equal to 1 where we define the equivalence interval for each TOST as $H_{01} \rightarrow \gamma \leq -1$ and $H_{02} \rightarrow \gamma \geq 1$. To check the robustness of our results, we also use two additional equivalence levels: 0.50 and 0.75.

To consider a thorough analysis of our main findings, we use both the null hypothesis significance test (NHST) and the equivalence test (ET).¹⁵

Table 8 provides the summary of results from the equivalence test. Panel A shows that in the lab, paying 1 out 10 yield to *relevant* differences with respect to hypothetical decisions, while using real payments to *equivalent* results. In both field studies (panel B and C), equivalence tests suggest that both treatments (real and BRIS) yield equivalent results to the hypothetical payment scheme.

Table 8: Equivalence testing results

	Equivalence level					
	0.5		0.75		1	
	<i>P</i> - value	Conclus.	<i>P</i> - value	Conclus.	<i>P</i> - value	Conclus.
<i>Panel A: Lab</i>						
T_B	0.514	RD	0.182	RD	0.033	TD
T_R	0.078	E	0.008	E	0.000	E
<i>Panel B: Nigeria</i>						
T_B	0.015	E	0.000	E	0.000	E
T_R	0.003	E	0.000	E	0.000	E
<i>Panel C: Honduras</i>						
T_B	0.000	E	0.000	E	0.000	E
T_R	0.000	E	0.000	E	0.000	E

P-value refers to the maximum p value obtained from the two one-side test. Conclus. shows the conclusion considering both test NHST and ET. The baseline is the group making decisions with Hypothetical money (T_H).

All in all, from the equivalence test we can summarize the following findings:

Result 8: *Hypothetical and real measures yield to equivalent results in both the lab and field.*

Result 9: *BRIS and Hypothetical payment mechanisms yield to equivalent results **only** in the field.*

¹⁵According to these tests, there are four possible outcomes in the analysis: *i*) the observed effect can be statically equivalent ($-\gamma_L < \gamma < \gamma_U$) and not statistically different from zero (Equivalence or *E*); *ii*) statistically different from zero but not statistically equivalent (Relevant Difference or *RD*); *iii*) statistically different from zero and statistically equivalent (Trivial Difference or *TD*); and *iv*) neither statistically different from zero nor statistically equivalent (Undetermined or *U*).

7 Conclusion

This paper shows a systematic study of the impact of monetary incentives in the elicitation of risk preferences using Holt and Laury (2002). We cover a lab and two field experiments with different subject pools. The first one with university students, the second with parents living in rural areas and the third one with parents living in urban areas.

Our study answers a simple question: *Can we elicit risk preferences in the field without using monetary incentives and without generating inequality among participants?* We show in this study that this is possible. Lab experimentalists, who have widely used monetary incentives in the lab when eliciting such measures can confidently trust the measures that their field counterparts collect, in most of the cases, using hypothetical payments; and field experimentalists can now trust that these measures are consistent across payment schemes (i.e., real, probabilistic and hypothetical) and that this can be done faster (and therefore cheaper) by using a short version of the Holt-Laury approach.

Our concrete findings are summarised in here:

- The proposed short version (HL5 – 5 choices) of the Holt-Laury approach generates the same results as the original long version (HL10 – 10 choices).
- Using the short version (HL5) in the lab and field, produces the same risk preference profiles (in terms of consistency and safe choices) regardless of the payment scheme (hypothetical or real payment).
- When comparing the HL5 measures, hypothetical (no payment) and real payment schemes generate the same consistency and risk levels, that is, they yield equivalent measures.
- When comparing the HL5 measures, hypothetical (no payment) and BRIS scheme generate the same consistency and risk levels in the field, while in the lab the measurement of risk preferences display higher levels of risk loving attitudes under the probabilistic scheme.

Academic scholars concerned about the inequality generated among participants (when using lotteries) or between participants and non-participants (when offering monetary incentives to some, but not all) may want to consider our reduced version of the hypothetical HL to measure risk preferences. This approach is faster (and cheaper) and does not create any asymmetries among participants. Field experimentalists who not only aim at reducing data collection costs, but also at minimising feelings of unfairness among experimental participants, should consider hypothetical payment schemes to minimise potential frictions between participants and experimenter, or among winners and losers.

References

- Al-Ubaydli, O., List, J. A., and Suskind, D. L. (2017). What can we learn from experiments? understanding the threats to the scalability of experimental results. *The American Economic Review*, 107(5):282–86.
- Amador, L., Brañas-Garza, P., Espín, A. M., Garcia, T., and Hernández, A. (2020). Cognitive abilities and risk-taking: Errors, not preferences.
- Andersen, S., Harrison, G. W., Lau, M. I., and Rutström, E. E. (2008). Eliciting risk and time preferences. *Econometrica*, 76(3):583–618.

- Apestequia, J. and Ballester, M. A. (2018). Monotone stochastic choice models: The case of risk and time preferences. *Journal of Political Economy*, 126(1):74–106.
- Apestequia, J., Ballester, M. Á., Gutierrez, A., et al. (2019). Random models for the joint treatment of risk and time preferences. Technical report.
- Banerjee, A. V., Chassang, S., and Snowberg, E. (2017). Decision theoretic approaches to experiment design and external validity. In *Handbook of Economic Field Experiments*, volume 1, pages 141–174. Elsevier.
- Barreda-Tarrazona, I., Jaramillo-Gutiérrez, A., Navarro-Martínez, D., and Sabater-Grande, G. (2011). Risk attitude elicitation using a multi-lottery choice task: Real vs. hypothetical incentives. *Spanish Journal of Finance and Accounting/Revista Española De Financiación Y Contabilidad*, 40(152):613–628.
- Crosetto, P. and Filippin, A. (2013). The ‘bomb’ risk elicitation task. *Journal of Risk and Uncertainty*, 47(1):31–65.
- Eckel, C. C. and Grossman, P. J. (2008). Forecasting risk attitudes: An experimental study using actual and forecast gamble choices. *Journal of Economic Behavior & Organization*, 68(1):1–17.
- Etchart-Vincent, N. and l’Haridon, O. (2011). Monetary incentives in the loss domain and behavior toward risk: An experimental comparison of three reward schemes including real losses. *Journal of risk and uncertainty*, 42(1):61–83.
- Exadaktylos, F., Espín, A. M., and Branäs-Garza, P. (2013). Experimental subjects are not different. *Nature Scientific Reports*, 3:1213.
- Friedman, B. (1974). Risk aversion and the consumer choice of health insurance option. *The Review of Economics and Statistics*, 56:209–214.
- Gneezy, U. and Potters, J. (1997). An experiment on risk taking and evaluation periods. *The Quarterly Journal of Economics*, 112(2):631–645.
- Holt, C. A. and Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, 92(5):1644–1655.
- Holt, C. A. and Laury, S. K. (2005). Risk aversion and incentive effects: New data without order effects. *The American Economic Review*, 95(3):902–912.
- Katz, E. and Stark, O. (1986). Labor migration and risk aversion in less developed countries. *Journal of Labor Economics*, 4(1):134–149.
- Kühberger, A., Schulte-Mecklenbeck, M., and Perner, J. (2002). Framing decisions: Hypothetical and real. *Organizational Behavior and Human Decision Processes*, 89(2):1162–1175.
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*, 8(4):355–362. PMID: 28736600.
- Liu, E. M. (2013). Time to change what to sow: Risk preferences and technology adoption decisions of cotton farmers in china. *Review of Economics and Statistics*, 95(4):1386–1403.
- Miller, J. B. and Sanjurjo, A. (2018). Surprised by the hot hand fallacy? a truth in the law of small numbers. *Econometrica*, 86(6):2019–2047.

- Peters, J., Langbein, J., and Roberts, G. (2018). Generalization in the tropics—development policy, randomized controlled trials, and external validity. *The World Bank Research Observer*, 33(1):34–64.
- Schubert, R., Brown, M., Gysler, M., and Brachinger, H. W. (1999). Financial decision-making: are women really more risk-averse? *The American Economic Review*, 89(2):381–385.
- Weiss, Y. (1972). The risk element in occupational and educational choices. *Journal of Political Economy*, 80(6):1203–1213.
- Wellek, S. (2010). *Testing statistical hypotheses of equivalence and noninferiority*. Chapman and Hall/CRC.
- Wiseman, D. B. and Levin, I. P. (1996). Comparing risky decision making under conditions of real and hypothetical consequences. *Organizational behavior and human decision processes*, 66(3):241–250.

Appendix

Study 1: the lab experiment

In this section, we present the results from different specifications in the lab. Table A1 provide the regression results for consistency using different specifications. In all the specifications, the treatment coefficients are not statically significant. This suggest that the different payment schemes have no impact on consistency. These results are robust to different specifications.

Table A1. OLS estimation of payments impact on consistency.

VARIABLES	(1) Consistency	(2) Consistency	(3) Consistency	(4) Consistency
T_B	-0.009 (0.064) [0.885]	0.003 (0.062) [0.958]	-0.035 (0.073) [0.636]	0.003 (0.063) [0.959]
T_R	-0.052 (0.067) [0.434]	-0.053 (0.067) [0.435]	-0.063 (0.068) [0.355]	-0.053 (0.053) [0.317]
Constant	0.869*** (0.044) [0.000]	1.059*** (0.195) [0.000]	0.940*** (0.223) [0.000]	1.059*** (0.321) [0.001]
Observations	178	177	177	177
R-squared	0.004	0.036	0.095	0.036
Controls	N	Y	Y	Y
Session FE	N	N	Y	N
Session CE	N	N	N	Y
$T_B - T_R$	0.043	0.056	0.028	0.056
$p(T_B - T_R = 0)$	0.531	0.411	0.686	0.201

Note: the baseline or control group is the hypothetical group. Robust standard errors are presented in parentheses and p-values in brackets.

However, for risk aversion (see Table A2) we find that the BRIS treatment has an impact on risk aversion: people in this treatment tend to choose a lower number of safe choices than in the real incentives group.

Study 2: the field experiment in Nigeria

In this section, we present the results from different specifications in the Nigeria experiment. Tables B1 and B2 provide the regression results for consistency and risk aversion using different specifications. In all the specifications of both tables, the treatment coefficients are not statically significant. This suggests that the different payments scheme have no impact on consistency or risk aversion. This results are robust to different specifications.

Table A2. Impact estimation of payments on risk aversion.

VARIABLES	(1)	(2)	(3)	(4)
	# safe choices	# safe choices	# safe choices	# safe choices
T_B	-0.540** (0.261) [0.038]	-0.553** (0.257) [0.032]	-0.537** (0.274) [0.050]	-0.553*** (0.203) [0.006]
T_R	0.133 (0.209) [0.525]	0.156 (0.206) [0.449]	0.135 (0.220) [0.541]	0.156 (0.216) [0.469]
Observations	152	151	151	151
Controls	N	Y	Y	Y
Session FE	N	N	Y	N
Session clustered errors	N	N	N	Y
$T_B - T_R$	-0.673	-0.709	-0.672	-0.709
$p(T_H - T_B = 0)$	0.00632	0.00441	0.00630	0.00663

Note: the baseline or control group is the real incentives group. The table reports the marginal effect of negative binomial regressions. Robust standard errors are presented in parentheses and p-values in brackets.

Table B1. Estimation of the treatment payments effect on risk aversion

VARIABLES	(1)	(2)	(3)	(4)
	Consistency	Consistency	Consistency	Consistency
T_B	-0.029 (0.025) [0.251]	-0.029 (0.024) [0.244]	-0.004 (0.017) [0.835]	-0.026 (0.030) [0.388]
T_R	0.001 (0.021) [0.950]	0.003 (0.021) [0.892]	0.005 (0.017) [0.758]	-0.002 (0.009) [0.863]
Constant	0.974*** (0.015) [0.000]	1.028*** (0.055) [0.000]	1.044*** (0.056) [0.000]	0.971*** (0.071) [0.000]
Observations	360	360	360	360
R-squared	0.006	0.019	0.655	0.069
Controls	N	Y	Y	Y
Enum. FE	N	N	Y	N
Village FE	N	N	N	Y
Enum. CE	N	N	N	Y
$T_B - T_R$	-0.031	-0.031	-0.009	-0.025
$p(T_B - T_R = 0)$	0.222	0.213	0.614	0.382

Note: the baseline or control group is the hypothetical group. Robust standard errors are presented in parentheses and p-values in brackets.

Table B2. Impact estimation of payments treatments on risk aversion.

	(1)	(2)	(3)	(4)
	# safe	# safe	# safe	# safe
VARIABLES	choices	choices	choices	choices
T_B	-0.238 (0.214) [0.266]	-0.271 (0.215) [0.208]	-0.119 (0.160) [0.459]	-0.271 (0.215) [0.207]
T_R	-0.049 (0.203) [0.811]	-0.086 (0.204) [0.671]	-0.027 (0.163) [0.867]	-0.086 (0.194) [0.656]
Observations	347	347	347	347
Controls	N	Y	Y	Y
Enum. FE	N	N	Y	N
Village FE	N	N	N	Y
Enum. CE	N	N	N	Y
$T_B - T_R$	-0.189	-0.184	-0.0915	-0.184
$p(T_B - T_R = 0)$	0.361	0.371	0.584	0.351

Note: the baseline or control group is the real incentives group. The table reports the marginal effect of negative binomial regressions. Robust standard errors are presented in parentheses and p-values in brackets.

For time response (see Table B3) we find that T_H has an impact on risk time response: people in this treatment tend to be 15 seconds faster than in the real incentives group. However this effect is not robust to different specifications, especially when we took out from the regression the enumerator fixed effect.

Table B3. Estimation of the treatment payments effect on time response.

	(1)	(2)	(3)	(4)
VARIABLES	Time	Time	Time	Time
T_B	-0.679 (9.568) [0.943]	1.158 (9.827) [0.906]	5.243 (8.438) [0.535]	-0.722 (9.652) [0.941]
T_R	-4.319 (8.932) [0.629]	-2.506 (8.840) [0.777]	14.646* (8.388) [0.082]	-2.763 (10.251) [0.789]
Constant	58.877*** (6.304) [0.000]	88.674*** (25.670) [0.001]	70.150* (38.228) [0.067]	71.930** (28.998) [0.017]
Observations	360	360	360	360
R-squared	0.001	0.013	0.456	0.096
Controls	N	Y	Y	Y
Enum. FE	N	N	Y	N
Village FE	N	N	N	Y
Enum. CE	N	N	N	Y
$T_B - T_R$	3.640	3.665	-9.402	2.041
$p(T_B - T_R = 0)$	0.704	0.706	0.249	0.825

Note: the baseline or control group is the hypothetical group. Robust standard errors are presented in parentheses and p-values in brackets.

Study 3: the field experiment in Honduras

In this section, we present the results from different specifications in the Honduras experiment. Tables C1 provide the regression results for consistency using different specifications. Columns 3 and 4 shows that T_B is positive and significant ($p = 0.086$ and $p = 0.034$, respectively) when we controlled for school fixed effects. The estimated coefficient of T_H is not significant in all the specifications.

Table C1. Estimation of the treatment payments effect on consistency

	(1)	(2)	(3)	(4)
VARIABLES	Consistency	Consistency	Consistency	Consistency
T_B	0.027 (0.051) [0.595]	0.030 (0.051) [0.563]	0.047 (0.051) [0.364]	0.040 (0.041) [0.327]
T_R	-0.048 (0.056) [0.393]	-0.047 (0.058) [0.416]	-0.051 (0.059) [0.384]	-0.048 (0.042) [0.254]
Constant	0.782*** (0.037) [0.000]	0.690*** (0.123) [0.000]	1.028*** (0.196) [0.000]	0.820*** (0.187) [0.000]
Observations	359	356	356	356
R-squared	0.005	0.013	0.097	0.027
Controls	N	Y	Y	Y
Enum. FE	N	N	Y	N
School FE	N	N	N	Y
Clus. SE Enum	N	N	N	Y
$T_B - T_R$	0.0756	0.0765	0.0978	0.0875
$p(T_H - T_B = 0)$	0.171	0.177	0.0863	0.0362

Note: the baseline or control group is the real incentives group. Robust standard errors are presented in parentheses and p-values in brackets.

Table C2 provide the negative binomial regression results for risk aversion. None of the payment treatments are significant. This results are robust to different specifications.

Table C2. Impact estimation of payments treatments on risk aversion.

VARIABLES	(1)	(2)	(3)	(4)
	# safe choices	# safe choices	# safe choices	# safe choices
T_B	-0.193 (0.218) [0.377]	-0.167 (0.219) [0.445]	-0.151 (0.202) [0.456]	-0.112 (0.213) [0.599]
T_R	-0.241 (0.239) [0.312]	-0.187 (0.241) [0.438]	-0.230 (0.229) [0.317]	-0.186 (0.181) [0.304]
Observations	280	279	279	279
Controls	N	Y	Y	Y
Enum. FE	N	N	Y	N
School FE	N	N	N	Y
Clus. SE Enum	N	N	N	Y
$T_B - T_R$	0.0478	0.0199	0.0787	0.0739
$p(T_B - T_R = 0)$	0.844	0.936	0.737	0.745

Note: the baseline or control group is the real incentives group. The table reports the marginal effect of negative binomial regressions. Robust standard errors are presented in parentheses and p-values in brackets.

For time response (see Table C3) we find that the estimated coefficient of T_H is negative but not significant. However, in the last column (specification with school fixed effects and cluster error at enumerator level) this coefficient is significant at 10% level of significance. The estimated coefficient of T_B in all the specifications is not statically significant.

Table C3. Estimation of the treatment payments effect on time response.

VARIABLES	(1) Time (seconds)	(2) Time (seconds)	(3) Time (seconds)	(4) Time (seconds)
T_B	19.744 (15.733) [0.210]	22.523 (15.666) [0.151]	14.625 (13.376) [0.275]	25.528*** (9.128) [0.005]
T_R	14.751 (14.388) [0.306]	20.701 (14.268) [0.148]	11.549 (12.293) [0.348]	20.890** (9.790) [0.033]
Constant	170.772*** (10.143) [0.000]	88.296** (35.646) [0.014]	235.495*** (46.936) [0.000]	79.977 (57.976) [0.168]
Observations	352	349	349	349
R-squared	0.005	0.032	0.376	0.081
Controls	N	Y	Y	N
Enum. FE	N	N	Y	N
School FE	N	N	N	Y
Clus. SE Enum	N	N	N	Y
$T_B - T_R$	4.993	1.822	3.076	4.638
$p(T_B - T_R = 0)$	0.752	0.905	0.830	0.697

Note: the baseline or control group is the real incentives group. Robust standard errors are presented in parentheses and p-values in brackets.