RedNIE
Red Nacional de Investigadores en Economía

The Adventure of Running Experiments with Teenagers

Antonio Alfonso-Costillo (Universidad Loyola)

Pablo Brañas-Garza (Universidad Loyola)

Diego Jorrat (Universidad Loyola)

Pablo Lomas (Universidad Loyola)

Benjamin Prissé (Universidad Loyola)

Mónica Vasco (Universidad Loyola)

María J. Vázquez-De Francisco (Universidad Loyola)

DOCUMENTO DE TRABAJO N° 214

Febrero de 2023

Citar como:
Alfonso-Costillo, Antonio, Pablo Brañas-Garza , Diego Jorrat , Pablo Lomas, Benjamin Prissé , Mónica Vasco y María J. Vázquez-De Francisco (2023). El Estrés y Ansiedad por COVID-19. The Adventure of Running Experiments with Teenagers. *Documento de trabajo RedNIE N°214.*

The adventure of running experiments with teenagers*

Antonio Alfonso-Costillo[†], Pablo Brañas-Garza[†], Diego Jorrat[†‡], Pablo Lomas[†§], Benjamin Prissé[†], Mónica Vasco [†], María J. Vázquez-De Francisco[†§]

November 13, 2022

Abstract

Economists are increasingly interested in how to conduct experiments with teenagers. This paper evaluates whether different methodological factors impact the answers of teenagers to standard experimental tasks on measuring time preferences, risk preferences, cognitive abilities and financial abilities, among others. Results show: i) the recruitment process matters depending on whether the school includes the experiment as an institutional activity or the teachers led the process particularly for their class; the dropout rate reduced significantly from the first to the third experimental wave, when the school was responsible for organizing the experiment; ii) hypothetical payments elicits similar results than monetary payments; iii) adding visual elements to the experiment's interface improves the quality of answers; and iv) the type of electronic device on which subjects answer the tasks does not influence results, while administrating the experiment by school teachers does affect the answers. We conclude by giving three suggestions to researchers interested in conducting experiments with teenagers: first, run the experiment as a school-programmed activity; second, it is not necessary the use of real payments which increases the cost and complicates the recruitment; and third, integrate visual components to the task.

Keywords: developmental decision-making; field experiments; economic preferences; teenagers.

JEL-codes: C91, D81.

*This research was supported by the Spanish Ministry of Economy and Competitiveness (PID2021-126892NB-I00), Excelencia-Junta (PY-18-FR-0007) and Agencia Andaluza de Cooperación Internacional para el Desarrollo (AACID-0I008/2020).

[†]*LoyolaBehLab*, Universidad Loyola Andalucía, Spain.

[‡]Corresponding author. Email: dajorrat@gmail.com. Address: Av. de las Universidades, s/n, 41704 Dos Hermanas, Sevilla.

[§]ETEA Foundation-Development Institute, Universidad Loyola Andalucía, Spain.

1 Introduction

There is an increasing interest in children and teenagers among economists. The recent work of List et al. (2022) provides advice on how to run experiments with them, viewing this new research field as extremely promising. This interest of economists for non-adult people originates in the findings that undesirable human behavior is frequently linked with higher discount rate for both adults and children.

Chabris et al. (2008) showed that time discount rate was the best explanatory variable for a range of personal characteristics and behaviors in adults, such as higher Body Mass Index (BMI), higher smoking probability and less physical exercising. These findings gave more generality to previous research, which had linked higher time discount rate with addictive behavior, such as Kirby et al. (1999) did with heroin consumption, and Dixon et al. (2003) with pathological gambling. In addition, Reynolds (2006) provides a review of papers that link drug consumption with impatience.

These results were later extended to children and teenagers by Sutter et al. (2013) for alcohol and cigarettes consumption, higher Body Mass Index (BMI), higher disciplinary referrals at school and smaller savings. These findings interrogated the relationship between time preferences and behavioral outcomes among teenagers.

The economic literature provided further evidences that higher time discount rate is correlated with worse behavioral outcomes. Castillo et al. (2011) found that boys are more impatient than girls, and that black children are more impatient than white children. They also showed that being more patient decreases the likeliness of disciplinary referrals the following year by 14%. Castillo et al. (2018) later showed that more impatient teenagers are less likely to graduate from high school. Andreoni et al. (2019) found that children are initially impatient but become more patient with age. They also showed that black children are more impatient than white or Latin American children, and that schooling opportunities are not associated with time preferences. However, Horn et al. (2022) did not observe significant gender differences in patience using a sample of 1088 students from 53 classes in Hungary.

These evidences suggest the value of interventions on teenagers' financial education. Bruhn et al. (2013) and Alan and Ertac (2018) investigated about increasing patience on middle school students. Lührmann et al. (2018) showed that financial education increases the quality of intertemporal decision-making and decreases narrow bracketing in high school students. Such interventions aimed at guiding recipients' behavior to better directions during the formative years, have the potential to improve lifetime outcomes. Recent literature suggests that it could also benefit their offspring, as stated by Samek et al. (2021), that time preferences of parents are positively related with those of their children, and by Stoklosa et al. (2018), that impatience and present bias of parents are positively related with children's obesity likeliness. These findings point at the desirability to generate interventions to create virtuous circles (or break vicious ones).

Measuring risk preferences is as important as measuring time preferences (see Andreoni and Sprenger (2012)). Economic events do not only occur at different time periods. They also have different probabilities of occurring. It explains why Tymula (2019) recently contributed to this debate suggesting that being observed modifies teenager's choices occurring at different moments. She elicited both time and risk preferences, showing that teenagers are more present-biased, make more early choices and are more inconsistent under peer observation. She also found that teenagers are risk-averse, a finding shared with other authors that also illustrates diverse phenomena concerning risk-aversion. Eckel et al. (2012) measured risk preferences, finding that high school students with more low-income peers are more risk averse. They also showed that students in smaller classes or with more qualified educators have more moderate levels of risk-aversion. Additionally, taller and non-white individuals are more risk-seeking. They also found that girls are more risk-averse as Horn et al. (2022). Castillo et al. (2018) also showed that being more risk-averse reduces the likeliness of disciplinary referral and increases the likeliness of completing high school, even after controlling for irrationality, cognitive abilities, social environment and past behavior. Andreoni et al. (2020) elicited risk preferences, finding positive correlation between higher cognitive abilities with more risk-taking, and higher executive functions with more risk-aversion. Recently, Jørgensen et al. (2022) found that comparing younger with older Danish children, risk taking significantly increases with age for boys and marginally significantly for girls. And closely related with this, girls are less likely to compete when controlling for individual ability, confidence, risk aversion,

stereotypical beliefs, and interaction with the opposite gender for children performing above the class average. This might be linked with gender differences concerning career choices and labor market outcomes.

The goal of this paper is to present the results of different experimental design and data collection methods to measure economic preferences in teenagers using a multi-dimensional research platform. We indeed modified the experimental design between waves monitoring the results obtained at all times. Our goals were to identify the experimental design that provides the best quality of results and to shed more light on different data collection methods, which may be best suited to teenagers.

First, we tested whether recruiting subjects through agreements with teachers or institutionally with the schools[1] yields different attrition levels. Second, we tested whether using hypothetical payments rather than real payments influence results. Third, we studied whether using visual versions of the experimental tasks (as in Prissé (2022)) improved the quality of results. Fourth, we analyzed whether responding to the experiment on different electronic devices affects results; and finally, we tested whether administrating the experiment through university staff or teachers yields different answers.

The data were collected in three waves. For the first and the second waves, we used a first version of the booklet composed of standard tasks from the economic literature especially adapted to teenagers: MPLs to measure time preferences (in different formats), the Holt and Laury (2002) task (HL) to measure risk preferences, the Cognitive Reflection Test (CRT) of Frederick (2005) using the adaptation of Thomson and Oppenheimer (2016), and a test composed of three questions to measure financial abilities (Fin). For the third wave we used a revised version of the booklet updated with "visual" versions of the MPL-time and the HL tasks, a modified version of the CRT, and the Delavande test (DL) of Estepa et al. (2021) inspired by the original work of Delavande and Kohler (2009) to measure the ability of teenagers to understand probabilities.

All the participants of the three waves in this research were recruited in Spain and they were from grades 1 to 4 (12 to 16 years old) of secondary school. It is necessary to clarify the differences among waves in terms of recruiting and experimental design. During the first wave, the experiment was conducted online due to the Covid-19 pandemic by launching a questionnaire in Lime Survey. Subjects were recruited by contacting their teachers directly or by sending direct private invitations to their parents through WhatsApp. We obtained a sample of 1075 participants. We used monetary incentives in half of the sample to engage participation, with a peer-to-peer draw where the odds of winning real money were 1 in 20 among participants[2]. The experimental payment was between 10 and 35 euros and the average was 20 euros.

The second and the third wave were conducted as lab-in-the-field experiment, since sessions were run in the classrooms. We used an application tailored to students that added greater control over data privacy: SAND[3]. In both waves, researchers agreed with schools' administrations to include this activity as part of the school tutoring program. We obtained a sample of 564 and 959 participants, respectively. It is important to note that while in the first wave the payoffs were real, in the last two waves incentives were hypothetical.

Regarding the data collection methodology, we found five main results. First, recruiting adolescents by school agreements and including the experiment as a regular class activity reduces attrition significantly. Second, paying subjects with hypothetical monetary incentives elicits results similar to paying real monetary payoffs. Third, using different visual elements yields different results, because subjects become more consistent and less patient when pictures are included in the experimental design. Fourth, using different electronic devices to fill the experiment does not affect results, except if subjects answer using their mobile phones, since they do it more rapidly. Fifth, administrating the experiment by University staff rather than school teachers reduces consistency in risk, makes subjects less patient and decreases the number of correct answers in cognitive tasks. We conclude that our large-scale study provides useful insights and sheds light on how to conduct experiments with teenagers.

[1] The first method requires subjects to provide a parental consent. The second does not have such requirement.

[2] Between-subject Random Incentivized System (BRIS) payment.

[3] Acronym for Social Analysis and Network Data. Platform offered by Kampal company (https://www.kampal.com/)

The structure of this paper is as follows: Section 2 presents the recruitment process and subjects characteristics; Section 3 describes the experimental tasks; Section 4 presents the research questions; Section 5 presents the results; and Section 6 concludes with a discussion.

2 Recruitment process and subjects characteristics

The study was approved by the Ethical Committee of Universidad Loyola Andalucía and the entire experiment was pre-registered in AsPredicted[4]. We obtained a total sample of 2598 subjects throughout three waves of data collection.

In the first wave, we recruited participants by contacting teachers, academic coordinators and psycho-pedagogical teams of secondary schools in Andalusia and Madrid. We also directly reached our target group using WhatsApp messages addressed to their parents with a link to the survey. All participants above 14 years old signed a consent form. Students under 14 years old were asked to provide an informed consent from their parents to participate in the experiment. We obtained a sample of 1075 adolescents. Due to the Covid-19 pandemic lock-down, the experiment was conducted using a self-administrated online questionnaire programmed in Lime Survey[5]. We used real incentives with half of the sample, and paid one subject out of twenty by randomly selecting them to win the money. The other half were paid with hypothetical money. Assignment to hypothetical / real payments was not random.

The recruitment process was different in the second and third waves. We signed an agreement with school directors to integrate the experiment as part of their pedagogical curriculum and to run it as a class activity. It removed the need of parental consent for subjects below 14 years old and also made the experiment more scalable. This allowed us to reach 564 adolescents in the second wave and 959 in the third wave. Students read the instructions and navigated throughout the questionnaire in different screens to complete the survey. As we mentioned above, the experiment was conducted in a platform called SAND, which permitted greater control over data privacy. All responses were anonymous and subjects were paid with hypothetical payoffs, since schools would not accept to run experiments with real money.

Table 1 summarizes the recruitment process and the progressive dropping out of participants.

Table 1: Attrition of subjects

	First Wave		Second Wave		Third Wave	
Agreement with schools	No		Yes		Yes	
Class activity	No		No		Yes	
	N	Ratio	N	Ratio	N	Ratio
Potential sample	1075	100.00%	564	100.00%	959	100.00%
No parental consent	135	12.56%	-	-	-	-
First Screen drop-out	118	10.98%	53	9.40%	0	0%
Not age of interest	34	3.16%	7	1.24%	9	0.94%
Time preferences drop-out	90	8.37%	2	0.35%	6	0.63%
CRT & Fin drop-out	13	1.21%	20	3.55%	5	0.52%
Risk preferences drop-out	6	0.56%	6	1.06%	1	0.10%
Others tasks drop-out	26	2.42%	5	0.89%	9	0.94%
Survey	83	7.73%	36	6.38%	-	-
Gender: "Does not want to answer"	-	-	-	-	20	2.09%
Gender: "Other"	-	-	-	-	7	0.73%
Final Sample - Response rate	570	53.02%	435	77.13%	902	94.06%

[4]See at: https://aspredicted.org/blind.php?x=af3rw7

[5]Several papers showed that online subjects display the same behavior in economic games as traditional participants in lab experiments. For example, Prissé and Jorrat (2022) showed that online environment (vs Lab) does not influence response time, consistency and answers of subjects in time and risk preferences. Additionally, Jorrat (2021) obtained substantial cooperation in the Prisoner's Dilemma with online subjects, and also increased cooperation when priming it.

In the first wave, we observed that 10.98% participants decided to abandon immediately the survey at the first screen. We also saw that asking subjects under 14 years old to provide a parental consent reduced the sample by 12.56%. We additionally lost 3.16% of the sample for not being in the ages of interest. In the second wave, we had 9.40% of subjects giving up at first screen. Moreover, in the third wave all subjects started the experiment and we only lost 0.94% subjects because they were not in the age of interest.

Concerning attrition during the experimental session, in the first wave subjects dropped out of the experiment by 10.98% in the survey[6], 8.37% in the time preferences task, 1.21% in the CRT task and financial skills questions, 0.56% in the risk preferences task and 2.42% in other tasks. We observed that attrition was smaller during the second wave, since we lost 0.35% of the subjects during the time preferences task, 3.55% during CRT and financial skills questions, 1.06% during the risk preferences task, and 0.89% during others tasks. In the third wave, attrition was still reduced: 0.63% of participants dropped out during the time preferences task, 0.52% during the CRT task and financial skills questions, 0.10% in the risk preferences task and 0.94% in the subsequent tasks. Finally, we also lost participants during the final control questions. In the first wave, we lost 7.73% in the participant's characterization section (see "Survey" in Table 1). In the second wave, we lost 6.38% of the subjects in this part and 1.24% of them were not in the age of interest. In the third wave, we lost 2.09% of the subjects who refused to disclose their gender and we lost 0.73% of the subjects who considered their gender as "Other". We therefore removed 2.82% of the subjects from the sample since we did not know their gender despite having their answers.

In conclusion, 53.02% of all subjects completed the experiment in the first wave, 77.13% in the second wave and 94.06% in the third wave. Henceforth, we can conclude that signing an agreement with schools to run the experiment dramatically reduced attrition, and running the experiment as a class activity further reduced attrition by removing drop-outs at the first screen.

We finally reached 1907 individuals who completed the entire experiment, complied with age criteria and reported male or female gender. We compensated the large attrition in the first wave by including in the analysis each task completed by subjects even if they later dropped out, which was possible because we had their control variables from the initial survey. Therefore, we could retrieved 13 individuals from the time preferences task who dropped-out in the CRT and financial questions, and 6 individuals from the time preferences task, CRT and financial questions who dropped-out in the risk preferences task. We did not retrieve individuals in the second and third waves because we asked for control variables in the final questions, meaning that we could only include in the analysis subjects who entirely completed the experiment. Thus, we had 1926 observations for the time preferences task and 1913 observations for the cognitive abilities questionnaires[7]. Table 2 summarizes the socioeconomic characteristics of the final sample participants.

[6] In the first wave, participant's characterization was introduced before the tasks, while in the second and third waves it was introduced after the tasks.

[7] One subject finished the CRT, but not the financial questions. We therefore have $n = 1914$ observations in the CRT task.

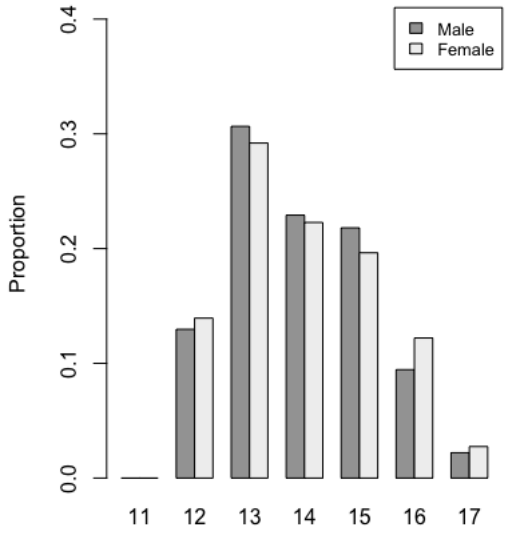
Table 2: Socioeconomic characteristics of subjects

Variable	n	Mean	Median	Min	Max
Female	1926	0.495	0	0	1
Age	1926	13.92	14	12	17
Repeater	1926	0.217	0	0	1
Grade 1	1926	0.242	0	0	1
Grade 2	1926	0.315	0	0	1
Grade 3	1926	0.233	0	0	1
Grade 4	1926	0.210	0	0	1
School: Public	1926	0.523	1	0	1
School: Semi-Private	1926	0.342	0	0	1
School: Private	1926	0.134	0	0	1
Province: Cadiz	1926	0.510	1	0	1
Province: Cordoba	1926	0.196	0	0	1
Province: Madrid	1926	0.227	0	0	1
Province: Malaga	1926	0.004	0	0	1
Province: Seville	1926	0.056	0	0	1
Province: Other	1926	0.007	0	0	1

The mean age was 13.92 years old (13 years and 11 months), 49.5% were female and 21.7% were repeaters. Regarding school characteristics, we found a fair distribution of the sample with 52.3% of participants from public schools, 34.2% from semi-private (not elite) schools and 13.4% from private schools. Regarding the geographic location of individuals, 22.7% were recruited in Madrid, 5.6% in Seville, 19.6% in Cordoba, 51% in Cadiz and the remaining 0.01% in Malaga or others provinces.

Figure 1 explores the composition of the sample by gender. We can observe that the distribution is similar across ages and gender and a two-sided Kolmogorov-Smirnov test does not reject the equality of distribution ($D = 0.032$, $p = 0.720$).

Figure 1: Distribution of participants by age and gender



3 Experimental tasks

Participants always completed the tasks in the same order. They first answered the Multiple Price List (MPL) task of Prissé (2022) to measure time preferences. We used five different treatments for this task. Subjects then completed the Cognitive Reflection Test (CRT) of Thomson and Oppenheimer (2016) to measure cognitive abilities and a financial abilities (Fin)

test adapted to teenagers. Then participants answered the risk preferences task of Holt and Laury (2002). Additionally, participants also answered in the third wave the Delavande test (DL) of Estepa et al. (2021) inspired by Delavande and Kohler (2009) to measure teenagers' ability to understand probabilities[8]. Each task is detailed hereafter.

Time preferences

The time preferences task included six decisions in all the three waves. We ran five treatments. In the last two options, we introduced two different visual elements. Figures 2a, 2b and 2c provide some examples.

1. MPL: subjects were asked to introduce either 0 euros or 10 euros using a piggy bank (n_{MPL}= 188). The total amount of money to introduce was 10 euros.
2. MPL-Cont: subjects were asked to introduce 0 euros, 1 euro, 2 euros, ..., 10 euros using a piggy bank (n_{VCTP}= 255). Subjects might choose any number between 0, 1... and 10 euros. The total amount of money to distribute was 10 euros.
3. MPL-Video: text instructions were replaced by identical video instructions describing the task (n_{Video}= 146).
4. MPL-Choice: subjects clicked on a button to indicate in which piggy bank they wanted to allocate the ten coins. (n_{Choice}=435).
5. MPL-Gift: subjects clicked on a button to indicate which gift of indicated monetary value they wanted to choose. (n_{Gift}=902).

During the first wave, individuals were randomly assigned to MPL or MPL-Cont, except students of one school who were assigned to the MPL-Video treatment. In each decision of these three treatments, participants were given ten coins of 1 euro each and were asked to allocate (typing the number) them between two options presented as piggy banks: if they decided to introduce the money into the piggy bank on the left, they could get the money at the early date of tomorrow; if they decided to introduce the money into the piggy bank on the right, they could receive the money at a later date of one week. The amount of money corresponding to the 10 coins in the early date (left) is always 1 euro, and the corresponding amount of money in the later date (right) increases from decision to decision: 1 euro, 1.2 euros, 1.4 euros 1.6 euros, 1.8 euros and 2 euros. Figure 2a provides an example of the decision screen for the first decision.

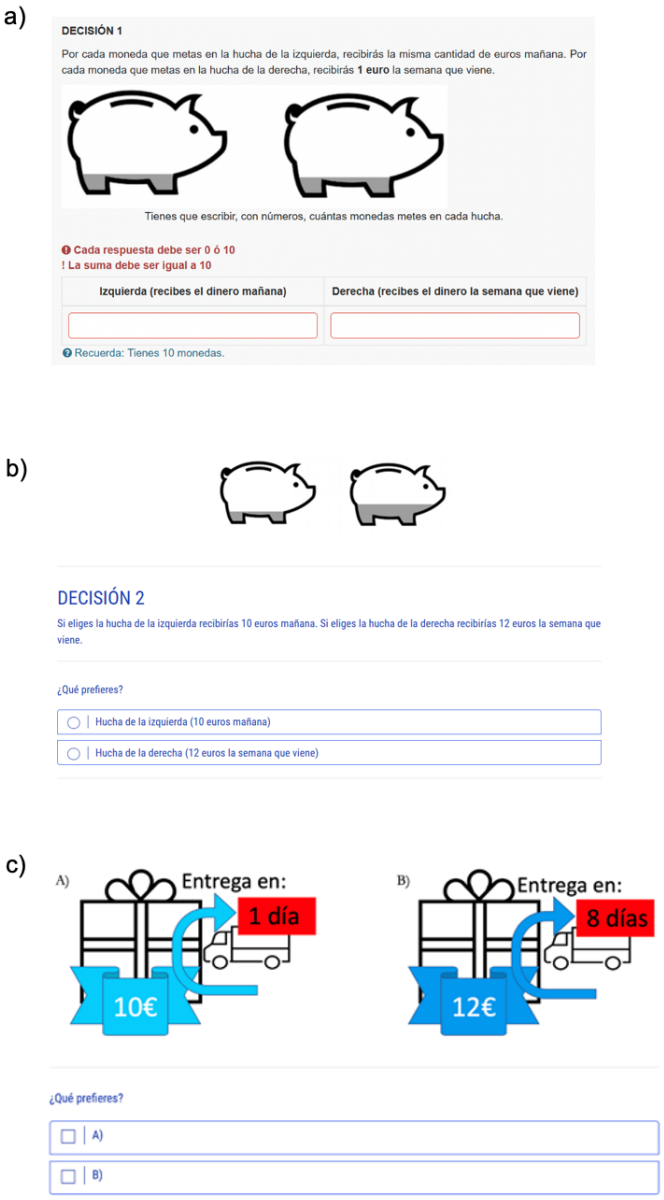
For all MPL treatments, the piggy bank on the right allowed subjects to visualize the increase in the interest rate, as shown in Appendix in Figure A1. Below the piggy banks, a warning message reminded that "**!** The sum must be equal to 10" (MPL-Cont), or "**!** Each reply must be 0 or 10" (MPL).

During the second wave, subjects answered the MPL-Choice treatment. Figure 2b provides an example of a decision screen in the MPL-Choice treatment. Individuals should click to choose between options. Subjects chose their desired allocation by clicking through the options rather than typing their answers.

During the third wave, participants answered the MPL-Gift treatment. Figure 2c provides an example of the decision screen in the MPL-Gift treatment. Monetary amounts were represented by a gift with a blue ribbon indicating their value. Figure A2 shows the ribbon darkening proportionally to the increase in the interest rate. The delay before receiving the payment is explained using a delivery van rather than text explanations.

[8]We analyze data of the Delavande test in Brañas-Garza et al. (2022b).

Figure 2: Time preferences: screens decisions by treatments

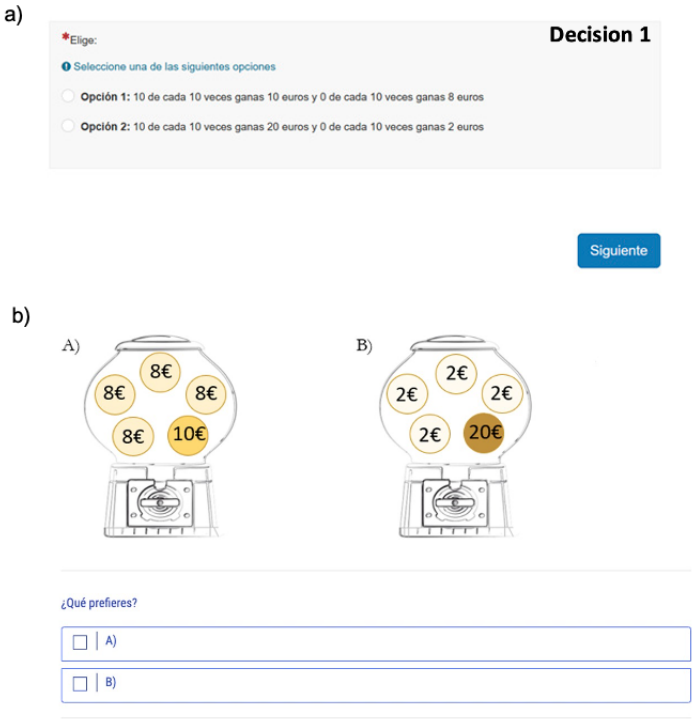


Risk preferences

Risk preferences were measured by a modified version of the Holt-Laury task. In the first two waves, participants were asked to make eleven decisions between two paired lotteries where p_h was the probability to obtain the highest payoff, and p_l was the probability to obtain the lowest payoff. The first decision was taken with probabilities p_h=0.0 and p_l=1. Then p_h increased by 0.1 in each following decision. Lottery A is initially better than Lottery B, until p_h becomes sufficiently high and it reverses. Because Lottery A is less risky than Lottery B, participants might continue picking Lottery A.

The trial at which they switch to Lottery B gives an interval of estimated values for their risk-aversion parameter. Because inconsistency in the Holt-Laury task is usually high, we expected teenagers to face serious problems in this task. We therefore added the (p_h=0, p_l=1) trial that is not present in the standard Holt-Laury task, which starts at (p_h=0.1, p_l=0.9) to get an additional measurement testing the consistency of subjects. Figure 3a displays an example of the screen for the first decision.

Figure 3: Risk preferences: decision screens by treatments



Due to the poor performance (huge inconsistency) in waves 1 and 2, we modified the visualization of the risk preferences task in the third wave. We reduced the number of trials to six, with p_h=0 in the first trial and then increasing by 0.2 in each of the subsequent decisions until it equals 1.

Additionally, we introduced a visual support using a gumball machine to help subjects understand the concept of probabilities. Figure 3b displays an example of a screen for the second decision. The safe lottery is represented on the left and the risky lottery on the right. The low outcome is lightly colored and the high outcome is darkly colored. Figure A3 and Figure A4 respectively show the safe and the risky lotteries and how gumballs represent the increase in the probability of the highest outcome.

We had two treatments in the risk preferences task:

1. HL: subjects chose eleven times between a safe and a risky lottery with a high outcome probability of p_h increasing from 0 to 1 by 0.1 increments.
2. Gumball: lotteries were replaced by gumball machines. Subjects chose six times with p_h increasing from 0 to 1 by 0.2 increments.

Cognitive Reflection Test and Financial abilities

We used two complementary tasks to study cognitive abilities: the Cognitive Reflection Test (CRT) adapted to teenagers, and a financial numeracy test (Fin) composed of three mathematical questions related to basic operations and interest rates.

The questions were as follows:

CRT1: If you're running a race and you pass the person running in second place, what place are you in? (reflective: second; intuitive: first).

CRT2: Emilia's father has three daughters. The first two are named April and May. What is the name of the third daughter? (reflective: Emilia; intuitive: June)

CRT3: A farmer has 15 sheep and all but 8 died. How many are left? (reflective: eight; intuitive: seven).

CRT3(n): In a library, the number of books doubles every month. If the library takes 48 months to fill, how long would it take to fill it halfway? Indicate with a number. (reflective: 47 months; intuitive: 24 months)

Fin1: If there are 5 people who hold the winning ticket of a lottery and the price to share is two million euros, how much money would each person receive? (correct: 400000).

Fin2: Imagine that you have 100 euros in a savings account and the annual savings interest rate is 2%. If you maintain the money on the account for 5 years, how much money will you have at the end of the 5 years? (correct: More than 102).

Fin3: Imagine that you have 100 euros in a saving account. The account has an annual interest rate of 10%. How much money will you have in the account after two years? (correct: 121)

In the first and second waves, we used the questions of Thomson & Oppenheimer (2016). In the third wave, we replaced the CRT3 question by the CRT3(n) question adapted from Frederick (2005), because some subjects counted the dead sheep rather than the alive ones, making the answers ambiguous. Another reason was that some participants had a very high score in this question (80.96%) and we wanted to increase the variability of answers.

CRT and Financial skills questions were displayed in two different screens. CRT questions were presented in random order, and then subjects answered Fin questions in random order. In addition, subjects were given 3 euros for each correct answer in the first wave, but in the second and third waves there was no reward for participants.

4 Research questions

We have 5 main questions in this research:

- Q1: How can we succeed in recruiting teenagers for experiments?
- Q2: Do hypothetical payments provide similar results as real payments?
- Q3: Do different visualization modes yield similar results in experiments?
- Q4: Does the use of different electronic devices provide similar outcomes?
- Q5: Does the presence of experimenters bias the results?

We will answer questions Q1, Q2 and Q3 using data from waves 1, 2 and 3. To respond the last 2 questions (Q4 and Q5) we use data from wave 3 only, since the comparison here is cleaner.

We used a propensity score matching methodology to answer Q2, and estimated a linear regression model to answer questions 3 to 5. For the empirical analysis of these 4 questions (Q2 to Q5), we used the consistency in time and risk preferences (*ConsTime* and *ConsRisk*, respectively) and the number of future and risky choices (*NumFut* and *NumRisk*) as dependent variables. We also used the number of reflexive options (*NumCRT*) in the CRT task and the number of correct answers in the financial skills questionnaire (*NumFin*) as outcome variables.

5 Results

5.1 How can we succeed in recruiting teenagers for experiments? (Q1)

The main characteristics of each wave were as follows (see also Table 1:

- Wave 1 was conducted online and it was not part of any school activity.
- Wave 2 was conducted within the school premises but it was not included as a school activity.
- Wave 3 was also conducted at the school and it was scheduled as an internal activity.

Table 3 shows that the recruitment process matters. The percentage of subjects who completed the experiment varies dramatically in each wave (53.02%, 77.13%, 94.06%, respectively).

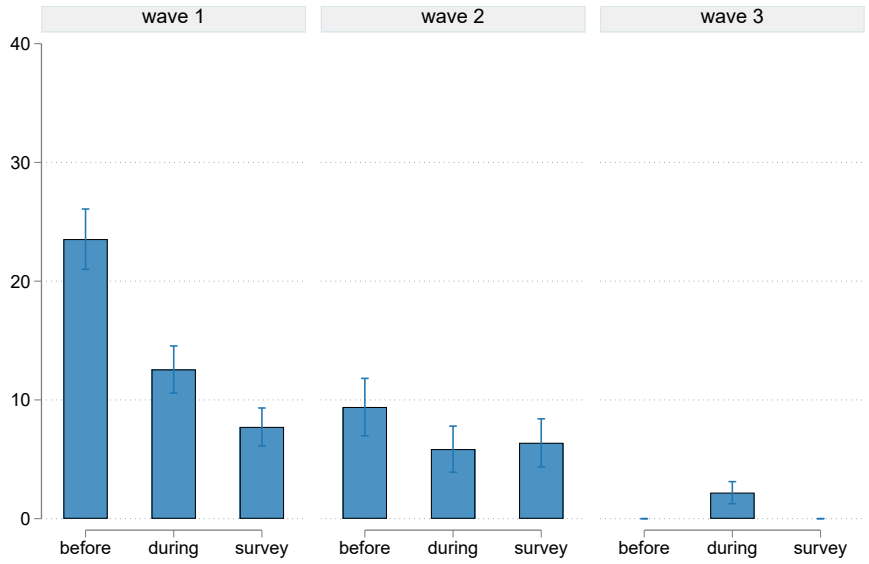
Table 3: Dropout of subjects in the different waves

Drop-out	First Wave	Second Wave	Third Wave
Before (1st-screen+parental consent)	23.54%	9.40%	0%
During	12.56%	5.85%	2.19%
Survey	7.73%	6.38 %	0%
Response rate	53.02%	77.13%	94.06%

Note: Response rate is equal to the final sample. Besides, we are not considering subjects whose age is not in the interval of interest. Moreover, we do not include subjects who did not answer gender or who answered "Other", because this is not considered attrition.

Figure 4 explores how attrition evolves *before* and *during* the experiment across waves.

Figure 4: Attrition before and during the experiment



Before includes subjects who did not start the survey for any reason (not show-up, not present consent, etc.). *During* reflects those participants who abandoned the experiment at any point. We also add *Survey* attrition, which includes those individuals who did not fill the survey entirely. As observed, the three types of attrition decrease significantly from wave to wave, having a small attrition of 2.19% in wave 3.

From Table 3 and Figure 4 we can conclude:

Result 1: Agreements with schools to run experiments as a class activity reduces attrition enormously.

5.2 Do hypothetical payments provide same results as real payments? (Q2)

In this section, we investigate the effect of payments on outcomes. This question is critical especially with children and adolescents. Besides the obvious monetary costs, the use of monetary incentives requires special parental consent and additional requirements from ethics committees. In addition, recent papers evidenced that hypothetical and real monetary incentives yield the same results in lab and field samples (see Brañas-Garza et al. (2021) for risk preferences and Brañas-Garza et al. (2022a) for time preferences), but these studies have been conducted with adults only. Thus the impact of the payment method with teens remains open. We shed some light to this discussion by comparing Hypothetical (H) and BRIS (B) payment using a between-subject design.

It is worth mentioning that treatments were not randomly assigned across the sample. Thus

we selected a sub-sample of adolescents from waves 1 and 2 that were similar in terms of school characteristics, age and gender. We applied a propensity score matching (PSM) methodology following Abadie and Imbens (2006, 2016) to generate a control group similar to the group of subjects who made the experiment with real payoffs. The resulting sample was made of 782 adolescents (n_B=390 and n_H=392).

Figure 5 presents the nearest-neighbor matching point estimates of the *H vs B* difference for consistency and for subjects' choices in time and risk preferences tasks. We display the 95% confidence interval estimated with robust Abadie-Imbens standard errors and control for age, gender and school type (semi-private or private).

As can be observed, hypothetical payment marginally increases consistency in the time preferences task by 7% ($p = 0.089$) and significantly increases consistency in the risk preferences task by 11% ($p = 0.006$) compared to the BRIS group. Overall, we found no effects of hypothetical payments on the number of future allocations ($p = 0.691$) and risky choices ($p = 0.887$).

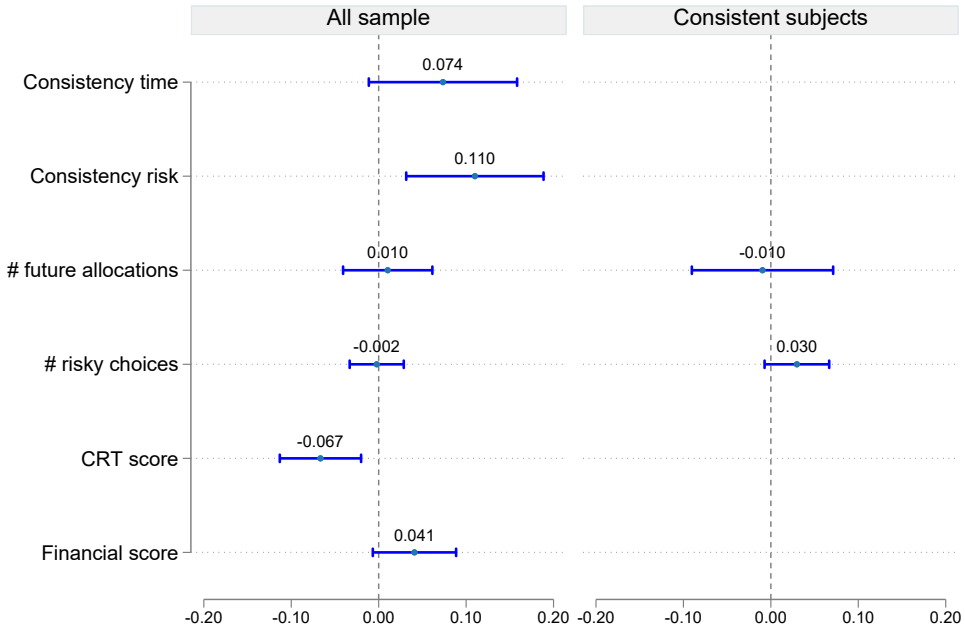
In the right-side of Figure 5 we consider only consistent subjects. We found the same results on the number of future allocations ($p = 0.816$) and risky choices ($p = 0.114$). We therefore conclude:

Result 2a: Hypothetical payments (vs monetary) have no effect on the elicitation of both time and risk preferences.

However, Figure 5 also shows that hypothetical payments might have an impact on the elicitation of cognitive abilities: Indeed, we observe that hypothetical payments reduce CRT score[9] ($p = 0.005$) and marginally increase financial abilities score ($p = 0.092$).

Result 2b: Hypothetical payments (vs monetary) impact negatively on CRT performance.

Figure 5: Effect of hypothetical payments (vs BRIS) on outcomes



Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. Control group is BRIS. *Consistency Time* and *Consistency Risk* outcomes are equal to 1 for subjects who did not do multiple-switching in time and risk tasks, and 0 otherwise. *# future allocations*, *# risky choices*, *CRT score* and *Financial score* are normalized using the min-max standardization process.

[9]This result is quite unexpected: the meta-analysis of CRT by Brañas-Garza et al. (2019) shows that incentives do not matter. However a recent review of the same data shows some effects Yechiam and Zeif (2022).

Besides the obvious implications for running experiments with teenagers - lower costs, smaller frictions with IRB, disclosure of private information, etc. - Result 2 contributes to the emerging literature that discusses whether monetary incentives are really necessary.

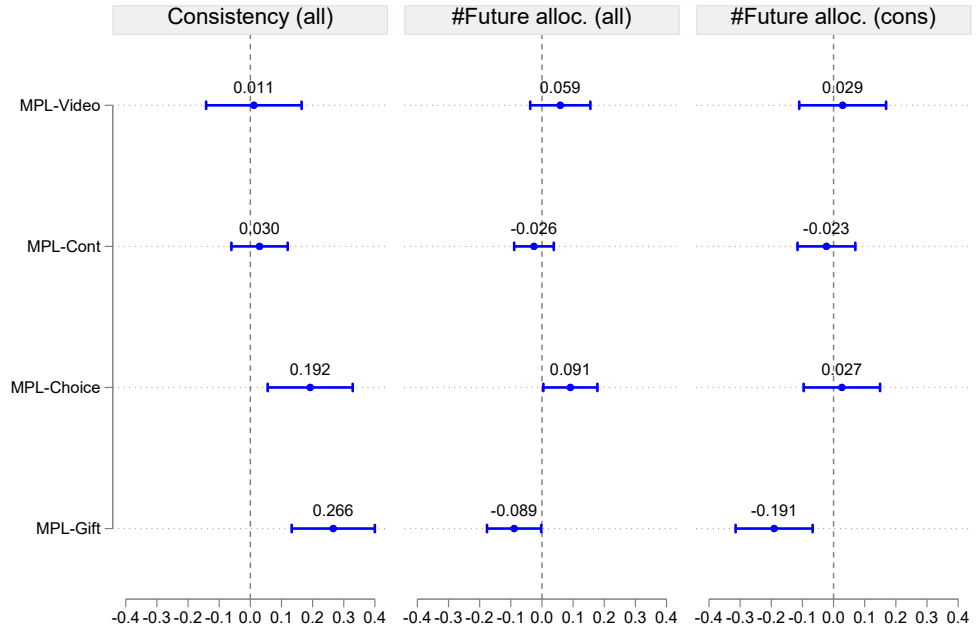
5.3 Do different visualization modes yield similar results in experiments? (Q3)

In this section, we compare whether the use of different visual elements affects time and risk preferences elicitation. We regressed the outcome variables (*consistency* and $\#\ future\ allocations$) on different dummy variables that represented the different visualizations (treatments) used in the tasks. We controlled for gender, age and school fixed effects.

Time preferences

Figure 6 displays a summary of results shown in Table A1. It illustrates the effect of the different visual elements on the performance in time preferences task, with the MPL condition as the reference category. We can observe that consistency in the task increased in $MPL - Choice$ ($p = 0.006$) and $MPL - Gift$ ($p < 0.001$). Additionally, we see that $MPL - Choice$ increased the number of future allocations ($p = 0.039$) but this effect disappeared for consistent subjects ($p = 0.671$). We found different results for $MPL - Gift$, since the number of future allocations decreased for all subjects ($p = 0.044$) and consistent subjects ($p = 0.003$). Hence, $MPL - Gift$ made subjects less patient.

Figure 6: Effect of visual elements in the time preferences task



Note: Control group is MPL. Consistency is equal to 1 for subjects that do not commit multiple-switching in the time task, and 0 otherwise. #future alloc. is equal to the number of future allocations normalized using the min-max standardization process. (all) reflect the entire sample while (cons) means consistent subjects only.

These results suggest that visual elements influence the performance in the task. We observed that $MPL - Video$ and $MPL - Cont$ did not affect results, while $MPL - Choice$ increased the consistency. We obtained the best results with $MPL - Gift$,

which largely increased the level of consistency. But it also decreased the number of future allocations for all subjects and consistent ones. Since a gift might represent a temptation hard to resist for teenagers, this result might suggest that time preferences of teenagers are sensible to the illustrations used.

Result 3a: The use of visual elements increases consistency but makes subjects less patient.

Risk preferences

Figure 7 illustrates the results of OLS estimations about the effect of using a figure of a *Gumball* machine. We ran these regressions without controlling for clusters since we detected multicollinearity between clusters and *Gumball*.

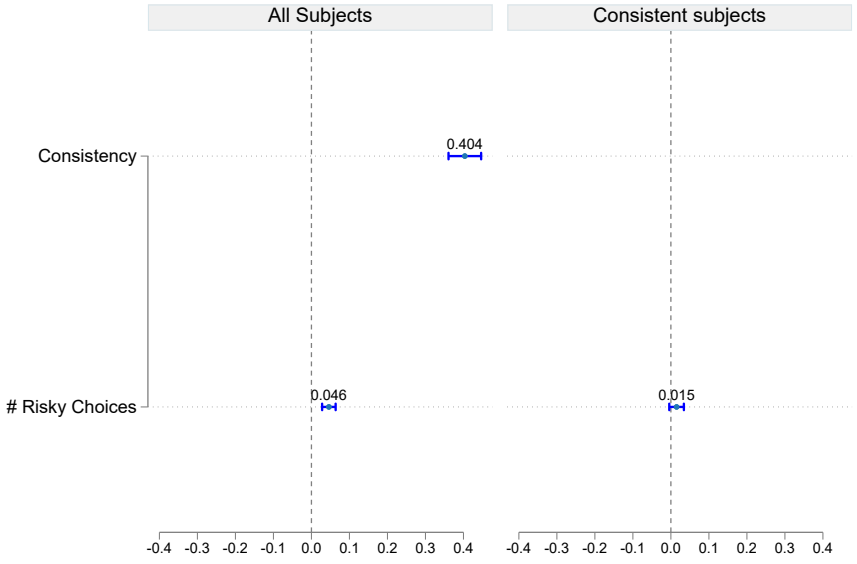
We also observed that using the gumball increased the consistency notably by an estimated 40.4% ($p < 0.001$). These results also suggest that it slightly increased the number of risky choices for all subjects by 0.05 ($p < 0.001$), but this effect disappeared when we consider only consistent subjects ($p = 0.127$). Table A2 of the Appendix shows the regression results.

The visual elements largely increased consistency in the risk preferences task. It could also have affected the elicited preferences, but the different number of trials in the HL and Gumball treatments do not allow to conclude whether it is visualization what influenced elicited preferences.

We can conclude:

Result 3b: The use of visual elements increases consistency and has no impact on the number of risky choices.

Figure 7: Effect of visual elements on the risk preferences task



Note: Control group is HL. Consistency is equal to 1 for subjects that do not commit multiple-switching in the time task, and 0 otherwise. *#Risky choices* is equal to the number of decisions where the risk lottery was choose for each subject, normalized using the min-max standardization process. (all) reflect the entire sample while (cons) means consistent subjects only.

5.4 Does the use of different electronic devices provide similar outcomes in experimental tasks? (Q4)

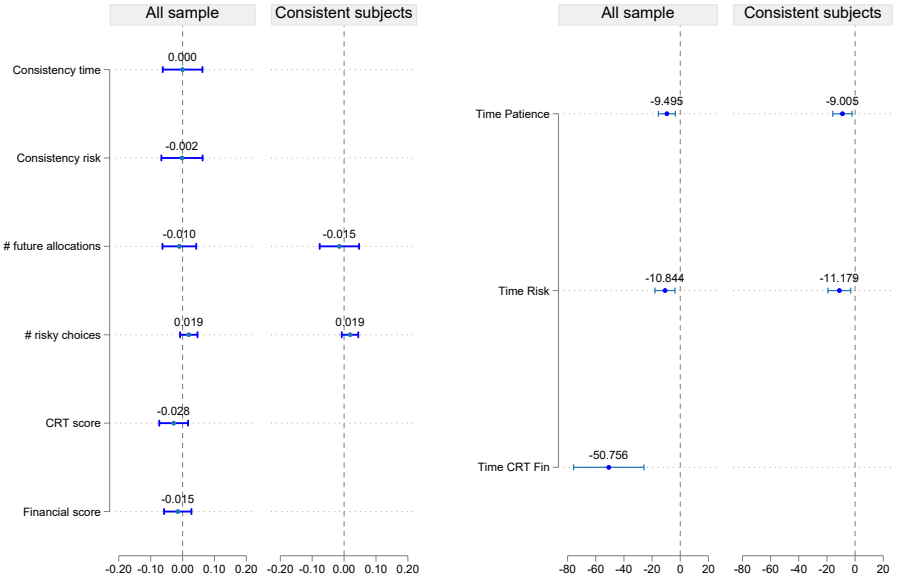
The goal of this section is to investigate whether answering on a mobile phone versus answering on a more standard experimental device, like a computer or a tablet, has any effect on the performance in experimental tasks. Mograbi (2022) studied this question in the lab and found no significant differences when using different devices in the case of risky choices, but there was significantly more present bias when using a smartphone than using a computer.

We only used data from the third wave because the experiment was on-site for all the subjects (not on-line). In this case, third wave, we could also study response time because all subjects answered on the same device and platform.

We obtained a sample of 898[10] individuals: 22.16% of them answered the questionnaire on a mobile phone rather than on the more standard computer (21.94%) or an electronic tablet (55.90%) provided by experimenters[11]. The results display the comparison between subjects answering on a mobile phone (treatment) and subjects answering on other devices (control), since this visualization is the most familiar to participants. We conducted a regression of the different outcome variables on the treatment variable, controlling for age, gender and school fixed effects.

Figure 8 summarizes the results of Table A3 presenting the linear regressions on the effect of *Mobile* phone on the outcome variables in the time preferences task, the risk preferences task and the abilities questionnaires. We observed that *Mobile* decreased response time in time and risk preferences tasks for all sample ($p = 0.002$ and $p = 0.003$) and consistent subjects ($p = 0.010$ and $p = 0.007$).

Figure 8: Effects of mobile phone on answers and response time



Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. Control group is Computer and Tablets platforms. *Consistency Time* and *Consistency Risk* outcomes are equal to 1 for subjects that did not do multiple-switching in time and risk tasks, and 0 otherwise. $\#future\ allocations$, $\#risky\ choices$, $CRT\ score$ and *Financial score* are normalized using the min-max standardization process. Time measure is expressed in seconds.

[10] Missing 4 outliers in time preferences and 3 in risk preferences
[11] 4 (0.44%) subjects answered on a different device

Additionally, the use of mobile phones also decreased response time in the cognitive and financial abilities questionnaires ($p = 0.000$). We found no other effect of using a mobile phone on the answers.

We conclude in this section:

Result 4: The use of mobile phones does not impact experimental outcomes and reduces response time.

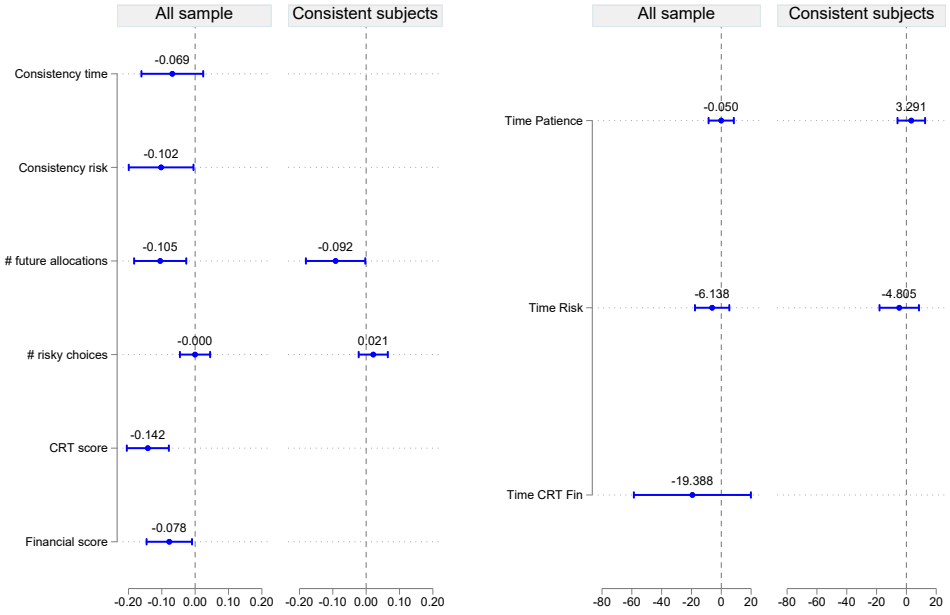
5.5 Does the presence of experimenters bias the results? (Q5)

In this final section, we explore the potential effect of the presence of experimenters in the experimental session. As in the previous section, we only used data from the third wave. Experimenters were always present to help subjects, but the experimenters were not always the same. In 28.49% of the cases, teachers administered the questionnaire. In the remaining 71.51% of the cases the questionnaire was administered by university staff[12].

We investigated whether administering the questionnaire by university staff influenced results. The reference group was conformed by subjects who filled the questionnaire in the presence of their teachers. As before, we conducted a regression of each outcome variable on the treatment variable ($Univstaff$), controlling for age, gender and school fixed effect.

Figure 9 summarizes the results of Table A4 displaying the estimation of the effect of $UnivStaff$ on the outcome variables.

Figure 9: Effects of university staff on answers and response time



Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. Control group is the one directed by the teachers. *Consistency Time* and *Consistency Risk* outcomes are equal to 1 for subjects that did not do multiple-switching in time and risk tasks, and 0 otherwise. *# future allocations*, *# risky choices*, *CRT score* and *Financial score* are normalized using the min-max standardization process. Time variables are defined in seconds.

[12] 13.08% of individuals did the experiment administered by two authors of this paper (Alfonso and Montero) and 58.43% by research assistants. Additional analysis not presented in this paper shows that both management type similarly influenced results.

We can observe that the presence of university staff reduced consistency in risk preferences task ($p = 0.039$), while it made subjects less patient in time preferences task for both samples ($p = 0.009$ for all subjects and $p = 0.044$ for consistent subjects). Regarding the different outcomes on response time, the presence of university staff had no effect.

Additionally, we observed a significant reduction of CRT and Financial score ($p < 0.001$ and $p = 0.026$). This significant decrease might suggest that teachers sometimes helped teenagers, while university staff tried to ensure the veracity of responses.

We conclude that:

Result 5: The presence of university staff impacts teenagers performance: decreases consistency (risk), CRT and Financial scores and increases impatience.

6 Discussion

In this paper, we analyze the potential effects of aspects related to the recruitment process and to the task design of traditional experimental tasks performed by teenagers. As suggested in the introduction, capturing better and quality information about children and adolescents competences and skills might help educators and educational professionals to identify current and prevent future problems, miss-behaviors or undesirable situations, such as worse health, addictions, lower academic levels or worse professional decisions.

First, we studied how different recruitment processes affected the quality of data in terms of response rate and attrition. Second, we analyzed the effect of hypothetical incentives (vs BRIS payment method) on the responses. Third, we investigated the effect of the introduction of visual elements in the tasks. Fourth, we analyzed whether responding using different devices such as mobile phones, tablets or computers influenced the results. Fifth, we studied whether administering experiments by school teachers or university staff had an impact on participants' responses.

Our findings suggest that: i) Signing agreements with schools to run experiments as a class regular activity reduces attrition enormously and improves the response rate.

ii) Hypothetical payments elicit similar time and risk preferences than BRIS payments. Besides, we did not find an adverse effect of hypothetical incentives on consistency. This conclusion is in line with the evidence found in lab and field experiments with adult population in Brañas-Garza et al. (2021), Brañas-Garza et al. (2022a). Overall, we showed that the use of real incentives on experiments with teenagers is not needed to collect better data. However, we found that measuring CRT with real incentives increases the number of reflective options, as it is suggested by Yechiam and Zeif (2022).

iii) Using visual elements in experimental tasks is relevant. We found that it reduced the response time and increased the consistency in time and risk preferences tasks, in line with previous literature (see Harrison and Rutström (2008) for risk elicitation). This finding suggests that individuals understand the tasks better when they can see a familiar object. These results also suggest that visual elements could influence the elicited preferences since using a gift increased the number of early choices whereas a gumball machine increased the number of risky choices. Future work should further investigate visual support elements, particularly those that are represented with tempting goods as it was in our experimental design.

iv) In sharp contrast to Mograbi (2022), our results show that the use of mobile phones does not have an impact on outcomes. Even better, it makes subjects complete the entire questionnaire faster. This result is of particular relevance because allowing students to use their own mobile phones would largely simplify the logistic.

v) Letting teachers administer the experiment (instead of university staff) might be relevant. Their presence increases the number of future allocations, the number of correct answers in the CRT and Financial questionnaires. A natural question is whether teachers "helped" students during the experiments. However, this question should be further investigated using a proper randomized design where the presence of the teacher is the treatment. In this line, it would also be relevant to study the effect of the neutrality of experimenters.

Although we were not able to randomize all possible treatments concerning the recruitment process and the experimental design, we consider that our large scale experiment provides valuable insights on how to conduct experiments with teenagers. We can summarize our findings in three advises for researchers.

The first advice is to run experiments as an activity which is part of the regular school program since it reduces attrition considerably.

The second advice is that hypothetical payments are reliable (or at least give similar results to real payments) and subjects can respond to the experiment on their mobile phones. This recommendation would reduce logistics costs.

The final advice is to present questions where individuals must choose between options and to include visual elements. The former increases the quality of answers, while the latter increases the quality of the results by improving the quality of choices in the time preferences task and reducing all types of inconsistencies in the risk preferences task.

References

Abadie, A., & Imbens, G. W. (2006). Large sample properties of matching estimators for average treatment effects. *Econometrica*, *74*(1), 235–267.

Abadie, A., & Imbens, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, *84*(2), 781–807.

Alan, S., & Ertac, S. (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy*, *126*(5), 1865–1911.

Andreoni, J., Di Girolamo, A., List, J. A., Mackevicius, C., & Samek, A. (2020). Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. *Journal of Economic Behavior and Organization*, *179*, 729–742.

Andreoni, J., Kuhn, M. A., List, J. A., Samek, A., Sokal, K., & Sprenger, C. (2019). Toward an understanding of the development of time preferences: Evidence from field experiments. *Journal of Public Economics*, *177*, 104039.

Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *The American Economic Review*, *102*(7), 3357–76.

Brañas-Garza, P., Estepa, L., Jorrat, D., Orozco-Olvera, V., & Rascon Ramirez, E. G. (2021). To pay or not to pay: Measuring risk preferences in lab and field. *Judgment and Decision Making*, *16*(5), 1290–1313.

Brañas-Garza, P., Jorrat, D., Espín, A. M., & Sánchez, A. (2022a). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *forthcoming*.

Brañas-Garza, P., Jorrat, D., Prissé, B., & Vázquez-De Francisco, M. J. (2022b). Strong stability in preferences and tiny gender differences among teenagers: Evidence from secondary schools. *Mimeo*.

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, *82*, 101455.

Bruhn, M., de Souza Leão, L., Legovini, A., Marchetti, R., & Zia, B. (2013). The impact of high school financial education: Experimental evidence from brazil. *World Bank Policy Research Working Paper*, (6723).

Castillo, M., Ferraro, P. J., Jordan, J. L., & Petrie, R. (2011). The today and tomorrow of kids: Time preferences and educational outcomes of children. *Journal of Public Economics*, *95*(11-12), 1377–1385.

Castillo, M., Jordan, J. L., & Petrie, R. (2018). Children's rationality, risk attitudes and field behavior. *European Economic Review*, *102*, 62–81.

Chabris, C. F., Laibson, D., Morris, C. L., Schuldt, J. P., & Taubinsky, D. (2008). Individual laboratory-measured discount rates predict field behavior. *Journal of Risk and Uncertainty*, *37*(2), 237–269.

Delavande, A., & Kohler, H.-P. (2009). Subjective expectations in the context of hiv/aids in malawi. *Demographic Research*, *20*, 817–874.

Dixon, M. R., Marley, J., & Jacobs, E. A. (2003). Delay discounting by pathological gamblers. *Journal of Applied Behavior Analysis*, *36*(4), 449–458.

Eckel, C. C., Grossman, P. J., Johnson, C. A., de Oliveira, A., Rojas, C., & Wilson, R. K. (2012). School environment and risk preferences: Experimental evidence. *Journal of Risk and Uncertainty*, *45*(3), 265–292.

Estepa, L., Jorrat, D., Orozco-Olvera, V., & Rascon Ramirez, E. G. (2021). Beans vs slider: Eliciting probabilities in the field. *Mimeo*.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*(4), 25–42.

Harrison, G. W., & Rutström, E. E. (2008). Risk aversion in the laboratory. *Research in Experimental Economics*, *12*(8), 41–196.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *The American Economic Review*, *92*(5), 1644–1655.

Horn, D., Kiss, H. J., & Lénárd, T. (2022). Gender differences in preferences of adolescents: Evidence from a large-scale classroom experiment. *Journal of Economic Behavior and Organization*, *194*, 478–522.

Jørgensen, L. K., Piovesan, M., & Willadsen, H. (2022). Gender differences in competitiveness: Friends matter. *Journal of Behavioral and Experimental Economics*, 101955.

Jorrat, D. (2021). Recruiting experimental subjects using whatsapp. *Journal of Behavioral and Experimental Economics*, *90*, 101644.

Kirby, K. N., Petry, N. M., & Bickel, W. K. (1999). Heroin addicts have higher discount rates for delayed rewards than non-drug-using controls. *Journal of Experimental Psychology: General*, *128*(1), 78.

List, J. A., Petrie, R., & Samek, A. (2022). How experiments with children inform economics. *Journal of Economic Literature*, forthcoming.

Lührmann, M., Serra-Garcia, M., & Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, *10*(3), 309–32.

Mograbi, E. (2022). Decision-makers are more impulsive on smartphones than on computers. *Journal of Behavioral and Experimental Economics*, *100*, 101916.

Prissé, B. (2022). Visual continuous time preferences: Lab, field and high schools. *Mimeo*.

Prissé, B., & Jorrat, D. (2022). Lab vs online experiments: No differences. *Journal of Behavioral and Experimental Economics*, *100*, 101910.

Reynolds, B. (2006). A review of delay-discounting research with humans: Relations to drug use and gambling. *Behavioural Pharmacology*, *17*(8), 651–667.

Samek, A., Gray, A., Datar, A., & Nicosia, N. (2021). Adolescent time and risk preferences: Measurement, determinants and field consequences. *Journal of Economic Behavior and Organization*, *184*, 460–488.

Stoklosa, M., Shuval, K., Drope, J., Tchernis, R., Pachucki, M., Yaroch, A., & Harding, M. (2018). The intergenerational transmission of obesity: The role of time preferences and self-control. *Economics and Human Biology*, *28*, 92–106.

Sutter, M., Kocher, M. G., Glätzle-Rützler, D., & Trautmann, S. T. (2013). Impatience and uncertainty: Experimental decisions predict adolescents' field behavior. *The American Economic Review*, *103*(1), 510–31.

Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, *11*(1), 99–113.

Tymula, A. (2019). An experimental study of adolescent behavior under peer observation: Adolescents are more impatient and inconsistent, not more risk-taking, when observed by peers. *Journal of Economic Behavior and Organization*, *166*, 735–750.

Yechiam, E., & Zeif, D. (2022). Revisiting the effect of incentivization on cognitive reflection: A meta-analysis. *Journal of Behavioral Decision Making*, *forthcoming*.

Appendix

Figure A1: Illustration of increasing interest rates with piggy banks

Figure A2: Illustrations of increasing interest rates with ribbons

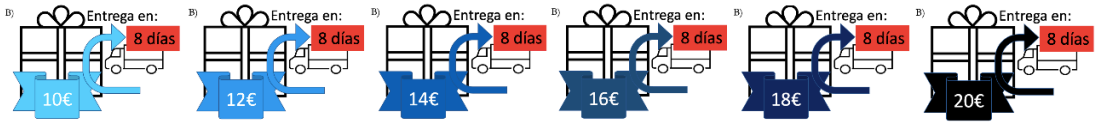


Figure A3: Lottery A used in the experiment

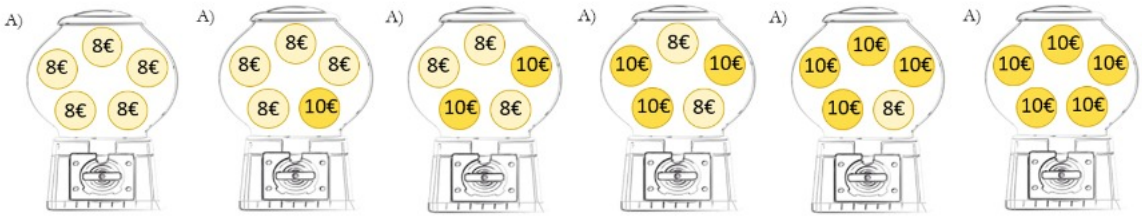
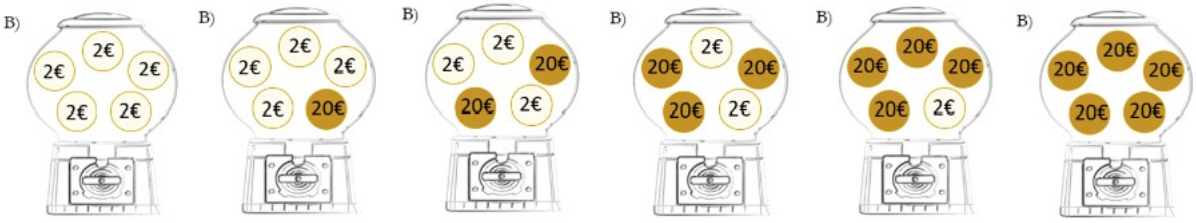


Figure A4: Lottery B used in the experiment



Note: Outliers were not considered for any of the following models.

Table A1: Regressions on the effect of interface (visual elements) on time preferences task

	(1) Consistency	(2) #Future alloc.	(3) #Future alloc. (cons)
MPL-Video	0.011	0.059	0.029
	(0.078)	(0.049)	(0.071)
	[0.885]	[0.232]	[0.681]
MPL-Cont	0.030	-0.026	-0.023
	(0.046)	(0.032)	(0.047)
	[0.521]	[0.430]	[0.628]
MPL-Choice	0.192***	0.091**	0.027
	(0.070)	(0.044)	(0.063)
	[0.006]	[0.039]	[0.671]
MPL-Gift	0.266***	-0.089**	-0.191***
	(0.068)	(0.044)	(0.063)
	[0.000]	[0.044]	[0.003]
Constant	0.249*	0.788***	0.963***
	(0.132)	(0.095)	(0.120)
	[0.059]	[0.000]	[0.000]
Observations	1,925	1,925	1,448
R-squared	0.057	0.072	0.100
Controls	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table A2: Regressions on the effect of interface on risk preferences task without controlling for clusters

	(1) Consistency	(2) # risky choices (all)	(3) # risky choices (cons)
Gumball	0.404***	0.046***	0.015
	(0.022)	(0.009)	(0.010)
	[0.000]	[0.000]	[0.127]
Constant	0.505***	0.577***	0.598***
	(0.121)	(0.052)	(0.052)
	[0.000]	[0.000]	[0.000]
Observations	1,909	1,909	1,128
R-squared	0.177	0.017	0.003
Controls	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table A3: Regressions on the effect of mobile phones on time preferences task, risk preferences task and abilities questionnaires

	(1) TimePat	(2) ConsTime	(3) NumFut	(4) TimePatC	(5) NumFutC	(6) TimeRisk	(7) ConsRisk	(8) NumRisk	(9) TimeRiskC	(10) NumRiskC	(11) TimeCRTFin	(12) NumCRT	(13) NumFin
Mobile	-9.495***	0.000	-0.010	-9.005***	-0.015	-10.844***	-0.002	0.019	-11.179***	0.019	-50.756***	-0.028	-0.015
	(3.073)	(0.032)	(0.027)	(3.480)	(0.031)	(3.623)	(0.033)	(0.014)	(4.133)	(0.013)	(12.722)	(0.023)	(0.022)
	[0.002]	[0.999]	[0.709]	[0.010]	[0.639]	[0.003]	[0.956]	[0.164]	[0.007]	[0.158]	[0.000]	[0.225]	[0.488]
Constant	69.717***	0.813***	0.791***	71.643***	0.764***	123.102***	0.724***	0.516***	112.419***	0.557***	263.059***	0.375***	-0.185*
	(14.812)	(0.136)	(0.117)	(16.929)	(0.133)	(14.971)	(0.142)	(0.060)	(16.435)	(0.059)	(61.693)	(0.101)	(0.099)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.060]
Observations	898	898	898	742	746	895	898	898	721	723	896	898	898
R-squared	0.019	0.000	0.007	0.016	0.008	0.043	0.002	0.015	0.042	0.013	0.030	0.010	0.060
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table A4: Regressions on the effect of university staff on time preferences task, risk preferences task and abilities questionnaires

	(1) TimePat	(2) ConsTime	(3) NumFut	(4) TimePatC	(5) NumFutC	(6) TimeRisk	(7) ConsRisk	(8) NumRisk	(9) TimeRiskC	(10) NumRiskC	(11) TimeCRTFin	(12) NumCRT	(13) NumFin
Univstaff	-0.050	-0.069	-0.105***	3.291	-0.092**	-6.138	-0.102**	-0.000	-4.805	0.021	-19.388	-0.142***	-0.078**
	(4.316)	(0.047)	(0.040)	(4.713)	(0.045)	(5.863)	(0.049)	(0.023)	(6.735)	(0.022)	(19.989)	(0.032)	(0.035)
	[0.991]	[0.147]	[0.009]	[0.485]	[0.044]	[0.295]	[0.039]	[0.985]	[0.476]	[0.339]	[0.332]	[0.000]	[0.026]
Constant	67.259***	1.094***	1.213***	57.994**	1.138***	150.893***	1.123***	0.494***	135.972***	0.455***	336.952***	0.953***	0.137
	(21.657)	(0.231)	(0.198)	(24.127)	(0.227)	(27.071)	(0.243)	(0.111)	(30.593)	(0.107)	(100.096)	(0.159)	(0.169)
	[0.002]	[0.000]	[0.000]	[0.016]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.001]	[0.000]	[0.417]
Observations	898	902	902	746	750	899	902	902	725	727	900	902	902
R-squared	0.013	0.002	0.014	0.013	0.013	0.048	0.006	0.021	0.046	0.019	0.027	0.032	0.068
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Additional analysis of the interface (visual elements) effect

In this section, we display additional analysis regarding the effect of interface on subjects' choices in both time and risk preferences tasks. These findings complement this paper's main results. We saw that the interface (visual elements) used influences the number of future choices made by subjects. Therefore, we further investigated how the interface in the time preferences task influences allocations of subjects by analyzing choices of consistent subjects. Regarding the risk preferences task, we saw that the interface does not influence subjects' choices. We further studied this line of analysis by documenting whether the interface of the risk preferences task influences the type of inconsistencies made.

We need new explanatory variables to perform this analysis. We used as dependent variables in the time preferences task *Present*, *Sophisticated* and *Future* describing the type of choices made by consistent subjects. These variables respectively define subjects allocating everything to earliest period, subjects switching from early to later allocations, and subjects allocating everything to later period. We also use as dependent variables in the risk preferences task *DomFirst*, *SwitchBack* and *DomLast* describing the type of inconsistencies made in this task. These variables respectively define subjects doing a dominated choice in the first decision, subjects switching from risky to safe choices, and subjects doing a dominated choice in the last decision.

Table A5 shows the OLS regression results regarding choices of consistent subjects in time the preferences task according to the interface used. We can observe that it does not affect the proportion of subjects allocating everything to present, but it does affect the proportion of sophisticated subjects and subjects allocating everything to the future. Indeed, we see that $MPL - Video$ decreased the proportion of sophisticated subjects ($p = 0.069$) and increased the proportion of subjects allocating everything to the future ($p = 0.019$). We also see that $MPL - Cont$ marginally increased the proportion of sophisticated subjects ($p = 0.081$) and that $MPL - Choice$ did the same with a relatively large magnitude ($p = 0.001$). Finally, we can see that $MPL - Gift$ largely increased sophisticated answers ($p < 0.001$) and decreased the proportion of subjects allocating everything to future ($p = 0.016$). These results suggest that in the time preferences task $MPL - Video$ decreases while $MPL - List$ and $MPL - Gift$ increase the quality of results.

Regarding the risk preferences task, Table A6 shows the results from the OLS regressions. We obtain further evidences that *Gumball* increases the quality of results in the risk preferences task. We observe that *Gumball* decreased the number of dominated choices in the first decision by 9.3% ($p < 0.001$), largely reduced by 37.7% the number of subjects switching back to the safe lottery after choosing the risky lottery ($p < 0.001$) and also reduced the number of dominated choices in the last decision by 18.5% ($p < 0.001$). These results suggest that visualization decreased all type of inconsistencies in the risk preferences task. We conclude that visualization increases the quality of data in both the time preferences and the risk preferences tasks.

Table A5: Regressions on the effect of interface (visual elements) on choices of consistent subjects in the time preferences task

	(1) Present	(2) Sophis	(3) Future
MPL-Video	-0.027	-0.123*	0.161**
	(0.043)	(0.068)	(0.068)
	[0.537]	[0.069]	[0.019]
MPL-Cont	0.007	0.071*	-0.049
	(0.033)	(0.041)	(0.040)
	[0.827]	[0.086]	[0.224]
MPL-Choice	-0.062	0.226***	0.029
	(0.040)	(0.066)	(0.059)
	[0.121]	[0.001]	[0.624]
MPL-Gift	0.057	0.345***	-0.135**
	(0.042)	(0.064)	(0.056)
	[0.175]	[0.000]	[0.016]
Constant	-0.064	-0.129	0.442***
	(0.091)	(0.148)	(0.117)
	[0.484]	[0.383]	[0.000]
Observations	1,925	1,925	1,925
R-squared	0.050	0.109	0.058
Controls	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table A6: Regressions on the effect of interface (visual elements) on errors in the risk preferences task

	(1) DomFirst	(2) SwitchBack	(3) DomLast
Gumball	-0.088***	-0.369***	-0.185***
	(0.018)	(0.022)	(0.016)
	[0.000]	[0.000]	[0.000]
Constant	0.205**	0.490***	0.120
	(0.096)	(0.120)	(0.089)
	[0.033]	[0.000]	[0.181]
Observations	1,909	1,909	1,906
Adjusted R^2	0.012	0.151	0.074
Controls	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Additional analysis of the electronic device effect

This section provides additional results regarding the effects of the electronic device used in the experiment. In the main text, we compared mobile phones with other devices. We now directly compare subjects doing the experiment on a *Computer* and subjects doing the experiment on an electronic *Tablet* with subjects answering the experiment on a mobile phone.

Table A7 shows the results. We can observe that computers marginally increased response time in the risk preferences task for all subjects ($p = 0.073$) and consistent subjects ($p = 0.062$). Regarding electronic tablets, response time increased in the time preferences task for all subjects ($p = 0.012$) and for consistent subjects ($p = 0.025$), as well as in the cognitive and financial abilities questionnaires ($p = 0.012$). Concerning answers, we only observed that *Computer* increased the number of risk choices ($p = 0.031$), but the effect was small and disappeared with consistent subjects.

Table A7: Regressions on the effect of computer and tablets on the time preferences task, risk preferences task and abilities questionnaires

	(1) TimePat	(2) ConsTime	(3) NumFut	(4) TimePatC	(5) NumFutC	(6) TimeRisk	(7) ConsRisk	(8) NumRisk	(9) TimeRiskC	(10) NumRiskC	(11) TimeCRTFin	(12) NumCRT	(13) NumFin
Computer	16.245	-0.035	-0.057	16.735	-0.070	23.999**	-0.022	0.080**	24.657*	0.043	37.339	-0.027	0.066
	(13.410)	(0.079)	(0.064)	(15.338)	(0.076)	(11.842)	(0.082)	(0.037)	(13.133)	(0.034)	(32.267)	(0.051)	(0.054)
	[0.226]	[0.660]	[0.378]	[0.276]	[0.356]	[0.043]	[0.789]	[0.029]	[0.061]	[0.200]	[0.248]	[0.597]	[0.223]
Tablet	26.261**	-0.014	0.009	27.423**	0.016	8.581	-0.015	0.012	4.798	0.011	40.670**	-0.017	0.012
	(10.413)	(0.038)	(0.033)	(12.148)	(0.038)	(5.913)	(0.039)	(0.019)	(4.906)	(0.018)	(15.970)	(0.030)	(0.026)
	[0.012]	[0.712]	[0.787]	[0.024]	[0.677]	[0.147]	[0.701]	[0.528]	[0.328]	[0.527]	[0.011]	[0.574]	[0.636]
Constant	79.138**	1.809***	1.312***	99.414**	1.096***	175.706***	1.720***	0.504***	168.377***	0.354**	327.001**	1.770***	1.148***
	(39.135)	(0.353)	(0.308)	(45.769)	(0.362)	(38.801)	(0.379)	(0.162)	(43.054)	(0.158)	(147.673)	(0.226)	(0.227)
	[0.043]	[0.000]	[0.000]	[0.030]	[0.003]	[0.000]	[0.000]	[0.002]	[0.000]	[0.026]	[0.027]	[0.000]	[0.000]
Observations	898	898	898	746	746	897	898	898	722	723	896	898	898
Adjusted R^2	0.022	0.005	0.014	0.018	0.010	0.026	0.003	0.020	0.054	0.020	0.030	0.066	0.103
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Additional regressions not displayed in this paper for concision show that computers and tablets have no effect on the type of choices in the time preferences task neither on the type of errors made in the risk preferences task. We conclude that the only effect of computers and electronic tablets is an increase in response time.

Additional analysis of the University staff presence effect

In this section, we display additional results about the effect of administering the experiment by university staff. We study the effect of university staff on the type of choices made by consistent subjects in the time preferences task and the type of inconsistencies made in the risk preferences task. Table A8 shows the results: administering the experiment by university staff has an effect on the elicited time preferences of consistent subjects. We see that the presence of university staff decreased the number of sophisticated subjects ($p = 0.005$) with a relatively high statistical magnitude (17.3%). This phenomenon is associated with an increase in the proportion of subjects allocating everything to the present ($p = 0.010$) but nothing to the future ($p = 0.957$). However, we observed that subjects made the same type of inconsistencies in the risk preferences task regardless of the experimenters. We conclude that university staff decreased the quality of the data in the time preferences task. Because the first task responded was the time preferences task, a potential interpretation is that teachers helped students to immediately concentrate in the experiment.

Table A8: Regressions on the effect of University staff on choices in time and risk preferences tasks

	(1) Present	(2) Sophis	(3) Future	(4) DomFirst	(5) SwitchBack	(6) DomLast
$Univstaff$	0.106***	-0.173***	-0.002	0.045	0.064	0.027
	(0.041)	(0.061)	(0.035)	(0.041)	(0.046)	(0.025)
	[0.010]	[0.005]	[0.957]	[0.277]	[0.165]	[0.277]
Constant	-0.415**	1.045***	0.464**	-0.024	-0.008	0.055
	(0.201)	(0.299)	(0.188)	(0.192)	(0.230)	(0.115)
	[0.039]	[0.001]	[0.014]	[0.901]	[0.972]	[0.633]
Observations	902	902	902	902	902	902
R-squared	0.016	0.021	0.015	0.002	0.004	0.008
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Robust standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1