Journal of Behavioral and Experimental Economics 106 (2023) 102048

Contents lists available at ScienceDirect

Journal of Behavioral and Experimental Economics

journal homepage: www.elsevier.com/locate/jbee

The adventure of running experiments with teenagers[☆]

Antonio Alfonso[a], Pablo Brañas-Garza[a], Diego Jorrat[a,*], Pablo Lomas[a,b], Benjamin Prissé[c], Mónica Vasco[a], María J. Vázquez-De Francisco[a,b]

[a] LoyolaBehLab, Universidad Loyola Andalucía, Spain
[b] ETEA Foundation-Development Institute, Universidad Loyola Andalucía, Spain
[c] Singapore University of Technology and Design, Singapore

ARTICLE INFO

JEL classification:
C91
D81

Keywords:
Developmental decision-making
Field experiments
Economic preferences
Teenagers

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 and cognitive abilities, among others. Results show: (i) the recruitment process matters. When the school includes the experiment as an institutional activity, the dropout rate is observed to be significantly reduced; (ii) hypothetical payments elicit similar results to monetary payments; (iii) adding visual elements to the experiment's interface improves the quality of answers; and iv) although the type of device has no effect on the results, who administers the experiment does have an effect. We conclude by giving three suggestions to researchers interested in conducting experiments with teenagers: first, run the experiment as a school-programmed activity; second, the use of actual payments is not necessary, which increases the cost and complicates the recruitment; and third, integrate visual components to the task.

1. Introduction

There is an increasing interest in children and teenagers among economists. Studying the decision-making in children and adolescents allow us to understand the developmental causes of anomalous behavior in adults and to propose interventions at early stages to improve future outcomes (Brocas & Carrillo, 2020a). Since laboratory experiments in Economics have been predominantly conducted using university students, the study of children and adolescents brings new methodological challenges: we need to adapt protocols, instructions and experimental tasks to make them understandable to teenagers and children to acquire reliable measures. Therefore, it is crucial to provide guidelines and insights on how to adapt experimental protocols to obtain reliable information from the decisions of children and adolescents.

This paper is related to relatively new contributions that provide guidelines and advice on how to run experiments with children. Brocas and Carrillo (2020a) provide some modifications of standard procedures regarding attention, incentives, presentation, child development and heterogeneity. In line with this research, List et al. (2023) provide 10 tips for pulling off experiments with children, including factors such as taking into account child competencies, causal identification, and logistical issues related to recruitment and implementation. We build on these contributions by analyzing the results of different experimental designs and data collection methods to measure risk and time preferences in teenagers using a multi-dimensional research platform. We also focus on the consistency of these measurements.

The economic literature provides robust evidence that time preferences are relevant in different domains. In the field of health behavior, 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. In education, the evidence suggests that subjects with a high level of patience are less likely to receive disciplinary referrals in school and more likely to graduate from high school (Castillo et al., 2011, 2019). These results were later extended to children and teenagers by Sutter et al. (2013) for alcohol and cigarette consumption, higher Body Mass Index (BMI), higher disciplinary referrals at school and smaller savings. These findings

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

* Correspondence to: Av. de las Universidades, s/n, 41704 Dos Hermanas, Sevilla, Spain.
E-mail address: dajorrat@gmail.com (D. Jorrat).

https://doi.org/10.1016/j.socec.2023.102048
Received 13 November 2022; Received in revised form 11 May 2023; Accepted 1 June 2023
Available online 13 June 2023
2214-8043/

interrogated the relationship between time preferences and behavioral outcomes among teenagers.

These pieces of evidence 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 & Sprenger, 2012). Economic events do not only occur at different time periods. They also have different probabilities of occurring. Also, a large literature in experimental economics has studied the role of risk preferences in explaining life outcomes: For example, Dohmen et al. (2011) find that risk measures are significantly related to several behaviors as holding stocks, being self-employed, participating in sports, and smoking. In the same line, 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. This evidence led economists to study the determinants of individual differences in risk attitudes. 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. Andreoni et al. (2020) elicited risk preferences, finding a positive correlation between higher cognitive abilities with more risk-taking, and higher executive functions with more risk-aversion.

All the above evidence suggests that preferences matter. Therefore, understanding how malleable are these preferences at different ages and also how gender shapes them, is crucially important to inform policy and design efficient interventions.[1] Regarding these topics, the literature shows that patience increases with age for younger subjects (7–11 years old) but there are no significant changes for older guys (12–18 years old) (Alfonso et al., 2023; Bettinger & Slonim, 2007; Sutter et al., 2019); while for risk preferences evidence suggests no significant effect of age (Alfonso et al., 2023; Eckel et al., 2012; Sutter et al., 2013, 2019).[2] Regarding gender differences, the results are mixed. Eckel et al. (2012) found that girls are more risk-averse as Horn et al. (2022) while Alfonso et al. (2023) do not find significant gender differences. The same mixed effects of gender on time preferences were found in the literature (Alfonso et al., 2023; Horn et al., 2022; Sutter et al., 2013, 2019).[3] While all these results might be linked with gender differences concerning career choices and labor market outcomes, further research with high statistical power and reliable information about these preferences for children and teenagers should be done.

Another concern among economists is whether or not children/adolescents can make consistent decisions in terms of the transitivity of choices. Harbaugh et al. (2001) were the first to study this topic and found that at an early age (7 and 11 years old) children are able to make choices (choosing between foods or toys, or Good domain) that obey the generalized axiom of revealed preferences (GARP). Brocas et al. (2019) extend this analysis to others economics domains. They found that in the Good and Social domains consistency improves significantly with age; while in the Risk domain participants were not nearly as consistent as in the other two domains, perhaps partly as a consequence of a still underdeveloped working memory system. Other works also contributed to the literature on strategic decision-making and social preferences of children and teenagers. Brocas and Carrillo (2020c, 2021b) adapted level-k games to children and demonstrated that young children as young as 5 years old can play a dominant strategy, with performance improving until reaching a plateau in late childhood. Results also suggest that subjects either understand all dimensions of the problem or none of them, except the first step of dominance. Brocas and Carrillo (2020b) showed that elementary school subjects are selfish, middle and high school subjects are generous if it is non-costly to them, and university students implement the costly but socially efficient allocation. Brocas and Carrillo (2021a) suggested an explanation by showing that children and teenagers lie in their favor but truthfully report for others, pointing at a higher valuation of their payoffs than others' payoffs. Because young teenagers are more spiteful against others, it provides additional evidence that part of the developmental process is learning to take others into account. Brocas and Carrillo (2022) studied mixed strategies of teenagers using a non-zero sum hide-and-seek game in which hiders must signal a potential hideout to seekers. They showed that the choices of subjects are governed by the powerful (erroneous) heuristic of playing and indicating the highest payoff location. Summarizing, these papers suggested that teenagers are able of strategic decision-making but lack sophistication when doing so, possibly because they are too self-centered and have difficulties conceiving situations beyond their interest.

The goal of this paper is to present the results of different experimental designs 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[4] yields different attrition levels. Second, we tested whether using hypothetical payments rather than monetary 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.

[1] Sutter et al. (2019) provide the first survey on the topic.

[2] There are a few exceptions. For patience, Andreoni et al. (2019) found that children are initially impatient but become more patient with age; while Jørgensen et al. (2022) found that in older Danish children, risk-taking significantly increases with age for boys and marginally significantly for girls.

[3] Scheres et al. (2014) suggest that these differences in results are partly related to task differences and warns of the limitations in terms of the ecological validity that tasks used by economists suffer when measuring patience in children and adolescents. They suggest that temporal discounting (TD) should be viewed as reflecting a trade-off between the ability to wait (self-control), and affective processes such as the sensitivity to the immediacy of the small reward and sensitivity to the magnitude of the delayed reward. The relative contribution of these two effects will thus depend on the specific parameters chosen in the TD task. In this line, Reynolds and Schiffbauer (2004) propose the use of experiential discounting task (EDT), where participants experience certain choice consequences *during* the measurement procedure. While in these tasks, the reward magnitudes are necessarily small (typically up to $1) and delay duration relatively short (up to 60 s), have been proven to be useful in examining individual differences in children and adolescents (Scheres et al., 2010).

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

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 LimeSurvey. 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.[5] 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.[6] 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 correlates with lower attrition levels. Second, paying subjects with hypothetical monetary incentives elicits similar results 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.

The structure of this paper is as follows: Section 2 presents the recruitment process and subject 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.[7] 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 informed consent from their parents to participate in the experiment. We obtained a sample of 1075 adolescents. Due to the Covid-19 pandemic lockdown, the experiment was conducted using a self-administrated online questionnaire programmed in LimeSurvey.[8] We used monetary 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/monetary payments was not random.

The recruitment process was different in the second and third waves. We run the experiments in the schools and we signed agreements with school principals to integrate the experiment as part of their pedagogical curriculum and to run it as a class activity. It removed the need for 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. The experiment was conducted on a platform called SAND, which permitted greater control over data privacy. Students read the instructions and navigated through the questionnaire on different screens to complete the survey. Subjects were paid with hypothetical payoffs since schools would not accept running experiments with real money. Despite these similarities in the experimental design, there were some differences that may reduce the comparability between both waves: (i) were run in different locations; and (ii) in the second wave teachers administered the questionnaire while in the third wave was administered by university staff or teachers.[9]

Table 1 summarizes the main features and differences of the experimental design across the three waves.

The response rate and the dropout during the experiment were different across waves. In the first wave, we observed that 10.98% of participants decided to abandon immediately the survey at the first screen. We also saw that asking subjects under 14 years old to provide 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 the 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.

Regarding the attrition during the experiment, in the first wave subjects dropped out of the experiment by 7.73% in the survey,[10] 6.60% 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

5 Between-subject Random Incentivized System (BRIS) payment.

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

7 See at: https://aspredicted.org/blind.php?x=af3rw7.

8 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.

9 In concrete, 645 observations in the third wave (71.51%) were collected by university staff, while the rest were done by the school teachers.

10 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.

Table 1
Experimental design across waves.

	First wave	Second wave	Third wave
Online vs In site experiment	Online	In site	In site
Class activity	No	Yes	Yes
Location	Andalusia and Madrid	Madrid	Andalusia
Software	LimeSurvey	SAND	SAND
Payment	Half monetary/ Half hypothetical	Hypothetical	Hypothetical
Experiments administrators	Not applicable (online)	Teachers	Researchers/Teachers
Time Preferences treatment	MPL/ MPL-Cont/ MPL-Video*	MPL-Choice	MPL-Gift
Holt-Laury treatment**	HL	HL	Gumball
Electronic devices used	M/C/T***	M/C/T	M/C/T

Note: *Assignment to treatments in the time preference task was random in the first wave. **While in Holt-Laury task only visual aids were included in the Gumball design, for the Time Preferences task the visualization and the different designs shown were refined and improved, having visual aids in MPL-Cont, MPL-Video, MPL-Choice and MPL-Gift. ***M/C/T is referring to Mobile phones, Computers and Tablets.

Table 2
Summary statistics of the variables.

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
Consistency Time	1926	0.752	1	0	1
# Future Choices	1926	0.640	0.666	0	1
CRT score	1913	0.636	0.666	0	1
Finance Score	1912	0.431	0.333	0	1
Consistency Risk	1906	0.592	1	0	1
# Risky Choices	1906	0.552	0.545	0	1

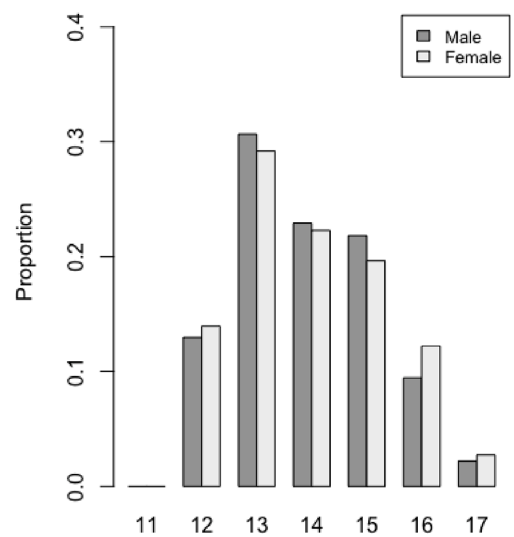

Fig. 1. Distribution of participants by age and gender.

control questions. In the first wave, we lost 7.73% in the participant's characterization section (see "Survey" in Table A.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. Table A.1 summarizes the dropout before and during the experiment.

Considering the response rate in the three waves, we reached a total of 1926 individuals who completed the entire experiment, complied with the age criteria and reported their age and gender. However, data are missing in at least one decision for some subjects in wave 1 for different tasks: 20 subjects do not complete at least one decision of the risk preference task, 13 the CRT and 14 the financial questions (see the following section for further detail on the tasks). In the second and third waves, we asked for age and gender at the end of the questionnaire, meaning that we could only include in the analysis subjects who entirely completed the experiment (see Table 2).

Table 2 shows summary statistics of the data. 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 other provinces.

Fig. 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 ($D = 0.032$, $p = 0.720$).

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.[11] Each task is detailed hereafter, except Delavande test, which will be part of future research.

Time preferences

The time preferences task included six decisions in all 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.

[11] We analyze data of the Delavande test in Alfonso et al. (2023).

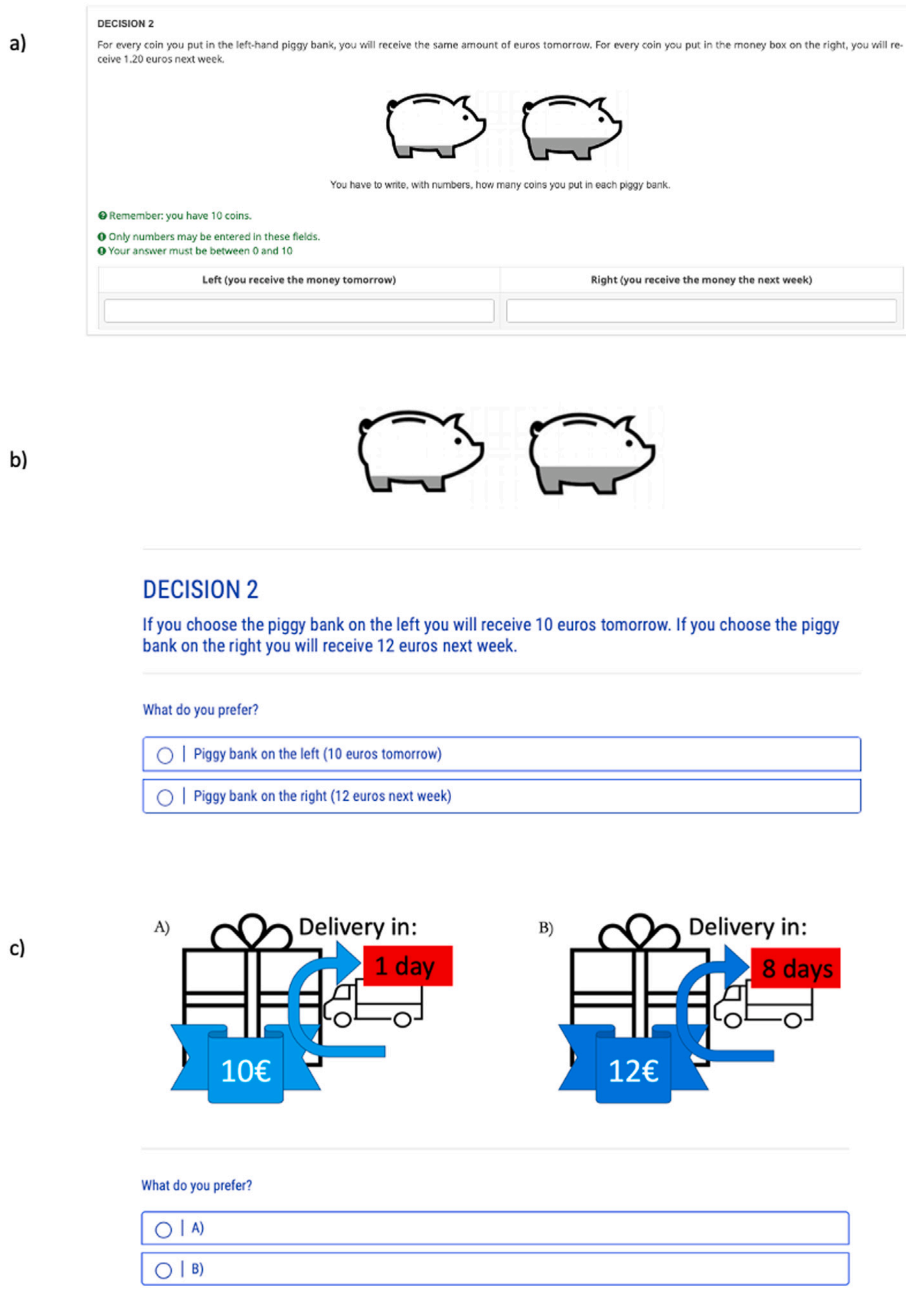

Fig. 2. Time preferences: screens decisions by treatments.

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 on the early date (left) is always 1 euro, and the corresponding amount of money on 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. Fig. 2a provides an example of the decision screen for the first decision.

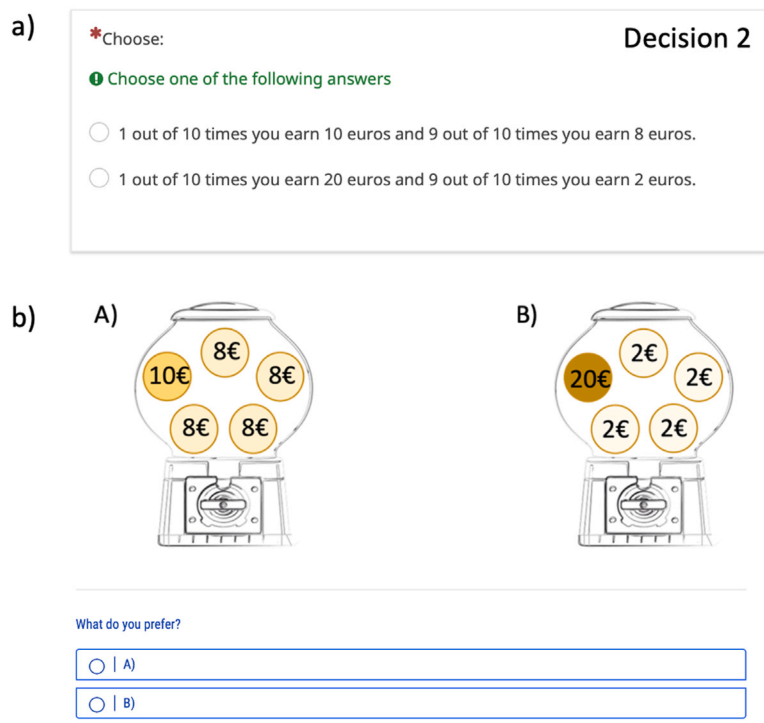

Fig. 3. Risk preferences: decision screens by treatments.

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 Fig. A.1. 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. Fig. 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. Fig. 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. Fig. A.2 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.

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. Fig. 3a displays an example of the screen for the second decision.

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 visual support using a gumball machine to help subjects understand the concept of probabilities. Fig. 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. Fig. A.3 and Fig. A.4 respectively show the safe and the risky lotteries and how gumballs represent the increase in the probability of the highest outcome.

Then, 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 for 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 are 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 with 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 on 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 (same location).

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. In all the regression we controlled for the number of reflexive options (*NumCRT*) in the CRT task, the number of correct answers in the financial skills questionnaire (*NumFin*), age, gender and school fixed effects.

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: 54.79% of all subjects completed the experiment in the first wave, 77.13% in the second wave and 94.06% in the third wave. However, the differences between wave 1 and wave 2 should be taken with caution, as the former was online while the latter was done at the school. Comparability between waves 2 and 3 is more straightforward, as both were done on-site.

Fig. 4 explores how attrition evolves *before* and *during* the experiment across waves. While we cannot establish causality, 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 dropouts at the first

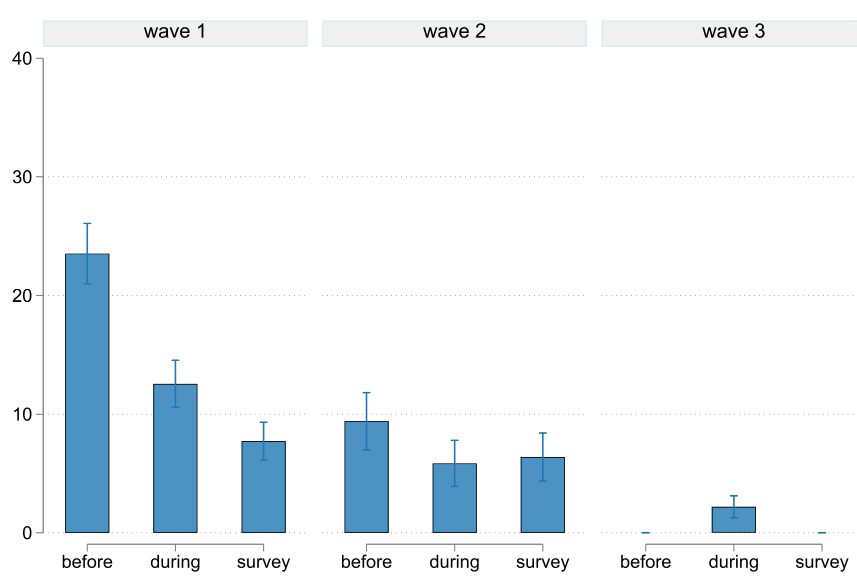

Fig. 4. Attrition before and during the experiment.

Table 3
Dropout of subjects in the different waves.

Dropout	First Wave		Second Wave		Third Wave	
	N	%	N	Ratio	N	%
Potential sample	1075	100.00%	564	100.00%	959	100.00%
Not age of interest	34	3.16%	7	1.24%	9	0.94%
Before (1st-screen+parental consent)	253	23.54%	53	9.40%	0	0%
During	126	11.72%	33	5.85%	48	2.19%
Survey	93	8.65%	36	6.38%	–	0%
Response rate	589	54.79%	435	77.13%	902	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.

screen. For more precise information about the dropout rate in the different waves, see Table A.1 in Appendix.

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 out the survey entirely. As observed, the three types of attrition decrease significantly from wave to wave, even if we compare wave 2 and 3 (both waves were online) attrition *during* the experiment decreases by 62.5% in wave 3.

From Table 3 and Fig. 4 we can conclude:

Result 1: Agreements with schools to run experiments as a class activity (on site and by experimenters as lab experiments) are important to reduce attrition.

5.2. Do hypothetical payments provide the 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., 2023 for time preferences), but these studies have been conducted with adults only. Thus the impact of the payment method on teens remains open. We shed some light on 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, in fact Table A.2 in the Appendix shows that both groups were different in terms of CRT. Thus we selected a sub-sample of adolescents from waves 1 and 2 that were similar in terms of school characteristics, age, gender and two measures of cognitive abilities: CRT and financial (or maths) score. 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 736 adolescents (n_B = 387 and n_H = 349).

Fig. 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).

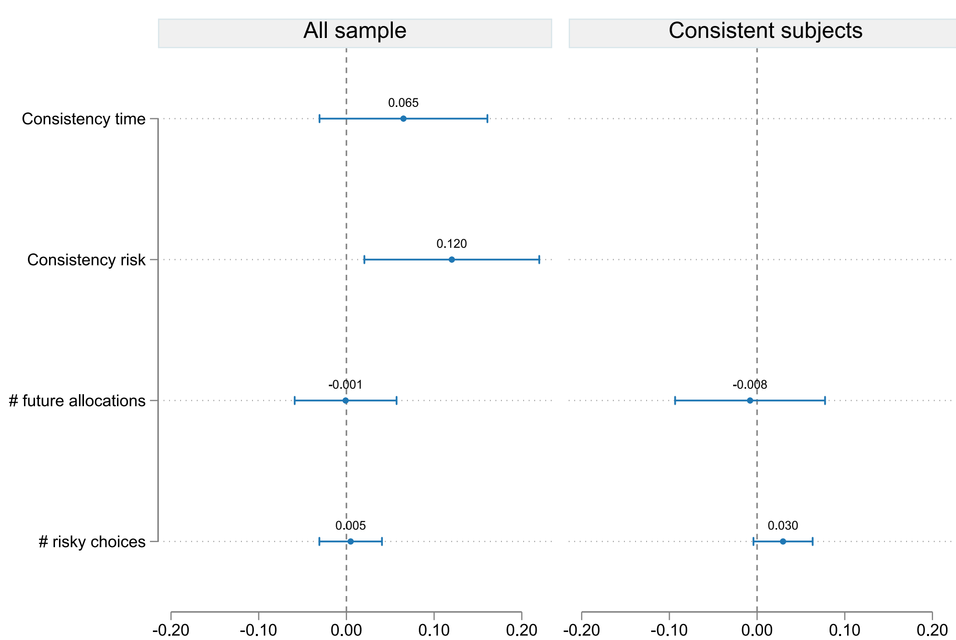

Fig. 5. Effect of hypothetical payments (vs BRIS) on outcomes.
Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. The control group is BRIS. *Consistency Time* and *Consistency Risk* outcomes are equal to 1 for subjects who did not do multiple-switching in time and risk tasks, and 0 otherwise. *# future allocations* and *# risky choices* are normalized using the min–max standardization process.

As can be observed, hypothetical payment 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 consistency in the time preferences task ($p = 0.183$), the number of future allocations ($p = 0.691$) and risky choices ($p = 0.887$).

In the right-side of Fig. 5 we consider only consistent subjects. We found the same results on the number of future allocations ($p = 0.855$) but hypothetical payments marginally increase the number of risky choices by 3% ($p = 0.085$).[12] We therefore conclude:

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

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 preference elicitation. We regressed the outcome variables ($consistency$ in risk and time preferences and the number of future and risky choices) on different dummy variables that represented the different visualizations (treatments) used in each task. As before, subjects were not randomly assigned to the different treatments, so the results shown in this section reflect correlations. In fact, Tables A.3 and A.4 of the appendix show that there are significant differences in terms of age, female, CRT and financial scores across the different visualizations treatments in both tasks, so we controlled for these variables and school fixed effects.

[12] Following the same methodology, we also test the effect of hypothetical payments on the elicitation of cognitive abilities: Indeed, we observe that hypothetical payments reduce CRT score. 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 (2023). ($p = 0.005$) and marginally increase financial abilities score ($p = 0.092$).

Time preferences

Fig. 6 displays a summary of results shown in Table A.5. It illustrates the effect of the different visual elements on the performance in the time preferences task, with the MPL condition as the reference category. We can observe that consistency in the task increased in $MPL{-}Choice$ ($p = 0.004$) and $MPL{-}Gift$ ($p < 0.001$). Additionally, we see that $MPL-Choice$ increased the number of future allocations ($p = 0.014$) but this effect disappeared for consistent subjects ($p = 0.394$). We found different and mixed results for $MPL - Gift$, since the number of future allocations decreased for consistent subjects ($p = 0.045$), but it is not significant for all the samples ($p = 0.395$).

These results suggest that visual elements influence the performance of 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 consistent subjects. 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, especially for consistent subjects.

Result 3a: The use of visual elements is associated with a higher consistency, and for consistent subjects with a lower number of future choices.

Additional regression analyses are presented in Table A.11 of Appendix A.2. These results suggest that $MPL - Choice$ and $MPL - Gift$ improve the quality of the results in terms of increasing the percentage of subjects switching from early to later assignments.

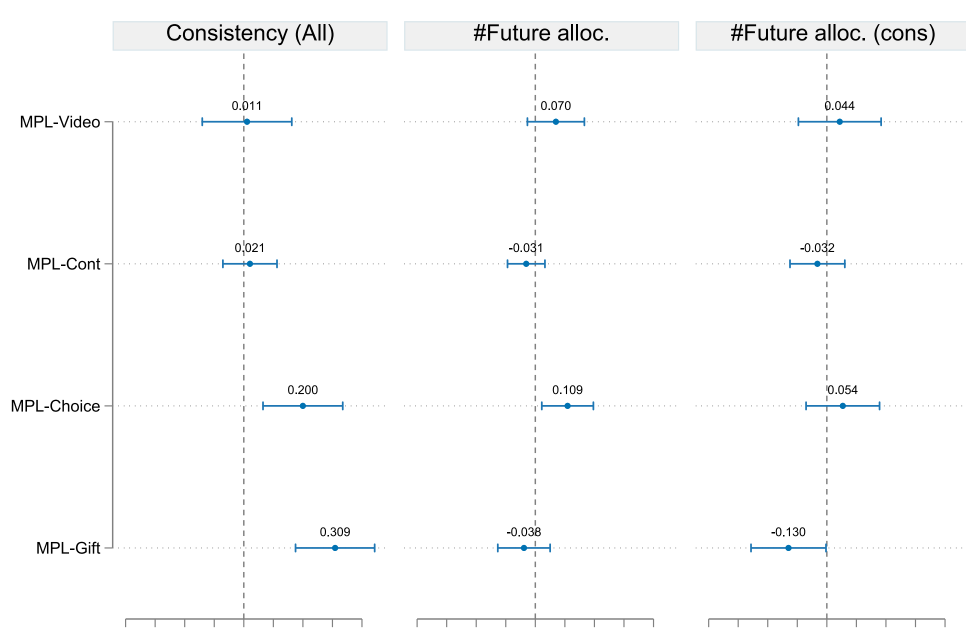

Fig. 6. Effect of visual elements in the time preferences task.
Note: Control group is MPL. Consistency is equal to 1 for subjects that do not commit multiple-switching in the time task, and 0 otherwise. #future alloc. is equal to the number of future allocations normalized using the min–max standardization process. (all) reflect the entire sample while (cons) means consistent subjects only.

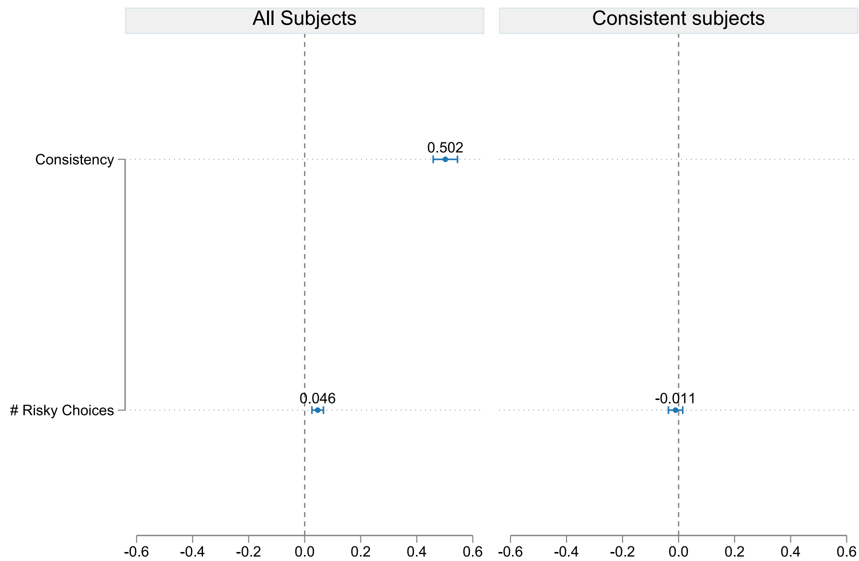

Fig. 7. Effect of visual elements on the risk preferences task.
Note: Control group is HL. Consistency is equal to 1 for subjects that do not commit multiple-switching in the time task, and 0 otherwise. $\#Risky\ choices$ is equal to the number of decisions where the risk lottery was chosen for each subject, normalized using the min–max standardization process. (all) reflect the entire sample while (cons) means consistent subjects only.

Risk preferences

Fig. 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* variable.

We also observed that using the gumball increased the consistency notably by an estimated 50.2% ($p < 0.001$). These results also suggest that it slightly increased the number of risky choices for all subjects by 4.6% ($p < 0.001$), but this effect disappeared when we consider only consistent subjects ($p = 0.408$). Table A.6 of the Appendix shows the regression results.

The use of visual elements is largely associated with a higher consistency rate 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 us to conclude whether it is visualization what influenced elicited preferences.

We can conclude:

Result 3b: The use of visual elements is associated with higher consistency and has no influence on the number of risky choices.

In Table A.12 of Appendix A.2 we present additional regression analyses. We obtain further evidence that *Gumball* increases the quality of results in the risk preferences task, reducing all types of inconsistencies.

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 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 online). In this case, we could also study response time because all subjects answered on the same device and platform.

We obtained a sample of 898[13] 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.[14] The results display the comparison between subjects answering on a mobile phone (treatment) and subjects answering on computer or tablet 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, cognitive abilities (CRT and financial score) and school fixed effects. Recall that subjects were not randomly assigned to the different devices and Table A.7 of the appendix reflects that both groups were different in terms of age, gender and cognitive abilities. So, the results should be interpreted as simple correlations.

Fig. 8 summarizes the results of Table A.8 presenting the OLS coefficient estimation of $Mobile$ phone on the outcome variables. We observed that $Mobile$ is negative and significant for the time response in both tasks, considering all the samples ($p = 0.002$ and $p = 0.003$) and only consistent subjects ($p = 0.011$ and $p = 0.010$). We found no other effect of using a mobile phone on the other outcomes.

We conclude in this section:

Result 4: The use of mobile phones is associated with a reduction in response time, while does not influence the experimental outcomes.

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,

13 Missing 4 outliers in time preferences and 3 in risk preferences.
14 4 (0.44%) subjects answered on a different device.

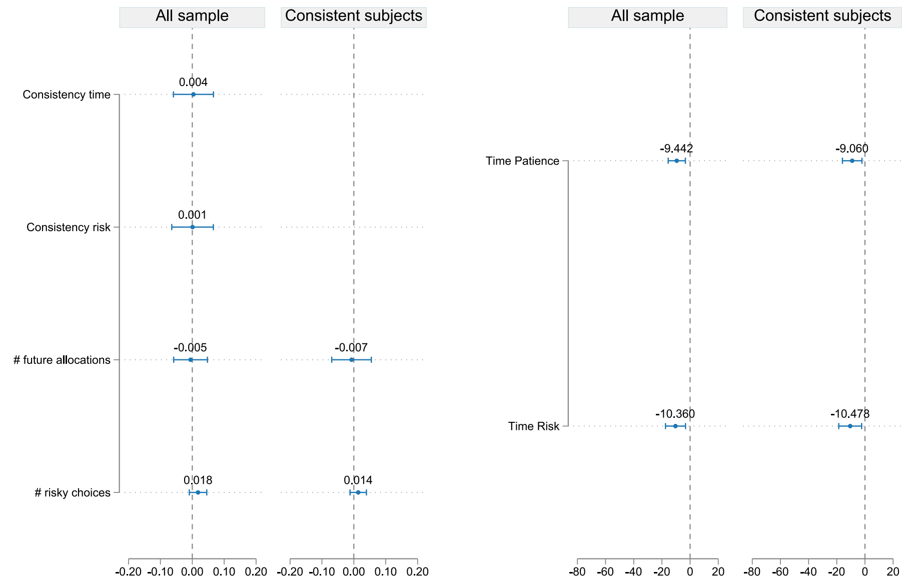

Fig. 8. Effects of mobile phone on answers and response time.
Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. The control group is Computer and Tablets platforms. *Consistency Time* and *Consistency Risk* outcomes are equal to 1 for 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.

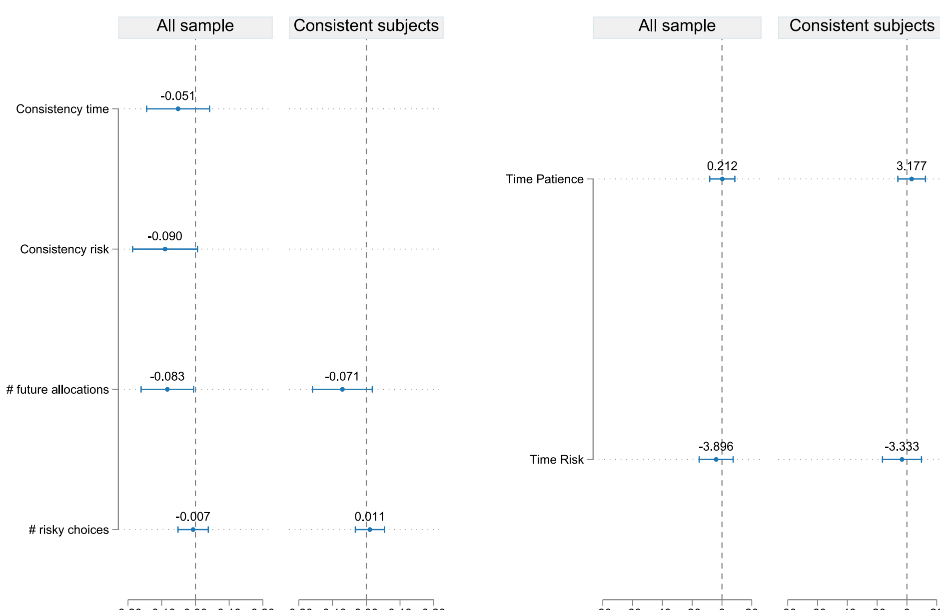

Fig. 9. Effects of university staff on answers and response time.
Note: Consistent subjects refers to individuals who did not do multiple-switching in time and risk preferences tasks. The control group is the one directed by the teachers. *Consistency Time* and *Consistency Risk* outcomes are equal to 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.

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.[15]

We investigated whether administering the questionnaire by university staff rather than by teachers, influenced results. The reference group was conformed of subjects who filled out the questionnaire in the presence of their teachers. As before, we conducted a regression of each outcome variable on the treatment variable ($Univstaff$). Since subjects were not randomly assigned to treatment, we controlled for age, gender, cognitive abilities (CRT and financial) and school fixed effect to enhance comparability. As before, results should be interpreted as correlational analysis, since Table A.10 shows that both groups differ in terms of age and cognitive skills.

Fig. 9 summarizes the results of Table A.9 displaying the estimated coefficient of $UnivStaff$ on the different outcome variables.

We can observe that the presence of university staff does not affect consistency in time and risk preferences tasks ($p = 0.280$ and $p = 0.066$, respectively). However, it is associated with subjects choosing fewer future allocations in the time preferences task for all the samples ($p = 0.036$), while it is not significant for consistent subjects ($p = 0.116$). Regarding the different outcomes on response time, the presence of university staff had no effect.

We conclude that:

Result 5: The presence of university staff does not impact teenagers performance in time and risk tasks.

In Appendix A.4 we present additional regression analyses. Table A.14 shows that administering the experiment by university staff increased one type of inconsistency (dominated choices in the first decision) and reduced another type (number of subjects switching back to the safe lottery after choosing the risky lottery). However, we observed that subjects made the same type of choices in the time preferences task regardless of the experimenters. We conclude that university staff has a null effect on the quality of the data in the time and risk preferences tasks.

6. Discussion

In this paper, we analyzed the influence of different 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 the competencies and skills of children and adolescents might help educators and educational professionals to identify current and prevent future problems, missbehaviors, or undesirable situations, such as worse health, addictions, lower academic levels, or bad 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 influence of the introduction of visual elements in the tasks. Fourth, we analyzed whether responding using different devices such as mobile phones, tablets or computers changed the results. Fifth, we studied whether administering experiments by school teachers or university staff had an impact on participants' responses.

[15] 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.

Our findings suggest that: (i) Signing agreements with schools to run experiments as a class regular activity (on site and carried out by experimenters as lab experiments) is correlated with lower attrition levels, improving 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 an adult population (Brañas-Garza et al., 2021, 2023). Overall, we showed that the use of real incentives in experiments with teenagers is not needed to collect better data.

(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 & Rutström, 2008 for risk elicitation). This finding suggests that individuals understand tasks better when the experiments has visual elements. This, could influence the elicited preferences since using a gift increased the number of early choices on consistent subjects. Future work should further investigate visual support elements using a proper randomized design and particularly, whether or not the use of visual aids that represent tempting goods affects elicited preferences.

(iv) In sharp contrast to Mograbi (2022), our results show that the use of mobile phones does not influence the outcome variables. Even better, it is associated with a reduction in the response time. 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 only increases the number of future allocations, but this effect vanishes when we consider only consistent subjects. A natural question is whether teachers "assisted" students during the experiment. 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 logistic costs.

The final advice is to present questions where individuals must choose between options and 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.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Pablo Branas Garza reports financial support was provided by Spanish Ministry of Economy and Competitiveness (PID2021 126892NB- I00). Pablo Branas Garza reports financial support was provided by Excelencia Junta de Andalucía (PY 18 FR 0007). Pablo Branas Garza reports financial support was provided by Agencia Andaluza de Cooperación Internacional para el Desarrollo.

Data availability

Data will be made available on request.

Appendix

A.1. Additional figures and tables

Fig. A.1. Illustration of increasing interest rates with piggy banks.

Table A.1
Dropout rate across different waves.

	First Wave		Second Wave		Third Wave	
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	71	6.60%	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	589	54.79%	435	77.13%	902	94.06%

Table A.2
Balance across hypothetical and BRIS treatments.

	H	B	Diff	p-value
Age	14.357	14.309	−0.048	0.510
Female	0.439	0.496	0.057	0.126
CRT	0.804	0.733	−0.071	0.001
Financial score	0.512	0.52	0.008	0.707

Table A.3
Balance across the different visual elements in the time preferences task.

	(1) Age	(2) Female	(3) CRT	(4) Financial
MPL-Video	−0.278***	−0.116**	0.022	−0.043
	(0.099)	(0.055)	(0.030)	(0.029)
	[0.005]	[0.035]	[0.469]	[0.141]
MPL-Cont	0.182**	−0.009	−0.023	0.056*
	(0.090)	(0.048)	(0.026)	(0.028)
	[0.043]	[0.852]	[0.387]	[0.050]
MPL-Choice	0.058	−0.037	−0.107***	−0.006
	(0.086)	(0.044)	(0.025)	(0.025)
	[0.500]	[0.398]	[0.000]	[0.809]
MPL-Gift	−0.997***	−0.028	−0.305***	−0.156***
	(0.077)	(0.040)	(0.021)	(0.023)
	[0.000]	[0.490]	[0.000]	[0.000]
Mean MPL	14.367***	0.527***	0.805***	0.502***
	(0.064)	(0.036)	(0.019)	(0.021)
	[0.000]	[0.000]	[0.000]	[0.000]
Observations	1,926	1,926	1,913	1,912
Adjusted R-squared	0.160	0.001	0.188	0.082

Note: $Mean\ MPL$ represents the mean in the baseline group (MPL). For each control variable, $MPL{-}Video$, $MPL{-}Cont$, $MPL{-}Choice$, $MPL{-}Gift$ represent the difference in the mean of that variable between that group and the baseline group.

Table A.4
Balance across the different visual elements in the risk preference task.

	HL	Gumball	Diff	p-value
Age	14.397	13.37	−1.027	0.000
Female	0.492	0.499	0.007	0.769
CRT	0.757	0.5	−0.257	0.000
Financial score	0.506	0.346	−0.160	0.000

Table A.5
Regressions on the effect of the interface (visual elements) on time preferences task.

	(1) Consistency	(2) #Future alloc.	(3) #Future alloc. (cons)
MPL-Video	0.011	0.070	0.044
	(0.077)	(0.049)	(0.071)
	[0.887]	[0.156]	[0.541]
MPL-Cont	0.021	−0.031	−0.032
	(0.047)	(0.032)	(0.047)
	[0.654]	[0.341]	[0.499]
MPL-Choice	0.200***	0.109**	0.054
	(0.069)	(0.044)	(0.063)
	[0.004]	[0.014]	[0.394]
MPL-Gift	0.309***	−0.038	−0.130**
	(0.068)	(0.045)	(0.065)
	[0.000]	[0.395]	[0.045]
CRT	0.094***	0.087***	0.093***
	(0.036)	(0.026)	(0.034)
	[0.010]	[0.001]	[0.006]
Financial score	0.132***	0.128***	0.105***
	(0.038)	(0.026)	(0.033)
	[0.000]	[0.000]	[0.002]
Constant	0.167	0.701***	0.863***
	(0.134)	(0.096)	(0.122)
	[0.211]	[0.000]	[0.000]
Observations	1,912	1,912	1,439
R-squared	0.069	0.094	0.115
Controls	Yes	Yes	Yes

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

Table A.6
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	0.502***	0.046***	−0.011
	(0.022)	(0.010)	(0.013)
	[0.000]	[0.000]	[0.408]
CRT	0.232***	0.004	−0.056***
	(0.035)	(0.017)	(0.021)
	[0.000]	[0.798]	[0.008]
Financial score	0.285***	−0.004	−0.028*
	(0.037)	(0.016)	(0.016)
	[0.000]	[0.802]	[0.086]
Constant	0.270**	0.575***	0.649***
	(0.121)	(0.054)	(0.054)
	[0.026]	[0.000]	[0.000]
Observations	1,909	1,909	1,128
R-squared	0.229	0.017	0.016
Controls	Yes	Yes	Yes

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

Table A.7
Balance across mobile and tablet groups.

	Comp./Tablet	Mobile	Diff	p-value
Age	13.432	13.176	−0.256	0.015
Female	0.479	0.573	0.094	0.020
CRT	0.509	0.467	−0.042	0.049
Financial score	0.355	0.317	−0.038	0.079

Fig. A.2. Illustrations of increasing interest rates with ribbons.

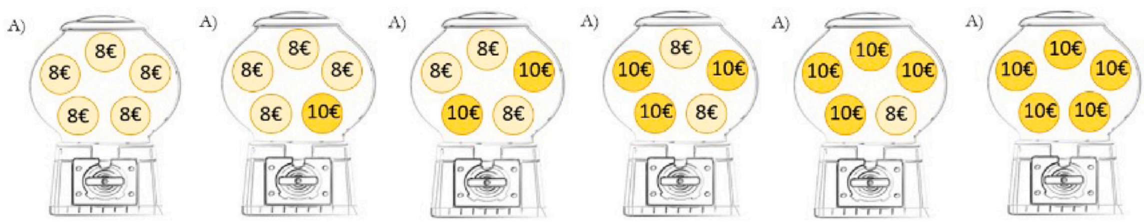

Fig. A.3. Lottery A used in the experiment.

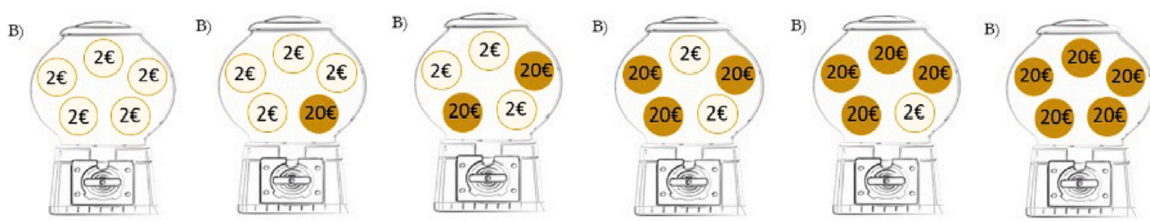

Fig. A.4. Lottery B used in the experiment. Note: Outliers were not considered for any of the following models.

Table A.8
Regressions on the effect of mobile phones on time preferences and risk preferences task.

	(1) TimePat	(2) ConsTime	(3) NumFut	(4) TimePatC	(5) NumFutC	(6) TimeRisk	(7) ConsRisk	(8) NumRisk	(9) TimeRiskC	(10) NumRiskC
mobile	−9.442***	0.004	−0.005	−9.060**	−0.007	−10.360***	0.001	0.018	−10.478**	0.014
	(3.102)	(0.032)	(0.027)	(3.532)	(0.032)	(3.628)	(0.033)	(0.014)	(4.127)	(0.013)
	[0.002]	[0.909]	[0.841]	[0.011]	[0.820]	[0.004]	[0.979]	[0.195]	[0.011]	[0.290]
gender	6.889***	0.009	0.031	4.728	0.041	15.881***	0.016	−0.003	16.649***	−0.006
	(2.666)	(0.025)	(0.022)	(3.004)	(0.025)	(2.944)	(0.027)	(0.011)	(3.313)	(0.011)
	[0.010]	[0.712]	[0.156]	[0.116]	[0.104]	[0.000]	[0.537]	[0.802]	[0.000]	[0.578]
age	−0.209	−0.004	−0.022**	−0.133	−0.018*	−3.178***	0.000	0.006	−2.401*	0.004
	(1.099)	(0.010)	(0.009)	(1.267)	(0.010)	(1.122)	(0.010)	(0.004)	(1.235)	(0.004)
	[0.849]	[0.685]	[0.011]	[0.916]	[0.072]	[0.005]	[0.969]	[0.158]	[0.052]	[0.292]
CRT	−0.352	0.069	0.098**	−0.887	0.125**	12.762**	0.025	−0.035	9.816	−0.077***
	(4.948)	(0.048)	(0.043)	(5.471)	(0.050)	(5.577)	(0.053)	(0.023)	(6.282)	(0.022)
	[0.943]	[0.157]	[0.024]	[0.871]	[0.013]	[0.022]	[0.636]	[0.125]	[0.119]	[0.000]
Financial score	4.349	0.110**	0.127***	−0.703	0.093*	8.316	0.133***	−0.024	8.660	−0.043**
	(4.895)	(0.051)	(0.041)	(5.548)	(0.048)	(5.753)	(0.050)	(0.022)	(6.476)	(0.021)
	[0.375]	[0.030]	[0.002]	[0.899]	[0.054]	[0.149]	[0.008]	[0.270]	[0.182]	[0.042]
Constant	70.659***	0.808***	0.778***	71.850***	0.734***	119.754***	0.740***	0.524***	109.810***	0.581***
	(14.479)	(0.139)	(0.118)	(16.586)	(0.135)	(15.396)	(0.144)	(0.060)	(17.003)	(0.058)
	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	894	898	898	742	746	895	898	898	721	723
R-squared	0.020	0.010	0.027	0.016	0.025	0.054	0.011	0.020	0.050	0.042
Controls	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 A.9
Regressions on the effect of university staff on time preferences and risk preferences task.

	(1) TimePat	(2) ConsTime	(3) NumFut	(4) TimePatC	(5) NumFutC	(6) TimeRisk	(7) ConsRisk	(8) NumRisk	(9) TimeRiskC	(10) NumRiskC
$Univstaff$	0.212	−0.051	−0.083**	3.177	−0.071	−3.896	−0.090*	−0.007	−3.333	0.011
	(4.298)	(0.048)	(0.040)	(4.713)	(0.045)	(5.802)	(0.049)	(0.023)	(6.671)	(0.022)
	[0.961]	[0.280]	[0.036]	[0.500]	[0.116]	[0.502]	[0.066]	[0.764]	[0.617]	[0.630]
CRT	−0.294	0.062	0.086**	−0.415	0.114**	11.761**	0.015	−0.032	9.225	−0.073***
	(5.073)	(0.049)	(0.043)	(5.633)	(0.050)	(5.494)	(0.052)	(0.023)	(6.169)	(0.022)
	[0.954]	[0.203]	[0.046]	[0.941]	[0.022]	[0.033]	[0.770]	[0.155]	[0.135]	[0.001]
Financial score	3.999	0.106**	0.121***	−1.162	0.089*	7.783	0.127**	−0.024	7.860	−0.042**
	(4.813)	(0.050)	(0.041)	(5.405)	(0.048)	(5.725)	(0.050)	(0.021)	(6.452)	(0.021)
	[0.406]	[0.035]	[0.003]	[0.830]	[0.064]	[0.174]	[0.011]	[0.265]	[0.224]	[0.045]
Constant	67.018***	1.020***	1.114***	58.372**	1.024***	138.632***	1.091***	0.528***	127.299***	0.521***
	(21.526)	(0.238)	(0.199)	(24.185)	(0.229)	(27.096)	(0.245)	(0.110)	(30.627)	(0.105)
	[0.002]	[0.000]	[0.000]	[0.016]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]	[0.000]
Observations	898	902	902	746	750	899	902	902	725	727
R-squared	0.014	0.011	0.031	0.013	0.028	0.057	0.014	0.025	0.052	0.045
Controls	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 A.10
Balance across university staff and teacher groups.

	Teachers	Univ. staff	Diff	p-value
age	15.012	12.716	−2.296	0.000
female	0.486	0.504	0.018	0.636
numcrt	0.549	0.481	−0.068	0.000
numfin	0.436	0.31	−0.126	0.000

A.2. 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 the 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 the earliest period, subjects switching from early to later allocations, and subjects allocating everything to the later period. We also use as dependent variables in the risk preferences task *DomFirst*,

Table A.11
Regressions on the effect of the interface (visual elements) on choices of consistent subjects in the time preferences task.

	(1) Present	(2) Sophis	(3) Future
MPL-Video	−0.037	−0.116*	0.165**
	(0.044)	(0.069)	(0.069)
	[0.399]	[0.089]	[0.017]
MPL-Cont	0.008	0.069	−0.056
	(0.033)	(0.042)	(0.040)
	[0.797]	[0.104]	[0.159]
MPL-Choice	−0.083**	0.248***	0.035
	(0.041)	(0.067)	(0.060)
	[0.044]	[0.000]	[0.563]
MPL-Gift	0.011	0.404***	−0.107*
	(0.043)	(0.066)	(0.058)
	[0.795]	[0.000]	[0.066]
CRT	−0.092***	0.141***	0.045
	(0.027)	(0.040)	(0.033)
	[0.001]	[0.000]	[0.174]
Financial score	−0.049*	0.063	0.117***
	(0.025)	(0.041)	(0.032)
	[0.051]	[0.122]	[0.000]
Constant	0.021	−0.250*	0.396***
	(0.092)	(0.149)	(0.119)
	[0.819]	[0.094]	[0.001]
Observations	1,912	1,912	1,912
R-squared	0.060	0.118	0.067
Controls	Yes	Yes	Yes

Robust standard errors in brackets
$^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.12
Regressions on the effect of the interface (visual elements) on errors in the risk preferences task.

	(1) DomFirst	(2) SwitchBack	(3) DomLast
Gumball	−0.130***	−0.458***	−0.241***
	(0.021)	(0.023)	(0.018)
	[0.000]	[0.000]	[0.000]
CRT	−0.098***	−0.193***	−0.145***
	(0.033)	(0.036)	(0.031)
	[0.003]	[0.000]	[0.000]
Financial score	−0.126***	−0.286***	−0.138***
	(0.031)	(0.036)	(0.029)
	[0.000]	[0.000]	[0.000]
Constant	0.295***	0.676***	0.251***
	(0.101)	(0.121)	(0.093)
	[0.003]	[0.000]	[0.007]
Observations	1,914	1,914	1,909
R-squared	0.032	0.201	0.107
Controls	Yes	Yes	Yes

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

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 A.11 shows the OLS regression results regarding choices of consistent subjects in the time preferences task according to the interface used. We can observe that it does not affect the proportion of subjects allocating everything to the 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 ($p = 0.089$) and increased the proportion of subjects allocating everything to the future ($p = 0.017$). We also see that $MPL-Choice$ decreased the proportion of subjects allocating everything to present ($p = 0.044$) and increased the proportion of sophisticated subjects ($p < 0.001$). Finally, we can see that $MPL-Gift$ largely increased sophisticated answers ($p < 0.001$) and decreased marginally the proportion of subjects allocating everything to the future ($p = 0.066$). These results suggest that in the time preferences task $MPL-Video$ decreases while $MPL-Choice$ and $MPL-Gift$ increase the quality of results.

Regarding the risk preferences task, Table A.12 shows the results from the OLS regressions. We obtain further evidence 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 13.0% ($p < 0.001$), largely reduced by 45.8% 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 24.1% ($p < 0.001$). These results suggest that visualization decreased all types 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.

A.3. 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 A.13 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.

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 nor 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.

Table A.13
Regressions on the effect of computer and tablets on the time preferences and risk preferences task.

	(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
***$p < 0.01$, **$p < 0.05$, *$p < 0.1$

Table A.14
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
$Univstaff$	0.050	0.053	0.016	0.084**	−0.140**	0.004
	(0.041)	(0.046)	(0.023)	(0.041)	(0.061)	(0.035)
	[0.222]	[0.254]	[0.494]	[0.037]	[0.022]	[0.904]
CRT	0.072*	−0.031	−0.044	−0.099**	0.151**	0.010
	(0.042)	(0.050)	(0.028)	(0.047)	(0.065)	(0.043)
	[0.086]	[0.527]	[0.122]	[0.038]	[0.021]	[0.813]
Financial score	−0.065	−0.094*	−0.061**	−0.098**	0.143**	0.061
	(0.042)	(0.048)	(0.026)	(0.044)	(0.064)	(0.038)
	[0.121]	[0.051]	[0.019]	[0.026]	[0.025]	[0.111]
Constant	−0.084	0.035	0.105	−0.308	0.882***	0.446**
	(0.194)	(0.232)	(0.113)	(0.200)	(0.304)	(0.188)
	[0.666]	[0.881]	[0.355]	[0.124]	[0.004]	[0.018]
Observations	902	902	902	902	902	902
R-squared	0.006	0.009	0.020	0.030	0.036	0.018
Controls	Yes	Yes	Yes	Yes	Yes	Yes

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

A.4. 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 A.14 shows the results: administering the experiment by university staff has an effect only on risk preferences. We see that the presence of university staff increased the number of dominated choices in the first decision (($p = 0.037$, see column 4) and reduced the number of subjects switching back to the safe lottery after choosing the risky lottery ($p = 0.022$, see column 5). However, we observed that subjects made the same type of choices in the time preferences task regardless of the experimenters. We conclude that university staff has a null effect on the quality of the data in the risk and time preferences tasks.

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.

Alfonso, A., Brañas Garza, P., Jorrat, D., Prissé, B., & Vázquez, M. (2023). The baking of preferences throughout high school. In *Mimeo*.

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 Behaviour 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*, Article 104039.

Andreoni, J., & Sprenger, C. (2012). Risk preferences are not time preferences. *The American Economic Review*, *102*(7), 3357–3376.

Bettinger, E., & Slonim, R. (2007). Patience among children. *Journal of Public Economics*, *91*(1–2), 343–363.

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. (2023). Paid and hypothetical time preferences are the same: Lab, field and online evidence. *Experimental Economics*, *26*, 412–434.

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

Brocas, I., & Carrillo, J. D. (2020a). Introduction to special issue "understanding cognition and decision making by children." studying decision-making in children: Challenges and opportunities. *Journal of Economic Behaviour and Organization*, *179*, 777–783.

Brocas, I., & Carrillo, J. D. (2020b). The development of social strategic ignorance and other regarding behavior from childhood to adulthood. *Journal of Behavioral and Experimental Economics*, *85*, Article 101524.

Brocas, I., & Carrillo, J. D. (2020c). The evolution of choice and learning in the two-person beauty contest game from kindergarten to adulthood. *Games and Economic Behavior*, *120*, 132–143.

Brocas, I., & Carrillo, J. D. (2021a). Self-serving, altruistic and spiteful lying in the schoolyard. *Journal of Economic Behaviour and Organization*, *187*, 159–175.

Brocas, I., & Carrillo, J. D. (2021b). Steps of reasoning in children and adolescents. *Journal of Political Economy*, *129*(7), 2067–2111.

Brocas, I., & Carrillo, J. D. (2022). The development of randomization and deceptive behavior in mixed strategy games. *Quantitative Economics*, *13*(2), 825–862.

Brocas, I., Carrillo, J. D., Combs, T. D., & Kodaverdian, N. (2019). The development of consistent decision-making across economic domains. *Games and Economic Behavior*, *116*, 217–240.

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.

Castillo, M., Jordan, J. L., & Petrie, R. (2019). Discount rates of children and high school graduation. *The Economic Journal*, *129*(619), 1153–1181.

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.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, *9*(3), 522–550.

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. In *Mimeo*.

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

Harbaugh, W. T., Krause, K., & Berry, T. R. (2001). GARP for kids: On the development of rational choice behavior. *American Economic Review*, *91*(5), 1539–1545.

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 Behaviour 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*, Article 101955.

Jorrat, D. (2021). Recruiting experimental subjects using WhatsApp. *Journal of Behavioral and Experimental Economics*, *90*, Article 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. (2023). How experiments with children inform economics. *Journal of Economic Literature 61*(2), 504-64.

Lührmann, M., Serra-Garcia, M., & Winter, J. (2018). The impact of financial education on adolescents' intertemporal choices. *American Economic Journal: Economic Policy*, *10*(3), 309–332.

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

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

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

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

Reynolds, B., & Schiffbauer, R. (2004). Measuring state changes in human delay discounting: an experiential discounting task. *Behavioural Processes*, *67*(3), 343–356.

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

Scheres, A., Sumiya, M., & Thoeny, A. L. (2010). Studying the relation between temporal reward discounting tasks used in populations with ADHD: a factor analysis. *International Journal of Methods in Psychiatric Research*, *19*(3), 167–176.

Scheres, A., Tontsch, C., Thoeny, A. L., & Sumiya, M. (2014). Temporal reward discounting in children, adolescents, and emerging adults during an experiential task. *Frontiers in Psychology*, *5*, 711.

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), 510–531.

Sutter, M., Zoller, C., & Glätzle-Rützler, D. (2019). Economic behavior of children and adolescents – A first survey of experimental economics results. *European Economic Review*, *111*, 98–121.

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

Yechiam, E., & Zeif, D. (2023). Revisiting the effect of incentivization on cognitive reflection: A meta-analysis. *Journal of Behavioral Decision Making*, *36*(1), e2286.